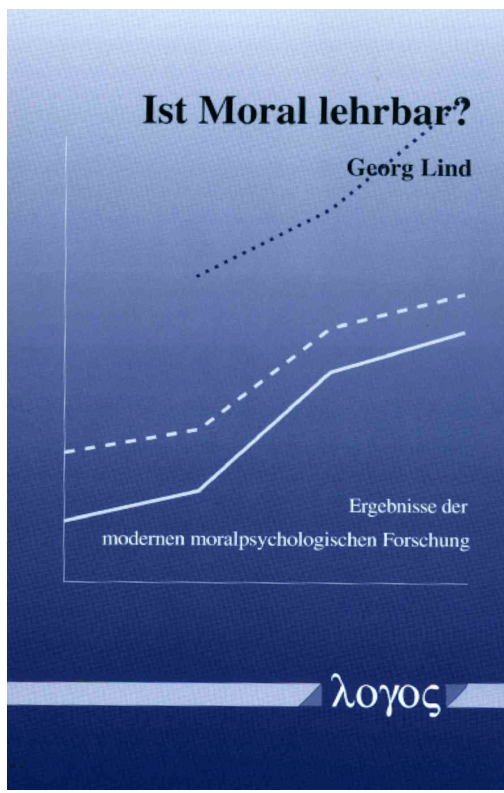
Georg Lind

Ist Moral lehrbar?

Ergebnisse der modernen moralpsychologischen Forschung



Ist Moral lehrbar?

Ergebnisse der modernen moralpsychologischen Forschung

Georg Lind

Zweite, ergänzte Auflage
2002, Berlin, Logos-Verlag

“Die einzig wahrhafte Kraft gegen das Prinzip von Auschwitz wäre Autonomie, wenn ich den Kantischen Ausdruck verwenden darf, die Kraft zur Reflexion, zur Selbstbestimmung, zum Nicht-Mitmachen.”

Theodor W. Adorno

Für
Gisela, Gregor, Antonio und Glenda

Inhalt

Vorwort zur zweiten Auflage 7

1 Zu einer Bildungstheorie der Moralentwicklung 15

2 Das „Zwei-Aspekte-Modell" des moralischen Urteilsverhaltens und der Moralentwicklung 37

3 Kohlbergs *Moral Judgment Interview* – Eine Kritik 69

4 Der *Moralisches Urteil-Test* – Ein experimenteller Fragebogen 83

5 Moralische Bildung versus moralische Reifung: Das Problem der Regression 99

6 Bildung versus Sozialisation I: Kann moralische Urteilsfähigkeit wie eine politische Einstellung simuliert werden? 119

7 Bildung versus Sozialisation II: Kulturelle Einflüsse auf die moralische Urteilsfähigkeit 133

8 Förderung der moralischen Urteilsfähigkeit durch Bildungsprozesse: Ein Interventionsexperiment 159

9 Bedingungen für den Effekt von Interventionen zur Förderung der moralischen Urteilsfähigkeit: Eine Meta-Analyse 191

10 Moralische Stereotype und pluralistische Ignoranz: Die Wahrnehmung moralischer Bedürfnisse von Schülern durch Lehrer 211

11 Bildungserfahrungen als Bedingung für moralische Entwicklung 241

12 Schlußfolgerungen 259

13 Literatur- und Quellenverzeichnis 273

Anhang: *Moralische Atmosphäre Fragebogen* (MAF) 303

Danksagung

An der Erstellung dieses Buches waren viele Personen beteiligt. Für die kritische Durchsicht verschiedener Manuskriptversionen danke ich Tino Bargel, Peter Dobbelstein-Osthoff, Dr. Monika Kuhn, Ottfried Lind, Leonore Link, Manfred Scholz, Christa Kolbert-Ramm, Johann-Ulrich Sandberger, Dr. Heinz Schirp, Karin Tritt. Redaktion und Lektorat übernahm Patricia Knoop.

Die hier geschilderten Untersuchungen wurden gefördert aus Mitteln der *Deutschen Forschungsgemeinschaft* (Bonn), der *Forschungsgruppe Gesellschaft und Region e.V.* (Konstanz), des *Kultusministeriums von Nordrhein-Westfalen*, von *EMNID* (Bielefeld) und der *Universität Konstanz*.

Vorwort zur zweiten Auflage[1]

Hier wird die Jahrtausende alte und heute mehr denn je aktuelle Frage zu beantworten versucht: Ist Moral lehrbar?

Bereits Sokrates, so berichtet Platon, warf diese Frage auf. Er fand keine eindeutige Antwort, erkannte jedoch, dass sie mit einer anderen, ebenso wichtigen Frage in Verbindung steht: Ist Moral eine Fähigkeit (oder nur eine Einstellung und Werthaltung)? Moral sei lehrbar, so verstehe ich seinen Dialog mit Menon, wenn sie eine "Tugend" oder, wie wir heute sagen würden, eine moralische Fähigkeit wäre. Auch das Umgekehrte sei wahr: Wenn sie eine Fähigkeit sei, dann sei sie auch lehrbar.

Viele Autoren verneinen beides, wobei sie dafür unterschiedliche Gründe angeben. Moral sei nicht lehrbar, sagen die einen, weil sie eine Frage angeborener Persönlichkeitseigenschaften (so genannter *traits*) sei. Man könne nur versuchen, durch Selektion die moralisch guten von den moralisch weniger guten Menschen zu trennen, um so dafür zu sorgen, daß die Moral zum Zug kommt oder doch wenigstens nicht ganz untergeht. Moral sei nicht lehrbar, sagen die anderen, weil sie keine Frage der Fähigkeit, sondern der *Werthaltung*, *Einstellung*, *Motivation* oder des *richtigen Bewußtseins* sei. Solche Dinge aber könne man nicht im herkömmlichen Sinne lehren oder unterrichten. Sie könne man nur durch Einsichtigmachen oder, wenn das nichts nutze, durch sozialen Zwang *verändern*.

Für Erziehung und Bildung ergeben sich aus diesen Vorstellungen über Moral klare Empfehlungen. Wenn Moral eine angeborene Eigenschaft der Menschen ist, dann ist moralische Erziehung zwecklos. Die Gesellschaft kann sich darauf beschränken, die Menschen in moralisch gute, weniger gute und schlechte zu klassifizieren und entsprechend dieser Klassifikation zu behandeln. Aus der Annahme, dass Moral bloß eine Werthaltung oder Einstellung oder ähnliches sei, leitet sich ab, dass moralische Bildung und Erziehung sich darauf beschrän-

1 Gegenüber der ersten Ausgabe dieses Buches unter dem Titel "Moral und Bildung" (im Asanger-Verlag, Heidelberg) wurden zahlreiche Änderungen eingefügt und viele Graphiken überarbeitet. Die zweite Auflage dieser Neuausgabe wurde vor allem dazu genutzt, Rechtschreibkorrekturen anzubringen und dieses Vorwort und die Literaturliste auf den neuesten Stand zu bringen. Die Seitennummerierung hat sich dadurch leicht verändert. Aus technischen Gründen wurde die alte Rechtschreibung beibehalten.

ken kann, dem Heranwachsenden die vorgegebenen moralischen Regeln der Gesellschaft einsichtig zu machen oder ihn auf andere Weise dazu zu bringen, dass er sie (zumindest verbal) akzeptiert. Solche Vorstellungen sind aber kaum vereinbar mit der Überzeugung, dass der Mensch bildbar ist, und mit den Prinzipien der Demokratie, die auf sozialen Zwang und Indoktrination verzichten. Beide Vorstellungen haben sich bislang auch kaum durch empirische Forschung untermauern lassen. Persönlichkeitseigenschaften existieren mehr in den Augen des Betrachters als in dem tatsächlichen Verhalten der Menschen (siehe u.a. Cohen, 1969). Auch moralische Einstellungen und Werthaltungen zeigen kaum einen systematischen Bezug zum Verhalten von Menschen (Hartshorne und May, 1929-30) und können, wenn nötig, in fast beliebiger Weise simuliert werden (Emler et al., 1983; siehe auch Kapitel 6). Diese Befunde besagen nicht, dass die genetisch bestimmte Ausstattung des Menschen und seine Werthaltungen und Einstellungen keine Bedeutung für sein Verhalten hätten. Sie weisen aber darauf hin, dass dies keine konsistenten und unveränderbaren Eigenschaften sind und dass die Ergebnisse von Persönlichkeitsskalen und Einstellungstests für die Frage der Lehrbarkeit der Moral wenig hilfreich sind. Seit Beginn des 19. Jahrhunderts haben moralpsychologische Studien vielfache Hinweise erbracht, daß Moral nicht bloß eine Frage der richtigen Einstellung oder Werthaltung, sondern vielmehr eine Frage der Fähigkeit ist. Der Berliner Gerichtspsychiater Levy-Suhl (1912) fand heraus, dass straffällige Jugendliche diesselben moralischen Werte haben, wie nichtstraffällige Jugendliche, und es also andere Gründe für unmoralisches Verhalten geben müsse. Dieser frühe Befund wurde nachfolgend vielfach bestätigt. Hartshorne und May (1929) zeigten in einer breit angelegten Experimentalstudie, dass betrügerisches Verhalten kaum an bestimmte moralische Einstellungen gebunden ist und auch in keinem messbaren Zusammenhang mit moralischer Unterweisung im Religionsunterricht steht. Piaget (1932/1973a) machte die Frage nach der Beziehung zwischen moralischem Urteil und Verhalten zum Gegenstand aufschlussreicher Untersuchungen über das kindliche Verhalten in Spielgruppen. Kohlbergs (1958) Studien zur Moralentwicklung bei Jugendlichen führte ihn zum dem Begriff der *moralischen Urteilsfähigkeit*, der, wie ich in diesem Buch zeigen werde, eine Brücke schlägt von individuellen moralischen Werthaltungen einerseits zum Verhalten andererseits (Kapitel 2). Diese Hinweise wurden jedoch lange Zeit wenig beachtet und werden auch heute noch vielfach in ihrer Bedeutung unterschätzt

oder falsch verstanden. Das sogenannte Urteils-Handlungs-Problem wird noch immer unter Ausschluss des Begriffs der Urteils*fähigkeit* und der dazu durchgeführten Forschung zu lösen versucht, was, wie nicht anders zu erwarten, nicht gelingen kann (Garz et al., 1999). Dabei sind moralische Fähigkeiten seit Anfang der sechziger Jahre expliziter Gegenstand der moralpsychologischen Forschung. Das ist vor allem das Verdienst des amerikanischen Psychologen und Pädagogen Lawrence Kohlberg. Kohlberg hatte als einer der ersten versucht, moralische Fähigkeiten zu definieren, zu messen und zu fördern. Diese bahnbrechende Forschungsleistung hat in der Wissenschaft und auch in der Öffentlichkeit viel Beachtung gefunden. Sie wird aber m.E. noch immer zu wenig gewürdigt, auch wenn, wie meine eigenen Forschung zeigen, einige von Kohlbergs Grundannahmen wie vor allem die Annahme einer invariant aufsteigenden Moralentwicklung sich heute nicht mehr halten lassen. Ein Verlust moralischer Urteilsfähigkeit ist möglich und bei vielen Menschen zu beobachten. Solche Regressionen gehen, wie es scheint, vor allem auf mangelhafte Lerngelegenheiten zurück (siehe vor allem Kapitel 5).

Ich bin auf Kohlbergs Arbeiten erstmals Anfang der siebziger Jahre, am Ende meines Studiums aufmerksam geworden. Sie übten sofort eine große Faszination auf mich aus. Ich war beeindruckt von der Tatsache, dass Kohlberg sich wissenschaftlich mit einem Thema auseinander setzte, das nach herrschender Lehrmeinung kaum "wissenschaftsfähig" schien. Besonders beeindruckt hat mich jedoch seine Annahme, dass Moral in einem hohen Grad eine Fähigkeit sei, die durch die Schaffung günstiger Lernbedingungen gefördert werden könne. Er nannte seine Theorie daher "kognitiv-entwicklungsorientiert". Ich ahnte damals, dass diese Annahme, sollte sie der Überprüfung durch die wissenschaftliche Forschung standhalten, eine große Bedeutung für die Frage der Moralerziehung im weitesten Sinne haben würde. Moral konnten sich viele von uns bis dahin nur als eine Frage der Einstellung und Werthaltung vorstellen. Wenn Kohlberg recht hätte, und moralisch gutes Verhalten wirklich eher eine Frage der Fähigkeit und Ausbildung und weniger der richtigen Haltung oder genetischen Anlage wäre, müssten wir Abstand nehmen von manchen traditionellen Methoden der Moralerziehung, die auf Einstellungs- und Wertwandel oder Selektion abzielen. Es gäbe auch keinen Grund mehr, sich unter Verweis auf den demokratischen Bildungsauftrag unserer Schulen der Aufgabe der Moralerziehung zu entziehen. Gerhard Portele (1978) bezeichnete das Problem der (traditionellen) Moralerziehung in der Demokratie als das "Du-sollst-das-wol-

len"-Paradoxon: Sozialer Zwang und demokratische Mündigkeit schließen einander aus und somit eigentlich auch Demokratie und (Moral-) Erziehung. Eine kognitiv-entwicklungsorientierte moralische Bildung könnte dieses Paradoxon auflösen: Das Ziel, die Fähigkeit zu fördern, gemäß universeller, vom einzelnen akzeptierter Moralprinzipien zu denken und handeln, ist nicht nur mit den Prinzipien einer demokratischen Gesellschaft vereinbar, sondern eine zentrale Aufgabe demokratischer Erziehung und Bildung.

An Kohlbergs Theorie wird zu recht kritisiert, dass sie sich – trotz seiner imposanten, 20jährigen Längsschnittstudie – nur auf eine relativ schmale Datenbasis beziehen kann, nicht zuletzt aufgrund der Tatsache, dass ihm für die Gewinnung von Daten ausschließlich ein sehr aufwendiges Interviewverfahren zur Verfügung stand. Durchführung und Auswertung dieses Verfahrens sind zeit- und kostenaufwendig. Zudem basiert die Auswertung – trotz der starken Standardisierung durch ein umfangreiches Auswertungshandbuch – stark auf subjektiven Einschätzungen und auf Vorgaben, die den Verdacht nähren könnten, die damit gewonnenen Daten seien zugunsten der Theorie verzerrt (siehe Kapitel 3).

Ich hatte durch meine langjährige Mitarbeit in der internationalen Längsschnittstudie des Konstanzer Sonderforschungsbereich 23 *Bildungsforschung* von 1973 bis 1986 die Gelegenheit, die Sokratische Hypothese und Kohlbergs Forschungshinweise mit Hilfe eines großen, objektiv-experimentell gewonnenen Datensatzes auf den Prüfstand zu stellen. Ergänzt wurde diese Forschung durch zahlreiche kleinere Studien und Experimente, die ich zusammen mit Studierenden oder Kollegen durchführen konnte. In den letzten zehn Jahren kamen vermehrt Studien aus anderen Kulturen und pädagogische Interventionsstudien hinzu. Anfangs drehten sich die meisten Studien um die Messung der Moral. Ich habe dieser Frage der Adäquatheit und Validität der Messung höchste Priorität eingeräumt. Was nützen uns gute Theorien und Praktiken, wenn die Datengrundlage nicht hinreichend valide ist und es dem einzelnen überlassen bleibt, an ihre Richtigkeit zu glauben oder eben nicht. Ich habe daher mehrere Jahre darauf verwandt, das von uns gewählte Verfahren zur Messung der moralischen Urteilsfähigkeit auf seine Validität hin zu prüfen (Lind, 1978; 1984; 1985a; 1995, 2002; Lind & Wakenhut, 1985), und habe auch bei Übersetzungen in andere Sprachen immer wieder darauf gedrungen, dieser Frage mehr Aufmerksamkeit und Energie zu widmen, als dies in den Sozialwissenschaften normalerweise der Fall ist. Übersetzte Versionen des MUT können nur

dann als valide und äquivalent gelten, wenn sie anhand von theoretisch fundierten Kriterien experimentell validiert werden (Lind, 2002). Nachdem Kohlberg und seine Gruppe den Begriff der moralischen Urteils*fähigkeit* mehr und mehr aus dem Blick verloren haben und ihr Interesse sich auf die Erfassung des moralischen *Bewusstseins* konzentrierte, ist, so scheint es, der MUT das einzige Verfahren, das moralische Urteilsfähigkeit erfassen kann. Andere Verfahren dokumentieren primär moralische *Einstellungen*, wie zum Beispiel der *Defining-Issues-Test* (DIT), dessen P-Wert ausdrückt, wie stark Personen postkonventionelle moralische Prinzipien über andere Formen der moralischen Argumentation bevorzugen. Nachdem sein Autor, Jim Rest, eine zeitlang den Anspruch erhob, mit dem DIT auch moralische Urteilsfähigkeit zu messen (Rest, 1979; 1986b), zog er diesen Anspruch indirekt zurück. In mehreren Experimenten wurde gezeigt, dass der Messwert leicht nach "oben" geschwindelt werden konnte (was bei Indikatoren für eine Fähigkeit nicht möglich sein sollte): "We [. . .] have eliminated the faking study from our set of the validity criteria" (S. 115). Der Fähigkeitsindikator des MUT, der C-Wert, hat sich in solchen Experimenten als "schwindelfest" gezeigt (siehe Kapitel 6; Lind, 2002/im Druck).

Diese Forschungslage verlangt uns, wie ich an anderer Stelle (Lind, 1985d; 1989a; 1989b; 1993; 1996a; 1996b; 2000a; 2000b; 2000c; 2002/im Druck) ausführlicher zu zeigen versuche, ein neues Verständnis von Moral und Demokratie ab. Dieses neue Verständnis haben bereits Piaget und Kohlberg in ihren frühen Schriften vorgedacht, auch wenn sie es nicht so eindeutig haben belegen können, wie uns heute dies mit Hilfe des MUT möglich ist. Erstens: Die Lösung moralischer Probleme hängt tatsächlich in ähnlicher Weise von erworbenen Fähigkeiten (und nicht nur von der richtigen Einstellung oder Motivation) ab, wie die Lösung anderer Arten von Handlungsproblemen. Die bisherige (und leider immer noch weit verbreitete) Vorstellung, Vorsätze und Wollen reichten aus, um moralisch (oder demokratisch) zu handeln, lässt sich nicht länger halten.

Zweitens: Moralische Fähigkeiten müssen – wie moralische Einstellungen – als eine *Eigenschaft* oder ein *Aspekt* des moralischen (Urteils-)Verhaltens begriffen werden und nicht als separate *Komponenten* oder Verhaltensdomänen. So wie man zum Beispiel an einem Ball die Farbe, Gestalt und Elastizität klar unterscheiden, aber nicht abtrennen und separat voneinander untersuchen kann, lassen sich auch die kognitiven und affektiven Aspekte des moralischen Ver-

haltens deutlich unterscheiden (und logisch unabhängig voneinander messen), aber nicht in verschiedene Verhaltensklassen trennen. Weil diese Einsicht immer noch zu wenig Beachtung findet, habe ich ihr ein eigenes Kapitel gewidmet.

Drittens: Dieses neue Verständnis hat für die Praxis der Erziehung und auch für die Gestaltung unseres Bildungswesens Konsequenzen. Es eröffnet uns vor allem neue Dimensionen der Pädagogik. Wir stehen als Lehrer und Eltern nicht mehr länger vor der Wahl zwischen einer totalen Abstinenz von Moralfragen auf der einen Seite und einer Werteindoktrination auf der anderen, wie das lange Zeit der Fall war, sondern sehen uns vor der – auch demokratisch legitimen – moralischen Aufgabe, die Fähigkeit der Heranwachsenden zu fördern, ihre eigenen moralischen Prinzipien im Handeln konsistent und differenziert anzuwenden. Welche pädagogische Energien diese Einsicht freisetzen kann, konnte ich in den achtziger Jahren erleben, als ich – mit Hilfe von Lawrence Kohlberg, Ann Higgins und Fritz Oser – in Nordrhein-Westfalen das schulpraktische Erprobungsprojekt *Demokratie und Erziehung in der Schule* (DES) initiiert und zusammen mit Fritz Oser, Wolfgang Althof und Sibylle Reinhardt wissenschaftlich begleitet habe (Lind & Raschert, 1987; Lind, 1990; Lind & Althof, 1992, sowie hier das Kapitel 8). Inzwischen habe ich vielfältig Gelegenheit gehabt, die schulpraktische Fruchtbarkeit der Fähigkeitstheorie der Moral zu studieren. Eine Schlüsselfunktion nimmt hier die von Blatt und Kohlberg vorgeschlagene Lehrmethode der *Moralisches Dilemma-Diskussion* ein, die ich in Kooperation mit Schulen und Studierenden weiter entwickelt habe (Lind, 1999c). Die Ergebnisse dieser Arbeiten dienen gegenwärtig der Konzeption von neuen Fortbildungsangeboten für Lehrer auf den Gebieten der Ethik, Demokratieerziehung[2] und Mentorik[3].

In der ersten Ausgabe von 1993, unter dem Titel *Moral und Bildung* im Asanger-Verlag erschienen, hat dieses Buch in Fachkreisen eine recht positive Resonanz gefunden. Es war bald vergriffen. Der Göttinger Moralpsychologe und Pädagoge Günter Schreiner bezeichnete die in dem Buch vorgelegten Forschungsergebnisse als „einen Meilenstein im moralpädagogischen Diskurs". Für Roland Wakenhut, Sozial- und Organisationspsychologe an der Katholischen Universität Eichstätt, liefert die hier zusammengefasste Arbeit "wichtige

2 Siehe http://www.uni-konstanz.de/ag-moral/

3 Siehe http://www.uni-konstanz.de/ag-moral/

Impulse [. . .] für die moralpsychologische Forschung, die nach dem Tode von Lawrence Kohlberg und der Auflösung des Harvard Center for Moral Education ihren Schwerpunkt nach Europa verlagert hat." Über die hier vorgeschlagene Methodologie zur „Messung" von kognitiven Strukturen im Bereich des moralischen Denkens und Handelns, dem „Experimental Questionnaire" (EQ) schreibt Peter H. Rossi, Soziologie-Professor an der University of Massachusetts, Amherst: „Ich bin erfreut, Ihre EQ-Methode kennen gelernt zu haben. Ihre Methode liegt viel mehr Theorie zugrunde als unserer Idee der faktoriellen Umfragemethode, und mit Ihrer Zustimmung werde ich einige Ihrer Ideen ausborgen." Dr. John M. Bynner (Professor an der Open University, Milton Keynes, UK) bemerkt dazu: "Ich denke, Ihre Argumente über Einstellungsmessung [ausführlicher beschrieben in Lind, 1985c; G.L.] sind äußerst überzeugend. Sie befinden sich in Übereinstimmung mit dem allgemeinen Anliegen, das Lamiell für 'ideothetisches Messen' macht." Seit der ersten Auflage sind weitere wichtige Untersuchungen durchgeführt worden, die sich auf die hier berichtete Forschung stützen und sich intensiv mit der in diesem Buch vorgestellten Bildungstheorie der Moralentwicklung auseinander setzen. Ich kann hier nur einige nennen: Untersuchungen zur Überprüfung der Validität des MUT (Heidbrink, 1995; 1996; Lind, 1996c; 2002; Rest et al., 1997; Biggs et al., 1999), zur Simulierbarkeit der moralischen Urteilsfähigkeit (Wasel, 1994; siehe hier *Kapitel 6*), zum Zusammenhang von Moral und politischem Aktivismus in internationaler Perspektive (Gross, 1997), zur moralischen Urteilsfähigkeit von medizinischem Personal (DuBois, 1997; Lind, 2000a), zur Abhängigkeit der moralischen Urteilsfähigkeit von Bildungsgelegenheiten bei Berufsschülern (Beck, 1995; siehe hier *Kapitel 5* und *11*), zum Zusammenhang von moralischer Urteilsfähigkeit und Gelegenheiten zur Verantwortungsübernahme (Lind, 1996a; Comunian & Gielen, 1999), zur Wirkung der Dilemma-Diskussion bei jugendlichen Strafgefangenen (Lind, 1994) und zur "pluralistischen Ignoranz" von Lehrern in bezug auf die moralischen Werthaltungen von Schülern (Lind, 1995). Außerdem wurden die hieraus gewonnenen Erkenntnisse für die Weiterentwicklung von pädagogischen Methoden zur Förderung von moralischen und demokratischen Kompetenzen herangezogen (Lind, 1999c). Der für unsere Forschung entwickelte Moralisches Urteil-Test (MUT)[4] liegt inzwischen in mehreren Sprachen vor. Die meisten Übersetzungen wurden inzwischen stren-

4 Die englische Bezeichnung ist *Moral Judgment Test* (MJT).

gen *Validierungstests* unterzogen (englisch, französisch, hebräisch, holländisch, italienisch, portugiesisch, spanisch, baskisch, türkisch, mazedonisch, tschechisch, finnisch). Weitere Versionen sind in Arbeit (spanisch, chinesisch, koreanisch). [Inzwischen 39 Sprachen; 2019]

Trotz der umfangreichen Datenlage (es gibt derzeit mehr als 40.000 individuelle Erhebungen mit dem MUT) und einer intensiven – und zum Teil sehr kritischen – Auseinandersetzung mit diesem Verfahren gibt es bisher keinen Befund, der die *Zwei-Aspekte-Theorie* und damit die Existenz eines Fähigkeitsaspekts der Moral grundsätzlich in Frage stellt. Im Gegenteil, diese Theorie und der auf dieser Basis entwickelte *Moralisches Urteil-Test* werden auf immer neue Weise bestätigt (Biggs et al. 1999). Ich hoffe, dass die in diesem Buch vorgetragene *Bildungstheorie* der Moral, die sich weitgehend auf Untersuchungen mit dem MUT stützt, dazu beiträgt, die Bedeutung des Fähigkeitsaspekts der Moral für den pädagogischen Alltag und die Bildungspolitik zu erkennen und über die notwendigen Konsequenzen nachzudenken.

Georg Lind Konstanz, Januar 2002

E-mail-Adresse: *Georg.Lind@uni-konstanz.de*

1 Zu einer Bildungstheorie der Moralentwicklung

Ist die Entwicklung der Moral in unseren Genen angelegt oder wird sie von der Gesellschaft mittels Druck und Vergünstigen erzwungen? Muss die Erziehung in der Familie und in den Schulen daher sich auf das Abwarten verlegen oder, folgt man der zweiten Sichtweise, auf das Druckmachen und Sanktionieren? Der zentrale Befund der hier vorgelegten Studien ist, dass keine der beiden Erklärungen zutrifft und daher auch beide Empfehlungen für die pädagogische Praxis unbegründet sind. Moralisches Verhalten und moralische Entwicklung ist vielmehr, so legen alle Untersuchungen nahe, das Ergebnis von Bildungsprozessen (oder dem Fehlen derselben).

Eng verbunden mit der empirisch nahegelegten "Bildungstheorie" der Moralentwicklung ist Zwei-Aspekte-Theorie des moralischen Verhaltens. Nach dieser reicht es nicht aus, moralisches Verhalten allein unter dem Gesichtspunkt moralischer Affekte (Einstellungen, Gefühle, Werthaltungen, Motive etc.) zu sehen. Hinzu kommen muss noch das Studium von kognitiven Aspekten oder Fähigkeiten (Lind, 1985a; 1985b; 1986a; 2002/1995). Würde es allein nach unserem Wollen und Fühlen, wären wir fast alle moralisch perfekt. Kein Mensch (oder zumindest kaum einer) möchte *unmoralisch* sein oder *ungerecht* handeln. Die meisten, wenn nicht gar alle Menschen reagieren empfindsam auf Ungerechtigkeit und Unmoral. Aber vom moralischen und Wollen ist of ein weiter mit vielerlei Schwierigkeiten bestückter Weg bis zum moralischen Handeln (siehe Levy-Suhl, 1912; Piaget, 1973a). Um unsere moralischen Ideale mit dem Verhalten in Übereinstimmung zu bringen, benötigen wir daher bestimmte Fähigkeiten, vor allem moralische Urteilsfähigkeit (Kohlberg, 1964). Ich halte dies für die wichtigste Entdeckung des 20. Jahrhunderts, zumal sie sich gegen mächtige Widerstände hat durchsetzen und behaupten müssen. Diese haben ihren Ursprung in dem theologischen Postulat, dass allein der gute Wille schon selig mache, und in der psychologischen Trennung von Affekt und Kognition in zwei verschiedene Verhaltensdomänen (Bloom et al., 1956).

Wichtigstes Ergebnis unserer Untersuchung ist also, dass beim moralischen Verhalten zwei gleichermaßen wichtige, untrennbare Aspekte zu unterscheiden sind, der affektive und der kognitive Aspekt. Der affektive Aspekt (unsere moralischen Ideale, Einstellungen und Motive) ist ein ebenso notwendiger und unverzichtbarer Bestandteil moralischen Verhaltens wie der kognitive, aber anders als letzterer unterscheiden sich hierin die Menschen nur wenig, so dass er uns

auch wenig darüber sagt, warum Menschen sich in moralisch relevanten Situationen so unterschiedlich verhalten. Der kognitive Aspekt (das sind vor allem die Fähigkeiten, die für die Lösung moralischer Konflikte notwendig sind) scheint derjenige zu sein, der bei verschiedenen Menschen oft ganz unterschiedlich ausgeprägt ist und zu ganz unterschiedlichem Verhalten führt, und für den Bildungsprozess von großer Bedeutung sind.

Diese Thesen stellen den Kern meiner *Bildungstheorie* der Moralentwicklung dar. Sie sind nicht neu. Sie gehen bis auf die Griechischen Philosophen des Altertums zurück. Sokrates hat sie bereits vorgedacht in seinem berühmten Dialog mit *Menon* über die Frage, ob den Tugend lehrbar sei. Wenn Tugend eine Fähigkeit ist, dann sei sie auch lehrbar, argumentiert Sokrates, und wenn sie eine Fähigkeit sei, dann sei sie auch lehrbar. Anfang dieses Jahrhunderts hat der Berliner Gerichtspsychiater Levy-Suhl (1912) die erstaunliche Feststellung gemacht, dass jugendliche Straftäter diesselben moralischen Werte haben wie nichtstraffällige Jugendliche.

