

10. 12. 1988

**Zur Erfassung der moralischen Atmosphäre in der Schule -
Theoretischen Grundlegung und erste Ergebnisse¹**

Georg Lind² und Leonore Link

Konstanz 1988

¹ Die vorliegende Arbeit wurde mit Mitteln des Landes Nordrhein-Westfalen gefördert. Die Urheberrechte für den "Moralische Atmosphäre Fragebogen" und die Rechte für die Nutzung des Fragebogens außerhalb des Projekts "Demokratie und Erziehung in der Schule" liegen bei den Autoren.

² Anschrift des Erstautors: Prof. Dr. Georg Lind, Universität Konstanz, 78457 Konstanz.
E-mail: georg.lind@uni-konstanz.de. Web: <http://www.uni-konstanz.de/ag-moral>

Zur Bedeutung der moralischen Atmosphäre in der Schule

Wenn wir uns mit der Struktur der moralischen Atmosphäre einer Schule befassen, setzen wir voraus, daß sie pädagogische Relevanz besitzt. Die pädagogische Forschung hat sich in den letzten Jahrzehnten vorwiegend mit unterrichtsbezogenen Lern- und Lehrtechnologien oder mit sozio-strukturellen Merkmalen von Schulsystemen befaßt. Das Schulleben als eine eigenständige soziale Kategorie aber wurde lange Zeit übersehen - übrigens im Gegensatz zur geisteswissenschaftlich orientierten Pädagogik, die die einzelne Schule und ihre Atmosphäre schon immer als wichtigen Faktor im Erziehungsprozeß betrachtet hat.

Die schulische Atmosphäre hat die die Bühne der pädagogischen Forschung gewissermaßen als Störenfried betreten - in Gestalt des "hidden curriculums", des versteckten oder unsichtbaren Lehrplans (vgl. JACKSON 1968). Das Interesse der pädagogischen Forschung an der moralischen Atmosphäre der Schule hat in den letzten Jahren deutlich zugenommen, und zwar im Zusammenhang mit drei sehr unterschiedlichen Fragestellungen: (1) der Frage nach den Bedingungen von Schulerfolg, (2) der Frage nach der "guten Schule" und (3) der Frage nach der Treatment-Validität bei pädagogischen Interventionen.

In neueren Studien hat sich gezeigt, daß das Klima oder das Ethos einer Schule mehr Unterschiede in den Lernleistungen der Schüler erklären kann als soziostrukturelle oder unterrichtstechnologische Merkmale. Beispielhaft hierfür ist die berühmt gewordene Studie von RUTTER, MAUGHAN, MORTIMORE & OUSTON (1980). In ihrer Zusammenfassung der Ergebnisse heben die Autoren hervor, daß das "Ethos" der einzelnen Schule, das heißt "die Grundstruktur ihrer Wertorientierungen, Einstellungen und Verhaltensmuster, die für die Schule *insgesamt* charakteristisch sind" (RUTTER et al. 1980, 211, unsere Hervorhebung), die bedeutsamste Einflußgröße für Lern- und Sozialisationseffekte war. Dies wird auch durch die Befunde der Konstanzer Schuluntersuchungen von Fend und seinen Kollegen und durch andere Studien bestätigt (vgl. u.a. FEND, DREHER & HÄNISCH 1980; FEND 1985; HÄNISCH 1985; STEFFENS 1985; DREESMANN 1986).

Auf die pädagogische Relevanz des Schullebens verweist die Rutter-Studie noch in einer weiteren Hinsicht. Ihr Titel, "Fünftehtausend (Schul-)Stunden", erinnert uns an die Tatsache, daß wir heute einen großen Teil unseres Lebens in der Schule verbringen. Der Qualität des Lebens in der Schule kommt für sich genommen große Bedeutung für den Einzelnen zu. Schüler lernen nicht nur für das Leben *außerhalb* der Schule, sondern auch für das Leben *in* der Schule. "Non vitae sed scholae" (nicht für das Leben sondern für die Schule!), heißt es in den Epistulae von Seneca, auch wenn es vielerorts sinnverdreht in Stein gemeißelt steht (zur neueren Diskussion über die "Qualität von Schule" vgl. die sechs Schulprofile von LIGHTFOOT 1983, und die Beiträge zu STEFFENS & BARGEL, 1987). So gesehen, hat die moralische Atmosphäre der Schule nicht nur den Status einer "unabhängigen Variable" in der pädagogischen Forschung, sondern auch den einer "abhängigen Variable". Sie ist ein Kriterium für den Erfolg erzieherischen Handelns und bildungspolitischer Maßnahmen.

Schließlich ist die moralische Atmosphäre in der Schule als Randbedingung für Qualität und Erfolg pädagogischen Handelns zu sehen (OSER & PATRY, 1986). Diese Randbedingung

schließt natürlich ein Bündel von Variablen oder besser ein bestimmte Konfiguration von Merkmalen ein wie z.B. Vertrauen und Kooperation zwischen Lernenden und Lehrenden, die eine wichtige Voraussetzung für das Lernen sind.

Was ist und wie mißt man die 'moralische Atmosphäre' in der Schule?

Wie läßt sich die moralische Atmosphäre in der Schule erfassen? Was sind die zentralen, was die peripheren Dimensionen? Was sind theoretisch und pädagogisch valide Indikatoren?

Die Beantwortung dieser Fragen ist schwierig. Als Indikator zur Bestimmung der moralische Atmosphäre könnte die Häufigkeit von Regelübertretungen an einer Schule, z.B. die Anzahl von Diebstählen gewählt werden. Kann dies als eine valide Messung der moralischen Atmosphäre angesehen werden? Ist es möglich, von der Diebstahlsrate an einer Schule auf ihre pädagogische Qualität zurückzuschließen? Das muß wohl bezweifelt werden. Wir wissen, daß die Delinquenzrate an Schulen stark von den sozio-ökonomischen Verhältnissen abhängt, aus denen sie ihre Schüler rekrutieren (vgl. RUTTER et al. 1980, 211). Sie kann der Schule weder angelastet noch, bei niedriger Rate, gutgeschrieben werden.