Aber die kognitive Bildungstheorie der Moral wurde bislang noch kaum empirisch-experimentell so belegt, dass sie gegenüber konkurrierenden und weitaus populäreren Theorien der Moralentwicklung bestehen kann, zum Beispiel neben der *Sozialisationstheorie*, wonach die Entwicklung der Moral das Ergebnis eines sozialen Anpassungsdrucks ist, oder der *Reifungs- oder Emergenztheorie*, nach der sich die moralische Urteilsfähigkeit mit dem Alter praktisch ohne äußere Einflüsse, quasi von alleine, als Ergebnis eines genetisch gesteuerten Reifungsprozesses entfaltet.

Nach der *Sozialisationstheorie* und der *Sozialen Lerntheorie* geschieht Moralentwicklung durch das Erlernen von Normen und Werten derjenigen Gruppe, der man angehört. Einer der "Väter" der Sozialisationstheorie war der französische Soziologe und Erziehungstheoretiker Emile Durkheim. Für ihn lebt der Mensch von Natur aus in einem instabilen Zustand, bedrängt von seinen Trieben, die es zu begrenzen und im Sinne der Gesellschaft umzulenken gilt. Dies bedarf, so Durkheim (1984/1902), des Einsichtigmachens sozialer Normen und Werte durch Zwang und Autorität. Ziel der Moralentwicklung ist nach dieser Theorie der Erwerb moralischer Einstellungen, an dem, in neueren theoretischen Entwürfen, kognitive Prozesse beteiligt sein können. Nicht thematisiert – und von einigen Vertretern dieser Theorie sogar als überflüssig angesehen – werden die kognitiven Fähigkeiten, die notwendig sind, um moralische

Werte oder Prinzipien in situationsgemäßes Verhalten zu übersetzen und die moralischen Konflikte zu lösen, die dabei auftreten können.

Nach der *Reifungstheorie* dagegen ist der Mensch von Geburt gut oder bekommt die Anlagen fürs Gut-sein mit; die Gesellschaft ist hingegen in diesem Denksystem die Wurzel des Bösen und stellt mit ihren Normen und Zwängen eine Behinderung der Moralentwicklung dar. Das Spektrum der Reifungstheoretiker der Moral ist vielfältig, dennoch gibt es einige gemeinsame Überzeugungen. “Alles ist gut, wie es aus den Händen des Schöpfers kommt; alles entartet unter den Händen des Menschen”, leitet einer der großen Vertreter dieser Denkschule, J.-J. Rousseau (1971/1762, S. 9), sein epochales Werk *Emile oder Über die Erziehung* ein. Er insistiert darauf, “dass der Mensch von Natur aus gut ist ..., [aber] die Gesellschaft den Menschen verdirbt und widernatürlich macht” (S. 241). Sigmund Freud und seine psychoanalytische Schule wiederholen die Rousseausche These in abgewandelter Form: Die herrschenden gesellschaftlichen Normen führten zur Unterdrückung menschlicher Triebe; diese Triebunterdrückung sei mitverantwortlich für unmoralisches Verhalten. Die Psychoanalyse vertraut daher darauf, dass, wie Arno Plack (1969) resümiert, “die menschliche Natur, wenn sie in Ruhe gelassen wird und nicht von klein auf verbogen, aus sich heraus ein Verhalten entwickelt, das dem Individuum wie der Gesellschaft am besten entspricht” (S. 345). Infolgedessen kann auch institutionalisierte Bildung keine positiven Effekte für die individuelle Moralentwicklung haben und kann es der Pädagogik, wie es von Braunmühl (1975) auf den Punkt brachte, nur um ihre eigene Abschaffung gehen.

Die *Kognitive Entwicklungstheorie* der Moralentwicklung von J. Piaget und L. Kohlberg nimmt unter den bisherigen Ansätzen eine Sonderstellung ein. Sie weist Elemente beider Theorien auf. “Die Entwicklung kognitiver Strukturen ist”, so Kohlberg (1984), “eher das Resultat von Vorgängen der Interaktion zwischen der Struktur des Organismus und der Struktur der Umwelt als das direkte Resultat der Reifung” (S. 8). Aber Richtung und Abfolge der Entwicklung sind, nach dieser Theorie, (genetisch?) festgelegt. Kohlberg nimmt unter Verweis auf Piaget an, dass die Stufen der Moralentwicklung „eine invariante Abfolge, Ordnung bzw. Aufeinanderfolge in der individuellen Entwicklung” bilden (S. 14). Während kulturelle Faktoren diese Entwicklung beschleunigen, verlangsamen oder anhalten können, können sie nicht ihre Abfolge ändern. Auch die Prognose der kognitiven Entwicklungstheorie ist optimistisch. Statt 'der Mensch *ist* gut', nimmt diese Theorie an, 'der Mensch *wird* gut im Verlaufe

seiner kognitiv-moralischen Entwicklung'. Zumindest wird er nicht schlechter. Aufgrund der inneren "Logik" der Moralentwicklung können soziale Einflüsse die Moralentwicklung nur beschleunigen oder verlangsamen, aber sie können sie nicht umkehren. Diese Annahme der *invarianten Sequenz* oder Nicht-Regression bildet einen zentralen Pfeiler der Theorie und beherrscht auch, im Gegensatz zu manchen ihrer Aussagen über die Rolle der sozialen Umwelt, die Forschung von Kohlberg und seinen Kollegen. Ihre Theorie ist also im Kern eine Reifungstheorie.

Zahlreiche Studien der letzten Jahre haben einerseits Kohlbergs (1969) scharfe Kritik sozialisationstheoretischer Ansätze wie auch naiv-reifungstheoretischer Ansätze untermauert und scheinbar auch Anhaltspunkte für eine Bestätigung der kognitiven Entwicklungstheorie gesammelt. Andererseits haben sie aber auch Hinweise dafür gefunden, dass einige dieser Bestätigungen auf unzulänglichen und teilweise voreingenommenen Forschungsmethoden bzw. Versuchsanordnungen beruhen (Kapitel 3). Hauptaufgabe der hier vorgelegten Untersuchung ist es daher, diese Mängel – auf der Basis der kognitiven Entwicklungstheorie – zu beheben. Das Ergebnis zeigt, dass auch Kohlbergs Theorie die moralische Entwicklung nicht adäquat beschreibt. Vielmehr muss allererst eine solche Theorie entwickelt werden, die wir oben als "Bildungstheorie" der Moralentwicklung bezeichnet haben. Dieses theoretische Fundament zu entwickeln war das Hauptanliegen meiner Untersuchung.

1.1 Praktische Bedeutung

Die vorliegende Untersuchung ist primär grundlagenorientiert; sie will eine Entscheidung zwischen drei alternativen Erklärungsansätzen herbeiführen. Das schließt nicht aus, dass diese Erklärungsansätze oder Thesen praktische Bedeutung haben, vor allem für pädagogische, bildungspolitische und sozialtherapeutische Maßnahmen. Im Gegenteil, die Frage nach der Rolle, die Bildungsprozesse für die Moralentwicklung haben, scheint von höchster Relevanz für die Ziele und Methoden jeder Moral- oder Sozialerziehung, ja für Erziehung überhaupt. Eine Bestätigung der Emergenzthese würde bedeuten, dass Kinder am besten zu moralisch reiferen Wesen werden, wenn man sie sich selbst überläßt und sie möglichst wenig erzieht. Tatsächlich haben viele "Reform-Pädagogen" und "Anti-Pädagogen" wie Alexander S. Neill (1969) in seiner berühmten

Summerhill-Schule ihre Methoden durch Hinweis auf die "natürliche" Entwicklung moralischer Fähigkeiten begründet. Die Bildungspolitik könnte sich, sofern diese Sicht empirisch belegbar wäre, auf die Bereitstellung von Freiräumen – im konkreten wie im übertragenen Sinne des Wortes – beschränken. Wenn Kohlbergs kognitive Entwicklungstheorie zutrifft, wäre das schlimmste, was bei fehlender Förderung passieren kann, eine Verlangsamung oder Stagnation der Moralentwicklung. Was aber, wenn diese Theorie nicht zutrifft, wenn sich die Moralentwicklung bei fehlender Förderung gerade umkehrt? Wir werden zeigen, dass dies entgegen der Kohlberg-Theorie tatsächlich der Fall ist. Wenn die pädagogische Stimulation durch allgemeinbildende Institutionen (in der Regel die Schule) wegfällt, nimmt die moralische Urteilsfähigkeit der Menschen deutlich ab (Kapitel 5).

Eine Bestätigung der Sozialisationstheorie würde das genaue Gegenteil bedeuten. Es würde heißen, dass die Erziehungsinstanzen Familie, Schule – ihr besonderes Augenmerk auf die Vermittlung von sozialen Normen und auf die Kontrolle ihrer Einhaltung legen müßten, um aus dem zunächst amoralischen Kind, ein moralisch kompetentes Mitglied der Gesellschaft zu machen. Der Heranwachsende brauchte die moralischen Normen nur insoweit zu verstehen, wie ihm dies das Einhalten derselben erleichtert (Durkheim 1976/1924; Brezinka 1988). Dagegen sagt die Bildungstheorie der Moralentwicklung voraus, dass es mehr bedarf als das Verstehen sozialer Normen und der Anpassung hieran, dass für die Moralentwicklung vielmehr, wie Piaget und Kohlberg annahmen, entwickelte kognitive Fähigkeiten wesentlich sind. Des weiteren sagt diese Theorie voraus, dass die Forderungen, die sich aus der Sozialisationstheorie ableiten lassen, weitgehend ins Leere laufen, da fast alle Kinder, selbst in sehr jungen Jahren, eine hohe Sensibilität für grundlegende moralische Prinzipien haben und diese ihnen nicht erst vermittelt werden müssen.[5] An was es Kindern – und auch vielen Jugendlichen und Erwachsenen – jedoch mangelt, ist die Fähigkeit, moralische Prinzipien in konkreten Situationen anzuwenden und moralische Konflikte zwischen widerstreitenden Prinzipien zu lösen.

Die Bildungstheorie impliziert daher einen dritten Weg der Erziehung. Sie besagt, dass Moral ein Aufgabenfeld von ähnlicher Struktur und mit ähnlichen Erfordernissen darstellt wie jedes andere Aufgabenfeld, dem wir uns im Leben

5 Neuere Untersuchungen zur Moralentwicklung im Kindheitsalter werden in Kapitel 2 referiert; s. auch Lind 1989a.

stellen müssen, und dass daher ganz ähnliche didaktische Bemühungen notwendig sind, um moralische Urteilsfähigkeit optimal zu fördern. Moralische Probleme können natürlich nicht mit denselben Methoden und Kompetenzen gelöst werden wie mathematische oder physikalische Probleme. Aber auch hier gilt: Es müssen a) Kenntnisse in lebenspraktischen Zusammenhängen ("moralisches Wissen") vermittelt und b) Möglichkeiten für die Erprobung dieser Kenntnisse an praktischen Aufgaben geschaffen werden. Wirksame Moralerziehung erfordert eine spezielle Didaktik, gut ausgebildete Pädagogen und entsprechende Lehrpläne. Erste Ansätze hierfür liegen vor.[6] Freilich muß hier auch an Kohlbergs Warnung vor dem "psychologischen Fehlschluß" erinnert werden. Eine psychologische Theorie mag noch so wahr, empirisch getestet und vollkommen sein, sie kann keinen umfassenden Handlungsplan für alle Akteure in einem Anwendungsfeld anbieten. Handeln und Entscheiden in lebensweltlichen Bezügen ist mehr als die Anwendung einer Theorie, es ist die Koordination einer Vielfalt von Wissensformen und Theorien unterschiedlicher Herkunft in einem bestimmten Augenblick in einer bestimmten Situation durch eine bestimmte Person. Es wäre falsch, die umfassende Handlungskompetenz, die Menschen, selbst Kinder, in solche Entscheidungen einbringen, durch eine einzige, monolithische Theorie über einen relativ kleinen Aspekt der Lebenswelt ersetzen zu wollen. Der Moralpsychologe ist immer auf die Zusammenarbeit mit den Betroffenen "im Feld", mit den Lehrern, Eltern und Kindern angewiesen.

Nach der Bildungstheorie der Moralentwicklung sind demokratisch-diskursive Prozesse nicht nur eine zusätzliche Bedingung für die Entwicklung der moralischen Urteilsfähigkeit, sondern deren unverzichtbare Voraussetzung (s. Kapitel 8 und 9). Die hier vorgelegten Ergebnisse untermauern die zentrale Bedeutung, die der moralischen Urteilsfähigkeit im individuellen Entwicklungsprozeß zukommt. Die Tatsache, dass ausschließlich "kognitive" Interventionen, die Denken und Diskussionen über moralische Konflikte zum Inhalt haben, einen signifikanten und nachhaltigen Effekt von beachtlicher Größe bewirken, spricht eher für die Gültigkeit der kognitiven Entwicklungstheorie als für Reifungs- und Sozialisationstheorien zur Erklärung der Veränderungen im Moralverhalten von Jugendlichen und jungen Erwachsenen. Der Befund, dass Bildungsinstitutionen, aber nicht die berufliche Umwelt, aufs Ganze gesehen einen stark

6 Power et al. 1989; Lind & Raschert 1987; Oser & Althof 1989.

fördernden Einfluß auf die Moralentwicklung haben, spricht für die Bedeutung von fähigkeitsfördernden Maßnahmen für die Moralentwicklung, und gegen Reifungs- oder Anpassungsthesen und deren theoretische und praktische Implikationen.[7] Entwicklungsfördernde moralische Diskurse können prinzipiell überall stattfinden, also auch in der Familie, im Freundeskreis und am Arbeitsplatz. Sie finden aber offenbar die günstigsten Voraussetzungen in allgemeinbildenden Schulen. Wenn die Moralentwicklung nicht durch Bildungsprozesse ausreichend stimuliert wird, stagniert die Entwicklung der moralischen Urteilsfähigkeit oder bildet sich sogar zurück. Jugendliche, die das Schulsystem früh verlassen und in die Berufswelt eintreten, in der weniger Raum für dialogische Prozesse ist, weisen eine deutliche *Abnahme* der moralischen Urteilsfähigkeit auf (Kapitel 5). Kohlberg und viele von ihm inspirierte Pädagogen haben gezeigt, dass die Schulen ihre Aufgabe, die Förderung moralisch-demokratischer Kompetenzen, besser als bisher erfüllen könnten, wenn sie sich gezielt und methodisch darum bemühen würden (Kapitel 8 und 9). Dass die Förderung solcher Kompetenzen heute, angesichts einer komplexen Umwelt mit ihren vielfältigen Problemen dringlich geboten ist, wird von Philosophen wie Jonas (1979), Böhler (1992) und Mittelstraß (1992) mit überzeugenden Argumenten untermauert.

Warum ist dann die Bedeutung von Bildungserfahrungen in der Moralpsychologie bislang fast gänzlich übersehen worden? Neben der Tatsache, dass weder in der Sozialisation noch in der Reifungstheorie Platz für die “Variable” Bildung ist, scheint mir der Grund darin zu liegen, dass Bildungserfahrungen heute so allgemein und für viele so selbstverständlich geworden sind, dass uns ihre Effekte kaum mehr bewusst werden. In vielen Gesellschaften besteht eine langjährige Schulpflicht, so dass die Effekte von Bildungserfahrungen heute nur noch schwer von rein altersbedingten (Reifung-) Prozessen unterschieden werden können. Was in Wirklichkeit bildungsbedingte Entwicklungen sind, wird häufig einer altersabhängigen Reifung zugeschrieben.

7 Diese These steht nicht im Widerspruch zu den Untersuchungen von Lempert (1988) und seinen Mitarbeitern, die bei den von ihnen interviewten Industriearbeitern Fortschritte in der Moralentwicklung gefunden haben. Auch am Arbeitsplatz können Bildungsgelegenheiten vorhanden sein.

1.2 Definition und Operationalisierung

Nachdem deutlich geworden ist, dass eine Klärung der empirischen Gültigkeit der drei hier behandelten Theorien der Moralentwicklung nicht bloß von akademischem Interesse ist, sondern auch von höchst praktischem, stellt sich die Frage, wie sich eine empirische bzw. experimentelle Bewertung herbeiführen läßt. Eine Voraussetzung dafür ist offenkundig, dass diese Theorien einen tatsächlichen, von jedem beobachtbaren Sachverhalt betreffen, so dass sich anhand intersubjektiver Prüfung ein Konsens über Richtigkeit oder Falschheit erreichen läßt. Die Forderung nach intersubjektiver Überprüfbarkeit klingt einfacher als sie ist. James B. Watson (1970/1924), ein Begründer der modernen Psychologie, sah sie dadurch erfüllt, dass sich jede Theorie "auf jene Dinge beschränkt, die beobachtet werden können, und dass nur bezüglich dieser Dinge Gesetze formuliert werden." Beobachtet werden kann, so Watson, "Verhalten – was der Organismus *tut oder sagt*" (S. 6). Nur wenn Theorien sich an diesen – trivial erscheinenden – Grundsatz halten, können sie intersubjektiv nachvollzogen und anhand empirischer Daten beurteilt werden. Leider wurde dieser methodologische Grundsatz von seinen frühen Vertretern, den *Behavioristen* ad absurdum geführt, da sie – weit über das Ziel hinausschießend – *a priori* festlegten, dass all jene psychologischen Phänomene nicht beobachtbar (und daher auch nicht existent) seien, die sich nicht mit den bislang bekannten – vor allem physikalischen – Meßverfahren messen ließen. Für viele Psychologen bedeutete dies eine Blockierung jeder Forschung auf zentralen Gebieten der Psychologie. Jene, die nicht ganz so weit gehen wollten wie Watson, sagen, dass psychologische Dispositionen nicht *direkt* beobachtbar, aber anhand von Indikatoren und bestimmten Zusatzannahmen erschließbar seien. Man spricht hier von "hypothetischen" oder "latenten Konstrukten". Aber diese Position steht wegen der ihr zwingend zugrundeliegenden psychologischen Prämissen, auf höchst unsicheren, wissenschaftlichen Grundlagen.[8]

Die methodologische Position unserer Untersuchung ist, dass psychische Dispositionen wie moralische Urteilsfähigkeit und moralische Einstellungen als manifeste Eigenschaften des Verhaltens begriffen und gemessen werden können (s. Kapitel 2 und 4). Nach dieser, als *methodologischer* Behaviorismus bezeichneten Position, muß keineswegs angenommen werden, dass moralische

8 Für eine ausführlichere Darlegung meiner Kritik siehe Lind, 1985a, und Kap. 3.

Prinzipien bloß latente Konstrukte sind, die nur unterstellt ("hypothetische Konstrukte") oder intuitiv erfaßt, aber nicht empirisch-experimentell nachgewiesen werden können. Es ist auch nicht notwendig, sie bloß als eine willkürliche Beschreibung bestimmter Verhaltensakte und daher als weitgehend selbstreferentiell aufzufassen (im Sinne der Aussage: "Die Person *A* ißt, also nenne ich sie hungrig").
Im Zentrum der vorliegenden Untersuchung steht der Begriff der "moralischen Urteilsfähigkeit" und seine Operationalisierung als Struktur des Urteilsverhaltens in moralisch relevanten Situationen. Nach Kohlberg (1964, S. 425) ist *moralische Urteilsfähigkeit*

> *"das Vermögen, moralische Entscheidungen und Urteile zu treffen, die moralisch sind (also auf moralischen Prinzipien gründen), und in Übereinstimmung mit diesen Urteilen zu handeln."*

Diese Definition steht in Kontrast zu zwei traditionsreichen Definitionen, von denen zu zeigen ist, dass sie weniger adäquat sind. Die eine definiert moralisches Verhalten als einen Akt, der – äußerlich – mit sozialen Konventionen und Normen übereinstimmt. Demnach wäre derjenige moralisch zu nennen, der eine als "moralisch" definierte Handlung durchführt oder eine "unmoralische" unterläßt. Diese Definition von Moral ist vor allem aus zwei Gründe unangemessen:

a) Eine isoliert betrachtete Handlung läßt fast nie einen eindeutigen Rückschluß auf die Moralität des Handelnden zu. Eine Handlung kann von anderen als "moralisch" betrachtet werden, ohne dass die Person moralisch sein wollte, oder sie kann allein dem Zweck dienen, von anderen als moralisch eingeschätzt zu werden, damit ein dahinter liegender – vielleicht unmoralischer – Zweck erfüllt wird. Jemand verschenkt zum Beispiel seinen Mantel an einen Bettler, weil er ihm nicht mehr gefällt (= nicht-moralisches Motiv), und sorgt dafür, dass die Presse über das Geschenk berichtet (= übergeordneter Zweck).
b) Die Definition von "moralisch" als die Übereinstimmung einer Handlung mit den sozialen Normen hat so viele Bedeutungen wie es sozio-kulturelle Normensysteme gibt. Ein und dieselbe Handlung müßte einmal als moralisch, ein andermal als unmoralisch beurteilt werden, je nachdem welche soziale Norm gerade als Kriterium herangezogen wird. Zum Beispiel das folgende Verhalten: Ein Schüler läßt einen anderen während einer Klas-

senarbeit abschreiben. Aus der Sicht des Lehrers ist das ein Verstoß gegen die Regeln der Schule, aus der Sicht des Schülers aber eine Erfüllung des Hilfegebots unter Freunden. Der Definition von Moral als einfache Normkonformität fehlen also zwei wichtige Bestimmungsstücke, nämlich dass der Akt auf moralischen Prinzipien des Individuums beruht und dass diese Prinzipien universalisierbar sind.

Die zweite, linguistische Definition von Moral als *Begründungen von Verhalten mit moralischen Begriffen* (= moralische Aussagen) unterstellt einen Bezug zum Handeln, kann ihn aber nicht belegen. So läßt sich zum Beispiel bei ihr nicht entscheiden, ob Aussagen mit moralischem Inhalt eine bestimmende Funktion für das Verhalten einer Person haben, oder ob sie nur Ausdruck einer intellektuellen Übung oder nachträglicher Rechtfertigung ("Rationalisierung") von Meinungen und Entscheidungen sind.[9] Moralische Aussagen bieten zwar einen Zugang zu den moralischen Dispositionen eines Menschen, aber für sich betrachtet ermöglichen sie keine gültige Messung der moralischen Urteilsfähigkeit. Erst der psycho-soziale Kontext, in welchem moralische Argumente benutzt werden, bestimmt ihre psychologische Realität. Ein Schauspieler, beispielsweise, der moralische Sätze rezitiert, wird nicht deshalb schon urteilsfähig genannt werden. Wenn sich aber eine Person in einer wichtigen Sache durch ein moralisches Argument überzeugen läßt, obwohl es ihrer eigenen Meinung entgegengesetzt ist, sind wir bereit zu unterstellen, dass bei ihr moralische Motive vorhanden sind und dass sie fähig ist, diese Motive auf die Steuerung ihres eigenen Tuns anzuwenden.

Die Probleme beider Ansätze werden in der Kohlbergschen Definition von moralischer Urteilsfähigkeit vermieden. Zum einen wird ein Handeln nur insoweit "moralisch" genannt, wie es nachweislich durch moralische Prinzipien determiniert ist. Zum anderen werden moralische Prinzipien (bzw. ihre verbale Repräs-

9 Kohlberg (1984) versuchte, mit Hilfe des Konstrukts des *Verantwortungsurteils* zwischen moralischen Aussagen und Handeln eine empirische Verbindung herzustellen. In einer exzellenten Untersuchung metaethischer Positionen der Psychologie hat Thomas Wren (1990) diesen Versuch verworfen, weil er einen infiniten Regreß der Urteile impliziert (wie kann die Brücke zwischen Verantwortungs*urteilen* und Handeln geschlagen werden?). Wren postuliert dagegen eine intrinsische Verbindung zwischen jedem *ernsthaft* geäußerten moralischen Argument und dem Handeln einer Person. Aber auch diese Hypothese der Ernsthaftigkeit kann im Einzelfall nicht geprüft werden und läßt somit das Problem ungelöst.

sentation in Form von Argumenten) nur soweit als Ausdruck dieser Fähigkeit angesehen, wie sie sich verhaltenswirksam erweisen. Moralische Urteilsfähigkeit läßt sich weder allein durch das Verhalten, noch allein durch das ihm zugrundeliegende Denken definieren, sondern immer nur durch beides. Nach der Kohlberg-Definition, auf die auch wir uns im großen und ganzen stützen, macht streng genommen die Frage keinen Sinn, wie Urteilen und Handeln *empirisch* zusammenhängen. Sie lassen sich unabhängig voneinander gar nicht definieren und messen.[10]

Dies zeigt, wie eng inhaltliche und methodologische Fragen miteinander verbunden sind. Ohne die Möglichkeit, das Konstrukt "moralische Urteilsfähigkeit" so zu messen, wie es theoretisch konzipiert ist, kann die Frage nicht empirisch entschieden werden, welche der Thesen zur Moralentwicklung zutreffend ist. Lange Zeit wurde Moral – in Übereinstimmung mit der Sozialisationsthese – nur als Einhaltung oder Abweichung von sozialen Normen gemessen (moralisch kompetent ist, wer die Normen kennt und einhält).[11] Piaget und Kohlberg aber haben als erste erkannt, dass der Kern der Moralität eine über das Erlernen sozial anerkannter Normen hinausgehende *kognitive* Fähigkeit darstellt und diese Erkenntnis in geeignete Meßinstrumente zu übersetzen versucht.

Die größte Aufmerksamkeit hat dabei Kohlbergs *Moral Judgment Interview* (MJI) erfahren, das in Kapitel 3 näher dargestellt wird. Wie dort gezeigt wird, erfüllt das MJI in seiner neuesten Revision voll die Kriterien der sogenannten klassischen Testtheorie, was ein wichtiges Ziel dieser Revision war. Aber die-

10 Davon zu unterscheiden ist die im Alltag interessierende Frage, ob sich aus der Kenntnis einer Verhaltensdisposition einer bestimmten Person (z.B. Herr X ist kein Betrüger) ihre Entscheidung in einer bestimmten moralischen Konfliktsituation vorhersagen läßt (Herr X nimmt kein Geld aus der Kasse, wenn er allein im Laden ist). Dies ist aber kein Urteil-Handlungs-, sondern ein Handlung-Handlungs-Problem, das sich prinzipiell empirisch untersuchen läßt. Praktisch aber sind aus ethischen und anderen Gründen nicht alle Handlungen der Forschung zugänglich. Nur als Literaturgestalt konnte Dostojewskijs *Raskolnikow* einen Menschen umbringen, um die Wirkung dieser Tat auf seine moralischen Gefühle studieren zu können. Wie eng die Grenzen des Experimentierens auf dem Gebiet der Moral sein können, zeigt die Diskussion um das *Milgram*-Experiment (Milgram 1974; für einen Überblick siehe Meeus & Raajmakers (1989).

11 Siehe die gründliche Kritik von Pittel & Mendelsohn (1966) an dieser Art der operationalen Definition und Messung von Moralität.

ser “Erfolg” hat sich, so soll gezeigt werden, negativ auf die theoretische Validität dieses Meßinstruments ausgewirkt. Zudem führen bestimmte Anweisungen des Handbuches (Colby, Kohlberg et al. 1987) zur Bevorzugung der Invarianztheorie, was einen fairen Theorievergleich erschwert. Diese Verzerrung zugunsten der Reifungstheorie wird zudem verstärkt durch die weit verbreitete Tendenz in der *Moral Judgment*-Forschung, das Bildungsniveau als Bedingungsfaktor zu übersehen und fast nur Jugendliche zu untersuchen, die sich in der Ausbildung befinden, bei denen also altersmäßige Reifung und Bildungserfahrungen zusammenfallen.

Zur Behebung dieser Probleme wird in dieser Untersuchung ein neuer Meßansatz angewandt, der *Moralisches Urteil-Test* (MUT), der eine unverzerrte Messung von Regressionen zuläßt (Kapitel 4). Der MUT orientiert sich zwar eng an Piagets und Kohlbergs *Theorie* der Moral, insbesondere an der Struktureigenschaft des moralischen Urteilverhaltens, aber nicht an deren Methoden. Die Operationalisierung der Struktur basiert vor allem auf den Methoden der psychologischen Einstellungsforschung (Likert-Skalen), der experimentellen Psychologie (multifaktorielle Designs), der *Judgment*-Skalierung und der – in Vergessenheit geratenen – *diakritischen Methode* zur Überprüfung von Dispositionshypothesen von Egon Brunswik.[12] Zum anderen wurden neben Jugendlichen, die lange in der Ausbildung verblieben, auch solche untersucht, die ihre Ausbildung früh beendeten, so dass der Einfluß von Alter und Bildungsniveau auf die Moralentwicklung – zumindest partiell – unabhängig voneinander studiert werden kann.

1.3 Kritik des Kohlberg-Ansatzes

Aufgrund des Kohlberg-Ansatzes wurden zwei wichtige Entdeckungen möglich: Erstens, fast alle Menschen anerkennen unabhängig von Geschlecht, Schicht oder Kultur moralische Prinzipien wie Gerechtigkeit und Hilfsbereitschaft. Nahezu jeder bevorzugt dabei – im Sinne der Kohlberg-Stufen – höhere Formen der moralischen Argumentation gegenüber niedrigeren (Rest 1973; Lind 1985a). Zweitens, die Fähigkeit, moralische Konzepte im eigenen Verhal-

12 Zur *Judgment*-Skalierung s. Torgerson 1967/1958; zur diakritischen Methode s. Brunswik 1955; auch Kapitel 4.

ten konsistent und differenziert anzuwenden, also das, was wir als moralische Urteilsfähigkeit bezeichnen, ist in jungen Jahren noch wenig ausgebildet und bedarf der Entwicklung. Kinder wenden moralische Prinzipien oft noch inkonsistent und undifferenziert an. Wenn ihr Bewusstsein mit der Verfolgung *eines* Ziels beschäftigt ist, haben andere Ziele kaum eine Chance. Auf den unteren Stufen der individuellen Moralentwicklung lassen sich Menschen noch selten von moralischen Argumenten beeinflussen, insbesondere wenn sie gerade versuchen, ihre Meinung durchzusetzen. Erst im Verlauf der Entwicklung lernen sie, das Engagement für eine Sache mit einem integrierten und differenzierten moralischen Urteil zu verbinden. Erst dann können sie mehrere Ziele, moralische und nicht-moralische bei einer Entscheidung gleichzeitig berücksichtigen und in ihrem Selbst integrieren.

Die vorliegende Untersuchung zeigt, dass der Kohlberg-Ansatz aber stark revisionsbedürftig ist, was das Verhältnis von Affekt und Kognition im moralischen Verhalten, die Definition von "Struktur" und, worauf bereits hingewiesen wurde, die Annahme der invarianten Sequenz betrifft, nach der gilt: a) *Moralische Affekte und Kognitionen dürfen nicht als separate Verhaltenskategorien, sondern müssen als Aspekte ein und desselben Verhaltens verstanden und gemessen werden*. In seiner ersten Arbeit hierzu, definierte Kohlberg (1958, S. 89) dieses Verhältnis ganz unterschiedlich: (A) als eine untrennbare Einheit, (B) als unterscheidbare, aber parallele Aspekte des moralischen Verhaltens, (C) als Stufe und Ebene der Moralentwicklung und (D) als moralische und logische Urteilsfähigkeit. Die Gleichsetzung von Stufen mit Affekten (Einstellungen) und Ebenen mit Kognitionen hat Kohlberg später zu Recht fallengelassen. Die Gleichsetzung des kognitiven Aspekts der Moral mit amoralischen, logischen Denkfähigkeiten war wohl eine Verlegenheitslösung, da ihm keine Methoden zur Verfügung standen, um kognitive Aspekte der Moral unabhängig von den affektiven zu konzeptualisieren und messen zu können. Es verbleiben die Definitionen (A) und (B). In etwas modifizierter Fassung bilden sie die Grundlage unseres Zwei-Aspekte-Modells der moralisch-kognitiven Entwicklung und unseres Meßansatzes (s. Kapitel 2 und 4). Mit dem Zwei-Aspekte-Modell lassen sich zentrale Widersprüche des Stufen- und Struktur-Begriffs bei Kohlberg auflösen. Kohlberg hat in seinen Stufen lange Zeit eine Verlängerung der Piaget-Phasen gesehen. Die Gleichsetzung von Stufe 1 und 2 mit Piagets Heteronomie und Autonomie war jedoch nicht haltbar, ebensowenig wie die Gleichsetzung von moralischer Autonomie mit Postkonventionalität.

Mit der Identifikation von Unterstufen (*A* und *B*) hat Kohlberg später die Piaget-Phasen unabhängig von seinen Stufen zu definieren versucht. Dieser Versuch hat aber die Probleme bei der Operationalisierung von Struktur und struktureller Ganzheit nur noch deutlicher hervortreten lassen. Wie lassen sich Unterstufen vorstellen, wenn Stufen eine strukturelle Ganzheit bilden?

b) *Die Struktur des moralischen Verhaltens darf nicht als ein bestimmter Verhaltensinhalt (auf die die Strukturdefinition im Auswertungshandbuch paßt) verstanden werden, sondern muß relational definiert und gemessen werden.* Struktur bedeutet üblicherweise die Menge der *Relationen*, die zwischen einzelnen Elementen bestehen. Für die Zuerkennung moralischer Motive ist zum Beispiel die relationale Information wichtig, wie *konsistent* eine Person ein moralisches Argument verwendet. *Konsistenz* kann man nicht messen, wenn man nur einzelne Akte einer Person isoliert für sich betrachtet, wie es die klassische Psychometrie tut, sondern nur, indem relationale Merkmale von Antwort*mustern* analysiert werden. In Kohlbergs Interviewmethode ist Struktur nicht relational, sondern *formal* definiert. Obwohl Kohlberg und seine Kollegen diese Definition von Struktur ständig zu präzisieren versucht haben, ist sie bis heute umstritten (Eckensberger & Reinshagen 1980; Lind 1985b; Keller 1990). Bemerkenswerterweise definiert Kohlberg in seinen theoretischen Arbeiten Struktur *relational*, nämlich als das manifeste *Muster* des individuellen Urteilsverhaltens. Eine relationale Definition von Struktur wird auch durch das Postulat der "strukturellen Ganzheit" nahegelegt, das Kohlberg durch die hohe Antwortkonsistenz seiner Befragten bestätigt sieht. Aber Kohlberg und seine Kollegen haben sich für eine a-strukturelle Methodologie (die sog. klassische Testtheorie) als Grundlage ihrer Forschung entschieden, was die Erhebung relationaler Informationen, wie Konsistenz, ausschließt (Kapitel 3).

c) Die Moralentwicklung ist gekennzeichnet durch einen relativ langsamen Aufbau von Fähigkeiten (statt einer schnellen Änderung von Einstellungen) infolge von Bildungsprozessen und einem mindestens ebenso langsamen Abbau dieser Fähigkeiten nach Wegfall von Bildungs- und Übungsmöglichkeiten. Moralentwicklung ist also nicht, wie Kohlberg annimmt, gekennzeichnet durch eine invariante Abfolge von Stufen; *moralische Urteilsfähigkeit entwickelt sich zurück, wenn sie nicht durch diskursive Prozesse und pädagogische Interventionen stabilisiert und gefördert wird* (s. Kapitel 5 und 11).

1.4 Die sozialpsychologische Perspektive der Moral Judgment-Forschung

Die Sozialpsychologie ist, so stellt Kumpf fest, "in den USA und auch hierzulande zusehends a-sozialer geworden, in dem Sinne, dass sie sich ganz dem Studium von Individuen zu- und von sozialen Phänomenen abwandte" (Kumpf 1990, S. 123). Dieser Vorwurf wird auch der kognitiven Entwicklungstheorie gemacht (s. Schreiner 1992), was aber nur zum Teil berechtigt ist. Schon früh hat Kohlberg (1969) versucht, individuelle Eigenschaften auf der einen Seite und Merkmale sozialer Institutionen auf der anderen, sowie ihr Aufeinandertreffen und ihr Zusammenwirken theoretisch miteinander in Beziehung zu setzen. Weniges davon konnte aber bislang in empirische Forschung übersetzt werden. Die vorliegende Untersuchung greift daher die sozialpsychologische Dimension verstärkt auf, indem Kommunikationsprozesse, Phänomene der sozialen Wahrnehmung, demokratische Interaktionen und Prozesse der sozialen Wahrnehmung – Stereotypisierung und pluralistische Ignoranz – in die Analyse einbezogen werden (Kapitel 8 und 10). Darüber hinaus knüpft sie, wie schon erwähnt, am klassischen Einstellungskonzept der Sozialpsychologie an bzw. baut dieses strukturell aus. Das Einstellungskonzept hat, wie Scott (1968, S. 208) zu Recht bemerkte, den unschätzbaren Vorteil, dass seine präzise Definition und Empiriebezogenheit seine Unangemessenheiten offenlegt. Vor Kohlberg wurde lange Zeit fast ausschließlich dieses Konzept verwendet, um die Moralität von Personen zu analysieren.[13] Einstellungsskalen liefern unmittelbares verbales Verhaltensmaterial. Sie sind daher nicht mit den Problemen (Halo-Effekt, Anker- und Referenzreiz-Problemen) der Selbstbeschreibungen oder Selbstdiagnosen in Persönlichkeitstests behaftet.[14] Das Manko des klassischen Einstellungskonzepts liegt darin, dass es keine Möglichkeit bietet, *strukturelle* Unterschiede festzustellen, so dass es zum Beispiel nicht möglich ist, undifferenzierte Einstellungen von neutralen Einstellungen zu unterscheiden (Scott 1968; Converse 1970). Durch die strukturelle Erweiterung des Einstel-

13 May & Hartshorne 1926; Hartshorne & May 1928; Krathwohl et al. 1964; Shaw & Wright 1967; zusammenfassend Pittel & Mendelsohn 1966.

14 Es liegt nahe, systematische Tendenzen von Menschen bei der Beurteilung anderer (R. Cohen 1969) auch dort zu vermuten, wo es um die Beurteilung der eigenen Persönlichkeit geht.

lungskonzept im Rahmen unseres *Zwei-Aspekte-Modells* ist dieses Problem weitgehend gelöst (Kapitel 2 und 4; Lind 1982). Das um das Strukturmerkmal erweiterte Einstellungskonzept ist zur Zeit offensichtlich die einzige Möglichkeit, für uns wichtige Thesen wie die der Regression, der Kulturspezifität oder der affektiv-kognitiven Parallelität adäquat zu überprüfen. Darüber hinaus erfüllt es die Ockham-Russelsche Forderung nach Intersubjektivität und nach theoretischer Transparenz und Sparsamkeit[15], vor allem im Vergleich zu sogenannten qualitativen Meßansätzen, die sich oft auf *ad hoc* formulierte, nur schwer nachvollziehbare Annahmen stützen.[16]

1.5 Ein Prozeßmodell der Veränderung der moralischen Urteilsfähigkeit

Der Aufbau der hier vorgelegten Arbeit orientiert sich an einem allgemeinen Prozeßmodell der Veränderung der moralischen Urteilsfähigkeit. Dieses Prozeßmodell spiegelt in groben Zügen die Grundannahmen wider, die wir hier auf den Prüfstand stellen wollen, und bildet gleichzeitig ein Resümee unserer Untersuchung (Abb. 1.1). Alle unsere Analysen gründen auf offen *beobachtbarem Verhalten* von Menschen und nicht auf prinzipiell unverifizierbaren ("hypothetischen") Konstrukten oder unbeweisbaren Annahmen über Motive von Menschen. Wo wir mentale Begriffe bei der Beschreibung des moralischen

15 Die Forderung des Wilhelm von Ockham "entia non sunt multiplicanda, sine necessitate" wird auch als "entia non sunt multiplicanda praeter necessitatem" überliefert. Frei übersetzt: Man solle die Zahl der Konstrukte nicht ohne Not vermehren. Der schneidenden Kritik wegen, mit der Ockham (1290-1350) diesen Satz gegen die scholastische Unart gebraucht hat, jeder in einer Theorie auftretenden Schwierigkeit mit der Erfindung immer neuer Konstrukte zu begegnen, ist er in der Philosophie auch als "Ockhams Rasiermesser" bekannt. Die moderne Version dieser Forderung stammt von B. Russel (1972): "Wenn in einer Wissenschaft alles interpretiert werden kann, ohne diese oder jene hypothetische Sache anzunehmen, dann gibt es keinen Grund sie anzunehmen" (S. 472; soweit nicht anders vermerkt sind alle fremdsprachigen Zitate in dieser Arbeit vom Verfasser übersetzt).

16 Kohlberg hat seine Skepsis gegen das Einstellungskonzept erst in jüngerer Zeit revidiert und anerkannt, daß mit strukturellen Einstellungstests wie dem *Moralisches-Urteil-Test* (MUT) wesentliche Verbesserungen erreicht werden können: Mit dem MUT, so Kohlberg (1985), "ist ein Auswertungsalgorithmus möglich, mit dem sich jedem Individuum ein reiner Strukturwert zumessen läßt" (S. xvii).

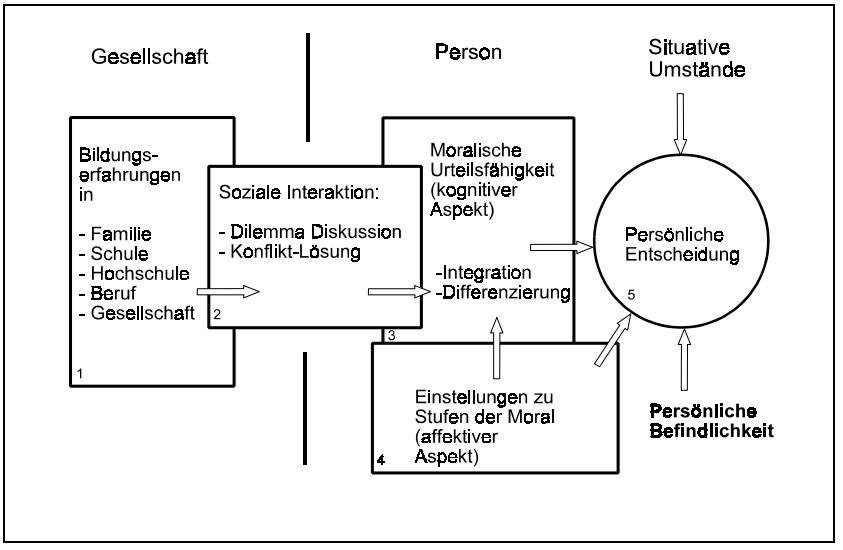


Abb. 1.1 Prozeßmodell

Verhaltens verwenden, versuchen wir, diese so präzise wie möglich als Eigenschaften des *Verhaltens* eines Individuums operational zu definieren, statt allein auf den unsicheren Grund unserer Intuition zu vertrauen. Aber wir verbinden mit unseren Untersuchungen an keiner Stelle den Anspruch, einen ganz bestimmten Verhaltensakt eines Menschen *vorherzusagen*, wie es schon so viele vergeblich versucht haben. Dieser Verzicht resultiert sowohl aus einer ethischen Selbstbeschränkung wie auch aus der Einsicht in die prinzipielle Unmöglichkeit eines solchen Unterfangens. Wir können nur davon ausgehen, dass die moralische Urteilsfähigkeit ebenso wie die moralischen Einstellungen einer Person einen wichtigen Einfluß auf ihre persönliche Entscheidung haben. Aber wir müssen ebenso davon ausgehen, dass diese Entscheidung, soll sie angemessen und gerecht sein, eben so sehr von – unvorhersehbaren – situativen Umständen und Befindlichkeiten abhängt. Die vorliegende Untersuchung beschränkt sich daher auf die moralische Urteilsfähigkeit und die Bedingungen ihrer Entwicklung. Dies schließt nicht aus, *dass* moralische Urteilsfähigkeit eine große Bedeutung für die Güte individueller Entscheidungen hat. Es spricht sogar vieles für diese Vermutung. Aber es soll dem Leser überlassen bleiben, ob er die vorliegende Evidenz dafür für ausreichend hält. Unser Modell zeigt daher die persönliche Entscheidung des Einzelnen als Leerstelle.

Die Kapitel dieser Arbeit können den Komponenten unseres Prozeßmodells zugeordnet werden. Die nachfolgenden Kapitel beginnen mit dem zentralen Aspekt der moralischen Urteilsfähigkeit (Abb. 1.1, Kasten 3). Kapitel 5 untersucht anhand des Regressionsphänomens die Reifungsthese. Die Kapitel 6 und 7 richten sich auf die Hypothesen, die aus der Sozialisationsthese folgen. Die folgenden Kapitel enthalten Teiluntersuchungen zu verschiedenen Aspekten und "Mechanismen" der Veränderung des moralischen Urteils in sozialen Interaktionen im Kontext von Bildungsprozessen, und zwar zur Möglichkeit der effektiven Veränderung des "moralischen Klimas" eines institutionellen Kontextes (Kapitel 8), zur Wirksamkeit sozialer Interventionen, die sich gezielt auf eine Stimulation der kognitiven Struktur des moralischen Urteilsverhaltens richten (Kapitel 9) und zur Genauigkeit der Wahrnehmung von "moralischen Tatsachen" im institutionellen Kontext Schule (Kapitel 10). Im Kapitel 11 wird abschließend die Frage untersucht, welchen Einfluß Bildungserfahrungen im Vergleich zu anderen Faktoren wie Sozialisation, Alter und Selektion auf die moralisch-kognitive Entwicklung haben.

1.6 Technische Informationen

1.6.1 Die Datenbasis

Die Hypothesen dieser Untersuchung werden an verschiedenen Arten von empirischen und experimentellen Daten überprüft. Der überwiegende Teil der Daten, über die hier berichtet wird, stammt aus Untersuchungen im Rahmen von größeren Forschungsprojekten oder eigenen Projekten. Die Hypothesen wurden aber auch an Ergebnissen anderer Untersuchungen überprüft, wofür diese in vielen Fällen neu berechnet oder durch geeignete Umrechnungen vergleichbar gemacht wurden. Für diese Berechnungen wurden meist die Originalberichte oder manchmal auch die Rohdaten herangezogen. In einigen Fällen wurden die Ergebnisse nur in eine graphische Darstellung übersetzt, um besser beurteilen zu können, welche These sie stützen. Auf weitere Einzelheiten und die Quellen wird jeweils an der entsprechenden Stelle verwiesen. Ein wesentlicher Teil der Datengrundlage dieser Arbeit stammt aus der *international vergleichenden*

Längsschnittstudie der Konstanzer Forschungsgruppe Hochschulsozialisation.[17] Hieraus wurden, je nach Fragestellung, verschiedene Teilgruppen analysiert (Kapitel 5, 7 und 11). Die Definition dieser Teilgruppen wird in den entsprechenden Kapiteln angegeben. Als weitere Datengrundlage für diese Untersuchung dienen ein *Laborexperiment* zur Frage der Simulation von moralischer Urteilsfähigkeit (Kapitel 6) und ein *Feldexperiment* zur Veränderung der moralischen Atmosphäre in der Schule (Kapitel 8). Letzteres wurde in Zusammenarbeit mit Fritz Oser (Fribourg) und dem Landesinstitut für Schule und Weiterbildung des Landes Nordrhein-Westfalen geplant und durchgeführt. Die Daten der Repräsentativ-Untersuchung über die moralische Urteilsfähigkeit bei Jugendlichen und jungen Erwachsenen (Kapitel 5) entstammen einer 1990 durchgeführten, unveröffentlichten Berufsbildungsstudie von EMNID, Bielefeld.

1.6.2 Zur statistischen Bewertung von Zusammenhängen

Die vorliegende Untersuchung ist relativ stark von methodologischen und forschungsmethodischen Überlegungen geprägt. Dies erwies sich als notwendig, da in diese Methoden unser ganzes Vorverständnis eingeschmolzen ist, mit dem wir uns dem Gegenstand unserer Untersuchung nähern. Sie sollen daher hier näher erläutert werden (s. Lind 1985a). Wenn wir unsere Forschung auf inadäquate Methoden, d.h. auf falsches Vorverständnis gründen, kann sie nicht viel erbringen. Unglücklicherweise glauben viele Sozialwissenschaftler, dass bestimmte, historisch gewachsene Methoden von absoluter Gültigkeit, und damit *der* Garant für Wissenschaftlichkeit und die Basis fachlicher Zugehörigkeit seien – womit sie jede kreative Forschung, die immer auch auf der Entwicklung neuer Methoden beruht, zum Erliegen bringen. Die wirkliche

17 Mitglieder der Forschungsgruppe waren neben dem Autor Tino Bargel, Barbara Dippelhofer-Stiem, Gerhild Framhein, Hansgert Peisert (Projektleiter), Johann-Ulrich Sandberger und Hans Gerhard Walter. Auf internationaler Ebene war die Studie Teil der vom Europäischen Koordinationszentrum für Sozialwissenschaftliche Forschung in Wien koordinierten, international vergleichenden Längsschnittstudie "Diplomés de l'Université: leur *Form*ation et leur Conception de la Vie" (kurz: FORM) unter der Leitung von Hansgert Peisert und Władisław Markiewicz. Vgl. Sandberger et al. 1982, Framhein & Langer 1984, Dippelhofer-Stiem & Lind 1987.

Basis kollegialer Zusammenarbeit ist das Interesse an einem gemeinsamen Thema. Die richtige Einstellung zur Wissenschaftlichkeit manifestiert sich in dem Ideal, möglichst informative oder, wie Popper (1968) sagt, "riskante" Hypothesen aufzustellen, und sie einem fairen und intersubjektiv nachvollziehbaren Test durch Beobachtungen zu unterziehen. Hier wird also nicht der methodologische Standpunkt des "anything goes" vertreten. Im Gegenteil, es wird verlangt, dass die verwendeten Methoden nicht nur eine gehaltvolle und objektive Prüfung von Theorien garantieren, sondern auch, dass sie eine valide Operationalisierung der Fragestellung sind. Im Rahmen unserer Untersuchung, in deren Mittelpunkt die Prüfung dreier Theorien über die Natur und die Bedingungen der Moralentwicklung steht, interessiert vor allem die Frage, ob und in welchem Maße diese Theorien bzw. die daraus ableitbaren Hypothesen im Einklang mit unseren Beobachtungen stehen. Diese Fragestellung unterscheidet sich sehr von solchen Studien, die mit dem Ziel durchgeführt werden, an einer kleinen Gruppe Daten zu gewinnen, die – mittels der Inferenzstatistik – auf größere Gruppen verallgemeinert werden können. Daher werden hier, wie der Statistiker Kruskal (1958, S. 815) vorschlägt, Untersuchungsgruppen nicht als Stichproben, sondern als Populationen behandelt. Maße der statistischen Signifikanz wie *t*, *F* oder χ^2 sichern nicht generell, wie ihr Name suggerieren mag, die "Bedeutsamkeit" oder "Verallgemeinerungsfähigkeit" eines Befundes.[18] Zudem gibt die Höhe der statistischen Signifikanz keinen Aufschluß über die Güte einer Theorie bzw. einer Vorhersage, da sie vor allem von der Größe der Stichprobe – also einer frei wählbaren Größe – abhängig ist. Statt, wie mancherorts vorgeschlagen, diesen Nachteil durch komplizierte Maßnahmen zur Sicherung der "Teststärke" (power) auszugleichen, erscheint es naheliegender, direkt Maße für die Effektstärke zu berechnen (Cohen 1988). Diese Maße stellen mathematisch gesehen nichts anderes dar als eine Standardisierung der Signifikanzmaße auf die Stichprobengröße. Es handelt sich hier im wesentlichen um Kennziffern wie den Korrelationskoeffizienten *r*, Varianzanteile oder *d*-Werte (Glass et al. 1978; Cohen 1988). Wir werden uns in dieser Arbeit in der Hauptsache auf *r* als Maß

18 Diese Maße geben nur an, inwieweit ein Befund repräsentativ für die Grundgesamtheit von Fällen ist, aus der eine Stichprobe ausgewählt wurde. Die Stichprobenbefunde sind nur dann repräsentativ, wenn die Gesamtheit genau definiert und die Auswahl streng nach Zufall geschehen ist (Hays 1963).

für die Stärke eines Effekts bzw. die Größe eines Zusammenhangs oder einer Differenz stützen (s. Kapitel 9). Aber auch Effektstärkemaße haben Nachteile. Sie sind, entgegen der oft benutzten Kennzeichnung als "Maße der praktischen Relevanz", nicht immer geeignet, praktisch relevante Aussagen über die Größe eines Effekts zu formulieren. Sie sind nicht unabhängig von zufälligen, für die Aussage irrelevanten Charakteristika der jeweiligen Untersuchungsgruppe. Ist zum Beispiel die Varianz einer Variablen, deren Effekt untersucht werden soll, in einem Sample zufällig sehr gering oder null, dann kann sich kein Effekt zeigen, auch wenn er vorhanden wäre. Eine naheliegende Möglichkeit der Lösung dieses Problems besteht darin, die absoluten Meßwertunterschiede, die in einer Studie registriert werden, *direkt* als Maß für die Effektstärke zu nehmen. Wenn zwei Meßwerte, z.B. 4,3 und 5,1, vorliegen, so ist der Effekt die Differenz, also 0,8. Hier wird man aber zurecht einwenden, dass die Zahl 0,8 nur sinnvoll interpretiert werden kann, wenn man z.B. weiß, wie groß der Wertebereich ist, den die Skala zuläßt. Bei einer Skala von 4 bis 6 stellt eine Differenz von 0,8 einen relativ großen Wert dar; bei einer Skala von –100 bis +100 aber einen sehr kleinen Wert. Die Bedeutung absoluter Differenzen hängt also ganz entscheidend von den Werten ab, die bei einer bestimmten Skala aufgrund ihrer Konstruktion theoretisch erreicht werden können. Einige Autoren wie beispielsweise Fend und seine Mitarbeiter (Fend et al. 1976) haben daher alle Skalen, die sie in einer Untersuchung verwendet haben, auf eine gleiche Länge (0 bis 20) umgerechnet und so eine Vergleichbarkeit hergestellt. Diese Idee in etwas abgewandelter Form aufgreifend, will ich zwei Schritte zur Erzielung einer standardisierten (von irrelevanten Gegebenheiten unabhängigen) Effektstärke vorschlagen: Erstens die Berechnung einer *relativen Differenz* (die absolute Differenz dividiert durch die absolute Skalenbreite) und zweitens eine Übereinkunft darüber, wie groß eine relative Differenz sein muß, um als "bedeutsam" oder "sehr bedeutsam" bezeichnet zu werden. Eine solche Übereinkunft wurde, soweit ich weiß, bislang von niemanden explizit getroffen, aber sie ist in der empirischen Forschung faktisch vorhanden. Die Durchsicht sehr unterschiedlicher sozialwissenschaftlicher Studien zeigt, dass zwar die Spannweite der Differenzwerte, die in der Literatur als *sehr bedeutsam* oder *hoch bedeutsam* bezeichnet werden, groß ist; wir fanden die Bezeichnung "hoch signifikant" für relative Differenzen von 1,5 bis 15 Prozentpunkte. Aber etwa drei Viertel aller Differenzwerte, die so bewertet wurden, lagen bei 8 bis 10 Prozentpunkten. Das heißt, in der Regel werden Differenzwerte, die eine

Größe von ca. 8 bis 10% der theoretisch möglichen Skalenbreite betragen, als "hoch signifikant" oder "sehr bedeutsam" bezeichnet. Folglich werde ich in dieser Untersuchung den Terminus "sehr bedeutsam" für Effekte verwenden, die gleich oder größer als 10% der Skala sind, mit der sie gemessen wurden. Eine relative Differenz von 5% soll bloß als "bedeutsam" bezeichnet werden. Mit diesen beiden Eckwerten haben wir zwei "praktikable" Kriterien für die Interpretation von Differenzwerten, wo die isolierte Bewertung von Effekten angesagt ist. In den Fällen, in denen ein direkter Vergleich zweier oder mehrerer Vorhersagen mit den Beobachtungen möglich ist, sind solche Interpretationshilfen ohnehin unnötig.

2 Das *Zwei-Aspekte-Modell* des moralischen Urteilsverhaltens

Unser Zwei-Aspekte-Modell des moralischen Urteilsverhaltens und der Moralentwicklung baut auf Piagets und Kohlbergs kognitiver Entwicklungstheorie auf, modifiziert dieses aber in wesentlichen Teilen. Dieses Modell legt offen, welche Vorstellungen unseren Beobachtungs- und Meßmethoden zugrunde liegen, anhand derer wir die Richtigkeit – oder Falschheit – der drei in Frage stehenden Theorien der Moralentwicklung prüfen wollen. Es wird hier ausführlich behandelt, weil die Frage, wie der Untersuchungsgegenstand definiert ist, d.h. was man sich darunter im Einzelnen vorstellt, von ganz entscheidender Bedeutung für den Aufbau und die Auswertung der Beobachtungs- oder Meßinstrumente ist, mit deren Hilfe geeignete Daten für die Beurteilung der Theorien gefunden werden sollen. Das Zwei-Aspekte-Modell wurde vor allem entwickelt, um einige grundlegende Probleme des Kohlberg-Ansatzes zu lösen. Wie im nächsten Kapitel gezeigt wird, besteht zwischen dem theoretischen Modell und der Meßmethode Kohlbergs (dem *Moral Judgment Interview*) ein Widerspruch, da mit seinen Beobachtungsdaten zentrale Vorhersagen seiner Theorie nicht oder nicht adäquat überprüft werden können. Dazu gehören die Vorhersage einer invarianten Entwicklungssequenz, auf die in Kapitel 5 näher eingegangen werden soll, und die Vorhersage einer *affektiv-kognitiven Parallelität*, die hier eingehend untersucht werden soll.

Was ist mit der "affektiv-kognitiven Parallelität" gemeint? Nur eine Definition oder auch ein empirischer Sachverhalt? Obwohl die Annahme einer Parallelität in vielen Arbeiten von Kohlberg, aber auch schon von Piaget auftaucht, haben beide sich dazu immer nur sehr vage geäußert. Kaum einer, so kritisiert William Damon (1983), hat bisher für die Beziehung zwischen Affekt und Kognition ein befriedigendes Konzept vorgeschlagen. Gertrud Nunner-Winkler kommt sogar zu dem Schluß, dass es keine Parallelität der Entwicklung beider Aspekte gibt. Die Befunde, schreibt Nunner-Winkler (1989a), lassen sich nur dann stimmig deuten, "wenn man die Piaget-Kohlbergsche These der kognitiv-affektiven Parallelität aufgibt und die unterschiedlichen Ergebnisse im Zusammenhang mit den jeweils verwendeten Erhebungsmethoden liest" (S. 593; auch Nunner-Winkler 1990a, S. 4).

In diesem Kapitel soll gezeigt werden, a) dass der Begriff der kognitiv-affektiven Parallelität die Grundlage für eine adäquate Definition des moralischen Verhaltens (und damit auch der moralischen Entwicklung) bietet, b) dass er die Hypothese eines engen korrelativen Zusammenhangs zwischen affektiven und kognitiven Aspekten des moralischen Urteilsverhaltens enthält, die sich empirisch gut bestätigen läßt, und c) dass er auch eine Hypothese über das entwicklungsmäßige Zusammenwirken beider Aspekte impliziert, die in Einklang mit anderen Forschungsbefunden steht.