Wichtiger wäre hier schon die Beantwortung der folgenden Fragen: Gelingt es der Schule, Normübertretungen zu verringern? Wie reagieren Mitschüler und Lehrer auf Normenverstöße? Was wird unternommen, um moralisches Bewußtsein in moralische Kompetenzen zu überführen? Und was trägt die Schule dazu bei, daß soziale Normen und Gesetze von den Einzelnen als gerecht und legitim erfahren werden können und Schüler Verantwortung für ihre Einhaltung übernehmen?³

Uns scheint, daß der Just Community-Ansatz von Kohlberg, Power, Higgins und anderen einen beachtenswerten Versuch darstellt, das Konzept der moralischen Atmosphäre theoretisch umfassend zu begründen, und damit eine wichtige Voraussetzung für ihre valide Messung erfüllt (vgl. KOHLBERG 1987; KOHLBERG & TURIEL 1978; POWER 1983; HIGGINS 1987). Dieser gibt wichtige Aufschlüsse über die pädagogische Zielsetzung und über die Möglichkeit einer schulpraktischen Umsetzung und Erprobung. (Zu den bisherigen Ergebnissen des Just Community-Ansatzes vgl. HIGGINS im Druck; HIGGINS, POWER & KOHLBERG 1984; für eine Diskussion des Ansatzes vgl. OSER 1981, LESCHINSKY 1985; LIND 1987, sowie die Beiträge in REGENBOGEN 1984, und in LIND & RASCHERT 1987).

Methodologische Überlegungen zur Messung der moralischen Atmosphäre

Die Messung der moralischen Atmosphäre in der Schule soll theoretisch fundiert sein, und sie soll schulpädagogischen Absichten dienen. Diese Zielsetzung impliziert ein *Messen im Sinne von Hypothesenprüfung* und im Gegensatz zum *Messen als Technik* (vgl. COOMBS, DAWES & TVERSKY, 1975, 45f.).⁴

³ In einer Untersuchung der Reaktion von Jugendlichen auf hypothetische Stehlsituationen fand JACKSON (1984), daß auch bei Delinquenten das Bewußtsein der Falschheit von Stehlen voll entwickelt war, daß es aber kaum Bedeutung für ihr Handeln hatte.

⁴ Es handelt sich hier um zwei Grundtypen des Messens. Sie zu unterscheiden ist wesentlich, um das Vorgehen bei der Konstruktion des Meßinstruments - hier der Moralische Atmosphäre Fragebogen - verstehen und entsprechend bewerten zu können.

Beim Messen als Technik handelt es sich um eine 2-stellige Relation zwischen Meßmethode und Empirie. Zur Bewertung

Die Messung der moralischen Atmosphäre der Schule dient einerseits der Überprüfung konstruktbezogener Hypothesen und andererseits als Kriterium für die Wirkung pädagogischer Interventionen. Der Meßwert reflektiert im Wesentlichen den Grad der Übereinstimmung der empirischen Beobachtungen mit psychologisch und pädagogisch begründeten Konstrukten (statt mit empirischen Mittelwertsnormen wie beim Messen als Technik).

Zur theoretischen Validierung solcher Meßwerte und zur Frage angemessener Gütekriterien gibt es zahlreiche, zum Teil noch kontrovers diskutierte Vorschläge (vgl. KOHLBERG 1984, 395ff.; BERK 1984), die hier weder vollständig aufgezählt, noch gar im Einzelnen erläutert werden können (siehe LIND 1982). Näher eingegangen werden soll auf das Gütekriterium der "kommunikativen Validität", das bislang nur ungenügend Beachtung fand.⁵

Meßwerte sind das Ergebnis einer Wechselwirkung zwischen dem Meßinstrument *und* dem Gemessenen, also nicht nur eine Funktion des einen oder des anderen. Reliabilität oder Validität sind daher genommen nicht Merkmale des Meßinstruments, sondern Merkmale der Kommunikation einer Person mit dem Forscher mittels Fragebogen, Test oder einer anderen Art von Meßinstrument. Dieser Sachverhalt ist keineswegs trivial. Wie bereits CAMPBELL (1963) anhand von Lernexperimenten zeigen konnte, hängen Meßwerte nicht nur von einem "wahren Wert" und einem zufallsverteilten Meßfehler ab, sondern noch von einem dritten, oft übersehenen Faktor. Diesen Faktor bezeichnet CAMPBELL als den Grad an "kategorialer Überlappung" zwischen Meßinstrument und Gemessenem, in unserem Zusammenhang: zwischen der Meßintention des pädagogischen Forschers und der Antwortintention des befragten Schülers. Da wir eine hohe kategoriale Überlappung nicht a priori als gegeben ansehen können, setzt die Interpretation von Meßwerten voraus, daß das Meßinstrument, das wir verwenden, sensibel für Nichtüberlappungen zwischen Antwortintention und Antwortinterpretation ist. Wir bezeichnen dies als seine "kommunikative Validität".

Um Meßwerte möglichst eindeutig interpretieren zu können, muß der Meßvorgang "kommunikative Validität" besitzen. Ein bekanntes Beispiel für mangelnde kommunikative Validität ist das Problem der Interpretation von mittleren Skalenwerten bei Einstellungsfragebogen (vgl. SCOTT 1968; LIND 1985). Auf was lassen solche Werte schließen? Auf eine "neutrale" oder "mittlere" Einstellung? Auf eine inkonsistente, noch nicht

der Güte der Methode (Reliabilität, Validität) werden empirische Daten herangezogen. Messen als Technik findet dort Anwendung, wo es das Hauptziel ist, auf ökonomische Weise Meßwerte zu erhalten. Die Meßwerte gelten dann als valide, wenn sie mit dem Kriterium, um das es eigentlich geht, hoch korrelieren. Eine theoretische Begründung ist hierfür nicht notwendig. Die Testergebnisse einer Person werden bestimmt, indem man anhand einer Referenzgruppe Normgrößen (z.B. empirische Mittelwerte) ermittelt und mißt, wieweit die Antworten einer Person davon abweichen. Auf diesen Grundtypus des Messens, der - leicht mißverständlich - auch als "normorientierte Testkonstruktion" bezeichnet wird, findet das Arsenal der empirischen Test- und Fragebogenanalyse Anwendung. Im wesentlichen sind dies Verfahren zur Bestimmung von Reliabilitäts- und Validitätsmaßen.

Demgegenüber handelt es sich beim Messen zur Überprüfung einer erziehungswissenschaftlichen Theorie oder eines pädagogischen Programms um ein 3-stelliges Relationsgefüge, das aus Meßmethode, Empirie und Theorie besteht. In diesem Meßansatz wird die Güte der Meßmethode nicht an Daten, sondern an einer Theorie geprüft. Damit Messungen als eine unabhängige Instanz zur Bestätigung oder Falsifikation der damit verbundenen Hypothesen dienen können, muß die verwandte Meßmethode theoretisch, d.h. unabhängig von den empirischen Daten, validiert werden. Grundlage der Bewertung einer Methode, die im Sinne von Messen als Hypothesenprüfung verwandt wird, ist also deren "theoretische Validität". Es geht hierbei um die Frage, wie gut die Methode mit den Vorgaben der Theorie übereinstimmt, die ihrerseits durch die Meßdaten empirisch validiert werden soll. Die empirische Validierung bezieht sich hier auf die Theorie und nicht, wie beim Messen als Technik, auf die Methode. In der erziehungswissenschaftlichen Forschung wird dieser Ansatz als "criterion-referenced test construction" (kriteriumsbezogene Testkonstruktion) bezeichnet (vgl. POPHAN 1978; BERK 1984; KLAUER 1986).