2.1 Das Problem der Erfassung von Verhaltensaspekten

Der Dissens zwischen beiden Standpunkten gründet in der Frage, wie man sich die Beziehung zwischen Affekt und Kognition vorstellen soll und wie man beides in der Forschung methodisch angemessen erfassen kann. Der schwierigste Teil dieses Problems besteht wohl darin, Affekt und Kognition als *Aspekte ein und desselben Verhaltens* zu begreifen und beide Aspekte so zu operationalisieren, dass aus ihnen nicht unversehens Namen für zwei ganz verschiedene Komponenten werden. Der Unterschied zwischen Aspekten und Komponenten wird in der Psychologie oft übersehen oder als unwichtig erachtet, obwohl er im Alltag geläufig und wichtig ist. Die Aspekte eines Gegenstands, zum Beispiel die Elastizität oder Rundheit eines Balls, lassen sich nicht als eigene Komponenten ansehen und vom Ball entfernen; Komponenten, beim Fußball beispielsweise der äußere Mantel und die "Blase", können ohne weiteres in Teile zerlegt und separat untersucht werden. Ähnlich gibt es verschiedene Eigenschaften, die für das Verhalten einer Person charakteristisch sind, die aber von ihr nicht materiell zu trennen sind. Man kann zum Beispiel die Vermutung, dass eine Person hilfsbereit, aber unzuverlässig ist, an ein und demselben Verhaltensausschnitt gleichzeitig prüfen. Es ist nicht nötig, oder gar zulässig, beide Verhaltenseigenschaften an ganz verschiedenen Verhaltensausschnitten in verschiedenen Kontexten zu erfassen. Denn die Hypothese impliziert ja auch, dass diese Person generell gerade in Hilfesituationen unzuverlässig ist: sie sagt schnell zu, zu helfen, merkt aber später, dass sie dazu gar

nicht in der Lage ist. Das Problem ist nicht auf die Moralforschung begrenzt.[19] In der Theorie wird es uns immer deutlicher, dass es nicht möglich ist, affektive Dispositionen unabhängig von kognitiven zu erfassen. Verschiedene Bereiche wie die Motivations-, die Emotions- und die Einstellungsforschung lehren uns, dass sich affektive Aspekte des menschlichen Verhaltens und Bewußtseins nur angemessen verstehen lassen, wenn man kognitive Aspekte hinzuzieht.[20] Ohne die Annahme kognitiv-struktureller Aspekte läßt sich beispielsweise das Problem der Vieldeutigkeit von Werten im mittleren Bereich von Einstellungsskalen nicht lösen. Während zum Beispiel in den Extrembereichen einfach von einem positiven oder negativen Affekt einer Person gegenüber der moralischen Norm "Du darfst nicht lügen" gesprochen werden kann, stehen uns bei Werten zwischen diesen Extremen drei ganz verschiedene Interpretationen zur Auswahl, zwischen denen wir keine empirisch begründete Entscheidung treffen können. Ein mittlerer Wert kann

(1) eine nicht oder nur schwach vorhandene Einstellung
(2) einen Einstellungskonflikt oder
(3) eine hochdifferenzierte Einstellung

gegenüber der moralischen Norm anzeigen.[21] Nur die erste Interpretation ist rein "affektiv". Die beiden anderen sind "kognitive" Interpretationen wie etwa: Eine Person erhält einen mittleren Wert, wenn sie meistens für die Einhaltung

19 Tesser und Shaffer (1990, S. 480) verzeichnen einen Trend in der Psychologie, die dreigeteilte Definition von Einstellungen zu verwerfen. Es bleibt allerdings unklar, wie ihre Eingrenzung der Psychologie auf physiologische Methoden, auf die Messung von Hautwiderständen u.ä., zu einer Lösung des Problems beitragen kann.

20 U. Neisser (1966): "Es ist offensichtlich, dass Kognition in allem beteiligt ist, was ein menschliches Wesen tun mag; dass jedes psychologische Phänomen ein kognitives Phänomen ist" (S. 4). Unterstützt wird diese Sichtweise u.a. durch Tolmans Theorie des latenten Lernens (1932), die auf L. Festinger zurückgehende kognitive Konsistenztheorie der Einstellung (McGuire 1966), die kognitive Motivationstheorie von Weiner (1976, S. 130 ff.), die Zwei-Faktoren-Theorie der Emotion von Schachter und Singer (1962) und die Theorie der kognitiven Bewältigung von psychischem Streß von Lazarus (1966).

21 Auf diese Vieldeutigkeit haben bereits Shaw & Wright (1967, S. 7) hingewiesen.

der Norm ist, aber in bestimmten Umständen Lügen als erlaubt oder, wenn es um Leib und Leben geht, sogar als geboten ansieht.[22]

Die "kognitive Wende" in der Psychologie war eine Reaktion auf die Sackgasse, in die das rein affektive Modell der Einstellung und der Persönlichkeit geführt hat. Aber, wie wir heute erkennen müssen, hat sie zu einer anderen, kognitiven Einseitigkeit geführt. Kognitive Prozesse können nicht definiert und noch weniger untersucht werden, ohne dabei Affekte, Einstellungen, Motive usw. zu berücksichtigen.[23] Anders als in der unbelebten Natur haben wir es in der belebten fast immer mit gerichteten, "teleologischen" Strukturen zu tun. Das menschliche Verhalten ist nicht durch Konsistenz "an sich" charakterisiert, sondern durch Konsistenz *in bezug* auf ein Ziel oder ein Prinzip. Konsistenz ist [hier] ein zweiwertiger Begriff. Während rein formale Strukturen, wie sie in der Welt der Physik und Chemie vorkommen (z.B. im Sinne der Konsistenz eines Stoffes), Anordnungen ohne eine dynamisch-affektive Bedeutung sind, beziehen sich *dynamische* Strukturen auf menschliche Handlungen, die eine Bedeutung besitzen (Lind 1985b, S. 30).

2.2 Grundzüge des *Zwei-Aspekte-Modells*

Um eine Bilanz aus diesen Ergebnissen der psychologischen Forschung zu ziehen, können wir ein Zwei-Aspekte-Modell von Affekt und Kognition formulieren. Im Kern geht es auf Piagets (1976) Erkenntnis zurück, dass "*jede Form des Verhaltens einen energiehaften oder affektiven und einen strukturellen oder kognitiven Aspekt hat* [. . .]. Beide Seiten des menschlichen Verhaltens [sind]

22 Solche Antworten scheinen fortgeschrittene Stufen der moralisch-kognitiven Entwicklung anzuzeigen, wo die Rigorosität moralischer Prinzipien durch Überlegungen der Billigkeit gemildert wird (Bull 1969, S. 210; Piaget 1973a, S. 358).

23 So betont Neisser (1966), "dass viele kognitive Phänomene unverständlich bleiben müssen, wenn man nicht darauf achtet, was die beobachtete Person jeweils beabsichtigt". Neisser (1980) sah sich jüngst gar veranlaßt, vor einer übertriebenen "Kognitivierung" der Psychologie zu warnen. Posner (1976) gibt mehrere Beispiele, wie Einstellungen das Problemlöseverhalten von Personen beeinflussen (S. 171 ff.). Einer der Begründer der "Cognitive Science"-Bewegung, D.A. Norman (1980), weist daher mit Nachdruck darauf hin, dass "die menschliche Intelligenz aus mehr besteht als aus reinen kognitiven Systemen und dass die Wissenschaft der Kognition es sich nicht leisten kann, diese anderen Aspekte zu ignorieren" (S. 4).

	Aspekte-Modell	Komponenten-Modell
Affekt-Kognition-Beziehung	Unterscheidbare, aber untrennbare Eigenschaften eines Verhaltensmusters (Parallelität)	Getrennte Komponenten, repräsentiert durch verschiedene Verhaltensklassen
Affekt	Inhalt des Verhaltens, Moralische Motive (Stufen), andere Gewohnheiten	"Affektives" Verhalten, Handlungsstrukturen, moralische Kompetenz
Kognition	Struktur des Verhaltens: Integriertheit und Differenziertheit des Urteils	"Kognitives" Verhalten, Wahrnehmungsstrukturen, moralisches Urteil
Verhalten	Interaktion mit moralischen Situationen	Normenkonformes Verhalten, moralische Performanz

Abb. 2.1 Gegenüberstellung von Aspekte- und Komponenten-Modell

nicht aufeinander zurückführbar" (S. 7-8).[24] Das Zwei-Aspekte-Modell impliziert, wie Kohlberg (1958) in seiner ersten Untersuchung des moralischen Verhaltens und Urteilens schrieb, "*dass ein moralischer Akt oder eine moralische Einstellung weder anhand rein 'kognitiver' noch anhand rein 'motivationaler' Kriterien definiert werden kann*" (S. 16).[25] Es impliziert aber auch, dass

24 Meine Hervorhebung. Siehe auch Krathwohl et al. (1964). Für sie ist "in jedem affektiven Verhalten ein kognitives Element vorhanden oder impliziert" (S. 47). An dieser Einsicht kommt, wie H. Simon (1982) feststellte, auch die experimentelle Psychologie nicht vorbei, da alle "Maßnahmen zur Erzeugung von Affekten an kognitiven Prozessen ansetzen" (S. 338). Zajoncs (1980) vager Einwand, dass Objekte, um Affekte zu erzeugen, "sehr wenig kogniziert" werden müßten (S. 154), widerlegt nicht die Gültigkeit dieser Feststellung. Vielmehr unterstreicht sie, worauf Locke (1984) hingewiesen hat, die Notwendigkeit einer begrifflichen Klärung.

25 Hervorhebung von mir. Auch J. Loevinger (1976) betont die Notwendigkeit "einer Kombination des entwicklungstheoretischen Ansatzes mit einem gleichzeitigen Interesse an Motivation und kognitiven Aspekten der Persönlichkeit" (S. 101).

Aspekte logisch unabhängig voneinander, aber an ein und demselben Verhaltensmuster einer Person erfaßt werden können.[26]

Während die erste Implikation in der Moralforschung inzwischen weitgehend unbestritten ist, gibt es bei der zweiten Implikation, so scheint es, Schwierigkeiten, sie in der Forschung zu realisieren. Es herrscht – in der kognitiv-entwicklungstheoretischen Moralforschung wie in der Psychologie insgesamt – fast überall noch die Auffassung, dass Affekt und Kognition zwei irgendwie aufeinander bezogene, aber dennoch dinglich klar getrennte und daher an unterschiedlichen Verhaltensklassen beobachtbare Dispositionen darstellen. Und es gibt, von unserem Vorschlag für einen integrativen Meßansatz abgesehen, keine Ansätze, Affekt und Kognition als *Aspekte* des moralischen Urteilsverhaltens zu operationalisieren. Vielmehr gehen die meisten Untersuchungen im Bereich der Moralpsychologie (wie auch, was kaum mehr betont werden muß, in anderen sozialpsychologischen Forschungsgebieten) noch immer explizit oder implizit davon aus, dass sie getrennte *Komponenten* darstellen, die durch inhaltlich verschiedene Verhaltensklassen repräsentiert werden.[27] In Abb. 2.1 sind beide Modelle einander schematisch gegenübergestellt, um die Unterschiede zwischen dem Aspekte-Modell und dem Komponenten-Modell von Affekt und Kognition zu verdeutlichen. Aspekte sind wie Komponenten *voneinander un-*

26 Vgl. Lind 1978; 1985a; 1985b; 1986a, Lind und Wakenhut 1985. Bereits Moers (1930) hatte erkannt, dass es wegen der inneren Verbindung zwischen den kognitiven und affektiven Aspekten der Moral "niemals möglich ist, sie getrennt zu erfassen" (S. 458). Doch konkrete Methodenvorschläge ist Moers schuldig geblieben. Ähnlich bezeichnen Villenave-Cremer & Eckensberger (1986) den affektiven und den kognitiven Aspekt des moralischen Urteils als zwei "Seiten der Medaille" (S. 203), aber operationalisieren beide Aspekte als zwei völlig verschiedene Verhaltensklassen.

27 Ein explizites Komponenten-Modell der Moral findet sich u.a. bei Boesch 1984, Kohlberg & Candee 1984, Döbert & Nunner-Winkler 1978, Eckensberger 1984, Rest 1984, Walker 1986. Beispielhaft sei die Untersuchung von Villenave-Cremer & Eckensberger (1986) erwähnt, in der Kognition als moralisches Urteil im Sinne von Kohlberg und Affekt als Bewältigungsstrategie im Sinne von Haan (1977) operationalisiert wird. Die Problematik dieser Gleichsetzung zeigt sich bei dem Versuch, "Affekt" als "affektive Genauigkeit" zu präzisieren. Diese wird definiert als die Genauigkeit, mit dem die Versuchsperson ihre eigenen Affekte und die Affekte anderer Personen kennt (S. 198). Genauigkeit ist aber in erster Linie ein Hinweis auf kognitive Prozesse und nicht auf Affekte. Allenfalls sehr indirekt und unter Zuhilfenahme von unbeweisbaren Vermutungen läßt sich eine Verbindung zu Affekten herstellen.

terscheidbar, aber nicht wie diese trennbar; sie repräsentieren inhaltliche und strukturelle Eigenschaften *eines* Verhaltens und meinen nicht wie Komponenten inhaltlich verschiedenes Verhalten. Schließlich ist moralisches Verhalten im Aspekte-Modell definiert als jene Interaktion mit moralischen Situationen, die durch bestimmte moralische Inhalte und Strukturen gekennzeichnet sind, während im Komponenten-Modell moralische Performanz gewöhnlich als Übereinstimmung des Verhaltens mit bestimmten äußeren Normen definiert ist.[28] Im Rahmen des Zwei-Aspekte-Modells sind Aussagen wie etwa "moralisches Urteilen hängt von kognitiven Prozessen ab" oder "kognitive Prozesse gehen dem moralischen Urteilen zeitlich voran" gegenstandslos, da solche Aussagen ein Komponenten-Modell implizieren.

Während das Aspekte-Modell durch die Ergebnisse psychologischer Forschung und Theoriebildung nahegelegt wird, wird in Psychologie und Erziehungswissenschaft weiterhin das *Zwei-Komponenten-Modell* bevorzugt, manchmal ausdrücklich, wie bei Rest (1986), zumeist aber nur implizit durch die Wahl bestimmter Forschungsmethoden. Es scheint sogar, dass die Wahl des Komponentenmodells einer Not gehorcht. *Es gibt eine Vielzahl von Taxonomien und Meßansätzen, in denen der affektive und kognitive Bereich als isolierte Komponenten operationalisiert werden, aber selten ein Meßansatz, in dem versucht wird, beide als Aspekte ein und desselben Verhaltensbereichs zu erfassen.*

Das methodologische Problem wird durch verschiedene Einwände gegen die Parallelitätsthese im Sinne des Aspekte-Ansatzes deutlich. So argumentiert Rest (1984), dass Affekt und Kognition als untrennbare Aspekte des Verhaltens keine theoretisch klar unterscheidbaren Einheiten darstellen und sich ihre empirische Analyse erübrige. Rest übersieht hierbei aber, dass, wie Piaget aufwies, die Annahme der Parallelität nicht eine Identität, sondern nur eine Nichttrennbarkeit beider Aspekte meint. Dieser Einwand scheint auf einer Gleichsetzung oder Verwechslung der Begriffe "unterscheiden" und "trennen" zu beruhen. Würde die Unterscheidbarkeit von Eigenschaften immer auch implizieren, dass sie getrennt, an verschiedenen Verhaltensausschnitten erfaßt werden müßten, dann wäre es in der Tat unmöglich, den affektiven und den kognitiven Aspekt logisch unabhängig voneinander zu definieren.

28 Zur Notwendigkeit der Berücksichtigung von Situationen in der Messung psychischer Phänomene vgl. Wakenhut (1978).

Boesch (1984) wendet gegen die Parallelitätsthese ein, es sei schwierig, sich vorzustellen, dass affektive Schemata dieselben Eigenschaften haben wie kognitive Schemata. Dieser Einwand geht davon aus, dass es sowohl affektive wie auch kognitive Schemata gibt. Piaget (Piaget 1951, S. 220 ff.) legt eine solche Deutung durchaus nahe (Tomlinson-Keasey & Eisert 1981). Aber diese Deutung impliziert ein Komponenten-Modell, das sich in der Forschung als nicht realisierbar erwiesen hat. Schemata bezeichnen die kognitiv-strukturellen Aspekte des Verhaltens, Affekte ihren "energetischen" und orientierenden Aspekt. Letztere selbst wieder als Schemata zu bezeichnen, würde die Fruchtbarkeit dieses Konzepts ernsthaft gefährden. Wenn man, wie Piaget, ein Schema als den kognitiven oder strukturellen Aspekt des Handelns definiert und diesen klar von dem energetischen, affektiven Aspekt unterscheidet, dann ist ein "kognitives Schema" ein Pleonasmus und der Begriff "affektives Schema" eine unzulässige Begriffsverbindung ohne psychologischen Sinn.

Schließlich wird verschiedentlich darauf verwiesen, dass man bei der Prüfung der Parallelitätsthese Affekt und Kognition fast immer mittels separater Verhaltensausschnitte (Aufgabenklassen) operationalisiert und damit auch den kognitiven Aspekt als Inhalt und nicht als Struktur erfaßt hat. Dieser Einwand ist insoweit berechtigt, als er sich auf die Forschungsmethoden bezieht. Aber er ist falsch, insofern er unterstellt, in diesen Untersuchungen würde tatsächlich die Parallelität von *Aspekten* untersucht. Denn häufig handelt es sich in diesen Untersuchungen um die Analyse der *Décalage* zwischen zwei Komponenten, der moralisch-kognitiven Komponente auf der einen Seite und der Komponente logischer Denkfähigkeit auf der anderen.[29] Wie ein Meßansatz aussehen könnte, der affektive und kognitive Eigenschaften des Verhaltens gleichzeitig erfaßt, galt lange Zeit als ungelöstes Problem. "Im Fall der moralischen (Urteils-) Stufen ist nicht klar, was genau die psychische Struktur oder das System geistiger Operationen ist, das jedem Urteilsniveau unterliegt", resümieren Kuhn, Langer, Kohlberg und Haan (1977, S. 136). Diese Autoren haben beide Aspekte daher ersatzweise an völlig verschiedenen Verhaltensausschnitten (die Bearbeitung von Logikaufgaben für den kognitiven Aspekt und von Moraldilemmas für den affektiven Bereich) erfaßt. Aber genau diese Operationalisierung erweckt, wie Rest sagt, den falschen Eindruck, "dass die [. . .] Elemente ein separates Eigenleben führen" (Rest 1984, S. 25; auch Rest 1986a).

29 Vgl. Kuhn et al. 1977; Haan et al. 1982; Edelstein et al. 1984.

Bei unserem Meßverfahren, dem *Moralisches Urteil Test* (MUT) können hingegen affektive und kognitive Charakteristika des Verhaltens einer Person – in Übereinstimmung mit dem Zwei-Aspekte-Modell – an ein und demselben Verhalten einer Person erfasst werden (Kapitel 4). Mit diesem Verfahren können wir daher die Parallelitätsthese Piagets und Kohlbergs auch in operationaler Hinsicht adäquat präzisieren und die darin enthaltenen empirischen Annahmen direkt, ohne die fragwürdige Komponenten-Annahme prüfen, die einer getrennten Erfassung zugrunde liegt.

2.3 Drei Bedeutungen von affektiv-kognitiver Parallelität

Kohlbergs Formulierung der Parallelitätsthese lautet: "Affektive Entwicklung und Funktionen und kognitive Entwicklung und Funktionen sind keine getrennten Bereiche (distinct realms). 'Affektive' und 'kognitive' Entwicklung sind *parallel*; sie repräsentieren verschiedene Perspektiven und Kontexte bei der Definition struktureller Veränderung" (1984/1969, S.9). Kohlberg hat diese Formulierung nicht weiter präzisiert und in seinen späteren Arbeiten nicht mehr erwähnt, wohl auch wegen der Schwierigkeit, sie adäquat zu operationalisieren. Wenn wir von *der* Parallelitätsthese sprechen, so müssen wir berücksichtigen, dass die Annahme einer Parallelität zwischen dem affektiven und kognitiven Aspekt des moralischen Handelns und Urteilens im Rahmen der kognitiven Entwicklungstheorie drei unterschiedliche Bedeutungen hat, von denen eine rein begrifflich-methodologischer Natur ist und zwei empirisch testbare Hypothesen darstellen.

1. Bedeutung: Methodologische Nicht-Separierbarkeit

Grundlegend für den Aspekte-Ansatz ist das methodologische Postulat einer begrifflichen Parallelität: *Der kognitive und der affektive Aspekt des moralischen Verhaltens bilden eine Einheit und können theoretisch klar unterschieden, aber nicht dinglich voneinander getrennt werden.* Dieses Postulat ist zentral für die kognitive Entwicklungstheorie. Für Piaget (1976b) sind "Gefühlsleben und Erkenntnis [. . .] untrennbar, wenn auch unterschieden. Sie sind untrennbar, weil jeder Austausch zwischen Subjekt und Umwelt gleichzeitig eine

Strukturierung und eine Wertung voraussetzt, aber sie sind dennoch unterscheidbar, weil diese beiden Seiten des menschlichen Verhaltens nicht aufeinander zurückzuführen sind. Sogar in der Mathematik ist es unmöglich zu denken, ohne gleichzeitig bestimmte Gefühle zu empfinden, und umgekehrt gibt es kein Gefühlsleben ohne ein *Minimum* von Verständnis und Unterscheidung" (S. 8).[30]

Wenn Affekt und Kognition im moralischen Verhalten untrennbar sind, dann kann, wie Kohlberg (1958) feststellte, "eine moralische Handlung weder durch rein "kognitive" noch durch rein "motivationale" Kriterien definiert werden" (S. 16). In der Tatsache, dass dies häufig übersehen wurde, liegt nach Kohlberg der Hauptgrund dafür, dass sich kein systematischer Zusammenhang zwischen beiden Aspekten finden ließ. Die konzeptuelle Parallelität darf nicht mit der empirischen Hypothese verwechselt werden, dass zwischen beiden Aspekten eine *empirische* Beziehung wie zum Beispiel eine *Décalage* besteht. Décalage setzt die Definition von *separaten Komponenten* voraus; sie ist, wie Kohlberg (1977) betont, "die Verbreitung einer Struktur über Inhalte hinweg. Moralische Entwicklung ist aber nicht in sich selbst eine Anwendung einer kognitiven Struktur auf moralische statt auf kognitive Probleme und Inhalte" (S. 15). Ebensowenig darf, worauf Piaget hinwies, aus diesem Postulat der Nichttrennbarkeit gefolgert werden, dass beide Aspekte identisch oder aufeinander reduzierbar seien.

2. Bedeutung: Korrelative Parallelität

Die erste empirische Form der Parallelitätsthese wollen wir wie folgt formulieren: *Zwischen affektiven und kognitiven Aspekten des moralischen Handelns besteht eine enge, positive Korrelation*. Die korrelative Parallelität wird im Rahmen der kognitiven Entwicklungstheorie - zumeist implizit - als gegeben unterstellt. Sie wird vor allem in der Forschung vorausgesetzt, wo für beide Aspekte der kognitiv-moralischen Entwicklung gewöhnlich nur ein einziger Meßwert ermittelt wird, in den beide Aspekte vermischt eingehen. So gibt Kohlbergs *Moral Maturity Score* (MMS) zwei Aspekte wieder: a) den affektiven

30 Danach wäre es falsch, Piaget eindeutig zu den Befürwortern eines Komponenten-Modells zu zählen, wie Garz (1989, S. 112) dies tut.

Aspekt, d.h. welche moralischen Prinzipien, Einstellungen oder Motive eine Person als Gründe für eine Entscheidung am häufigsten akzeptiert oder selbst verwendet, und b) den kognitiven Aspekt, d.h., wie Kohlberg (1977) schreibt, den eine bestimmte Situation oder Reiz-Reaktionseinheit transzendierenden Aspekt des moralischen Denkens (S. 62-68). Einen hohen MMS-Wert erhält, wer a) Argumente einer *hohen* Stufe benutzt und b) dies konsistent (häufig) tut.[31] Ähnlich vermengt auch der P-Wert aus dem *Defining-Issues-Test* von Rest (1979) beide Aspekte. Der P-Wert zeigt gleichzeitig an, a) wie *stark* eine Person postkonventionelle moralische Argumente gewichtet (affektiver Aspekt) und b) wie *konsistent* sie diese Argumente präferiert (kognitiver Aspekt). Eine solche Vermischung des affektiven und des kognitiven Aspekts des moralischen Urteilens, die sich aus der Anwendung der klassischen Testtheorie erklärt, läßt sich nur rechtfertigen, wenn, was bislang nicht geprüft wurde, zwischen beiden Aspekten empirisch ein enger Zusammenhang besteht.

3. Bedeutung: Entwicklungsparallelität

Die zweite empirische Hypothese ist: *Der affektive und der kognitive Aspekt des moralischen Verhaltens stehen in einem engen entwicklungsmäßigen Zusamenhang, wobei die affektive Bindung an moralische Ideale eine Vorreiterfunktion hat für die Entwicklung des kognitiven Aspekts, d.h. der Integration, Differenzierung und Hierarchisierung des moralischen Urteilens (Entwicklungsparallelität*). Die korrelative Parallelität ist eine notwendige, aber keine hinreichende Bedingung für die Entwicklungsparallelität. Die Hypothese, dass beide Aspekte des moralischen Urteilens korrelieren, könnte sich sowohl auf die Annahme stützen, dass solche Unterschiede angeboren oder durch unterschiedliche soziale Milieus bedingt sind, als auch auf die Annahme, dass dieser Zusammenhang entwicklungsbedingt ist. Nach der kognitiven Entwicklungstheorie müßte man letzteres vermuten, nämlich, a) dass moralische Entwicklung nicht nur eine affektive, sondern auch eine kognitive Grundlage hat und b) dass der Erwerb einer affektiven Bindung an moralische Prinzipien eine notwendige Voraussetzung für die Entwicklung der kognitiven Struktur des moralischen

31 Für die sehr am MJI orientierten Meßverfahren von John Gibbs gelten dieselben Einwände (Gibbs et al. 1982).

Verhaltens ist. Es scheint, so Kohlberg (1984), dass "die aktuelle Verwendung der Stufen [. . .] durch zwei gegensätzliche sequentielle Ordnungen diktiert wird – die der Präferenz und die der Leichtigkeit –, wobei die Modalstufe des Individuums die bevorzugte Stufe darstellt, welche es ohne Schwierigkeiten benutzen kann" (S. 60).

Die Annahme einer Entwicklungsparallelität zwischen Affekt und Kognition wird durch den Befund gestützt, dass Heranwachsende höhere Stufen der moralischen Argumentation in der Regel schon früh bevorzugen, aber erst in einem viel späteren Alter fähig sind, diese Argumente selbst in einer Diskussion moralischer Dilemmas anzuwenden. Es scheint, dass moralische Werte und Ideen eine doppelte Funktion haben: a) sie geben dem aktuellen Verhalten eine Orientierung und b) sie tragen längerfristig dazu bei, die kognitiven Strukturen zu entwickeln, die notwendig sind, um divergierende Orientierungen in die eigene Persönlichkeit zu integrieren und um situationsangemessen bzw. differenziert zu handeln. Nur wenn das Individuum eine affektive Bindung an moralische Prinzipien besitzt, ist es für die moralische Problematik einer Entscheidung sensibel und kann Hinweise auf die moralischen Implikationen seiner Handlung richtig verstehen. Ein Prozeß der moralischen Reflexion kann nur in Gang kommen, wenn objektive moralische Konflikte auch subjektiv erlebt und verstanden werden.

Nunner-Winkler (1989a) hat auf der Basis ihrer Untersuchungen zum moralischen Urteil von Schulkindern gegen die These der Entwicklungsparallelität eingewendet, dass sich die affektive oder motivationale Komponente später entwickelt als die kognitive. Somit unterscheidet sie zwei Phasen der Entwicklung moralischer Konzepte:

Phase (1): "Schon früh erwirbt das Kind ein Wissen um die Geltung moralischer Regeln und ein angemessenes Verständnis ihrer Begründung: Nicht externe Sanktionen, sondern das Unterlassen von Unrecht und die Linderung von Not sind Geltungsgrund moralischer Normen. Allerdings nimmt das Kind Moral zunächst nur als bloßes Faktum zur Kenntnis – Turiel spricht von 'informational knowledge' [. . .]. Von Anfang an haben die Kinder ein genuin moralisches Verständnis der Geltung von Normen [. . .]."

Phase (2): "Erst in einem zweiten Lernschritt, der etwa zwischen dem 6. und 8. Lebensjahr vollzogen wird, beginnt das Kind dann, sich diese Regeln wirklich zu eigen zu machen - in psychoanalytischer Terminologie: zu verinnerlichen [. . .]. Erst dann [. . .] begreift [es], dass man moralische Regeln befolgen

wollen soll [. . .]. Dieser zweite moralische Lernprozeß bezieht sich auf moralisches Handeln im Kantischen Sinn: auf Handeln ‘nicht aus Neigung, sondern aus Pflicht’[. . .]. Dafür genügt es nicht, Gutes zu tun, wenn man dies will; vielmehr gilt es Gutes zu tun und Böses zu unterlassen, gerade auch dann, wenn man dies nicht spontan will [. . .]. Dies zu begreifen, ist jedoch ein mühsamer, ein schmerzlicher Lernprozeß” (S. 593-597).

Gegen diese Theorie gibt es aus der Sicht des Zwei-Aspekte-Ansatzes zwei Einwände: 1. Nunner-Winkler übersieht, dass sich der kognitive und affektive Aspekt operational nicht trennen lassen. 2. Die Verbindung zweier inkommensurabler Theoriensysteme wie das der kognitiven Entwicklungstheorie und das der Psychoanalyse bedingt eine Vieldeutigkeit wichtiger Begriffe. Als “Kognition” wird bei Nunner-Winkler das bezeichnet, was im Zwei-Aspekte-Modell als Einstellung oder Affekt etikettiert wird. Zwar fällt das Kennen von Normbegriffen unter die kognitiven Fertigkeiten, aber es wäre verfehlt, dieses Kennen bereits als (voll entwickelte) Kompetenz oder Kognition zu bezeichnen. “Kognitive Strukturen sind”, wie Kohlberg (1984) betont, “immer Strukturen (Schemata) von *Handlungen*” (S. 8). Diese Kennzeichnung wäre also nur gerechtfertigt, wenn sich nachweisen ließe, dass die Kinder bereits eine voll entwickelte moralische Urteilsfähigkeit besitzen, also dass sie a) diese Norm konsistent, auch gegen Widerstände oder Ablenkungsversuche einzuhalten versuchen, und b) sie diese Norm nicht rigide, ohne Rücksicht auf andere ebenso wichtige oder sogar wichtigere moralische Normen befolgen, sondern nach allgemeinen Prinzipien der Gerechtigkeit, Verhältnismäßigkeit und Humanität differenzieren.

Auch die zweite Phase in Nunner-Winklers Modell müßte vom kognitiv-entwicklungstheoretischen Standpunkt aus anders beschrieben werden. Es handelt sich danach nicht um eine Phase der Willensentwicklung, sondern um die Phase der Entwicklung genau jener kognitiven Fähigkeiten, die wir eben als Ausdruck kompetenter moralischer Urteilsfähigkeit beschrieben haben. Mit der kognitiven Entwicklungstheorie nehmen wir an, dass die Übersetzung von normativen Überzeugungen in Handlungen nicht durch mangelndes Wollen behindert wird, sondern durch einen Mangel an kognitiver Kompetenz, die von den Heranwachsenden erst nach und nach (mit äußerer Lernunterstützung) erworben wird.

Die Kompetenz-Performanz-Theorie trifft die entgegengesetzte Annahme, nämlich dass die moralische Kompetenz schon früh vorhanden ist und nur durch die

Entwicklung eines entsprechenden Willens zur Entfaltung im Handeln (Performanz) gebracht werden muß.[32] Sie ist aus der Perspektive des Zwei-Aspekte-Modells nicht haltbar: Sie steht in diametralem Gegensatz zur kognitiven Theorie der Moralentwicklung, auch wenn Kohlberg ihr gelegentlich zugeneigt hat. Sie ist empirisch bislang kaum belegt, sieht man von einigen hermeneutischen Fallstudien ab, die von den Befürwortern dieses Ansatzes eher illustrierend als beweisend berichtet werden. Und sie führt zum Teil zu unauflösbaren inneren Widersprüchen, da sie zwischen dem eigenen Willen der Person, der spontan und amoralisch ist, und einem moralischen Willen unterscheidet, den sich die Person im Laufe der Entwicklung zu eigen macht (oder machen soll). Macht die Person sich den moralischen Willen zu eigen, dann wird dieser definitionsgemäß zum eigenen, spontanen Willen; dieser aber ist amoralisch, weil eigen. Dieses Begriffsproblem mündet, wie Portele ausgeführt hat, in ein nichtlösbares Paradoxon der Moralerziehung: "Wie macht man das, dass [Kinder] lernen, zu wollen, was sie sollen?" (1978, S. 147). Das Kompetenz-Performanz-Modell läßt nur zwei Erziehungsmethoden als angemessen erscheinen: Autoritärer Zwang (Du sollst wollen!) oder therapeutische Behandlung (das Kind ist krank).[33] Im Rahmen der kognitiven Entwicklungstheorie kann dieses Paradoxon aufgelöst werden. Piaget und Kohlberg nehmen an, dass die Verinnerlichung des "sozialen Willens" am Anfang der Entwicklung steht (Phase der Imitation und Heteronomie) und nicht am Ende des Entwicklungsprozesses. Davon gehen auch neuere Ansätze in der Psychoanalyse aus, die als wesentliche Aufgabe der Entwicklung die Transformation

32 Döbert & Nunner-Winkler (1978) sowie Nunner-Winkler (1989a) haben diese Annahme auf den Punkt gebracht, sind aber keineswegs die einzigen Vertreter. Die Kompetenz-Performanz-Theorie wird inzwischen von vielen Autoren zitiert oder gar übernommen.

33 Die erste Alternative wird kaum vertreten, da sie allzu sehr gegen die liberale pädagogische Grundauffassung der meisten Autoren verstößt; sie ist aber nicht von der Hand zu weisen, wenn man sagt, was das Kind am moralischen Handeln hindere, sei nicht seine Unfähigkeit, sondern sein mangelnder Wille. Aber der Verweis auf "schmerzliche Prozesse" deutet an, dass Zwangsmaßnahmen in der Moralerziehung aufgrund der Theorie zumindest als verständlich angesehen werden. Wegen der Nähe vieler Vertreter der Kompetenz-Performanz-Theorie zur Psychoanalyse bietet es sich zumeist an, auf therapeutische Maßnahmen zu verweisen, um, wie man vermutet, jene vielfältigen Abwehrmechanismen auszuräumen, die Menschen am Moralisch-Sein hindern.

des (sozialen) Über-Ichs beim Kind in ein reifes Ich beim Erwachsenen ansehen.[34]

2.4 Eine Definition von Affekt und Kognition als Aspekte des moralischen Verhaltens

Nach dieser Diskussion konkreter Hypothesen können wir erneut versuchen, die kognitive Entwicklungstheorie so zu rekonstruieren, dass eine widerspruchsfreie und mit den empirischen Befunden weitgehend verträgliche Definition und Messung der für sie zentralen Begriffe Affekt und Kognition möglich ist. Dies ist die Voraussetzung für eine adäquate Meßmethode der Moralforschung und für die experimentelle Untersuchung der These einer korrelativen Parallelität und Entwicklungsparallelität.

2.4.1 Der affektive Aspekt der Moral

Nach Piaget (1951) drücken "die Affekte das Interesse und die Werte des Handelns aus, dessen Struktur die Intelligenz ist" (S. 221). In dieser Charakterisierung definieren wir, auch in Übereinstimmung mit der sozialpsychologischen Einstellungstheorie, den *affektiven Aspekt* des moralischen Verhaltens *als die im Verhalten sich zeigende, gefühlsmäßige Bindung an moralische Maximen, Regeln oder Prinzipien*. Affekte meinen ganz allgemein Verhaltensorientierungen, angefangen von sehr einfachen Gefühlen (*sentiments fundamentaux*) bis hin zu moralischen Motiven und Gerechtigkeitsvorstellungen (Boesch 1984, S. 174; Piaget 1976, S. 7). Ähnlich definiert die Einstellungstheorie den affektiven Aspekt als die Richtung und Stärke einer gefühlsmäßigen Reaktion gegenüber einem sozialen Gegenstand wie z.B. gegenüber einem moralischen Prinzip.[35] Beispielsweise wird der affektive Aspekt der Einstellung einer Person zum Prinzip "Aufrechterhaltung von Recht und Ordnung einer Gesellschaft" operationalisiert als der Grad der positiven oder negativen Reaktion auf ein

34 Siehe Heinz Hartmann 1972 und Mowrer 1972.

35 Allport 1935; Fishbein & Ajzen 1975; Krathwohl et al. 1964; Shaw & Wright 1967.

Argument, das auf dieses Prinzip Bezug nimmt, oder durch die Häufigkeit, mit der dieses Prinzip angeführt wird, um Handlungen zu rechtfertigen.[36]
Unter der Vielzahl von moralischen Regeln oder Prinzipien, an denen eine Person ihr Verhalten in konkreten Situationen orientiert, sind, wie Kohlberg und seine Mitarbeiter nachgewiesen haben, sechs Typen oder Stufen entwicklungstheoretisch von besonderer Bedeutung. Sie charakterisieren den Rahmen, innerhalb dessen eine Person in einem bestimmten Abschnitt ihrer Entwicklung eine ethisch begründbare, gerechte, allgemeingültige und logisch konsistente Lösung für Handlungskonflikte zu finden versucht. Nach Kohlberg (1984) ist jede der sechs Stufen durch bestimmte moralische Affekte oder Motive charakterisiert (siehe Abb. 2.2). Die Ordnungsfolge der Moralstufen wird normativ und empirisch begründet. Als normative Begründung kann angeführt werden, a) dass die "höheren" Stufen moralphilosophisch adäquater sind (Kohlberg 1971) und b) dass sie von den meisten Menschen präferiert werden, und zwar unabhängig von ihrem kulturellen Hintergrund (Lind 1984; auch Kapitel 7). Auf die empirische Begründung, dass die höheren Stufen schwieriger im konkreten Verhalten anzuwenden sind, und dass ihnen daher auch eine "kognitive" Ordnung zugrunde liegt, wird im Zusammenhang mit der Überprüfung der These einer korrelativen Parallelität noch einzugehen sein.

2.4.2 Der kognitive Aspekt

Der kognitive Aspekt des moralischen Verhaltens ist nach Kohlberg die Gesamtheit der "Beziehungen zwischen Ideen und Inhalten im Denken eines Individuums" (1977, S. 9). Er ist definiert als *Struktur*, *Muster* oder *Organisation* des moralischen Verhaltens in bezug auf bestimmte moralische Motive oder

36 Hier könnte eingewendet werden, dass "Richtung und Stärke einer gefühlsmäßigen Reaktion" nicht dasselbe ist, wie "der Grad der positiven oder negativen Reaktion". Das eine beschreibt einen inneren Zustand, das andere ein äußeres Verhalten. Es könnte aber auch eingewendet werden, dass das erste in einer "behavioristischen" Sprache dasselbe ausdrückt wie das zweite in einer "mentalistischen". In der Tat gebrauche ich beide Ausdrucksweisen so, dass sie ausgetauscht werden könnten. Es erscheint mir müßig, darüber zu streiten, welche Verbalisierung die wissenschaftlichere ist, wenn man sich darüber im klaren ist, dass (a) wir hier immer nur über (prinzipiell) Beobachtbares sprechen und (b) mentale Begriffe unabdingbar sind, um menschliches Verhalten adäquat beschreiben zu können.

Stufe 1	Das Motiv, eigene körperliche Schäden und Verletzungen ('Strafe') zu vermeiden.
Stufe 2	Das Motiv, Vorteile und Vergünstigungen ('Belohnungen') zu erlangen, auch wenn dabei gelegentlich Nachteile in Kauf genommen werden müssen.
Stufe 3	Das Motiv, Mißbilligung durch andere zu vermeiden bzw. ihre Anerkennung zu erreichen.
Stufe 4	Das Motiv, die Gesetze und Ordnung der Gesellschaft zu achten und zu ihrer Aufrechterhaltung beizutragen. Affektiv-kognitive Parallelität: Korrelationen zwischen der moralischen Urteilsfähigkeit (MUT) und den Einstellungen zu den sechs Moralstufen.
Stufe 5	Das Motiv, die Regeln freier, demokratischer Willensbildung (Mehrheitsprinzip etc.) zu beachten und die auf ihnen gründenden Verträge einzuhalten.
Stufe 6	Das Motiv, die Würde jedes einzelnen Menschen zu achten und sein Handeln an den universellen Prinzipien der Gerechtigkeit, Vernunft und Logik zu messen.

Abb. 2-2 Der affektive Aspekt: Stufenmodell moralischer Motive nach Kohlberg

Prinzipien. Diese Definition von Kognition als *Prozeß* ist kennzeichnend für die kognitive Entwicklungstheorie, gegenüber Ansätzen, welche die Kognition bloß als Wahrnehmung des Handelnden definieren. Nach Piaget und Kohlberg sind Kognitionen die Schemata des Verhaltens. “Die Strukturierung des Verhaltens bildet seine erkenntnismäßige Seite” (Piaget 1976, S. 8). Entwicklung im strukturellen Sinne meint dann also das Entstehen eines qualitativ neuen Verhaltens *(mode of behavior)*, “das neu ist hinsichtlich seiner Form oder Organisation und nicht einfach hinsichtlich der Elemente oder Informationen, die es beinhaltet” (S. 4).

Wie der affektive Aspekt ist auch der kognitive Aspekt keine Komponente, die vom Verhalten losgelöst ist, und auch keine latente Eigenschaft, die nicht direkt beobachtbar ist; vielmehr ist er das “beobachtbare Muster des Denkens”, das

sich in der Struktur und Organisation des moralischen Argumentierens manifestiert. Der kognitive Aspekt ist definitionsgemäß keine Ursache für moralisches Verhalten, das unabhängig von diesem Aspekt gedacht und erfaßt werden kann; er *ist* die Struktur des Urteilsprozesses, der sich im Verhalten ausdrückt.

Die kognitiv-strukturellen Eigenschaften des moralischen Verhaltens sind näher bestimmt durch die formalen Kriterien eines adäquaten moralischen Urteils; diese sind Universalität, innere Konsistenz und Unparteilichkeit (Kohlberg 1977, S. 71). Die kognitive Entwicklung der Moralität umfaßt die Integration und Differenzierung moralischer Einstellungen gleichermaßen. Konsistentes moralisches Verhalten meint nicht nur konsistentes Verhalten in bezug auf *ein* bestimmtes Prinzip (das wäre rigides Verhalten), sondern Verhalten, das alle in einer Situation relevanten Prinzipien zu berücksichtigen und zu integrieren versucht. Beispielsweise würde man von einer integrierten Einstellung gegenüber "Recht und Ordnung" nur dann sprechen, wenn dieses Prinzip konsistent auf eigenes wie auf fremdes Handeln angewendet wird und wenn es nicht nur "zufällig" mit der eigenen Meinung übereinstimmt. Differenziertes Urteilen ist notwendig, a) wenn z.B. eine Situation das Prinzip "Recht und Ordnung" gar nicht tangiert und Überlegungen auf "niedrigeren" Stufen angemessener sind, und b) wenn Recht und Ordnung den übergeordneten Prinzipien der Gerechtigkeit und Humanität widersprechen und daher ein Urteilen auf "höheren" moralischen Stufen erforderlich ist.

Bei der Entwicklung des kognitiven Aspekts lassen sich in Anlehnung an Piaget und Kohlberg drei Phasen unterscheiden: die Phase der Imitation, der Heteronomie und der Autonomie (siehe Abb. 2.3).[37] Es ist bemerkenswert, dass diese Phasen bei Piaget keine Gesamtstadien der Entwicklung des individuellen Denkens bezeichnen, sondern Stadien der strukturellen Entwicklung in bezug auf jede der sechs Stufen moralischer Affekte. Diese Phasen der kognitiv-moralischen Entwicklung sind Piaget (1973a) zufolge "als aufeinander folgende Phasen regelmäßiger Prozesse aufzufassen, die sich wie Rhythmen auf den übereinander liegenden Gebieten des Verhaltens und des Bewußtseins wiederholen" (S. 91).

37 Siehe vor allem Piaget (1973a) und Kohlberg (1984, S. 652 ff.).

1. *Moralische Imitation*

In dieser Phase wird das Verhalten in einer konkreten Situation von der Imitation und dem allgemeinen Brauch bestimmt. Das Kind ahmt moralisches Verhalten nach, ohne es voll zu verstehen. Auch wenn ihm moralische Prinzipien bekannt sind, sieht es keinen inneren Zusammenhang zwischen diesen und seinen konkreten Handlungen. Es fühlt keine Notwendigkeit, zwischen seinem gewohnheitsbestimmten Handeln und den moralischen Prinzipien, die es kennt und zuweilen auch präferiert und benutzt, eine Übereinstimmung herzustellen. Das, wie Dewey (1976/1898) es nannte, "habituelle Selbst" als das Produkt vergangener Erfahrungen, wird nicht mit dem "idealen Selbst", d.h. dem Ergebnis des reflektierenden Denkens, in Verbindung gebracht. In der Terminologie der Psychoanalyse könnte man auch sagen, Es und Über-Ich dominieren; das Ich, das die Integration beider leisten soll, spielt noch keine Rolle.

2. Moralische Heteronomie

In dieser Phase entsteht ein hohes Verantwortungsgefühl gegenüber der Einhaltung von Regeln, Ideen oder Prinzipien, die in einer bestimmten Lebenswelt gelten. Denken und Verhalten werden in fast monomanischer Weise von diesen bestimmt und tendieren daher zu Rigidität, Dogmatismus und im Extremfall zu Fanatismus. Das moralische Urteil in dieser Phase ist "situationsgebunden" und "normkonform" (Kohlberg 1977, S.68). In dieser Phase versucht das Individuum, eine Übereinstimmung zwischen seinem Handeln und konkurrierenden Prinzipien dadurch zu erreichen, dass es diese Prinzipien instrumentalisiert, d.h. zur nachträglichen Rechtfertigung statt zur kritischen Überprüfung seines Handelns heranzieht. Eine solche "Rationalisierung" des habituellen Verhaltens ist, wie Mowrer (1972) feststellte, charakteristisch für die Phase, in der Es und Über-Ich stark ausgeprägt sind, aber die vermittelnden Strukturen des Ichs noch wenig entwickelt sind.[38] Diese Phase ist durch hohe "Meinungskonformität" des Urteilens charakterisiert; d.h. moralische Argumente werden vorwiegend danach beurteilt, ob sie die eigene Meinung zu einer konkreten Entscheidung stützen oder ihr widersprechen (Keasey 1975). Piaget spricht hier von der Unfähigkeit der Person, zu "dezentrieren".

3. *Moralische Autonomie*

In dieser Phase gewinnt das Individuum eine reflektierende Distanz zu einzelnen Prinzipien und zur eigenen Meinung. Das moralische Argumentieren transzendiert nun isolierte Prinzipien und wird mehr "ausgleichend" (Kohlberg 1977, S. 68). Das Individuum wird fähig, unterschiedliche Prinzipien in einer Situation wahrzunehmen, zu

38 Für eine Diskussion der Verbindung von psychoanalytischer und kognitiv-entwicklungstheoretischer Perspektive vgl. Lind et al. (1985).

differenzieren und im Falle eines Konflikts zwischen ihnen zu entscheiden (hierarchische Integration). Moralische Prinzipien werden nicht mehr als extern erlebt, sondern als Teil der eigenen moralischen Identität. An die Stelle von Rationalisierung tritt Rationalität und Ich-Stärke. Eine Übereinstimmung zwischen rationaler Einsicht und habituellem Verhalten wird dadurch zu erreichen versucht, dass das gewohnheitsmäßige Handeln kritisch reflektiert und gegebenenfalls auch korrigiert wird. Ein wichtiges Kriterium für das Erreichen moralischer Autonomie ist die "moralische Determiniertheit" des Urteilsverhaltens. Darunter verstehen wir das Ausmaß, in dem eine Person bereit ist, auch moralische Argumente zu bedenken, die ihrer intuitiv gefaßten Meinung entgegenstehen, und ihre ursprüngliche Meinung dann zu revidieren, wenn sie den als wichtig erkannten moralischen Prinzipien widerspricht. In moralphilosophischer Sicht bedeutet diese Phase den Eintritt in die "ideale Argumentationsgemeinschaft". Sie ist Voraussetzung für konfliktauflösende Diskurse in werte-pluralistischen Gesellschaften.[39]

Abb. 2.3 Der kognitive Aspekt des moralischen Urteilsverhaltens: Phasen der Entwicklung

2.4.3 Die Definition moralischen Verhaltens im Zwei-Aspekte-Modell

Über die Definition des moralischen *Verhaltens* herrscht in den Sozialwissenschaften fundamentaler Dissens. Gemeint sind hiermit nicht Meinungsunterschiede bezüglich der Frage, ob ein bestimmtes Verhalten, zum Beispiel "Schwarzfahren" moralisch oder unmoralisch zu nennen ist, sondern die viel grundlegender Frage, ob bei der Definition – und der Messung – des moralischen Entwicklungsgrades einer Handlung die *individuellen* Absichten und Urteilsstrukturen des Handelnden zu berücksichtigen sind, oder nicht. Die Bedeutung dieser Frage wird am schärfsten von Pittel und Mendelsohn (1966) gesehen, die die "größte Unzulänglichkeit" und die Stagnation der Moralforschung auf das Versagen zurückführten, "wertende Einstellungen als subjektive Phänomene zu begreifen, deren Messung am besten unabhängig erfolgt von der

39 Apel 1990, S. 164, 284. Foerster (1918) stellte bereits die große Bedeutung des Übergangs von der Meinungskonformität zur autonomen Prinzipienorientierung für die Entwicklung von Demokratiekompetenz heraus: "Gerechtigkeit wird nur dann konkret verstanden, wenn man jene erste, elementarste Aufgabe des Staatsbürgers, das Nebeneinander von entgegengesetzt Denkenden und Wollenden sittlich zu organisieren, gründlich erfaßt und in die Selbsterziehung aufgenommen hat" (S. 14).

Sorge um die Beziehung dieser Einstellung zu konventionellen, normativen Standards moralischer Bewertung" (S. 34).
Oft besteht der Dissens nicht so sehr zwischen verschiedenen Autoren, sondern zwischen verschiedenen Rollen der Autoren selbst. Als Theoretiker sehen sie die Berücksichtigung individueller Absichten oft als unabdingbar an, als empirische Forscher erscheint ihnen diese Annahme oft nur als hinderlich. In dieser Rolle messen sie die Moralität eines Verhaltens zumeist an einem konventionellen Außenkriterium: Wird eine bestimmte sozial, religiös oder rechtlich festgelegte Norm verletzt, dann wird das Verhalten als "unmoralisch" deklariert. Dazu drei Beispiele:

1. Beispiel: Bezeichnenderweise erst am Ende ihrer Arbeit über (un-)moralisches Verhalten bei Kindern stellen Hartshorne und May (1928) fest, dass "die Essenz einer Handlung ihre Absicht ist.[40] [. . .] Es ist nicht die Handlung, die Betrug darstellt, sondern die *Beziehung* dieser Handlung zu den Absichten der Person und den Absichten ihrer Mithandelnden" (S. 377). Leider bleibt diese Einsicht bei Hartshorne und May (1928) ohne Konsequenz für ihre Meßmethoden und spielt auch keine Rolle für die Diskussion ihrer (vielbeachteten) Ergebnisse. In ihrer experimentellen Forschung wurde die Moralität des Verhaltens von Kindern allein danach bestimmt, ob es mit äußeren Kriterien übereinstimmte oder nicht. "Subjektive" und situative Momente des Verhaltens wurden bewusst aus der Analyse ausgeschlossen, weil, so die Autoren, "es keinen Fortschritt geben kann, solange nicht die offene Handlung beobachtet und, wenn möglich, gemessen werden kann, ohne Bezug zu nehmen auf ihre Motive" (S. 11).
2. Beispiel: Auch die Forschung, die sich des *Moral Judgment Interviews* bedient, zieht – entgegen der ihr zugrundeliegenden Theorie – zur Bestimmung der Moralität des (Urteils-)Verhaltens von außen gesetzte Regeln heran, statt individueller Motive und Prinzipien (s. Kapitel 3 und 7). Wie wir bereits in Kapitel 1 sahen, definiert Kohlberg (1964) moralische Urteilsfähigkeit als „das Vermögen, Entscheidungen und Urteile zu treffen, die moralisch sind (also auf moralischen Prinzipien gründen), und in Übereinstimmung mit diesen Urteilen

40 Im Original wird der innere Zusammenhang zwischen beiden lautmalerisch hervorgehoben: "The essence of an act is ist pretense."

zu handeln" (S. 425).[41] Diese Definition ist "universalisierbar" und nicht an einen bestimmten sozialen Standpunkt gebunden. Folgt man ihr, so muß die Moralität des Verhalten einer Person in ihrer Gesamtheit und in bezug auf ihre *eigenen* Prinzipien beurteilt und gemessen werden. Tatsächlich bedienen sich Kohlberg und seine Schüler in der Forschung einer Meßmethode, des *Moral Judgment Interviews* (MJI), in der die moralische Reife des Urteilsverhaltens an äußeren Kriterien, nämlich an dem Stufenmodell Kohlbergs, gemessen wird und nicht an dem Grad, mit dem die Person nach ihren eigenen moralischen Prinzipien handelt.

3. Beispiel: Ein Dauerthema der letzten Jahre ist das Verhältnis von moralischem Urteilen und moralischem Verhalten (Garz et al., 1999). Hierunter wird, wenn man genau hinsieht, die Frage verstanden, wie zwei methodisch verschiedene, moralische Einschätzungen des individuellen Verhaltens miteinander korrelieren, nämlich die Einschätzung mittels des MJI – oder eines anderen psychologischen Tests – und die Einschätzung mittels einfacher konventioneller Kriterien für Moralität. Wie hoch, so fragt beispielsweise Blasi (1980) in seinem Übersichtsartikel, ist der Zusammenhang zwischen den MJI-Werten, die Personen erreichen, und dem Maß an "Ehrlichkeit", "Hilfsbereitschaft" usw., das ihrem Verhalten von *außen*, ohne jeden Bezug zu Absichten und situativem Kontext, *zugeschrieben* wird. Nach dem Verständnis von Hartshorne und May ebenso wie dem von Kohlberg handelt es sich hier *nicht* um *moralisches* Verhalten, sondern um *normkonformes* Verhalten. Normkonformes Verhalten kann auch moralisch sein, muß es aber nicht. Demnach wird in der Forschung regelmäßig etwas anderes untersucht, als die Fragestellung impliziert. Dies mag daran liegen, dass es zumeist relativ einfach ist, festzustellen, ob ein Verhalten konform mit bestimmten Erwartungen ist, aber oft sehr schwer und manchmal sogar unmöglich, moralisches Verhalten oder Urteilen adäquat zu erfassen. Um Konformität mit sozialen Normen festzustellen, genügt es zumeist, einzelnes Verhalten zu beobachten. Die Ausprägung von Motiven und Urteilsstrukturen dagegen kann nur durch die systematische Beobachtung von ganzen Verhaltens*mustern* eindeutig und valide bestimmt werden.

Aber vielleicht liegt dem Urteils-Verhaltens-Problem noch ein weiteres Mißverständnis zugrunde. Auch Urteilen ist Verhalten. Obwohl dafür prinzipiell ver-

41 Siehe auch Kohlberg 1971; Eckensberger 1983.

schiedene Verhaltensklassen in Frage kommen, haben Piaget und Kohlberg in ihrer Forschung eine spezielle Verhaltensklasse, das Urteilsverhalten einer Person als Informationsquelle bevorzugt. Moralisches Urteilsverhalten meint, wie wir bereits in Kapitel 1 gesehen haben, keinen inneren, nicht beobachtbaren Prozeß, sondern beobachtbares Verhalten, das affektive und kognitive Eigenschaften hat. Moralisches Urteilen meint auch nicht ein Urteilen im "leeren Raum", sondern, wie schon angedeutet, die *Interaktion* einer Person mit moralisch relevanten, sozialen Objekten. Diese Interaktion zeigt sich ebenso im "spontanem" wie im "reaktivem" Verhalten einer Person. Die Messung oder Beobachtung selbst ist, wie Piaget (1973b) bemerkte, das "Ergebnis einer Interaktion zwischen Experiment und Realität" (S. 52). Der Psychologe muß mit einer Person interagieren, um ihr Verhalten wirklich zu verstehen.[42]

Diese Klärung der theoretischen Definition von "moralischem Verhalten" führt zu dem Zwei-Aspekte-Modell, das in Abb. 2.4 in seinen wesentlichen Teilen dargestellt ist. Es zeigt schematisch die Beziehung von Affekt, kognitiver Struktur und Verhalten, wobei die Darstellung von *zwei* moralischen Affekten nur andeuten soll, dass Verhalten in der Regel durch *mehrere* Motive gleichzeitig determiniert ist, dass also ein und derselbe Verhaltensakt auf mehrere Dispositionen hinweisen kann und daher selten eindeutige Schlüsse auf Inhalt und Struktur des moralischen Verhaltens zuläßt. Der Affekt *A* kann zum Beispiel für die Orientierung an den oben typologisierten, moralischen Motiven stehen, und Affekt *B* für das Motiv, an einer intuitiven Meinung zu einem Problem festzuhalten. Diese beiden Motivgruppen sind im MUT operationalisiert (Kapitel 4). Der *kognitive Aspekt* ist in diesem Modell als eine *relationale Struktur* definiert, und zwar als die Beziehungsgesamtheit zwischen verschiedenen Urteilshandlungen in bezug auf die Motive, an denen sie sich orientieren. Struktur ist in diesem Modell also kein weiteres Element, sondern die Beziehung zwischen Elementen. Während der affektive Aspekt in der Terminologie der Einstellungstheorie einen Inhalt, nämlich Richtung und Stärke der handlungsleitenden Einstellung meint, bedeutet der kognitive Aspekt den Grad von Konsistenz und Differenziertheit, mit welchem der jeweilige Affekt das Urteilsverhalten determiniert.

42 Siehe dazu schon Heider (1958, S. 39); über die Einbeziehung von Situationen in psychologische Messungen vgl. Wakenhut (1978).

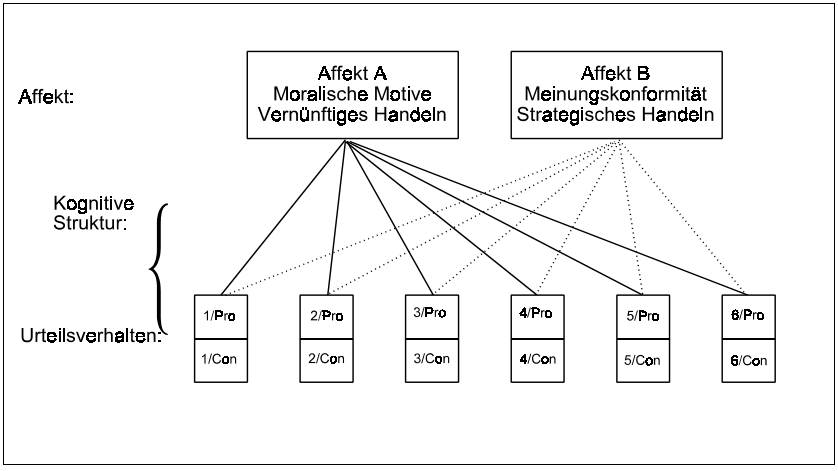


Abb. 2.4 Schematische Darstellung der Beziehungen zwischen Affekt, Kognition und Verhalten im Zwei-Aspekte-Modell

Das Zwei-Aspekte-Modell ist nicht begrenzt auf moralisches Verhalten, dessen Motive und Strukturen dem Handelnden *bewusst* sind. Eine solche Eingrenzung wäre nicht vertretbar. Wesentlich ist nur, dass das Urteilen einer Person tatsächlich von moralischen Prinzipien determiniert ist. Um festzustellen, ob und wie stark das Urteilsverhalten einer Person von *moralischen* Motiven determiniert ist, wird man fast immer auf verbales Material zurückgreifen müssen. Zur Gewinnung geeigneter (verbaler oder non-verbaler) Verhaltensdaten für die Einschätzung der moralischen Determiniertheit einer Person kann man, wie es Kohlberg und andere tun, die Person auffordern, ihre Beweggründe für ihr Verhalten und ihre Meinungen zu artikulieren. Der Forscher muß dann anhand eines detaillierten Auswertungsmanual entscheiden, wie der Inhalt und die Struktur dieser Begründungen auf der Meßskala einzuordnen sind (Colby, Kohlberg et al. 1987; Kapitel 3). Man kann der Person aber auch verbalisierte Begründungen mit der Aufforderung vorlegen, diese zu bewerten; auch daran läßt sich messen, inwieweit ihr Urteilsverhalten von (bestimmten) moralischen Überlegungen bestimmt ist. Ausgezeichnete Beispiele dafür, wie aus der reinen Beobachtung des verbalen Verhaltens von Kindern in genau festgelegten Situationen valide Schlüsse über ihr moralisches Urteil gezogen werden können, ohne dass die Kinder selbst Begründungen angeben, liefern die Studien von Piaget (1973a) und Nunner-Winkler (1989a). Eine ganz andere Frage ist,

Untersuchungsgruppe	Quelle	N
Studierende, 1. Semester (BRD 1977)	Forschungsgruppe Hochschulsozialisation	1288
Studierende, 5. Semester (BRD 1979)	dto.	812
Abiturienten (BRD, 1976)	dto.	516
Berufsschüler (CH)	HASMU-Projekt (Oser)	579
Straftäter (BRD)	Wischka (1980)	58
Berufsschüler (BRD 1979)	Heidbrink (1985)	256
Gymnasiasten, 8.-13. Klasse (BRD 1990)	Seitz (1991)	218

Anmerkung: Mit zwei Ausnahmen haben alle anderen Untersuchungen die Standardversion des MUT von 1977 verwendet. In der Abiturientenstudie wurde die Vorversion eingesetzt; in der Straftäter-Studie wurden die Antwortvorgaben vereinfacht (5er- statt 9er-Skala). Differenzen zu den in den Originalstudien berichteten Gruppengrößen sind durch fehlende Angaben bedingt. Es werden hier nur die Befragten gezählt, die den MUT ausgefüllt haben.