⁵ Ähnlich, aber in der Zielsetzung deutlich verschieden sind die Ansätze der "kommunikativen Validierung" und der "sozialen Validität" (Schuler & Stehle, 1983), bei denen die Akzeptanz der Meßergebnisse durch die Betroffenen im Vordergrund steht.

entwickelte Einstellung? Oder auf eine sehr differenzierte Einstellung?

Offenbar können wir nicht von einer validen Messung sprechen, wenn ein Meßwert so grundverschiedene Interpretationen zuläßt und er keine Anhaltspunkte dafür gibt, welche von mehreren möglichen Interpretation der Kommunikationsabsicht des Befragten angemessen ist. Dieses Problem wird durch die Tatsache verschärft, daß es für einen bestimmten Skalenwert keine allgemeine, für alle Personen gültige Interpretation geben kann und sich daher kein Fragebogen konstruieren läßt, der für jeden Befragten gleich valide ist. Wie wir aus dem Alltagsleben wissen, kann die gleiche Antwort bei verschiedenen Personen oder in verschiedenen Situationen eine andere Bedeutung haben. Die kommunikative Validität muß daher für jede Test-Person-Interaktion neu überprüft werden.

Nehmen wir beispielsweise an, die Lehrerin einer Klasse möchte das moralische Verständnis ihrer Schüler herausfinden. Sie schildert das Dilemma eines Arztes, der von einer totkranken Frau um Sterbehilfe gebeten wird. Aus Zeitungen hat sie zu diesem Thema eine Reihe unterschiedlicher Argumente gesammelt, die sie nun ihren Schülern zur Beurteilung vorlegen möchte. Dazu wählt sie zunächst zwei Argumente aus, beide gegen Sterbehilfe: (a) Der Arzt sollte das nicht tun, weil ein Arzt unter allen Umständen Leben, das das höchste Gut darstellt, erhalten muß. (b) Der Arzt sollte das nicht tun, weil er sonst bestraft wird. Ein Schüler meldet sich und sagt, das erste Argument finde er sehr gut. Was kann die Lehrerin aus dieser Antwort schließen? Ganz gewiß würde eine gute Pädagogin hierauf antworten, gar nichts! Sie weiß, daß die Zustimmung dieses Schülers eine Vielzahl von Interpretationen zuläßt. Es kann sein, daß der Schüler einfach nur geraten hat. Er kann das Argument auch gewählt haben, weil sie es zuerst nannte. Oder er ist aus Glaubensgründen grundsätzlich gegen Sterbehilfe und würde, bei Nachfrage, auch jedes andere Argument dagegen sehr gut finden, alle Pro-Argumente dagegen schlecht, und zwar unabhängig davon ob sie moralisch gut begründet sind oder nicht. Die Lehrerin wird deshalb versuchen, durch geeignetes Nachfragen herauszubekommen, welche Interpretation mit dem übereinstimmt, was für den Schüler ausschlaggebend war.⁶

Wie das Beispiel andeutet, darf ein Verhalten nicht isoliert betrachtet werden, sondern muß im Kontext der Struktur des Verhaltens einer Person interpretiert werden. Kommunikative Validität ist nur zu erreichen, wenn wir nicht einzelnes Verhalten, sondern *Verhaltensmuster* beobachten und erfassen. Nur durch die Messung struktureller Merkmale können wir erfahren, (a) *ob* zwischen Meß- und Verhaltensintention eine kategoriale Überlappung besteht, und (b) *zu welchem Grad* sie in einer bestimmten Person-Test-Interaktion vorhanden ist.

Das Moralische Atmosphäre-Konstrukt von POWER, HIGGINS & KOHLBERG

Bei der Entwicklung ihres Instruments zur Messung der moralischen Atmosphäre in der Schule gingen POWER, HIGGINS und KOHLBERG von der Überlegung aus, daß Moralität nicht nur eine individuelle Angelegenheit des einzelnen Schülers oder Lehrers ist, sondern eine soziale Tatsache. "Die Entwicklung moralischer Individuen", so POWER, "kann ohne die

⁶ Dieser, im Alltag geläufige Ansatz zur Diagnostik psychischer Strukturen anderer Personen war Vorbild für die Konstruktion des Moralische-Urteil-Tests (MUT) von Lind. Der Ansatz wurde dort allerdings zu einem systematischen, experimentellen Testdesign ausgearbeitet (vgl. LIND 1982; LIND & WAKENHUT 1985).

Entwicklung einer moralischen Gesellschaft niemals vollständig realisiert werden" (POWER 1981, 5). Moralität ist also auch ein Merkmal von Institutionen wie der Schule. Sie wird durch Inhalt und Form des in dieser Institution bestehenden sozialen Interaktionsgefüges bestimmt. Das Konstrukt 'moralische Atmosphäre' umfaßt also institutionelle *und* individuelle Handlungen und Sichtweisen sowie deren Zusammenwirken.

Uns liegen bisher zwei Fassungen des Moralische Atmosphäre Fragebogens von Power, Higgins und Kohlberg vor: Diejenige, die in den meisten Untersuchungen von den Autoren verwendet wurde, und eine stark revidierte Fassung, die gegenwärtig in den neuen Schulversuchen in New York zum Einsatz kommt. Die ältere Version des Moralische Atmosphäre Fragebogens ist aus dem Jahr 1984. Bereits bei dieser Version wurde versucht, soziologische und sozialpsychologische Theorien über Aufbau und Funktion von Gruppenstrukturen mit einer entwicklungstheoretischen Perspektive zu verbinden. Der, wie POWER (1983) es nennt, "Grad der Kollektivität", der an einer Schule besteht, und der Wert, den die Gruppe für die Schüler hat, bilden die Kernelemente des Fragebogens. Kollektivität ist dabei bestimmt durch das Ausmaß an Verbindlichkeit, das die gemeinschaftlichen Normen haben. In dieser Hinsicht lehnen sich die Autoren stark an die Theorie der Moralerziehung von EMILE DURKHEIM (1984/1902) an. Die auf ihn zurückgehende Gegenüberstellung von "Stufen der Normenankennung" und "Phasen ihrer Institutionalisierung", die ursprünglich Teil des Moralische Atmosphäre Fragebogens war, hat sich aus Gründen, denen wir hier nicht nachgehen können, nicht bewährt und ist in der neueren Fassung nicht mehr enthalten.

Daneben wurden eine Reihe von weiteren inhaltlichen und strukturellen Verbesserungen vorgenommen. Die Items, mit denen die Stufen der Normenankennung gemessen werden, wurden so verändert, daß sie heute der entwicklungstheoretisch begründeten Ordnung, die sie abbilden sollen, besser entsprechen. Eine weitere wichtige Verbesserung bildet die Auflösung einer in der ersten Version vorhandenen Konfundierung zwischen der wahrgenommenen Verpflichtung zur Normenbefolgung und ihrer Begründung. Die Perspektiven der Beurteilung - Ich, andere Schüler, Lehrer - und die Gründe der Normenbefolgung waren in den Fragebogenaussagen miteinander vermengt. In der neueren Version wird eine klare analytische Trennung vorgenommen, wofür unter anderem die Erfahrungen mit dem experimentellen Fragebogendesign genutzt wurden, wie es dem Moralischen-Urteil-Test zugrundeliegt (LIND 1982; LIND & WAKENHUT 1985).

Auch das Konzept der Verantwortlichkeit hat eine deutliche Präzisierung erfahren. Ausgehend von Argumentationen in Schulversammlungen wurde das Konzept der "kognitiven Verarbeitungsmodi" entwickelt. Dieses wiederum bildete die Grundlage der Formulierung stufenspezifischer Argumente für und gegen normenbezogenes Handeln, zu denen die Befragten Stellung beziehen sollen.

Schließlich wurden als wesentliches Element demokratischer Erziehung Fragen zur Partizipationsbereitschaft aufgenommen. Hier, so scheint uns, haben die Autoren noch zu wenig die Erfahrungen mit der Messung dieses Konstrukts z.B. in der Politischen Psychologie berücksichtigt. Wir werden der Partizipation einen eigenen Fragebogenteil (außerhalb des MAF) widmen.

Auch formal weist die neue Version des MAF bemerkenswerte Änderungen auf. Durch die

Ausarbeitung des Konzepts der moralischen Atmosphäre wurde es möglich, für die meisten Fragen Antwortalternativen zu formulieren und so aus offenen Fragen geschlossene zu machen. Dies garantiert eine größere Vollständigkeit der Daten und eine bessere Transparenz der Fragen für den Schüler. Unseres Erachtens erhöht dies auch deutlich die kommunikative Validität der Meßergebnisse und ihrer Interpretation. Um die Nachteile geschlossener Antwortmodi auszugleichen werden die Schüler ermuntert, die Fragen und Antwortvorgaben zu kommentieren.

Auch in der neuen Version des MAF besteht das Konstrukt der moralischen Atmosphäre aus einer Reihe von Teilkonstrukten. Es wird angenommen, daß diese sich teilweise parallel entwickeln, teilweise aber auch unterschiedliche Phasen oder Aspekte des Schullebens widerspiegeln. Erst als Ganzes gesehen erlauben sie eine pädagogische Einordnung oder Bewertung des Schulklimas.

Die theoretischen Teilkonstrukte der moralischen Atmosphäre werden in der neuen Version konkretisiert durch allgemeine Fragen zum Schulleben und durch fünf Fragebogenteile, in denen spezifische, moralrelevante Probleme in der Schule thematisiert werden:

1. Mogeln bei Klassenarbeiten
2. Prügeleien in der Schule
3. Stehlen
4. Schwänzen des Unterrichts
5. Hilfsbereitschaft der Schüler untereinander.

Mit dieser Auswahl sind häufig vorkommende pädagogische Problembereiche in der Schule abgedeckt. Im Gegensatz zur ersten Fassung, deren Themen stark auf die Versuchsschule, die Cluster-Schule, bezogen waren, ist mit der neuen Version eine breitere Anwendung möglich.

Im allgemeinen Teil des Fragebogens geht es darum, die schulische Umwelt zu erfassen, so wie sie sich den Schülern darstellt. Er gliedert sich in Fragen zum Ist-Zustand, zum subjektiven Problemerkennen und zur Bereitschaft zur Partizipation an der gemeinschaftlichen Behandlung dieser Probleme. Im einzelnen handeln die Fragen von dem Umgang der Schüler miteinander, den Gruppen- und Beziehungsstrukturen, der Normenakzeptanz und dem Vorkommen devianten und prosozialen Verhaltens.

Die themenspezifischen Teile des Moralische Atmosphäre Fragebogens realisieren weitgehend ein einheitliches, experimentelles Design. In jedem Teil werden anhand eines konkreten Problems sechs übergreifende Grunddimensionen der moralischen Atmosphäre in der Schule erfaßt:

1. die wahrgenommene Auftretenshäufigkeit des angesprochenen Problems (Mogeln, Hilfe geben, etc.),
2. die Haltung des Schülers gegenüber der hierfür relevanten moralischen Norm (das sogenannte "deontische Urteil"),
3. die individuelle Verhaltensabsicht zur Lösung des Problems (das "praktische Urteil"),
4. die individuelle Stufe der Normenannahme,
5. die wahrgenommene Stufe der Normenannahme durch die Mitschüler und Lehrer,
6. und die Stufe der individuellen Verantwortlichkeit.

Diese sechs Grunddimensionen lassen sich, wie noch zu zeigen ist, zu Strukturinformationen mit unterschiedlicher Komplexität verbinden.

Entwicklung der deutschen Version des MAF

Ein theoriebezogenes Meßinstrument sorgt durch die Bereitstellung von validen Informationen für die Weiterentwicklung eben jener Theorie. Gleichzeitig machen Änderungen der Theorie und Fortschritte unseres Wissens Modifikationen des Instruments notwendig. Wie schon erwähnt, haben POWER, HIGGINS & KOHLBERG bereits einige Probleme und Unklarheiten der ersten Fassung beseitigt. Bei dem Versuch, den Fragebogen für deutsche Verhältnisse zu adaptieren, ergab sich die Notwendigkeit, weitere Revisionen am Instrument vorzunehmen. Diese sind begründet

- in neueren Erfahrungen mit der theoriegeleiteten Fragebogenkonstruktion,
- in Differenzen zwischen dem deutschen und amerikanischen Schulwesen,
- in neueren Theorien über das moralische Urteilen und Handeln sowie über die Beziehung zwischen Individuum und Gruppe;
- schließlich war ein Kompromiß zwischen wissenschaftlichen Erkenntnisinteressen und schulpraktischen Erfordernissen zu finden, was vor allem - aber nicht ausschließlich - den Umfang des Fragebogens betrifft.