Abb. 2.5 Studien zur kognitiv-affektiven Parallelität des moralischen Urteilens

welche Rolle die Selbsteinschätzung von Motiven und Strukturen des moralischen Verhaltens einer Person für die Moralpsychologie spielt. Diese Selbsteinschätzung könnte nur dann als Grundlage der Analyse benutzt werden, wenn zwischen ihr und der empirisch fundierten Diagnose eine perfekte Übereinstimmung bestünde. Laien können, so scheint es, die MJI-Werte anderer Personen oft recht gut einschätzen.[43] Ihre Fähigkeit zur Einschätzung der eigenen MJI-Werte ist dagegen sehr begrenzt.[44] Noch schwerer scheint es Befragten zu fallen, ihre MUT-Werte einzuschätzen, da sie sich hierbei nicht an

43 Kohlberg (1984/1969, S. 69) berichtet immerhin von einer Korrelation in Höhe von $r = .58$ zwischen (alten) MJI-Werten und der Einschätzung derselben Personen durch *Peers*, gibt jedoch keine Details zu der Untersuchung an. Vgl. auch Kohlberg 1958, S. 75; Peck & Havighurst 1962, S. 183; Keasey 1971.

44 Arbuthnot (1979) berichtet von nur 17% richtiger Einstufungen durch die Befragten selbst.

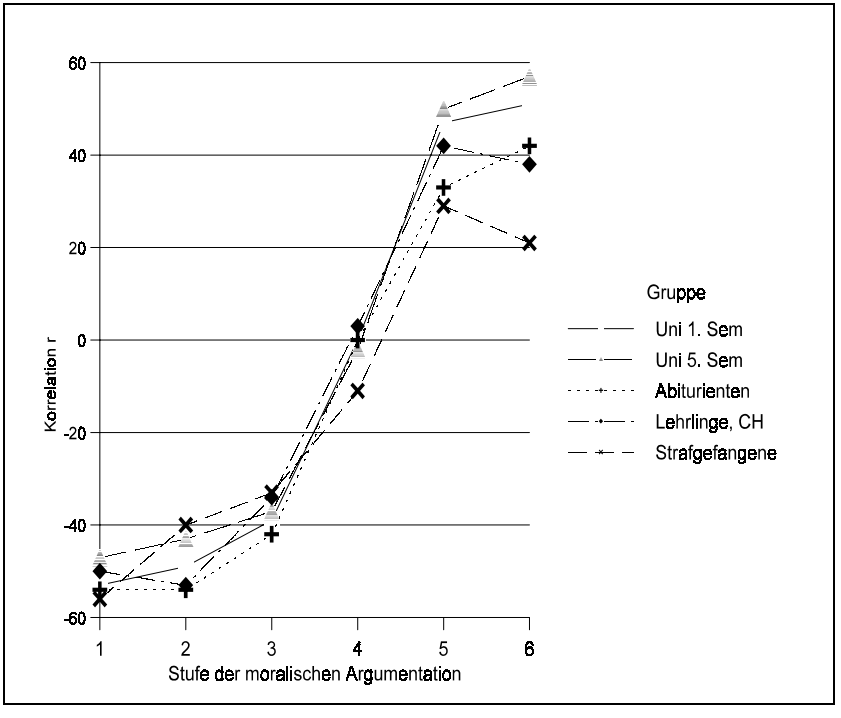


Abb. 2.6 Affektiv-kognitive Parallelität: Korrelationen zwischen der moralischen Urteilsfähigkeit (C-Wert im MUT) und den Einstellungen zu den sechs Moralstufen nach Kohlberg. (Der Übersichtlichkeit halber wurde hier nur eine Auswahl von Studien dargestellt)

stufentypischen Inhalten orientieren können (s. Kapitel 6). Aus diesen und anderen Gründen (s. Kapitel 4) ist es notwendig, die affektiven und kognitiven Aspekte des moralischen Urteils als manifeste Eigenschaften des moralischen Verhaltens zu definieren und ihre Existenz direkt, experimentell nachzuweisen, statt sie auf der Basis von Selbstdiagnosen einzuschätzen.

2.5 Eine empirische Studie zur korrelativen Parallelität

Aufgrund der Rekonstruktion von Affekt und Kognition auf der Grundlage des Zwei-Aspekte-Modells des moralischen Urteils läßt sich die Hypothese einer korrelativen Parallelität nun genauer formulieren: Wenn es zutrifft, dass die für jede Stufe typischen moralischen Perspektiven qualitativ verschieden sind, und

zwar derart, dass ihre Anwendung im Verhalten verschieden hohe Grade der kognitiv-strukturellen Entwicklung erfordert, dann ist zu erwarten, dass zwischen der Präferenz für höhere Stufen der moralischen Argumentation und der kognitiven Fähigkeit, von der eigenen Meinung zu abstrahieren und ein Dilemma konsistent moralisch zu reflektieren, ein positiver Zusammenhang besteht. Stimmt diese Hypothese, dann sollte die Korrelation zwischen der moralischen Determiniertheit des Urteilens (kognitiver Aspekt) und der Einstellung zu den Stufen der moralischen Argumentation (affektiver Aspekt) mit der Höhe der Stufe linear oder zumindest monoton zunehmen.
Die Falsifizierbarkeit dieser Hypothese ist hoch. Da es bei sechs Stufen,

$$6! \ (= 6 * 5 * 4 * 3 * 2 * 1) = 720$$

mögliche Rangfolgen der Korrelationen mit dem kognitiven Aspekt gibt, ist die a priori-Wahrscheinlichkeit, dass dieses Ergebnis in einer Untersuchung zufällig zustandekommt, $p = 1/(6!) = 0{,}00014$. Wie wir unten sehen werden, hat sich diese Hypothese – trotz der vergleichsweise geringen *a priori*-Wahrscheinlichkeit – in zahlreichen Studien bisher gut bestätigen lassen.
Die Parallelitätshypothese konnte adäquat erst mit dem *Moralisches Urteil Test* (MUT) geprüft werden, da nur mit diesem Instrument affektive und kognitive Aspekte des moralischen Urteilsverhaltens simultan, aber logisch unabhängig voneinander gemessen werden können. Durch die systematische Variation des situativen Kontextes im MUT ist es möglich, sehr differenziert auszuwerten, nach welchen Gesichtspunkten der oder die Befragte solche Argumente beurteilen. So ist es möglich festzustellen, ob und in welchem Ausmaß das Urteilsverhalten der Person von moralischen Überlegungen determiniert wird, anstelle von Meinungskonformität oder anderen Dingen. Dieses Ausmaß an Determiniertheit durch die moralische Stufe der Argumentation stellt den *kognitiven Aspekt* des moralischen Urteilens einer Person dar. Anders gesagt, es bezeichnet die Fähigkeit der befragten Person, die Argumente anderer nach moralischen, statt nach nicht-moralischen Gesichtspunkten wie zum Beispiel nach ihrer Meinungskonformität zu beurteilen.[45] Der *affektive Aspekt* meint im MUT die Einstellung der befragten Person zu den sechs Stufen der moralischen

45 Die C-Werte können theoretisch zwischen 0 % (nicht moralisch determiniert) und 100 % (völlig moralisch determiniert) variieren (s. Kapitel 4).

Orientierung, wie sie von Kohlberg typisiert wurden.[46] Die Hypothese der korrelativen Parallelität zwischen Affekt und Kognition wurde in mehreren Studien überprüft. Sieben dieser Studien sind in Abb. 2.6 widergegeben (für die Quelle und Gruppengrößen s. Abb. 2.5).

Die Korrelations-Hypothese wurde in allen Studien eindeutig bestätigt: Beide Aspekte des moralischen Urteilsverhaltens, der affektive und der kognitive, korrelieren in systematischer Weise miteinander.

Die Korrelationen zwischen der Einstellung zu jeder der sechs Argumentationsstufen und dem kognitiv-strukturellen Aspekt stimmen in allen Untersuchungen gut mit der theoretischen Vorhersage überein. Die meisten Korrelationen bilden eine monoton ansteigende Rangreihe.[47]

Es gibt nur geringe Abweichungen von der theoretisch vorhergesagten Rangreihe der Moralstufen, so dass kaum von einer Widerlegung der Hypothese gesprochen werden kann. Wegen der geringen *a priori*-Wahrscheinlichkeit dieser Ergebnisse aus fünf Untersuchungen dürfen diese Korrelationen sowohl der Höhe wie auch ihrer Struktur nach als Bestätigung einer bedeutsamen Hypothese angesehen werden. Eine weitere Bestätigung ergibt sich dadurch, dass bisher alle Replikationsstudien die korrelative Parallelität vollauf bestätigt haben.[48] Um einen Wert für den globalen Zusammenhang von affektiver und kognitiver Komponente zu erhalten, kann man eine multivariate Varianzanalyse der Polynomkontraste nach Bock (1975) durchführen.[49] Wir haben diese Analyse

46 Die Messung dieser Einstellungen erfolgt in klassischer Weise durch Summierung aller Antworten einer Person zu den Items, die zu einer Moralstufe gehören.

47 Dass die Rangreihe der Korrelationen am unteren und am oberen Ende der Stufenskala nicht linear verläuft, sondern S-förmig, ist vermutlich durch einen "Deckeneffekt" bedingt. Die absoluten Korrelationswerte sind dort bereits relativ hoch (zwischen $r = .50$ und .60).

48 Vgl. Schmied 1981; Wischka 1982; Lind 1985a, S. 185 ff.

49 Der affektive Aspekt wird dabei als der Grad der Akzeptabilität jeder der sechs Stufen der moralischen Argumentation operationalisiert und der kognitive Aspekt als der lineare, von Stufe 1 bis 6 ansteigende Kontrast der individuellen Antwortprofile. Ein lineares, steil von Stufe 1 nach Stufe 6 ansteigendes Urteilsprofil zeigt definitionsgemäß an, dass die Person zwischen den Argumentationsstufen stark differenziert, und zwar in der von der Hypothese erwarteten Richtung; ein flacher, ein nicht-monotoner oder ein abfallender Kontrast zeigen dagegen an, dass die Person in ihrem Urteilsverhalten nicht nach moralischen Gesichtspunkten differenziert bzw.

(Fortsetzung...)

für die beiden Studentengruppen getrennt durchgeführt. Das Ergebnis bestätigt den Eindruck, den Abb 2.6 vermittelt. Die lineare Korrelation zwischen dem Grad der Determiniertheit des Urteilsverhaltens durch moralische Überlegungen (kognitiver Aspekt) und dem Akzeptanzprofil auf der anderen Seite ist sehr stark (r = .60).[50]

2.6 Resümee

Wie in diesem Kapitel gezeigt wurde, steht das *Zwei-Aspekte-Modell* im Einklang mit theoretischen Grundpositionen von Piaget und Kohlberg. Es ermöglicht nur ein adäquates Verständnis der Piaget-Kohlberg-These der affektiv-kognitiven Parallelität, sondern auch deren empirische Überprüfung, indem es neue Wege der simultanen Messung beider Aspekte aufzeigt.

Affekt und Kognition dürfen nicht, wie häufig der Fall, als *Komponenten* des moralischen Verhaltens begriffen werden, sondern als *Aspekte*. Aus der Sicht des Zwei-Aspekte-Modells können Untersuchungen, die sich auf ein Komponenten-Modell stützen und Affekt und Kognition anhand verschiedener Verhaltensklassen beobachten, keine adäquaten Daten zur Prüfung der Parallelitätsthese liefern. Mit dem *Moralisches Urteil-Test* (MUT) ist belegt, dass eine logisch unabhängige Messung beider Aspekte an ein und derselben Verhaltensklasse möglich ist. Es konnte damit erstmals empirisch demonstriert werden, dass die beiden Aspekte hoch miteinander korrelieren (korrelative Parallelität). Im Lichte des Zwei-Aspekte-Modells stellen sich zudem bislang ungelöste, zentrale Problem der Moralpsychologe – wie die Urteils-Handlungs-Beziehung, die normative Begründung "höherer" Urteilsfähigkeit und das Verhältnis von affektiver und kognitiver Moralentwicklung – neu dar.

a) Das Problem der Urteil-Handlungs-Beziehung wurde bislang fast ausschließlich auf der Basis des Komponenten-Modells diskutiert. Das Komponenten-Modell, wonach Affekt und Kognition verschiedene Verhaltensbereiche markieren, ist aber, wie wir gesehen haben, inkompatibel mit dem

49(...Fortsetzung)
nicht in der von der Hypothese bestimmten Weise.

50 Die Bezeichnung der Korrelation als "sehr stark" entspricht den Bewertungskategorien von Cohen (1988). Siehe auch Kapitel 1.

"Theoriekern" (Lakatos) des Zwei-Aspekte-Modells. Urteilen und Handeln, Kompetenz und Performanz, lassen sich nicht unabhängig voneinander erfassen und als empirische Entitäten einander gegenüberstellen, da moralische Denkstrukturen immer nur als moralische Handlungsstrukturen begriffen werden können oder, wie Habermas (1985) feststellt, "die Welt moralischer Erscheinungen sich nur aus der performativen Einstellung von Interaktionsteilnehmern [erschließt]" (S. 17).[51] Das Zwei-Aspekte-Modell des moralischen Verhaltens ist demgegenüber kompatibel mit den Grundgedanken der kognitiven Entwicklungstheorie. Es macht auch möglich, zwischen affektiven und kognitiven Aspekten des moralischen Urteilsverhaltens zu unterscheiden und beide Aspekte, wie mit dem *Moralisches Urteil-Test* gezeigt wurde, an ein und demselben Verhaltensgesamt eines Individuums zu erheben. Damit ist vor allem der Streit obsolet geworden, der lange Zeit zwischen dem "behavioristischen", die psychologische Relevanz von Dispositionen leugnenden Standpunkt, wie er von Watson (1970/1924) und Skinner (1938) und ihren Anhängern vertreten wird, und (b) dem "kompetenztheoretischen" Standpunkt, wie er in Anlehnung an den linguistischen Ansatz von Chomsky (1973) vertreten wird, geherrscht hat. Beide Standpunkte sind Ausdruck eines methodologischen Problems, das sich im Rahmen der experimentellen Psychologie lösen läßt.

b) Im Zwei-Aspekte-Modell kann das normative Problem der Bewertung individueller Moralurteile besser gelöst werden. Gewöhnlich wird dieses Problem, wie Gordon Allport (1961) bemerkte, "externalistisch" gelöst, d.h., das Verhalten von Personen wird mit von außen angelegten Maßstäben beurteilt, ohne berücksichtigt wird, ob diese Maßstäbe generalisiert

51 Dieser Feststellung können sich auch andere Vertreter des Komponenten-Modells anschließen. So verweist Edelstein einerseits zwar, ein Komponenten-Modell voraussetzend, "auf die Vielzahl vermittelnder Kognitionen, welche die bindende Kraft moralischer Urteile angesichts unterschiedlicher Performanzbedingungen bestimmen" (S. 331), fügt aber andererseits hinzu, "dass Performanz nicht gleichbedeutend ist mit Handeln und Kompetenz nicht mit Urteil." Auch Garz (1989), ein Komponenten-Theoretiker, schließt sich dieser Auffassung an. Aber beide Autoren bleiben die Lösung dieses Widerspruchs schuldig.

werden können und die betroffene Person selbst diese Maßstäbe teilt.[52] Während im Bereich der Justiz und auch im Alltagsleben die Intentionen des Handelnden respektiert werden und bei der moralischen Beurteilung eine wichtige Rolle spielen (man denke nur an Differenzierungen wie fahrlässige Tötung, Totschlag, Mord im Affekt, vorsätzlicher Mord usw.), werden in der Psychologie Handlungen von Menschen oft nur nach einem willkürlichen, externalen Kriterium eingestuft. Auf dem Gebiet der Moralforschung hat sich dies dank universalistischer Ansätze – wie dem von Kohlberg – weitgehend geändert, auch wenn Kohlberg den Vorwurf, sein Meßinstrument sei ideologisch, kulturell oder sexistisch verzerrt, nicht hat völlig ausräumen können.[53]

c) Die Bestätigung der korrelativen Parallelität beider Grundaspekte des moralischen Urteilens beweist zwar nicht die These einer *Entwicklungsparallelität* von Affekt und Kognition, aber sie stützt sie. Aufgrund dieser Ergebnisse und denen von Turiel und von Nunner-Winkler können wir diese These auch noch weiter präzisieren.[54] Nunner-Winkler bestreitet eine Parallelitätsbeziehung zwischen Affekt und Kognition; sie nimmt eine phasenweise versetzte Entwicklung an, bei der die affektive Entwicklung der kognitiven nachfolgt. Diese Annahme enthält jedoch ungelöste Probleme. Nunner-Winkler definiert "Kognition" operational als die Fähigkeit eines Kindes, moralische Begriffe mehr oder weniger adäquat in Interviews zu verwenden und klammert andere Kriterien für moralische Urteilsfähigkeit (konsistent und differenziert zu urteilen und dieses auch angesichts widriger Umstände zu tun) aus. Diese Kriterien, die in unserem Modell eine große Rolle spielen, kommen in Nunner-Winklers (Komponenten-)Modell nur als nicht-empirische "Kompetenz" im Sinne von

52 In der ersten Auflage habe das Gegenteil zum externalistischen Ansatz als internalistisch bezeichnet. Das kann leicht mißverstanden werden. Die Formulierung "universalistisch" ist zutreffender.

53 Vgl. Simpson 1974, Sullivan 1977, Shweder 1982, Belenky & Gilligan 1983, Lind et al. 1987, aber auch Kapitel 7. Diese Einwände treffen natürlich auf alle Meßinstrumente zu, in denen beide Aspekte vermengt sind, so auch auf Rests (1979) *Defining Issues Test* (DIT) oder das *Socio-moral Reasoning Measure* von Gibbs et al. (1982).

54 Vgl. Turiel 1983, Nunner-Winkler 1990a; 1990b, Nunner-Winkler & Sodian 1988.

Chomskys Syntaxtheorie des idealen Sprechers vor.[55] Im Gegensatz dazu berücksichtigt das Zwei-Aspekte-Modell die Tatsache, dass in allen Menschen schon früh ein – mehr oder weniger stark ausgeprägter – moralischer Wille vorhanden ist, und dass die Unterschiede im moralischen Verhalten nicht durch unterschiedliches Wollen, sondern durch einen unterschiedlichen Grad der Entwicklung der moralischen Urteils*fähigkeit* bedingt sind. Die Annahme, dass die affektive Bindung an Moralprinzipien bereits früh vorhanden ist, wird m.E. durch die Tatsache belegt, dass Kinder (im Alter von 5 – 7 Jahren) schon moralische von bloß konventionellen Normen unterscheiden können (Turiel 1983), dass sie moralische Begriffe verwenden, um ihr Handeln zu rechtfertigen (Nunner-Winkler 1989a), und dass bereits Kleinkinder Aufforderungen nachkommen, die an ihr moralisches Gefühl appellieren (Lind 1989). Moralische Affekte können im Handeln einer Person nur dann voll zur Entfaltung kommen, wenn der "kognitive Apparat" dieser Person so weit entwickelt ist, dass sie moralische Prinzipien differenziert und integriert anwenden kann. Unsere Annahme einer kognitiv-moralischen Entwicklungsparallelität wird auch gestützt durch empirische Untersuchungen, die einen deutlichen Zusammenhang zwischen der Stufe der moralisch-kognitiven Entwicklung und moralischen Entscheidungen anzeigen[56], und durch die Tatsache, dass sich das moralische Urteil von Menschen durch geeignete pädagogische Methoden nachhaltig fördern läßt (Kapitel 8 und 9); besser wohl als durch Indoktrination oder sozialen Zwang.

55 Siehe Nunner-Winkler 1989a; 1990a. Auf die dadurch entstandene Erklärungslücke zwischen der Kenntnis moralischer Begriffe und dem moralischen Verhalten von Nunner-Winkler sind wir bereits oben eingegangen.

56 Vgl. den Übersichtsartikel von Blasi (1980). Neueren Datums ist die Arbeit von Gielen (1986), der einen deutlichen Zusammenhang zwischen der Stufe des moralischen Urteils und der Teilnahme an studentischen Protestaktionen zur Zeit der Studentenbewegung (1969) fand. Helkama & Ikonen (1986) berichten über hohe Korrelationen zwischen moralisch-kognitiver Entwicklung und verschiedenen Formen der sozialen Partizipation.

3 Kohlbergs *Moral Judgment Interview* – Eine Kritik

Kohlberg und seine Mitarbeiter stützten die Überprüfung der zentralen Annahmen der kognitiven Entwicklungstheorie der moralischen Entwicklung – weitgehend auf Untersuchungen mit dem von ihnen entwickelten *Moral Judgment Interview* (MJI; Colby, Kohlberg et al. 1987). Die in vielen Jahren entwickelte, hoch differenzierte Auswertungsmethode des MJI, das *Standard Issue Moral Judgment Interview and Scoring System*, so der Anspruch der Autoren, sei "eine theoriegestützte Erfassungsmethode" (S. 1). In diesem Kapitel soll untersucht werden, inwieweit das MJI diesem Anspruch theoretischer Validität genügt bzw. heute, nach vielfachen Revisionen besser als das ursprüngliche Verfahren gerecht wird, und für die Klärung der Frage nach den Bedingungen der Moralentwicklung, unserem zentralen Thema, geeignet ist.

3.1 Zum Problem der Bestimmung der Validität eines Meßinstruments

Die Richtigkeit oder Falschheit einer wissenschaftlichen Hypothese läßt sich selten, wenn überhaupt, einfach an Daten ablesen; die Daten oder Beobachtungen, an denen eine Theorie überprüft werden kann, müssen, wie der Wissenschaftstheoretiker Mittelstraß (1992, S. 16 ff) feststellte, zumeist erst hergestellt werden. Das bedeutet nicht, dass die Daten genau den Vorhersagen der Theorie angepasst werden (dies wäre nur durch eine Fälschung der Daten möglich). Es heißt vielmehr, dass die Wahl der Meß- und Auswertungsmethode, also die situativen Bedingungen, unter denen eine Person antwortet, und die Vorgehensweise bei der Zusammenfassung und Bewertung ihrer Antworten zu der zu prüfenden Theorie passen muß und nicht in Widerspruch zu ihr stehen darf. Erst wenn diese Voraussetzung erfüllt ist, erlauben die Meßdaten eine adäquate Überprüfung der Theorie. Sind diese Voraussetzung jedoch nicht erfüllt, sind die Meßdaten nicht theoretisch valide. Werden invalide Meßdaten zur Bewertung einer Theorie herangezogen, kann dies zu krassen Fehleinschätzungen dieser Theorie – zu einer falschen Bestätigungen wie einer falschen Widerlegung – führen. In der Psychologie wird die Frage nach der Validität eines Meßinstru-

ments (“Mißt das Instrument wirklich das, was es messen soll?”) zumeist rein empirisch zu bestimmen versucht, wozu die Korrelation der Meßwerte mit einem oder mehreren Außenkriterien herangezogen wird. Ergibt sich eine hohe positive Korrelation mit Indikatoren für dasselbe Konstrukt (und ergeben sich gleichzeitig niedrige Korrelationen mit Indikatoren für andere Konstrukte), dann sieht man die Validität des Meßinstruments als gegeben an. Dieses Vorgehen empfiehlt sich jedoch nicht für unsere Untersuchung. Hätten wir bereits ein zweifelsfrei valides Instrument zur Messung des moralischen Urteils, das als Außenkriterium dienen könnte, würde sich hier jede Diskussion erübrigen. Unser Problem besteht ja gerade darin, dass moralische Urteilsfähigkeit kein einfaches Konstrukt ist, das sich an einfachen Beobachtungen ablesen läßt, sondern dass es sich hier um ein komplexes Konstrukt handelt, das sich nur in einer sorgfältig hergestellten Beobachtungssituation valide messen läßt. Nicht jedes Verhalten, das wir im Alltag als Ausdruck von Moralität ansehen mögen, gewährt einen zuverlässigen und eindeutigen Einblick in die moralischen Fähigkeiten einer Person. Wir kommen hierauf zurück (S. 70). Wie schwer es ist, eine experimentelle Situation so zu konstruieren, dass das (Antwort-) Verhalten der befragten Person eine valide Messung moralischer Urteilsstrukturen ermöglicht, zeigt die lange Entwicklungsgeschichte des MJI und seiner Vorgänger.

3.2 Von der Intuition zur Wissenschaft

Eines der ersten Instrumente zur Erfassung moralischer Urteilsfähigkeit (“sittliche Reife”) stammt von Levy-Suhl (1912). Dieser hat in seinen Interviews mit straffälligen Jugendlichen herausgefunden, dass bestimmte moralische Werte immer wieder genannt wurden, und dass diese Werte recht gut den sittlichen Entwicklungsstand des Jugendlichen widerspiegelten. Erst in jüngster Zeit sind diese Arbeiten wiederentdeckt worden. Heidbrink (1989) hat in einer Reanalyse der Daten von Levy-Suhl gezeigt, dass sich diese Werte in eine Entwicklungsskala ähnlich der Kohlbergschen bringen lassen. Offenbar ohne Kenntnis dieser Vorarbeit hat Piaget (1973a/1932) seine “klinische Methode” entwickelt, um das moralische Urteil beim Kinde zu studieren. Bei der klinischen Methode handelt sich um eine Kombination von klassischen Verhaltensexperimenten in naturalistischen Kontexten mit der Befragung der handelnden Personen. Kohl-

berg (1958) hat für die Untersuchung von Jugendlichen und Erwachsenen Piagets Meßmethode zu einer reinen Befragungsmethode abgewandelt.

Im Kontext der Psychologie der 50er Jahre nahm sich Kohlbergs Meßmethode ebenso wie seine kognitiv-strukturelle Theorie des moralischen Urteils revolutionär aus. In dieser Zeit entwickelte sich die Psychologie "erratisch", wie Wolfle (1963) feststellte; sie war "erfüllt vom Streit darüber, was die wichtigsten Probleme und die stärksten Methoden" seien (S. v). Um die Ursache dieser Krise zu finden, setzte die *American Psychological Association* eine Forschergruppe ein, die als Kern des Problems jene Theorien identifizierte, die sich mit inneren psychologischen Prozessen oder intervenierenden Variablen befaßten (Koch 1963, S. 735). Dies kam einem Verdikt gegen Theorien wie die Piaget-Kohlberg-Theorie gleich. Verstärkt wurde die Neigung der Psychologie jener Zeit, sich von der Erforschung psychischer Dispositionen ganz zurückzuziehen. Moralität wurde weitgehend gleichgesetzt mit moralischem Begriffswissen ("Weiß das Kind, dass Stehlen verboten ist?") und moralischen Einstellungen ("Wie stark lehnt ein Kind Stehlen ab?"); beides wurde vom Bereich des "kognitiven Verhaltens" gänzlich getrennt.[57] Sofern von der Struktur der Persönlichkeit gesprochen wurde, handelt es sich um eine (z.B. mit einer Faktorenanalyse gefundenen) Eigenschaft einer Personen*gruppe*, aber nicht um die Disposition eines Individuums. Die strukturellen Aspekte des moralischen Verhaltens gerieten weitgehend aus dem Blick.

Um die moralische Struktur messen zu können, stützte sich Kohlberg anfänglich stark auf die "informierte psychologische Intuition" des Auswerters oder Schätzers (*rater*). Das größte Vertrauen hatte er in die Globaleinschätzung der Interview-Antworten anhand seines 6-Stufen-Schemas der Moralentwicklung (Kohlberg 1958; Oser 1988a). Dieses Vorgehen setzte voraus, dass die Schätzer mit Kohlbergs Theorie bestens vertraut waren. Da gemäß der Theorie die Schätzer, um die Antworten ihrer Probanden adäquat verstehen zu können, selbst eine mindestens ebenso hohe moralische Urteilsfähigkeit besitzen mußten wie diese, waren der Globaleinschätzung enge Grenzen gesetzt. Zudem war diese intuitive Meßmethode nur schwer in Einklang zu bringen mit der Grundforderung wissenschaftlicher Forschung nach vollständiger Wiederholbarkeit

57 Als Vertreter dieser Tendenz vgl. Hartshorne & May 1928, Krathwohl et al. 1964. Kritisch dazu Pittel & Mendelsohn 1966.

der Untersuchung[58] und allgemeiner Überprüfbarkeit der Ergebnisse: Die bei der Globaleinschätzung ablaufenden Deutungsprozesse könnten kaum exakt wiederholt werden; die Nachprüfung war nach der Theorie *a priori* auf die Leute beschränkt, die selbst nachgewiesen haben, dass sie mindestens eine ebenso hohe moralische Urteilsfähigkeit haben wie die Auswerter.

Kohlberg verteidigte diese "intuitive" Meßmethode mit dem Hinweis, dass die damals gängigen psychologischen Meßmethoden, vor allem die klassische Testtheorie, ungeeignet waren, um den kognitiv-strukturellen Aspekt der Moral adäquat zu erfassen (Kohlberg 1981). Dennoch erkannte er, dass, wie Colby, Kohlberg et al. (1987) schreiben, ein Wandel in "der Betrachtungsweise von Interpretation [von Daten] als einer Kunst zu der Betrachtungsweise von Interpretation als Wissenschaft" notwendig war (S. 40). Die Bemühung um eine mehr wissenschaftliche Dateninterpretation war aber – wohl bedingt durch das Fehlen von Alternativen und die heftige, um die Einhaltung der klassischen psychometrischen Regeln besorgte Kritik (Kurtines & Greif 1974) – hauptsächlich an jenen Regeln orientiert, die Kohlberg bis dahin als inkompatibel zu seiner Theorie ansah. Die Revision des MJI diente vor allem dem Ziel, "größere Objektivität und Reliabilität bei der Einstufung zu erreichen" (Colby, Kohlberg et al. 1987, S. 40; meine Übersetzung). *Dies* ist gut gelungen. Das modifizierte MJI ist, gemessen an den Standards der konventionellen Psychometrie, ein gutes Meßinstrument.[59] Wie berichtet wird, erreichen die Schätzer eine fast perfekte Übereinstimmung innerhalb einer Drittelstufe.[60] Aber konnte durch diese

58 Wenn die Ergebnisse nicht repliziert werden können, ist die geprüfte Theorie möglicherweise falsch, nicht aber das Experiment.