Dem Charakter eines ersten Berichtes entsprechend wollen wir hier nur auf den ersten und letzten Punkt eingehen. Das Problem, den Fragebogen auf einen akzeptablen Umfang zu reduzieren, steht in Verbindung zu den anderen Punkten. Nach der Auffassung von OSER (1982) und anderen bedarf die Operationalisierung der moralischen Atmosphäre durch POWER, HIGGINS & KOHLBERG einiger Ergänzungen. Gleichzeitig scheinen uns Teile der neuen Version verzichtbar zu sein, teils wegen ihrer Redundanz, teils wegen ihrer starken Alters- und Schulspezifität. Damit neben diesem Fragebogen noch andere Instrumente im Rahmen einer Doppelstunde - was wir aus organisatorischen Gründen für das obere Limit halten - eingesetzt werden können, wurde eine maximale Bearbeitungszeit von fünfundvierzig Minuten angestrebt. Wir haben uns deshalb auf zwei Themen - Mogeln und gegenseitiges Helfen - beschränkt. Ihre pädagogische Relevanz scheint uns am wenigsten von Schultyp, Alterstufe oder regionalen Besonderheiten abzuhängen. Darüberhinaus bestehen für beide Themen zwischen Schülern und Lehrern kontroverse Normenvorstellungen. Sie stellen also eine gute Grundlage für Diskussionen im Rahmen der Schule dar und können auf unterschiedlichen Stufen der Moral und Gemeinschaftlichkeit reflektiert werden. Gespräche mit Schülern zeigen beispielsweise, daß es in bestimmten Situationen auch gute Gründe für Mogeln bzw. gegen Helfen geben kann.

Zur Verbesserung der kommunikativen Validität von Fragebögen werden in der Literatur eine Reihe von Kriterien genannt (vgl. u.a. die einschlägigen Beiträge in BERK 1984), die wir aufgrund eigener Erfahrungen mit Fragebogenstudien um weitere ergänzt haben. Die von uns hierzu erstellte Liste erhebt dennoch nicht den Anspruch, vollständig zu sein.

Um die wechselseitige Verständigung mittels eines Fragebogens zu optimieren, sollten bei

der Revision folgende Anforderungen berücksichtigt - nicht aber rigide angewendet - werden:

- die Formulierung von möglichst einfachen Sätzen,
- die Vermeidung von Doppelverneinungen,
- die Vermeidung von Aussageverknüpfungen,
- die Orientierung am allgemeinen Sprachgebrauch, d.h. vor allem die Vermeidung einer kurzschlüssigen Übersetzung von psychologischen Fachbegriffen in Frage-Items,
- Vermeidung von "forced-choice" Vorgaben, d.h. die Ermöglichung von neutralen oder von Weiß-Nicht-Antworten,
- die Einbettung abstrakter Fragen (z.B. nach allgemeinen Werten) in einen situativen Kontext
- und die Vermeidung von Fragen, die forschungsethisch problematisch sind, wie z.B. Fragen über die Gesinnung anderer Personen.

Das letztgenannte Problem stellt sich in besonderer Weise für einen Fragebogen, der sich auf Gruppenstrukturen und Perspektivenübernahmen in einer sozialen Gemeinschaft bezieht. Wir können den Befragten um Auskunft über seine eigenen Werthaltungen bitten und auch darüber, wie er das Verhalten anderer wahrnimmt. Wenn er jedoch seine Einschätzung der Einstellung anderer beschreiben soll, dann drängen wir ihn ganz offensichtlich dazu, Projektionen oder Stereotype zu produzieren. Bei bestimmten Fragestellungen der pädagogischen Forschung kann hierin der Sinn der Befragung liegen. Die Erhebung der moralischen Atmosphäre in der Schule hat jedoch nicht nur eine wissenschaftsinterne Zielsetzung. Durch ihre Ergebnisse soll sie selbst dazu beitragen, das Schulklima zu verbessern. Mit dieser Intention unverträglich wäre Aufforderung, Stereotype über andere Menschen zu produzieren, zumal sie den Befragten als wissenschaftlich autorisiert erscheinen muß. Das würde weder der Förderung der Gemeinschaftlichkeit dienen, noch der Verbesserung eines demokratischen Umgangs miteinander. Die forschungsethische Problematik ist hier vielleicht weniger gravierend als im Milgram-Experiment, unseres Erachtens aber ebensowenig zu vernachlässigen.

In der Literatur kaum zu finden und auch nicht in der obigen Liste enthalten, ist das eingangs diskutierte Problem der theoretischen Validität des Meßinstruments. Mangelnde theoretische Validität kann ebenso wie die anderen angesprochenen Punkte die Interpretation der Ergebnisse einer Untersuchung stark beeinträchtigen. Das Problem ist zwar eng verbunden mit dem Stand der Theorieentwicklung, manifestiert sich aber besonders in der Formulierung der Fragebogenitems. Bei den einzelnen Teilen der amerikanischen Fassung des Moralische Atmosphäre Fragebogens ist dieses Problem, vergleicht man ihn mit anderen Fragebogen zum Schulklima, gut gelöst. Revisionsbedürftig erscheinen uns jedoch einzelne Teile, vor allem jene, die der Messung der Verantwortlichkeit dienen. Die Items der Version von POWER, HIGGINS & KOHLBERG lassen sich nur schwer den theoretisch formulierten Stufen zuordnen, was belegt, daß es gegenwärtig noch nicht gelungen ist, Verantwortung und Verantwortlichkeit theoretisch als ein konsistentes Entwicklungskonzept zu definieren.

Verantwortungsurteile

Nach Higgins, Power & Kohlberg (1984) bilden die Übernahme sozialer Verantwortung und die

darauf bezogenen Urteile eine Grunddimension der moralischen Atmosphäre einer Gemeinschaft. Hiernach sind Urteile über Art und Umfang personaler Verantwortung direkt bezogen auf das Wohlergehen anderer. Thematisiert und gegen Tendenzen zur Vermeidung persönlicher Unannehmlichkeiten abgewogen werden prosoziale Werte, wie gegenseitige Hilfe, Vertrauen und Teilnahme. Kohlberg und seine Kollegen postulieren, daß Verantwortungsurteile die subjektiven Handlungsentscheidungen leiten und über den Umgang miteinander die Organisation gemeinschaftlicher Lebensformen beeinflusst. Die durch Verantwortungsübernahme realisierte Gemeinschaftlichkeit bestimmt (zusammen mit dem Bemühen um Gerechtigkeit) nicht nur die in einer Gruppe bestehende moralische Atmosphäre. Sie beeinflusst nach Kohlberg auch die individuelle sozial-moralische Entwicklung durch die Erfahrung konkreter interaktiver Verhaltensweisen und des Raumes, der den einzelnen hier zur aktiven Mitgestaltung und Veränderung gewährt wird. Daraus resultierende Verantwortungsurteile legen die Verbindlichkeit fest, die subjektiv anerkannten moralischen Geboten zukommt. Schließlich vermitteln Verantwortungsurteile nach Kohlberg & Candee (1984) zwischen Urteilen (deontische Wahl) und Handeln mit dem Ziel einer zunehmenden Übereinstimmung zwischen beidem.