59 Die Reliabilität des MJI ist tatsächlich außerordentlich hoch. Innerhalb eines drei- bis sechswöchigen Zeitintervalls bleiben die Scores relativ stabil (die Test-Retest-Reliabilität liegt weit über 0,90). Die alternative Reliabilität wird mit 0,95 angegeben (S. 63-65). Die drei Reliabilitätsformen sind nicht völlig äquivalent; die Scores können sich bis zu einem Fünftel einer Stufe unterscheiden (S. 68). Der Meßfehler auf der Basis von Retestkorrelationen beträgt: 15,62 MMS-Punkte, berechnet nach der Formel (Colby, Kohlberg et al. 1987, S. 69):

$$\sigma_{mess} = \sigma * \sqrt{1 - r_{xy}}$$

60 Wegen der hohen Standardisierung der Auswertung benötigt die Ausbildung qualifizierter Schätzer nur kurze Zeit, trotz der nach wie vor geforderten Theoriekenntnis und Verstehensleistung. Bei den Schätzern handelte es sich um Universitätsstuden-

(Fortsetzung...)

Revision auch eine größere theoretische Validität des MJI, das heißt eine bessere Übereinstimmung zwischen Theorie und Methode erreicht werden?

3.3 Theorie: Ganzheitliche Betrachtungsweise des moralischen Urteilverhaltens

In Kohlbergs Theorie spielt die Struktur des moralischen Urteils eine zentrale Rolle.[61] Kohlberg geht davon aus, dass das moralische Urteil nicht nur affektive Aspekte oder moralische Einstellungen umfaßt, sondern auch kognitiv-strukturelle Aspekte. Die Veränderung der kognitiv-strukturellen Aspekte der Moral bedingt in jeder Entwicklungsstufe unterschiedliche Denkweisen und nicht nur die Änderung von moralischen Einstellungen; die moralische Entwicklung ist also qualitativ-strukturell und nicht nur quantitativ. Die strukturell definierten Stufen moralischer Denkweisen bilden eine invariante Sequenz der moralischen Entwicklung. Schließlich, so Kohlberg, sind kognitiv-moralische Entwicklungsstufen hierarchisch integriert: Die Stufen bilden eine Anordnung zunehmend differenzierter und integrierter Strukturen. Bemerkenswerterweise spricht Kohlberg – zumindest in den theoretischen Teilen seiner Arbeiten – von moralischen Strukturen im Sinne von *manifestem Verhalten* und nicht im Sinne von latenten moralischen Vorstellungen: "Was uns wichtig ist, ist, wie moralische Urteile zustandekommen, wenn [ein moralisches Prinzip] wirklich auf Wertkonflikte angewendet werden. Moral ist eine Sache von Wahl und Entscheidung. Es ist nicht bloß eine Frage des Gebrauchs von abstrakten Prinzipien wie Gerechtigkeit". Entsprechend versuchen er und seine Kollegen "nicht die abstrakten philosophischen Positionen einer Person zu messen, sondern vielmehr das, was wirksam ist, wenn diese Person einer wirklichen moralischen Entscheidung gegenübersteht" (Colby, Kohlberg et al. 1987, S. 58). Ein struktureller Meßansatz muß also die im tatsächlichen Verhalten wirksamen moralischen Dispositionen erfassen und nicht, wie etwa der phänomenologische Ansatz, die subjektive Wahrnehmung solcher Dispositionen durch die befragte

60(...Fortsetzung)
ten mit psychologischem Vorwissen, die einige Zeit mit dem MJI gearbeitet oder an einem viertägigen Auswertungsseminar teilgenommen hatten.

61 Siehe Kohlberg 1984; auch Lind 1985b, Oser 1981a und Kapitel 2.

Person. Im strukturellen Meßansatz wird die befragte Person zum Beispiel aufgefordert, ihre Meinung zu einem bestimmten moralischen Problem zu begründen; im phänomenologischen wird sie dagegen veranlaßt anzugeben, welche moralische Argumente sie in solchen Situationen typischerweise verwenden würde bzw. an welchen Werte sie sich gewöhnlich orientiert.[62]

Aus diesen Postulaten und Annahmen folgt, dass nicht jedes Verhalten oder jede Verhaltenseinheit geeignet ist für die Messung moralischer Dispositionen. Wenn eine Person mit einem moralischen Dilemma konfrontiert ist, sind drei Arten von Verhalten zu unterscheiden: die Meinung einer Person über die richtige Lösung eines Dilemmas, ihre Begründung für diese Meinung und die moralische Bedeutung dieser Begründung. Davon sind die ersten beiden Verhaltenseinheiten nicht für die Erfassung moralischer Dispositionen geeignet: 1. Vom Verhalten als *Meinung zu einer bestimmten Entscheidung* kann nicht auf moralische Urteilsstrukturen und Stufen geschlossen werden. Eine solche Entscheidung ist gewöhnlich von sehr vielen persönlichen und situativen Merkmalen bestimmt und daher ambivalent. Im bekannten "Heinz-Dilemma" beispielsweise soll der oder die Befragte sich entscheiden, ob Heinz, um das Leben seiner kranken Frau zu retten, eine sehr teure Medizin stehlen soll, die er sich finanziell nicht leisten kann. Diese Entscheidung ist stark situationsgebunden. Schon geringe Modifikationen der Situation (z.B., wenn die Medizin von der Krankenkasse bezahlt wird oder wenn der Befragte die Perspektive von Heinz oder seiner Frau einnimmt) können die befragte Person zu einer ganz anderen Antwort veranlassen, ohne dass sich ihr moralisches Urteilsniveau verändert hat. Zwischen konkreten, existentiellen Entscheidungen und einem allgemeinen

62 Beide Meßansätze sind "subjektiv" insofern in beiden das Subjekt involviert wird. Aber es wäre falsch und irreführend deshalb beide als phänomenologisch zu bezeichnen, wie es Kohlberg unter dem Einfluß seiner Habermas-Rezeptionen zuletzt tat und viele seiner Schüler tun. Dieses Mißverständnis scheint auch viel zu der Verwirrung um das sogenannte Urteil-Handlungs-Problem beizutragen. Wird nämlich das moralische Urteilen als eine vom Verhalten unabhängige Entität begriffen, die nicht direkt gemessen, sondern nur aus den Berichten der betreffenden Person erschlossen werden kann, stellt sich in der Tat die Frage, inwieweit diese Berichte mit dem tatsächlichen Urteilsverhalten korrelieren. Will er aber diese Frage empirisch klären, kommt der radikal-phänomenologische Meßansatz in eine Zwickmühle. Wenn eine direkte Erfassung moralischer Dispositionen nicht möglich ist, läßt sich die Frage nämlich prinzipiell nicht empirisch beantworten; würde man zugeben, dass eine direkte Erfassung möglich ist, würde dieser Meßansatz sich selbst obsolet machen.

moralischen Prinzip kann es kaum eine einfache Beziehung geben, weil oft mehr als dieses eine Prinzip beachtet werden muß, und weil die Abwägung der Folgen notwendiger Bestandteil einer moralischen Entscheidung ist.

2. Auch Verhalten als *Verwendung oder Bewertung moralischer Argumente* ist keine adäquate Basis für die Erschließung moralischer Dispositionen. Wenn eine Person Argumente benutzt, die linguistisch als moralisch klassifiziert werden können (z.B. Ehrlichkeit, Gerechtigkeit), dann kann davon ausgegangen werden, dass die Person diese moralischen Begriffe kennt und wertschätzt. Aber das bloße Äußern oder Bewerten moralischer Argumente zeigt noch keine bestimmte moralische Urteilsfähigkeit an. Sagte ein Befragter beispielsweise: "Es ist falsch, wenn Heinz die Medizin stiehlt, um das Leben seiner Frau zu retten, *da es gegen das Gesetz ist*", kann dieses, wie Kohlberg (1985) feststellt, nicht eindeutig einer Stufe der moralischen Argumentation zugeordnet werden: "Der Gebrauch der Norm des Gesetzes ist auf jeder Stufe zu finden" (S. xvi). Das bloße Erwähnen von moralischen Normen besagt noch nicht, ob eine Person gewillt und in der Lage ist, in konkreten Entscheidungssituationen moralische Normen und Werte adäquat anzuwenden. Wenn jemand zu einer Begründung gedrängt wird, wird er vielleicht nur deshalb moralische Prinzipien benutzen, um seine Wahl nachträglich zu rechtfertigen oder zu "rationalisieren", wie ein Psychoanalytiker das bezeichnen würde.

3. Nur die Beobachtung von *ganzheitlichen Verhaltensmustern* kann uns in die Lage versetzen, die Existenz und Ausprägung moralischer Dispositionen einer Person valide zu messen. Gemäß der kognitiven Entwicklungstheorie besteht die konstitutive Komponente des moralischen Verhaltens einer Person nicht bloß darin, dass sie moralische Prinzipien anwendet, sondern auch und besonders darin, wie konsistent und differenziert sie sich bei ihren Entscheidungen daran orientiert. Für Kohlberg und seine Kollegen muß sich die Messung des moralischen Urteils daher im Kern auf "die Analyse beobachtbarer Denkmuster [konzentrieren], wie sie sich in den Antworten einer Person im Moralinterview offenbaren" (S. 4).

3.4 Meßmethode: Zerlegung des Verhaltens

Das Hauptziel der Revision des *Moral Judgment Interviews* war eine Verbesserung im Sinne der klassischen Testtheorie, die davon ausgeht, dass in jedem Verhaltensakt die zu messende Disposition repräsentiert ist. Dies versuchten die Autoren zu erreichen, indem "klare und konkrete Stufenkriterien spezifiziert und die Entwicklungsabfolge eines jeden moralischen Begriffs innerhalb jeder Stufe und global, für die allgemeinen Stufen-Strukturen definiert wurde" (Colby & Kohlberg 1987, S. 40).

Wie ist das Vorgehen beim MJI? Ähnlich Piagets (1973a) Vorgehen, wird im MJI die befragte Person mit einem moralischen Dilemma konfrontiert, d.h. mit einer kurzen Geschichte, in der sich zwei oder mehrere moralische Prinzipien gegenüberstehen. Die Person wird gebeten, eine Wahl zwischen diesen Prinzipien zu treffen. Indem die Entscheidung zwischen zwei Lösungen auch eine zwischen verschiedenen moralischen Prinzipien ist, stimuliert das Dilemma die Person, sich mit der moralischen Seite statt nur mit der "technischen" Seite eines Problems zu beschäftigen. Durch gezieltes Nachfragen, durch "Warum-Fragen", wird die befragte Person angehalten, ihre Meinung mit Argumenten zu begründen, und durch die Konfrontation der Person mit variierenden Rahmenbedingungen für das Dilemma, mit neuen Dilemmas und mit gegnerischen Meinungen und Sichtweisen konfrontiert. Unter anderem wird gefragt: "Weshalb hast Du dich für die eine Handlungsalternative entschieden? Was sind deine Gründe? Was rechtfertigt diese Entscheidung?" etc. Die Qualität dieser Nachfragen ist, worauf Colby, Kohlberg et al. (1987) zu Recht hinweisen, "folgenschwer für die Erfassung der moralischen Urteilsstufe" (S. 151). Das gezielte Nachfragen ist notwendig, um eine gute Auswertbarkeit der Antworten zu erreichen, um die am weitesten entwickelte Stufe des moralischen Argumentierens hervorzulocken, zu der die Person fähig ist, und um überprüfen zu können, in welchem Maße die Person fähig ist, bei ihren Entscheidungen moralische Prinzipien anzuwenden (S. 5, 58, 61, 169).

Trotz der großen Bedeutung, die die Autoren der Interviewsituation zumessen, wird diese im Handbuch nicht systematisch behandelt. Sie basiert weiterhin größtenteils auf Intuition und Alltagsverständnis (S. 153-158). Die Instruktionen bleiben oft vage. Beispielsweise wird der Interviewer dazu angehalten, gezielt nachzufragen, solle jedoch nicht zu oft "Warum"-Fragen stellen (S. 54). Die bei jedem Dilemma gestellten Standardfragen berühren systematisch die

zwei Hauptthemen (Issues) des jeweiligen Dilemmas. Allerdings bleibt es dem Interviewer überlassen, die im Dilemma enthaltenen Normkonflikte auszuloten.[63] Noch bedauerlicher ist, dass die Konfrontation mit Gegenargumenten im neuen MJI kaum mehr angesprochen wird. Während Kohlberg (1958) früher mit dieser Technik die Festigkeit der moralischen Überzeugungen der Befragten getestet hatte, dient sie heute nur noch dazu, ausreichende Mengen an linguistischem Material für die Auswertung zu gewinnen (S. 151). Argumente, die von den Probanden zugunsten der Gegenmeinung vorgebracht werden, erhalten – sofern sie überhaupt noch in die Einstufung der individuellen Urteilsstruktur einbezogen werden – im neuen Auswertungssystem eine geringere Gewichtung als die Argumente, die sie zur Begründung ihrer eigenen Meinung vorbringen. Die Last der Identifikation der strukturellen Eigenschaft des individuellen Urteilsverhaltens liegt beim MJI somit ganz auf der Auswertungsseite, die daher einer eingehenden Betrachtung bedarf. Die Auswertung der Interviewurteile erfolgt in drei Schritten. Erster Schritt: Identifizierung der adäquaten Meßeinheit. Zweiter Schritt: Zuordnung eines Stufenscores. Dritter Schritt: Kalkulation eines globalen und durchschnittlichen Scores für das Individuum über alle Meßeinheiten hinweg. Wie werden die Meßeinheiten identifiziert? Man muß laut Handbuch zunächst bei einem Issue (z.B. dem Issue "Leben" im „Heinz-Dilemma") anfangen und alle Antworten gemäß der moralischen *Norm* und dem moralischen *Element* klassifizieren.[64] Die daraus resultierende „Issue-x/Norm-x" Elementkombination, von den Autoren auch als "Interview judgment" (IJ) bezeichnet, dient beim MJI als Grundlage für die hermeneutische Analyse. IJs werden nach den Issues klassifiziert, denen sie zugeordnet werden. Die "Rules for issue classification" (S. 163-165) liefern die tatsächlichen Meßeinheiten. Sind die Meßeinheiten identifiziert, können sie durch das sogenannte "matching"-Verfahren einer Kohlberg-Stufe zugeordnet werden, indem man die Antworten der befragten Person jener Stufe zuordnet, für die im Handbuch die ähnlichsten Beispielantworten zu finden sind. Das Handbuch gibt für dieses *matching* fünf Entscheidungshilfen: 1. Für die vorläufige, globale Einschätzung aufgrund aller Antworten (Handbuchanweisung: "Lies erneut alle Antworten durch, die unter einem Thema eingeordnet wurden, und denke dar-

63 Colby, Kohlberg et al. 1987, Band 2, S. 4 und 56.

64 Das Handbuch erläutert eingehend, was unter Normen und Elementen verstanden wird, und gibt für beides Beispiele.

über nach, was sie bedeuten"). 2. Für die Frage der Auswertbarkeit ("Liefern die *interview judgments* Argumente, hält sie der Befragte für valide und sind sie präskriptiver Natur?"). 3. Für die Einschätzung der Wertzuweisung und -transposition ("Rekonstruiere die Bedeutung von verstümmelten Worten und zerstückelten Sätzen"). 4. Für die Verteilung der IJs auf die *criterion judgments* (CJs) der richtigen Stufen. 5. Schließlich für die "strukturelle Bewertung" der vorgeschlagenen Zuordnungen, mit der "jene oberflächlichen Zuordnungen ausgemerzt oder mit einem 'Veto' versehen werden, für welche die Stufe des *criterion judgments* nicht die wahre strukturelle Bedeutung des Interviewmaterials widerspiegelt" (Colby, Kohlberg et al. 1987, S. 165-177).
Kohlberg und seine Koautoren werten die Interviewantworten im MJI also primär nach den Vorgaben der klassischen Testtheorie aus, indem sie das gesamte Interview in Meßeinheiten zergliedern, diese Einheiten isoliert einstufen und sie dann zu einem Gesamtwert addieren. Diese zergliedernde Analyse versuchen sie, durch eine hermeneutische ("wahre strukturelle Bedeutung") abzusichern. Es ist aber fraglich, ob die hermeneutische Analyse dieses leisten kann, zumal in diesen Teil der Auswertung vielfältige theoretische Annahmen des Auswerters einfließen können, die die Neutralität der Daten bezüglich der zentralen Postulate der Kohlberg-Theorie in Frage stellen.

3.5 Das methodologische Dilemma des MJI

Hierin besteht das methodologische Dilemma des revidierten MJI. Auf der einen Seite soll das MJI strukturelle Eigenschaften des individuellen Urteilsverhaltens (wie Inkonsistenz und Differenziertheit) messen, auf der anderen Seite werden (gemäß der klassischen Testtheorie, an der sich die Revision des MJI orientiert hat) solche strukturellen Informationen zur Bewertung des Meßinstruments benötigt, so dass sie nicht mehr für die Einstufung der Urteilsstruktur zur Verfügung stehen. Es wird anerkannt, dass es zwischen den Antwortmustern von Befragten strukturelle Unterschiede gibt, aber diese Unterschiede werden nicht als Basis für eine strukturelle Einschätzung der individuellen Urteilsfähigkeit herangezogen. So wird zum Beispiel die Tatsache, dass einige Personen für nicht gewählte Issues weniger Argumente oder niedrigere Stufenwerte produzieren als für die gewählten (S. 161), durch die "*Issue classification rules*" von der Analyse ausgeschlossen. Unterschiede innerhalb eines Issues (Antwort-

inkonsistenzen) werden durch die "Lower-stage inclusion rules" weitgehend nivelliert. Die Konstrukteure gehen zwar von der Annahme aus, dass das moralische Urteilen eine strukturelle Einheit bildet, also keine wesentliche Variation zwischen den Urteilen einer Person auftritt. Aber in einer Studie, so berichten Colby, Kohlberg et al. (1987), konnten mit dem MJI-Verfahren nur 29 von 188 Personen einer "reinen" Stufe der moralischen Urteilskompetenz zugeordnet werden (S. 136-139).[65] Mischstufen stellen ein weiteres Problem für den Versuch dar, klassische Testtheorie und kognitive Entwicklungstheorie miteinander zu verbinden. Auf der einen Seite, könnte man argumentieren, sind gemischte Stufenklassifikationen durch unzureichende Reliabilität („Meßfehler") bedingt. Wenn Stufen der moralischen Urteilskompetenz strukturelle Einheiten darstellen, dann sollte eine Person für jedes Issue oder jedes Dilemma den gleichen Score erhalten. Die Annahme der unzureichenden Reliabilität würde zwar die Hypothese der strukturellen Einheit retten, doch würden damit die häufig berichteten Teilstufen (und die Unterschiede, die kleiner als eine Stufe sind) bedeutungslos. Auf der anderen Seite können wir Mischstufen als Stufenübergänge begreifen. In diesem Falle wäre aber Kohlbergs Annahme der strukturellen Ganzheit nicht mehr haltbar. Die Idee der strukturellen Ganzheit verträgt sich nicht mit dem Konzept von Mischstufen. Das bestätigt wieder einmal, dass man den sprichwörtlichen Kuchen nicht gleichzeitig essen und behalten kann. Die beobachtete Konsistenz (oder Varianz) der Forschungsdaten läßt sich nicht gleichzeitig dem Individuum und dem Meßinstrument zuordnen. Man kann nicht versuchen, die Struktur des individuellen Urteilsverhaltens mit einer Methodologie zu erfassen, die diese Struktur als Fehler des Meßinstruments betrachtet. Entweder nimmt man an, die intra-individuelle Verhaltensvariation sei ein Ausdruck des Meßfehlers des Instruments, oder sie sei ein Ausdruck bestimmter kognitiver Strukturen des Befragten. Die klassische Testtheorie entschied sich für die erste Annahme (Lumsden 1976). Die kognitive Entwicklungstheorie legt eher die zweite Betrachtungsweise nahe, da sonst keine Möglichkeit bestünde, Urteils*strukturen* anhand von objektiven Beobachtungsdaten nachzuweisen.

65 Meine Berechnung. Da keine Angaben über die Berechnungsgrundlagen der "Stage Usage Percentage" gemacht werden, habe ich die "Issue-Einheit" verwendet.

3.6 Resümee

Das *Moral Judgment Interview* (Colby, Kohlberg et al. 1987) ist eine einzigartige Interview-Methode, und dennoch ist es vom Standpunkt der kognitiv-strukturellen Theorie der Moralentwicklung zu kritisieren. Seine theoretischen Grundlagen und seine Auswertungsregeln sind sehr transparent, was diese Methode vielen anderen Interview-Methoden überlegen macht, die allein auf globalen Einschätzungen beruhen. Erst diese Transparenz der MJI-Methode ermöglicht eine differenzierte Kritik des Verfahrens, aus der sich konstruktive Hinweise für eine Verbesserung ergeben. Auch ist das MJI gemessen an den Kriterien der klassischen Testtheorie hoch 'reliabel' und daher geeignet für psychologische Forschungs- und Diagnostikfragen, die eine hohe Reliabilität im Sinne dieser Testtheorie erfordern. Aber gerade wegen der Ausrichtung des MJI an der klassischen Testtheorie ist die Validität des MJI eingeschränkt. Die auf Zergliederung des individuellen Verhaltens angelegte Auswertung des MJI macht die Struktur der individuellen Urteilsmuster für den Forscher *unsichtbar*. Da das Auswertungsverfahren des MUT auf der Klassischen Testtheorie basiert, basiert es (implizit) auch auf der Annahme, dass die Probanden sich bei der Beurteilungen eines moralischen Dilemmas *alle gleichermaßen* an moralischen Normen (wenn auch auf verschiedenen Stufen) orientieren. Gleichzeitig erlaubt das Verfahren es nicht, diese Annahme empirisch zu prüfen. Dennoch gibt es, wie wir in den nachfolgenden Kapitel zeigen werden, vielfältige Hinweise dafür, dass diese Annahme falsch ist, das heißt, dass sich die Menschen sehr stark unterscheiden können hinsichtlich der Frage, ob und wie stark moralische Dispositionen ihr Urteilsverhalten determinieren. Art und Grad dieser Unterschiede lassen sich sogar aufgrund theoretischer Überlegungen vorhersagen. Es scheint, dass die Auswertung des MJI auf Annahmen beruht, die viel mehr mit klassischen Testtheorie als mit der kognitiv-strukturellen Theorie der Moralentwicklung in Einklang stehen. Für die Überprüfung solcher Annahmen benötigen wir also ein Instrument, das ohne diese Annahmen auskommt und die valide Messung von Inhalt *und* Struktur des moralischen Urteilsverhaltens ermöglicht. Man muß damit vor allem messen können, *ob* und wie stark eine befragte Person sich in der Diskussion über Probleme an moralischen Prinzipien orientiert. Das *Ausmaß der Determination des individuellen Urteilsverhaltens durch moralische Orientierungen* ist der entscheidende Gradmesser für die individuelle moralische Urteilsfähigkeit. Die Erfassung dieser Fähigkeit ist denn

auch die zentrale Meßabsicht des *Moralisches Urteil-Tests*, der im nächsten Kapitel vorgestellt wird.

Colby, Kohlberg et al. (1987) sehen die Validität des MJI durch die Tatsache bestätigt, dass die MJI-Daten die Hypothesen der invarianten Entwicklungssequenz und der strukturellen Ganzheit bestätigen. Wie wir gesehen haben, wirft diese Feststellung einige schwer lösbare Probleme auf, da das MJI aufgrund seines Auswertungsdesigns *voreingenommen* (biased) gegenüber der Prüfung dieser Hypothesen ist. Die Auswertungsanleitung des MJI setzt an vielen Stellen die Richtigkeit der Annahme einer invarianten Entwicklungsabfolge und strukturellen Ganzheit voraus. Die Einstufung der Probanden mit dem MJI soll laut Kohlberg (1976) so erfolgen, "dass das Kriterium der aufeinander folgenden Bewegung eingehalten wird" (S. 47). Somit begünstigen die mit dem MJI gewonnenen Daten in einem schwer abschätzbaren Ausmaß bereits *a priori* die Bestätigung beider Postulate. Aber auch wenn man annehmen könnte, dass die Messung des moralischen Urteilens mit dem MJI unverzerrt und unvoreingenommen ist, verbleibt das Problem, dass die Erfüllung der beiden von Kohlberg genannten Kriterien (invariante Sequenz und strukturelle Ganzheit) keine hinreichende Bedingung für die Validität dieses Meßverfahrens darstellen. Es gibt viele Maße (z.B. Gewicht und Höhe, mathematische Fähigkeiten), die beide Kriterien erfüllen, aber dennoch keine validen Maße des moralischen Urteils sind. Damit ein Meßinstrument die *moralische* Urteilsfähigkeit valide erfassen kann, muß die darin gestellte Aufgabe für die Befragten tatsächlich eine *moralische* Aufgabe sein. Die Kohlberg-Gruppe hält dies im Falle des MJI offensichtlich für gegeben und widmet sich dieser Frage daher nicht explizit.

Implizit lassen sich beim MJI die darin enthaltene *moralische Aufgabe* und die für ihre Bewältigung notwendige Fähigkeit rekonstruieren. Bei der Diskussion und Lösung eines moralischen Dilemmas müssen die Befragten vielfältige Erwartungen und Werte koordinieren. Sie müssen sich für eine von zwei Handlungsalternativen entscheiden, die beide moralische Probleme aufwerfen: Hatte Heinz, der Protagonist im gleichnamigen Dilemma im Kohlberg-Interview, die Medizin stehlen dürfen, um seine todkranke Frau zu retten, oder durfte er keinen Diebstahl begehen? Hätte er es auch für einen anderen Menschen tun müssen? Ist diese Rechtfertigung generalisierbar? Was wäre, wenn die befragte Person selbst die Bestohlene wäre? Offenbar sind diese Fragen nicht leicht zu beantworten. Moralisch wenig entwickelte Personen kommen wohl kaum über

die ersten Fragen des Interviews hinaus. Viele finden es schwierig, wenn nicht gar unmöglich, sich zum Beispiel auf eine rationale Diskussion über Ansichten einzulassen, die ihrer eigenen Meinung widersprechen. In diesem Sinn stellt das *Moral Judgment Interview* von Kohlberg offenbar eine Aufgabe für die befragte Person dar. Aber im Auswertungshandbuch wird dieses für die Validität wichtige Thema an keiner Stelle angesprochen. So bleiben viele Fragen unbeantwortet. Welche Überlegungen haben zur Auswahl der Dilemmas geführt, die das MJI enthält? Erfordern alle MJI-Dilemmas eine Begründung auf der im Sinne von Kohlberg höchsten Stufe des moralischen Urteils, oder lassen sich einige der dort vorgelegten moralischen Probleme auch auf einer niedrigeren Stufe adäquat lösen, bzw. wäre es adäquater zu versuchen, sie auf einer tieferen Stufe zu lösen? Einige Dilemmas im MJI scheinen durchaus auf einer tieferen Stufe als auf Stufe 6 adäquat lösbar. Schließlich die Frage: Worin besteht moralische Urteilsfähigkeit? In der Wahl einer möglichst hohen Stufe zur Rechtfertigung einer einmal gefaßten Meinung? Oder besteht sie nicht eher in der Fähigkeit, die eigene Meinung abzuändern, wenn man erkennt, dass diese nicht in Einklang mit den eigenen moralischen Überzeugungen steht?

4 Der *Moralisches Urteil-Test* - Ein experimenteller Fragebogen

Um die zentralen Annahmen der kognitiv-strukturellen Theorie der Moralentwicklung und besonders um annahmen bezüglich der Existenz und der Lehrbarkeit moralischer Fähigkeiten prüfen zu können, haben wir in den siebziger Jahren ein neues Instrument, den *Moralisches Urteil-Test* (MUT) entwickelt. Das neue Meßinstrument sollte drei Kriterien erfüllen: Es sollte mit der Theorie kompatibel, objektiv (replizierbar) und leistungsabfordernd sein. Es muß *theoriekompatibel* sein, damit die Meßwerte überhaupt als Kriterium für die Richtigkeit oder Falschheit von Vorhersagen dienen können. Daten von einem invaliden Meßinstrument besitzen für die Prüfung einer Theorie keinen Aussagewert. Im Rahmen unserer Untersuchung heißt die Forderung nach Validität vor allem, dass das Konstruktions- und Auswertungsdesign des Meßinstruments die Erfassung der moralischen Urteilsfähigkeit genau in der Weise garantiert, wie sie definiert ist. Wie wir in Kapitel 3 gesehen haben, ist diese Forderung insbesondere in bezug auf die kognitiv-strukturellen Eigenschaften des moralischen Urteilsverhaltens offenbar kaum oder gar nicht einzulösen, wenn man sich auf klassische psychologische Meßmethoden stützt.

Objektiv muß das Meßinstrument sein, damit die hier berichteten Ergebnisse auch von anderen Auswertern überprüft werden können. Damit die individuelle moralische Urteilsfähigkeit auf eine intersubjektiv überprüfbare Weise festgestellt werden kann, sollte das Instrument Meßwerte liefern, ohne dass dafür intuitive Einschätzungen durch die befragte Person selbst oder durch die Auswerter notwendig sind. Die Beteiligung von Schätzern ist kein prinzipieller Einwand gegen die Validität eines Meßinstruments; selbst in den Naturwissenschaften werden Meßwerte häufig durch geschulte Schätzer ermittelt. Aber das Angewiesensein auf Schätzer stellt hohe Anforderungen an die Klarheit der Auswertungsregeln und an die Kompetenz der Schätzer und sie begrenzt aus ökonomischen Gründen sehr stark die Überprüfungsmöglichkeiten durch die Forschung. Die Beteiligung der Befragten bei der Einschätzung ihres eigenen Verhaltens hingegen stellt eine prinzipielle Einschränkung der Meßvalidität dar, da deren Selbstwahrnehmung mit der zu messenden moralischen Urteilsfähigkeit mehr oder minder hoch korrelieren kann, und wir daher nicht wissen, in welchem Ausmaß der Meßwert einerseits die moralische Urteilsfähigkeit oder

andererseits die Art der Selbstwahrnehmung widerspiegelt. Auf den MJI trifft nur der erste Einwand zur Objektivität des Verfahrens zu. Der zweite Einwand trifft auf Verfahren zu, die von den Befragten verlangen, dass er oder sie ihr eigenes moralisches Urteil einschätzen soll ("Wie würden *Sie* selbst in einer solchen Situation urteilen ...?"). Die hier verlangte Selbsteinschätzung ist, wie wir aus der psychologischen Forschung wissen, in erheblichem Ausmaß durch "Response sets", soziale Erwünschtheit und "Halo-Effekte" beeinflußt.