Aus diesen theoretischen Vorüberlegungen ergeben sich folgende Hypothesen:

1. In Abhängigkeit zur moralischen Atmosphäre in Schülergruppen unterscheiden sich Klassen hinsichtlich (a) der Bedeutsamkeit, die prosozialen Werten beigemessen wird und (b) der Differenziertheit, mit der Argumente zur Verantwortung beurteilt werden.
2. Die Ausprägungen beider Komponenten der Verantwortungsurteile beeinflussen die individuelle Handlungsmotivation.

Higgins, Power & Kohlberg (1984) beschreiben 6 qualitativ unterschiedliche Modi der Verantwortung, die in einer hierarchischen Ordnung zueinander stehen sollen. Die Entwicklung der Verantwortung soll gekennzeichnet sein durch:

- den Übergang von einer extrinsischen zu einer intrinsischen Handlungsmotivation;
- die Ausdehnung des Personenkreises für den Verantwortung im Sinne der Förderung des Wohlergehens besteht, von Freunden hin zu allen Mitgliedern der Gruppe
-einer zunehmenden Anerkennung der Gemeinschaft als "Wert an sich".

Im Moralischen-Atmosphäre-Fragebogen wird das Konstrukt 'Verantwortung' folgendermaßen erfaßt: Bezogen auf ein vorgegebenes Dilemma, in dem es darum geht, einem unbeliebten Mitschüler zu helfen, beurteilen die Schüler je 6 Argumente für und gegen ein Helfen hinsichtlich ihrer Akzeptabilität auf einer Skala von -2 bis +2. Diese Werte bilden die Datenbasis für eine intraindividuelle Varianzanalyse mit der Bestimmung der Determinationskoeffizienten für die beiden realisierten Designfaktoren:

1. Richtung der Argumente (für oder gegen Helfen)
2. Modi der Verantwortung (6 qualitativ verschiedene Urteilsstrukturen).

Zusätzlich wurden die beiden Intra-Klassen-Korrelationen berechnet, die die Quadratsumme des jeweiligen Faktors direkt in Beziehung setzen zur Restvarianz und so einen um den Effekt es

anderen bereinigten Wert darstellen.

Erste Ergebnisse zur revidierten Fassung der Verantwortungs-Fragenbogens im MAF liegen aus 1987 durchgeführten Befragungen vor. Daran nahmen insgesamt 219 Schüler aus 9 Klassen unterschiedlicher Jahrgänge und Schulen teil. Für die Analyse haben wir hieraus zwei 8. Klassen aus zwei verschiedenen Gymnasien ausgewählt. Die Schüler in den beiden Klassen sind sich hinsichtlich Altersstruktur und Geschlechtszusammensetzung weitgehend ähnlich, so daß hieran der Einfluß des Klassenklimas besonders schön illustriert werden kann. In Tabelle 1 sind die Strukturwerte des Verantwortungsurteils der Schüler beider Klassen gegenübergestellt

(hier: Tabelle 16 a) (ohne MUT-Daten)

Die Unterschiede zwischen beiden Klassen sind angesichts der geringen Stichprobengrößen beachtlich. Während sich die Schüler der Klasse 8A stärker an der Richtung der Argumente und weniger an deren Qualität orientieren, verweisen die Ergebnisse der anderen Klasse auf ein dazu spiegelbildliches Urteilsverhalten.⁷

(hier: Tabelle 16 b) (ohne MUT-Daten)

Zur Prüfung der theoretisch postulierten Verbindung zwischen den Verantwortungsurteilen und dem Ausmaß, in dem moralische Werte handlungsleitend werden, wurden die Angaben zu 5 im Instrument enthaltenen Fragen herangezogen. Inhaltlich beziehen sie sich auf Einstellungen gegenüber der subjektiv empfundenen Verbindlichkeit prosozialer Verhaltensweisen und demokratisch orientierten Entscheidungsverfahren. Die Korrelationskoeffizienten zwischen den Angaben zu sozial- und verantwortungsbezogenen Verhalten und den Strukturmaßen der Verantwortungsurteile sind in Tabelle 3 aufgeführt. Sie zeigt, daß die Betroffenheit für die Konsequenzen eigenen Handelns am stärksten mit einer Orientierung an der Richtung der Argumente einhergeht. Die positive Korrelation wird auf dem 1 %-Niveau signifikant. Mit abfallender Höhe folgen: der Wunsch über Probleme gemeinsam zu diskutieren und die erklärte Bereitschaft, zu helfen. Für alle diese Items ist die Korrelation mit der Determination der Verantwortungsurteile durch die Qualität der Argumente negativ. Dies ändert sich lediglich für das anderen gewährte Recht auf Mitbestimmung, wo der Zusammenhang positiv und auf dem 5 %-Niveau signifikant ist.

Zusammengefaßt ergibt die Datenlage folgendes Bild:

- die Werte in den Strukturmaßen der Verantwortungsurteile variieren in Abhängigkeit zur Klassenzugehörigkeit der Schüler und
- die größten Unterschiede bestehen hier hinsichtlich der Bedeutsamkeit, die prosozialen

⁷ Bei einer einfaktoriellen Varianzanalyse mit Klassenzugehörigkeit als unabhängiger und den Strukturmaßen der Verantwortungsurteile als abhängigen Variablen beträgt die durch den Faktor "Pro/Con" erklärte Varianz 7,25 % und ist auf dem 5 %-Niveau signifikant, die dazugehörige Intra-Klassen Korrelation sogar auf dem 1 %-Niveau. Das heißt, die beiden Klassen unterscheiden sich am auffälligsten hinsichtlich der Bedeutung, die die Schüler prosozialen Werten, hier der gegenseitigen Hilfe, beimessen. Die Differenzen in der Ausrichtung der Urteile nach der Qualität der Argumente verschwinden bedingt durch die niedrigere Restvarianz der Klasse 8A für den Intra-Klassen-Korrelationskoeffizienten. Der Einfluß dieses Faktors ist bereinigt um den des anderen, für beide Klassen nahezu gleich.

Handlungen entgegengebracht wird.