Mit dem Kriterium "*leistungsabfordernd*" soll sichergestellt werden, dass das Meßinstrument nicht nur irgendeine kognitive Orientierung erfaßt, sondern eine bestimmte Fähigkeit, nämlich die moralische Urteilsfähigkeit. Damit die Bearbeitung eines Tests eine Leistung darstellt, muß dieser eine Aufgabe enthalten, deren Bewältigung eine Fähigkeit erfordert. Im Unterschied zu gewöhnlichen Leistungstests, bei denen die Richtigkeit der Lösung zumeist aufgrund objektiver Merkmale von außen festlegt ist, haben wir es bei moralischen Dilemmas mit Aufgaben zu tun, für die es keine "objektive", von außen festlegbare Lösung gibt (Pittel & Mendelsohn 1966). Kohlberg (1971) hat daher vorgeschlagen, die Qualität des individuellen Urteils nicht an der gewählten Entscheidung zu messen, sondern an der Art der gegebenen Argumente, und hat dafür gewichtige philosophische Gründe angeführt. Im MJI wird daher festgelegt: Je höher das Argumentationsniveau einer Person ist, desto größer ist ihre moralische Urteilsfähigkeit. Diese Festlegung ist aber auch ein externes Kriterium und ist zu rigide, um kulturellen Verschiedenheiten bei der Definition einer adäquaten Argumentationsweise und Unterschieden in der Anforderungsstruktur verschiedener Dilemmatypen gerecht werden zu können. Dilemmas, die in unserem Kulturkreis normalerweise eine moralische Diskussion auf Stufe 6 des Kohlberg-Schemas erwartet lassen, stellen in einer anderen Kultur vielleicht gar keine Stufe-6-Dilemmas dar und umgekehrt (siehe Kapitel 7). Auch ist zu erwarten, dass manche der zur Messung verwendeten Dilemmas auf tieferen Moral-Stufen adäquat diskutiert werden können.

Um diesen Einwänden gerecht zu werden, definieren wir im MUT die moralische Urteilsfähigkeit einer Person – ohne Berücksichtigung der von ihr gewählten Stufe der Argumentation – als das Ausmaß, mit dem sie sich bei ihren Begründungen (oder bei der Bewertung von Begründungen) *konsistent* an moralischen Prinzipien – anstatt zum Beispiel an Meinungskonformität – auf einer beliebigen Stufe orientiert.

Bei dieser Definition ist das Stufenkriterium nicht mehr extern bestimmt. Die befragte Person bestimmt dieses Kriterium selbst, indem sie angibt, welche Stufe des moralischen Urteils sie zur Lösung des Dilemmas bevorzugt. Extern definiert ist nur noch das Kriterium der Urteilskonsistenz, das allein in der psychologischen Meßabsicht der Untersuchung begründet ist, aber nicht in philosophischen oder kulturellen Festlegungen. Unsere Definition stellt eine Präzisierung von Kohlbergs (1964, S. 425) Definition der moralischen Urteilsfähigkeit dar, der diese als das "Vermögen" bestimmte, "*Entscheidungen und Urteile zu treffen, die moralisch sind (also auf moralischen Prinzipien gründen), und in Übereinstimmung mit diesen Urteilen zu handeln.*"

4.1 Konsistenz: Meßfehler oder Strukturmerkmal?

Die Konstruktion eines validen Instruments zur Messung der moralischen Urteilsfähigkeit setzt, wie in Kapitel 3 deutlich wurde, die Lösung eines methodologischen Dilemmas voraus (siehe oben, S. 74). Zur Lösung dieses Dilemmas ist einerseits die Frage zu beantworten, wie relationale Eigenschaften der individuellen Urteilsstruktur zu interpretieren sind: Ist Urteilskonsistenz oder -inkonsistenz, wie die klassische Testtheorie voraussetzt, ein Merkmal des Meßinstruments (als "Reliabilität" oder "Meßfehler") selbst oder, wie aufgrund der kognitiven Entwicklungstheorie angenommen werden müßte, ein Merkmal der befragten Person? Zum anderen muß, falls die zweite Interpretation zutreffend ist, eine Forschungsmethode gefunden werden, die eine Messung von relationalen Eigenschaften des moralischen Urteilsverhaltens ermöglicht. Wenden wir uns zunächst der ersten Frage zu.

Wie bereits in Kapitel 3 argumentiert wurde, besteht für die Interpretation von Konsistenz des moralischen Urteilsverhaltens als Eigenschaft des Meßinstruments aus kognitions-psychologischer Sicht wenig Plausibilität. Die Piaget-Kohlberg-Theorie legt vielmehr eine Deutung formaler Antwortmerkmale als Manifestationen der Organisation oder Struktur des moralischen Urteils nahe. Wir brauchen uns bei dieser Frage aber nicht mit Plausibilitätsüberlegungen zufrieden geben. Aus der kognitiv-strukturellen Sichtweise folgen testbare Hypothesen, die, sofern sie bestätigt werden können, die impliziten psychologischen Voraussetzungen der klassischen Testtheorie auch empirisch widerlegen würden.

Erste Hypothese: Menschen variieren nicht nur bezüglich ihrer moralischen Einstellungen, sondern auch hinsichtlich des Ausmaßes, mit dem diese ihr Urteilsverhalten bestimmen. Das heißt, das Urteilsverhalten von Menschen unterscheidet sich nicht nur hinsichtlich der – positiven oder negativen – Einstellung zu der moralischen Orientierung, die gemessen werden soll, es unterscheidet sich auch danach, wie stark diese Einstellung überhaupt das Urteilsverhalten determiniert und welche anderen moralischen oder außermoralische Einstellungen es determinieren. Das Urteilsverhalten unterscheidet sich demnach auch hinsichtlich formaler Eigenschaften wie Konsistenz und Inkonsistenz, Integriertheit und Differenziertheit. Es bestehen zwischen ihnen nicht nur, wie die klassische Meßtheorie annimmt, Richtungs- und Ausprägungsunterschiede, sondern auch Strukturunterschiede, die sich empirisch nachweisen lassen.
Zweite Hypothese: Insofern es solche Strukturunterschiede gibt, stehen sie in einem engen Zusammenhang mit der Komplexität der anzuwendenden moralischen Orientierung und damit zur Schwierigkeit des zu lösenden moralischen Dilemmas (siehe Kapitel 2).
Dritte Hypothese: Diese Strukturunterschiede spiegeln strukturelle Entwicklungsunterschiede wider, d.h. es sind Unterschiede, die in der kognitiv-moralischen Struktur des Individuums begründet liegen und durch geeignete Maßnahmen, wie beispielsweise durch schulische Allgemeinbildung, beeinflußt werden können.
Wir haben diese Hypothesen mit dem *Moralisches Urteil-Test* umfassend überprüft.[66] Ohne diese Befunde im einzelnen zu diskutieren, kann man feststellen, dass sie die Interpretation der kognitiven Entwicklungstheorie in allen Punkten bestätigen. Unsere Studien zeigen erstens eine hohe individuelle Urteilsvarianz (Lind 1985a), zweitens einen systematischen Zusammenhang zwischen der individuellen Urteilsvarianz und der Höhe der sogenannten "Test-Reliabilität" oder "internen Konsistenz", einen starken systematischen Zusammenhang zwischen der Höhe der Moralstufe im Sinne der Kohlbergschen Moralstufen und der Urteilskonsistenz (s. Kapitel 2) und drittens einen systematischen und in allen Untersuchungen bestätigten Zusammenhang zwischen der Höhe der Urteilskonsistenz und dem schulischen Bildungsniveau (Kapitel 11). Mit anderen Worten: diese Befunde widersprechen durchweg der Annahme, dass Konsistenz und Varianz des Urteilsverhaltens ein Merkmal der Messung (also bloß ein

66 Siehe Lind 1978, 1985a; Lind & Wakenhut 1985; dazu auch Kapitel 11.

Ausdruck des Meßfehlers) sind. Vielmehr bestätigen sie die Theorie, dass die individuelle Urteilsvarianz ein Strukturmerkmal der untersuchten Person widerspiegelt. Damit ist auch die Annahme bekräftigt, wonach moralische Dispositionen keineswegs bloß hypothetisch oder überflüssig sind, sondern konkret und empirisch überprüfbar.

4.2 Die Meßintention des MUT

Die Meßintention des MUT basiert auf der eingangs erwähnten Definition der moralischen Urteilsfähigkeit als *das Vermögen, moralische Entscheidungen auf der Grundlage von (selbstgewählten) moralischen Prinzipien zu treffen (statt z.B. auf Vorurteilen oder Konformität), und in bezug auf diese Prinzipien konsistent und unparteiisch zu handeln*. Mit dem MUT sollen die beiden in dieser Definition angesprochenen Aspekte gemessen werden: erstens an welchen Prinzipien bzw. Moralstufen eine Person ihr Urteil orientiert (= affektiver Aspekt) und zweitens wie stark ihre Argumentationen bzw. Bewertungen von Argumenten von den selbstgewählten Prinzipien determiniert sind (= kognitiver Aspekt). Die *affektiven Aspekte* des moralischen Urteils sind die individuellen moralischen Einstellungen, d.h. die moralischen Prinzipien, an denen der einzelne sein Urteilsverhalten ausrichtet. Wie beim MJI konzentriert sich auch beim MUT die Meßabsicht auf die sechs Stufen moralischer Orientierungen oder Perspektiven, die Kohlberg vorgeschlagen hat und die wir oben, im Kapitel 2 ausführlich dargestellt haben.[67] Bezüglich der kognitiven Aspekte, die mit dem MUT gemessen werden können (Lind 1978), steht das Ausmaß der Determination des Urteilsverhaltens durch moralische Prinzipien im Mittelpunkt. Die Determination kommt in der Konsistenz des Verhaltens *in bezug* auf diese Prinzipien zum Ausdruck: Wenn sich eine Person bei der Beurteilung von Argumenten *konsistent* an der moralischen Qualität dieser Argumente orientiert,

67 Kohlberg hat zwar in der revidierten Fassung des MJI darauf verzichtet, die Stufe 6 im Auswertungshandbuch (Colby, Kohlberg et al. 1987) aufzuführen, aber nicht darauf, sich mit ihr als einer realistischen Perspektive für die Lösung eines moralischen Dilemmas auseinanderzusetzen (Kohlberg et al. 1986). Auch wenn es nicht viele Menschen geben sollte, die alle präsentierten Dilemmas auf Stufe 6 adäquat lösen können, so ist schon lange bekannt, dass viele eine positive Einstellung zu dieser Stufe des moralischen Argumentierens haben (Rest 1973).

dann sagen wir, sie zeigt eine hohe moralische Urteilsfähigkeit. Wenn sie sich statt dessen bei der Bewertung von vorgetragenen Argumenten konsistent an der Übereinstimmung dieser Argumente mit ihrer eigenen (vorgefaßten) Meinung orientiert, dann erreicht sie nur einen niedrigen Urteilsfähigkeitswert im MUT.

Der MUT mißt beide Aspekte *gleichzeitig*, ohne sie jedoch zu vermengen, wie es das MJI und alle anderen bisher bekannten Verfahren (der DIT von Rest, das SRM von Gibbs u.a.m.) tun. Der affektive und der kognitive Aspekt des moralischen Urteilens sind logisch unterscheidbar, aber nicht dinglich trennbar. Das heißt, sie können nicht an verschiedenen Verhaltensweisen in verschiedenen Situationen beobachtet werden; sie können nur gleichzeitig, an ein und demselben Verhaltens- oder Urteilsmuster erfaßt werden. Sie sind eben zwei *Aspekte* oder Eigenschaften des individuellen Urteilens und keine *Komponenten* (s. Kapitel 2).

Für den kognitiven Aspekt allein reservieren wir den Begriff der *moralischen Urteilsfähigkeit*, da es wohl ohne Anstrengung oder Erfahrung möglich ist, hohe Stufen des moralischen Urteilens zu präferieren oder zu benutzen (weshalb der affektive Aspekt *keine* Fähigkeit darstellt), aber wohl nicht, sich konsistent an moralischen Prinzipien zu orientieren. Allerdings ist die Konsistenz des Urteils allein kein hinreichender Indikator für moralische Urteilsfähigkeit. Zum einen muß das Urteilsverhalten konsistent *in bezug* auf moralische Prinzipien sein, um als Ausdruck *moralischer* Urteilsfähigkeit gelten zu können. Einem Urteilsverhalten, das konsistent in bezug auf andere Dispositionen ist wie zum Beispiel in bezug auf die eigene Meinung zur Lösung eines Dilemmas, kann das Prädikat *moralisch* offensichtlich nicht zugesprochen werden. Zum anderen muß das Meßexperiment aus einer Aufgabe bestehen, da wir sonst nicht in der Lage sind, Konsistenz als Ausdruck einer *Fähigkeit* zu werten. Wie beim MJI und anderen Verfahren wird die oder der Befragte mit mehreren (im Standard-MUT mit zwei) Dilemmas konfrontiert. Hierdurch kann festgestellt werden, ob eine Person moralisch konsistent urteilt, das heißt immer wieder die Argumente einer bestimmten Stufe gebraucht oder akzeptiert, um ihre Meinung zu begründen bzw. abzustützen. Aber diese Konsistenz kann auch Ausdruck von Gewohnheit, Habitus, Rigidität oder ähnlichem sein. Damit sich beim Bearbeiten des MUT auch die Fähigkeit zum moralischen Urteilen erweisen kann, wird bei diesem Meßverfahren die befragte Person mit Argumenten konfrontiert, die zu ihrer eigenen Meinung konträr sind, d.h. die ihrer Meinung widersprechen.

Wenn eine Person eine moralische Position akzeptiert oder übernimmt, die ihrer eigenen vorgefaßten Meinung widerspricht, und diese Meinung im Lichte dieser Position neu bewertet und möglicherweise sogar revidiert, so hat sie zumeist unsere persönliche Hochschätzung. Das ist so, weil, wie der Philosoph Wellmer (1986) es formuliert, wir in der Art, wie Personen zu Argumenten stehen, die gegen ihre (vorgefaßte) Meinung sprechen, ein Zeichen für den Grad ihrer *Vernunftgeleitetheit* sehen. Es entspricht offenbar unserer allgemeinen Erfahrung, dass Menschen gewöhnlich dazu tendieren, an ihrer vorgefaßten Meinung festzuhalten und möglichst viele Gründe dafür suchen, ein solches Beharren zu rechtfertigen, dass sie nur selten fähig sind, sich mit Einwänden auseinanderzusetzen, die ihre Meinung in Frage stellen, und noch weniger in der Lage sind, sich von einer Gegenmeinung überzeugen zu lassen. Wer dies hingegen vermag, den schätzen wir hoch, weil solche Personen dadurch zeigen, dass sie moralische Argumente nicht benutzen, um vorgefaßte Meinungen nachträglich zu begründen, sondern dass sie ihr Urteil wirklich an eigenständigen moralischen Prinzipien auszurichten vermögen.

Wir sagen "vermögen", da diese Personen offenbar eine Fähigkeit besitzen, die andere nicht besitzen, nämlich die Fähigkeit, ihr Verhalten durch Vernunft und Überlegung leiten zu lassen, statt bloß von Intuition und Gewohnheit. Diese Fähigkeit schätzen wir vor allem deshalb, weil bei Meinungskonflikten nur durch sie eine vernünftige, diskursive, gewaltfreie Einigung ermöglicht wird. Wenn sie fehlt, läßt sich ein Konflikt nur gewaltsam, d.h. durch einen Kampf lösen, bei dem eine Meinung dominiert und die andere unterdrückt wird, ohne dass geklärt ist, welche Meinung nun die moralisch richtigere oder vernünftigere ist. Selten wird dieser Kampf auf Meinungen beschränkt, sondern richtet sich direkt gegen die Vertreter dieser Meinung. So resultieren aus der Unfähigkeit zum vernunftgeleiteten Verhalten nicht selten gewaltsame Auseinandersetzungen mit dem Ziel der Unterdrückung oder Vernichtung anderer Menschen. Eine humanitäre, demokratische Gesellschaft ist daher auf dem Prinzip der wechselseitigen Achtung und der friedlichen Konfliktlösung gegründet, die von jedem Bürger verlangt, dass er oder sie versucht,

gegensätzliche Meinungen über moralische, soziale oder politische Probleme durch vernünftige Überlegung und demokratischen Diskurs abzuwägen.[68]
Dass wir berechtigt sind, hier von einer besonderen Fähigkeit zu sprechen, ergibt sich bereits aus der bisherigen Forschung. So hat bereits Keasey (1974) empirisch belegt, dass die "Meinungsübereinstimmung" von Argumenten für das faktische Urteilsverhalten vieler Menschen eine überragende Bedeutung hat, und auch, dass die Entwicklung der Fähigkeit, die eigene Meinung im Lichte moralischer Prinzipien zu revidieren, offenbar an Entwicklungsprozesse gebunden ist. Meinungskonformität wird auch von Döbert und Nunner-Winkler (1978) als wesentlicher Faktor des psychischen "Apparats" bezeichnet. Bereits Piaget (1928) hat in seinem klinischen Interview seine Probanden mit Argumenten konfrontiert, die gegen deren jeweilige Meinung standen, um dadurch die Festigkeit ihrer kognitiven Struktur zu prüfen. Auch Kohlberg (1958) bediente sich anfänglich dieser Technik (*probing*), um den Grad der moralischen Determination des Urteilsverhaltens durch moralische Prinzipien zu prüfen (vgl. Eckensberger & Reinshagen 1980), auch wenn sie in der revidierten Fassung von Kohlbergs MJI (s. Kapitel 3) diese Funktion nicht mehr hat.

4.3 Experimenteller Fragebogen zur Erfassung psychologischer Dispositionen

Im MUT spielen meinungskonträre Argumente die Rolle eines besonderen Kriteriums für die Klärung der Frage, ob eine befragte Person sich tatsächlich an moralischen Prinzipien orientiert oder ob sie diese nur benutzt, um die eigene Meinung nachträglich zu stützen. Anlage und Auswertung des MUT sind so konzipiert, dass nur diejenige Person einen hohen Wert auf der Skala für moralische Urteilsfähigkeit erhält, die Argumente auch dann konsistent nach ihrer moralischen Qualität beurteilt, wenn diese der eigenen Meinung zur

68 Ein wesentliches Element einer auf diesen Einsichten fußenden Demokratie-Erziehung ist daher die systematische Förderung der Fähigkeit, in Dilemma-Situationen Argumente der Gegenseite zu verstehen und sich mit ihnen auseinanderzusetzen. Dabei ist es wichtig, wie Döbert (1987, S. 509) schreibt, "darauf zu insistieren, daß gerade Argumente, die der eigenen Position nicht entsprechen, stark gemacht und verteidigt werden". Vgl. auch Kohlberg 1987, Lind 1987, Oser und Althof 1992; auch Kapitel 8.

Lösung eines Dilemmas widersprechen. Dies ist die Zielsetzung, aber wie wird das Ausmaß der moralischen Determiniertheit des Urteilsverhaltens im MUT konkret gemessen?

Um diesen kognitiven Aspekt des individuellen Urteilsverhaltens unter variierenden Bedingungen (Dilemmatyp, Meinungskonformität) messen zu können, haben wir uns des Konzepts des multivariaten Experiments (statt der klassischen Testtheorie) bedient. Um hervorzuheben, dass es sich dabei nicht um ein Reiz-Reaktions-Experiment im Labor handelt, sondern um eine Befragungssituation zur Messung individueller Dispositionen, habe ich diese Methode als Experimentellen Fragebogen bezeichnet (Lind 1982). Angeregt wurde dieses Konzept durch einige methodische Ansätze in der Psychologie, die teilweise in Vergessenheit geraten sind oder in ihrer Bedeutung nicht voll erkannt wurden, so zum Beispiel Brunswiks (1955) diakritische Methode. Brunswik hat diese Methode vorgeschlagen, um ein altes Problem der psychologischen Diagnostik zu lösen: Wie kann auf experimentellem Wege unter mehreren möglichen Dispositionen jene gefunden werden, die ein Verhalten tatsächlich bestimmt? Brunswik hat seinen Vorschlag dem Vorgehen beim Experiment entlehnt, in dem mehrere unabhängige Variablen in einem geeigneten Forschungsdesign so kombiniert werden (orthogonales Design), dass ihr Einfluß auf eine bestimmte abhängige Variable unabhängig voneinander geschätzt werden kann. Brunswiks diakritische Methode unterscheidet sich aber von der üblichen Verwendung experimenteller Designs, insofern er vorschlug, nicht die untersuchten Variablen zu variieren (das ist bei Dispositionen auch schlecht möglich), sondern die externen Reize. Als Beispiel führt er folgendes Gedankenexperiment an: Ein Vogel fliegt von einem Punkt los in eine bestimmte Richtung. Will er den Mast oder den Baum anfliegen, die beide in einiger Entfernung vor ihm liegen? Wie ließe sich herausfinden, was die Absicht oder das Motiv des Vogels ist, bevor er sein Ziel erreicht? Solange Mast und Baum in einer Linie zur Position des Vogels liegen, lassen sich im Verhalten des Vogels keinerlei Anhaltspunkte für die Richtigkeit der einen oder der anderen Vermutung finden. Eine reine Beobachtung führt hier nicht zum Ziel. Der Experimentator kann nur dann eindeutige Hinweise im Verhalten des Vogels erkennen, wenn er das eine Objekt, hier den Mast, so versetzt, dass Vogel, Baum und Mast nicht mehr in einer Linie stehen. War es die Absicht des Vogels, den Mast anzufliegen, müßte er jetzt seine Flugrichtung ändern, womit auch für den Experimentator klar wäre, wohin der Vogel wirklich fliegen will.

Vielfach wird angenommen, dass mit experimentellen Designs nur Hypothesen über die Wirkung von Reizen geprüft werden können, weswegen der Einsatz von Experimenten in der Psychologie fast nur auf psycho-physische Fragestellungen begrenzt ist. Wie Brunswiks Anwendung der diakritischen Methode in bezug auf eine dispositionelle Fragestellung aber zeigt, können mit experimentellen Methoden auch Hypothesen über die Wirkung bestimmter Dispositionen im Verhalten untersucht werden, und zwar nicht nur global für mehrere Personen, sondern individuell für eine bestimmte Person. In einem Reiz-Experiment muß immer das zu untersuchende Objekt vom Experimentator variiert werden. Wenn Kausalhypothesen der Form "wenn *A* dann *B*" geprüft werden, muß "*A*" systematisch variiert werden, um den Kausalzusammenhang testen zu können. Anders liegen die Dinge, wenn eine Dispositionshypothese der Form "das Verhalten einer Person *P* ist durch die – positive oder negative – Einstellung zu "*E"* determiniert" geprüft wird. Da Dispositionen Verhaltenskonstanten sind, muß, um ihre Vorhandensein in einer bestimmten Situation feststellen zu können, die Umwelt von *"P"* systematisch in bezug auf *"E"* variiert werden. (Die in der psychologischen Einstellungsforschung übliche Annahme, dass Dispositionen selbst über längere Zeit konstant sind, ist hierfür nicht notwendig.)

Dispositionen können nicht dinglich separat vom betreffenden Verhalten erfaßt oder unabhängig davon definiert werden. Dennoch ist die Frage nach der Adäquatheit von Dispositionshypothesen nicht trivial. Wenn die Vermutung, dass das Urteilsverhalten einer Person von der Einstellung "*E*" determiniert ist, nicht zutrifft, dann wird dieses Verhalten durch einen bestimmten Wert auf einer Skala, welche die Einstellung zu *E* mißt, nicht richtig beschrieben. Oder, einfacher gesagt, der Psychologe mißversteht in diesem Fall seinen Probanden. Eine vermutete Disposition ist zunächst ein hypothetisches Konstrukt, aber sie bleibt nur solange hypothetisch, bis sie empirisch als existent nachgewiesen ist.

4.4 Das Design des Moralisches Urteil-Test (MUT)

Der Moralisches Urteil-Test ist eigentlich ein komplexes Verhaltensexperiment, mit dem für ein Individuum die Existenz und die Urteilsrelevanz von bestimmten Dispositionen nachgewiesen werden kann. Der Test wurde daher nicht nach den (eine solche Fragestellung ausschließenden) Standards der

psychologischen Testtheorie entwickelt, sondern auf der Grundlage des Konzepts des experimentellen Fragebogens (*EF*; Lind, 1982). Hierdurch ist es möglich festzustellen, an welchen der folgenden drei Faktoren (bzw. an welchen Kombinationen dieser Faktoren) sich einzelne Personen orientieren, wenn sie über Argumente für und gegen bestimmte Entscheidungen oder Verhaltensweisen urteilen:

a) An der moralischen Qualität der vorgegebenen Argumente, die verschiedene moralische Orientierungen oder Stufen im Sinne Kohlbergs repräsentieren (wir nennen diesen Faktor kurz Stufe)
b) An der Übereinstimmung oder Nicht- Übereinstimmung der vorgegebenen Argumente mit der Meinung des Probanden (= Faktor Pro-Contra oder Meinungskonformität)
c) An dem situativen Kontext der Argumentation (= Faktor Dilemma)[69]

Diese drei Faktoren bilden im Design des MUT die sogenannten unabhängigen Variablen. Es sind die Variablen, die zum Zwecke der kontrollierten Beobachtung konstruiert wurden, die also unabhängig vom Verhalten der befragten Person sind. Bei der Konstruktion des MUT wurde zudem sichergestellt, dass diese Variablen auch untereinander unabhängig oder orthogonal sind, d.h., dass ihre Ausprägungen nicht miteinander korrelieren: Die Argumente zu den Dilemmas im MUT wurden so ausgewählt, dass jede Ausprägungskombination vorkommt. Sie bilden, technisch gesprochen, ein 3-faktorielles (6 x 2 x 2) Design. Der Faktor Stufe hat sechs Ausprägungen (die sechs Kohlberg-Stufen), der Faktor Meinungskonformität zwei Ausprägungen (Pro- und Contra-Argumente) und der Faktor Dilemma im Standard-MUT hat auch zwei Ausprägungen (das Diebstahl-Dilemma und das Sterbehilfe-Dilemma). Die abhängige Variable im MUT ist die subjektive Beurteilung der Akzep-

69 Die dilemma-spezifische Argumentationsstruktur einer Person wird durch die sogenannte *Interaktion* der Faktoren "Dilemma" *und* "Qualität" (oder Stufe) angezeigt, *nicht*, wie häufig zu lesen ist, allein durch den Faktor Dilemma.

tabilität der vorgegebenen Argumente auf einer Bewertungs-Skala, die gewöhnlich von -4 bis +4 reicht.[70]

Die Darbietung des MUT sieht so aus, dass die befragte Person Verhaltensdilemmas dargeboten bekommt, zu der sie Stellung beziehen soll. Im Sterbehilfe-Dilemma wird beispielsweise gefragt: Halten Sie es eher für richtig oder eher für falsch, dass der Arzt der sterbenden Frau auf deren Wunsch eine todbringende Spritze verabreichte? Die Antwort kann die Person auf einer Bewertungsskala von - 3 (ganz falsch) bis +3 (ganz richtig) geben. (Manchmal wird auch eine Skala verwendet, die zwei Punkte länger oder kürzer ist.) Mit ihrer Antwort gibt die Person ihre Meinung zum Dilemma an. Sie spricht sich zum Beispiel offen für oder gegen die Sterbehilfe aus. Diese Meinung legt auch fest, welche der nachfolgenden Argumente für und gegen Sterbehilfe aus der Sicht der befragten Person Pro- und Contra-Argumente darstellen. Ist eine Person gegen Sterbehilfe, dann stützen die Gegen-Argumente im MUT ihre Meinung, sind also meinungskonforme Argumente. Nach diesem Teil werden gemäß dem Design des MUT der befragten Person jeweils sechs Argument für und gegen die Entscheidung des Protagonisten in der Dilemma-Geschichte präsentiert. Diese Argumente muß sie, wie oben beschrieben, auf einer Antwortskala bewerten.

Die kleinste Auswertungseinheit des MUT ist das gesamte Urteilsmuster einer Person und nicht einzelne Bewertungen, die eine Person gibt. Ausgewertet werden sowohl der affektive Aspekt wie der kognitive Aspekt des moralischen Urteilsverhaltens einer Person. Bei dem affektiven Aspekt werden gewöhnlich Maße für die Einstellung einer Person zu jeder der sechs moralischen Argumentations-Stufen nach Kohlberg gebildet. Dies geschieht in klassischer Weise durch Summierung der Einzelurteile der befragten Person oder, was gleichwertig ist, durch die Bildung von Mittelwerten.

Für den kognitiven Aspekt der moralischen Urteilsfähigkeit können mehrere, eng miteinander verwandte Maße gebildet werden, die ich an anderer Stelle (Lind 1978; 1985a) ausführlich diskutiert habe. Das am meisten verwendete Maß ist der Grad der Determination des Urteilsverhaltens einer Person durch moralische Überlegungen (kurz C-Wert, früher auch Determination Stufe

70 Manchmal werden auch Bewertungsskalen von -2 bis +2 verwendet, zumeist dann, wenn anzunehmen ist, daß die befragte Personengruppe mit der differenzierteren Skala überfordert wäre.

genannt). Der C-Wert ist definiert als der Anteil der Varianz des Faktors Stufe an der Varianz des gesamten individuellen Urteilsmusters (multipliziert mit 100). Der so ermittelte Meßwerte C hat ein Minimum von 0 (dieser Wert bedeutet, dass die befragte Person sich beim Beurteilen von Argumenten überhaupt nicht an den moralischen Prinzipien orientiert) und ein Maximum von 100 (dieser Wert bedeutet, dass sie sich ausschließlich an moralischen Gesichtspunkten orientiert). Mit der Konstruktion dieses Meßwertes ist festgelegt, dass einer Person genau dann eine hohe moralische Urteilsfähigkeit zugeschrieben wird, wenn sie Argumente nicht nur unabhängig von ihrer Übereinstimmung mit der eigenen Meinung, sondern auch unabhängig von dem jeweiligen Dilemmatyp beurteilt. Obwohl auch bei Kohlbergs Meßmethode einer Person dann eine höhere moralische Urteilsfähigkeit zugeschrieben wird, wenn ihr Urteil vom Dilemmatyp unabhängig ist, kann aus theoretischer Sicht in der Differenzierung des Urteils nach dem Dilemmatyp ein Anzeichen für höhere moralische Urteilsfähigkeit gesehen werden. Tatsächlich haben wir mit