Beide Ergebnisse zusammen können als ein Indikator für die unterschiedliche moralische Atmosphäre innerhalb der beiden Schülergruppen gewertet werden. Aufgrund des mit $N = 2$ sehr kleinen Stichprobenumfangs kann an dieser Stelle jedoch nicht abgeschätzt werden, inwieweit den berichteten Effekten umfassendere Gültigkeit zukommt.

Darüber hinaus konnten bedeutsame Zusammenhänge zwischen der Struktur der Verantwortungsurteile und einzelnen Aspekten sozialbezogener Handlungsmotivationen aufgezeigt werden. Dabei scheint die Sorge um das Wohlergehen anderer, Beziehungen und das daraus entstehende Gefühl der Verantwortung eine starke affektive Komponente zu haben. Hier wäre es sicher wichtig, die bisherigen Ergebnisse durch inhaltlich differenziertere Fragen zur Handlungsmotivation zu ergänzen, um ein klareres Bild über die Zusammenhänge zwischen Verantwortungsurteilen und moralischem Handeln zu erhalten.

Beispiel einer Beantwortung

Den Abschluß sollen zwei Beispiele bilden. Im ersten geht es darum zu zeigen, wie einzelne Fragen im Moralische Atmosphäre Fragebogen auf den Dimensionen Selbst- und Fremdurteil zu strukturellen Informationen verbunden werden können. Als zweites wollen wir anhand des Themas "Mogeln bei Klassenarbeiten" nochmals beispielhaft darstellen, welche Bedeutung Urteilsstrukturen für das Verständnis von Verhalten haben können.

Eines der wesentlichsten Ergebnisse der ersten Forschungen zur moralischen Atmosphäre in der Schule betrifft Unterschiede bei der Beurteilung der Hilfsbereitschaft ihrer Mitschüler, zwischen Schülern von "Gerechten Gemeinschaft Schulen" und denen aus "normalen" Schulen.

In den Experimental-Schulen zeigte sich eine, wie HIGGINS et al. (1984) es nannten, geringere "pluralistische Ignoranz". Mit diesem Terminus wird die Tatsache bezeichnet, daß Schüler ein falsches Bild von den Handlungsabsichten ihrer Mitschüler haben. Zwar war in *beiden* Schultypen bei den Schüler die Bereitschaft, anderen zu helfen, häufig vorhanden, aber nur in der Just Community Schule zeigten die Schüler diesbezüglich auch ein wechselseitiges Vertrauen (80% vs. 40%).

Interessanterweise verhielten sich die beiden Schüler im Alter von 15 und 17 Jahren, denen wir den MAF zu einer ersten Erprobung vorgelegt hatten, in ähnlicher Weise wie die amerikanischen Schüler der "normalen" Schulen. Beide bejahten die moralische Verpflichtung, daß man auch einem unbeliebten Mitschüler in einer Notlage helfen soll, waren sich aber über ihr eigenes Verhalten im Unklaren und glaubten, daß nur wenige ihrer Mitschüler bereit wären zu helfen.

Der methodisch interessante Aspekt an der "pluralistischen Ignoranz" ist, daß sie nicht durch die isolierte Betrachtung von Einzelfragen oder durch die Bildung von Summenwerten ermittelt werden kann, sondern - als strukturelle Information - nur über die Analyse von ganzen Antwortmustern.

Das zweite Beispiel, auf das wir hier näher eingehen wollen, betrifft den Problemkreis "Mogeln bei Klassenarbeiten". Im Gegensatz zu US-amerikanischen Schülern, bei denen Betrügereien bei Prüfungen gegen den Ehrenkodex verstoßen, ist für deutsche Schüler gerade

dieses Thema ein Punkt, wo ihre Normenvorstellungen mit denen der Lehrer in Konflikt geraten. Für die meisten Lehrer stellt abschreiben oder abschreiben lassen bei Klassenarbeiten einen eindeutigen Verstoß gegen schulische Normen dar. Im Gegensatz dazu sehen viele Schüler das Mogeln durchaus als legitim an. Insbesondere die Weigerung, einen Mitschüler abschreiben zu lassen, gilt oft als unsolidarisch und als ein Verhalten, das gegen die Gemeinschaft gerichtet ist. Es verbietet sich also, Mogeln a priori als Indikator für "unmoralisches Verhalten" zu werten.

Weil sich dieses Problem für Abschreiben und für Abschreiben lassen unterschiedlich darstellen kann, enthält der Moralische Atmosphäre Fragebogen zu beiden Themen eigene Fragen. Zusätzlich wurde ein (neues) Dilemma zum Abschreiben-lassen aufgenommen, in dem die Dimensionen "Hilfe für einen Freund" und "gerechte Leistungsbeurteilung" miteinander konfliktieren. Darin wird nach dem eigenen Verhalten und nach der Beurteilung seiner Richtigkeit gefragt.

Die Antworten unserer beiden Schüler bestätigen, daß ein und dasselbe Verhalten durchaus unterschiedlich beurteilt werden kann. Bei der Beurteilung des Mogelns differenzieren die Schüler nach situativen Bedingungen. Abschreiben lassen wird nicht durchweg positiv beurteilt, z.B. dann nicht, wenn der, der bei einem abschreiben will, offenkundig faul war.

Es reicht also nicht aus, allein das Verhalten der Schüler zu beobachten, um Rückschlüsse auf die moralische Atmosphäre in der Schule zu ziehen. Die Anerkennung schulischer Normen durch die Schüler setzt voraus, daß diese die Bedeutung der Normen für die Aufrechterhaltung der Gemeinschaft verstehen. Sie setzt aber auch ein Verständnis für Regelverstöße und deren Gründe voraus. Moralische Atmosphäre gründet auf wechselseitigem Verstehen, auf der Übernahme der Perspektive des anderen. An diesem Kriterium wird sich auch die Güte des Moralische-Atmosphäre-Fragebogens erweisen müssen.

Literatur

- BERK, R.A., (Hrsg.) (1984). *A Guide to Criterion-Referenced Test Construction*. Baltimore: John Hopkins University Press.
- CAMPBELL, D.T. (1963). Social attitudes and other acquired behavior dispositions. In KOCK, S. (Hrsg.). *Psychology: A study of a science*. New York: McGraw-Hill, 94-172.
- COOMBS, C., DAWES, R. & TVERSKY, A. (1975). *Mathematische Psychologie*. Weinheim: Beltz.
- DREESMANN, H. (1986). Zur Psychologie der Lernumwelt. In WEIDENMANN B. & KRAPP, A. (Hrsg.). *Pädagogische Psychologie*. München: Urban & Schwarzenberg, 449-491.
- DURKHEIM, E. (1984). *Erziehung, Moral und Gesellschaft*. Frankfurt: Suhrkamp (Original 1902).
- FEND, H. (1985). *Theorie der Schule*. München: Urban & Schwarzenberg.
- FEND, H., DREHER, E. & HÄNISCH, H. (1980). Auswirkungen des Schulsystems auf Schulleistungen und soziales Lernen. *Zeitschrift für Pädagogik*, 26, 673-698.
- HÄNISCH, H. (1985). Was ist eine 'gute' Schule? Empirische Forschungsergebnisse und Anregungen für die Schulpraxis. (Heft 7). Soest: Arbeitsberichte zur Curriculumentwicklung, Schul- und Unterrichtsforschung.
- HIGGINS, A. (1987). Ist moralisch-demokratische Erziehung in der Schule möglich? Schulpraktische Erfahrungen in den USA. In LIND, G. & RASCHERT, J. (Hrsg.). *Moralische Urteilsfähigkeit. Eine Auseinandersetzung mit Lawrence Kohlberg über Moral, Erziehung und Demokratie*. Weinheim: Beltz.
- HIGGINS, A., POWER, C. & KOHLBERG, L. (1984). The relationship of moral atmosphere to judgments of responsibility. In KURTINES, W. & GEWIRTZ, J. (Eds.). *Morality, moral behavior and moral development*. New York: Wiley, 74-106.
- JACKSON, Ph. (1968). *Life in the Classroom*. New York: Holt, Rinehart & Winston.
- JACKSON, M.S. (1984). The measurement of simulated stealing behavior and moral judgment - A technique. *Australian Psychologist*, 19, 263-269.
- KLAUER, K.J. (1986). Forschungsmethoden der Pädagogischen Psychologie. In WEIDENMANN, B. & KRAPP, A. (Hrsg.). *Pädagogische Psychologie*. München: Urban & Schwarzenberg, 73-104.
- KOHLBERG, L. (1984). The meaning and measurement of moral judgment. In KOHLBERG, L. (Hrsg.). *Essays on moral development*. Vol. II. New York: Harper & Row, 395-425.
- KOHLBERG, L. (im Druck). Moralische Entwicklung und demokratische Schule. In LIND, G. & RASCHERT, J. (Hrsg.). *Moralische Urteilsfähigkeit. Eine Auseinandersetzung mit Lawrence Kohlberg über Moral, Erziehung und Demokratie*. Weinheim: Beltz.
- KOHLBERG, L. & TURIEL, E. (1978). Moralische Entwicklung und Moralerziehung. In PORTELE, G. (Hrsg.). *Sozialisation und Moral*. Weinheim: Beltz, 13-80.
- LESCHINSKY, A. (1985). Werterziehung und Schulentwicklung. *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*, 5, 89-107.
- LIGHTFOOD, S. (1983). *The good high school. Potraits of character and culture*. New York: Basic Books.

- LIND, G. (1982). Experimental Questionnaires: A new approach to personality research. In KOSSAKOWSKI, A. & OBUCHOWSKI, K. (Hrsg.). *Progress in psychology of personality*. Amsterdam: North Holland Publishing, 132-144.
- LIND, G. (1985). Attitude change or cognitive-moral development? How to conceive of socialization at the university. In LIND, G., HARTMANN, H.A. & WAKENHUT, R. (Hrsg.). *Moral development and the social environment. Studies in the psychology and philosophy of moral judgment and education*. Chicago: Precedent Publishing Inc., 173-192.
- LIND, G. (1987). Die "Gerechte Gemeinschaft-Schule" - Eine Verbindung von Erziehung und Demokratie in der Schule. *Deutsche Schule, 1*, 4-12.
- LIND, G. & WAKENHUT, R. (1985). Testing for moral judgment competence. In LIND, G., HARTMANN, H.A. & WAKENHUT, R. (Hrsg.). *Moral development and the social environment. Studies in the psychology and philosophy of moral judgment and education*. Chicago: Precedent Publishing Inc., 79-105.
- LIND, G. & RASCHERT, J. (Hrsg.) (1987). *Moralische Urteilsfähigkeit. Eine Auseinandersetzung mit Lawrence Kohlberg über Moral, Erziehung und Demokratie*. Weinheim: Beltz.
- OSER, F. (1981). *Moralisches Urteil in Gruppen, soziales Handeln, Verteilungsgerechtigkeit*. Frankfurt: Suhrkamp.
- OSER, F. & PATRY, J.-L. (1986). Interventionsstudien für sozial-kognitive Kompetenz: Beispiele und theoretische Überlegungen zur Transformations-Validität. *Unterrichtswissenschaft, 14*, 254-268.
- POPHAM, W.J. (1978). *Criterion-referenced measurement*. Englewood Cliffs: Prentice-Hall.
- POWER, C. (1981). Moral education through the development of the moral atmosphere of the school. *The Journal of Educational Thought, 15*, 4-19.
- POWER, C. (1983). Die moralische Atmosphäre in der Schule: Eine Analyse demokratischer Versammlungen. In LIND, G., HARTMANN, H.A. & WAKENHUT, R. (Hrsg.). *Moralisches Urteilen und soziale Umwelt*. Weinheim: Beltz, 123-136.
- POWER, C. (1986). Demokratische und moralische Erziehung in einer großen öffentlichen High School. In OSER, F., FATKE, R. & HÖFFE, O. (Hrsg.). *Transformation und Entwicklung*. Frankfurt: Suhrkamp, 297-323.
- REGENBOGEN, A. (Hrsg.) (1984). *Moral und Politik - Soziales Bewußtsein als Lernprozeß*. Köln: Pahl-Rugenstein.
- RUTTER, M., MAUGHAN, B., MORTIMORE, P., & OUSTON J. (1980). *Fünftehtausend Stunden: Schulen und ihre Wirkung auf die Kinder*. Weinheim: Beltz.
- SCOTT, W.A. (1964). Attitude measurement. In LINDSEY, G. & ARONSON, E. (Hrsg.). *Handbook of Social Psychology*. Vol. 2. Reading MA: Addison-Wesley, 204-273.
- SCHULER, H. & STHELE, W. (1983). Neuere Entwicklungen des Assessment Center-Ansatzes - beurteilt unter dem Aspekt der sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie 27*, 33-44.
- STEFFENS, U. (1987). Gestaltbarkeit und Qualitätsmerkmale von Schule aus Sicht der empirischen Schulforschung. In U. STEFFENS & T. BARGELE (Hrsg.), *Qualität von Schule*. Heft 1, S. 21-40. Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung.
- STEFFENS, U. & BARGEL, T. (Hrsg.) (1987). *Qualität von Schule*. Heft 1 bis 4. Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung.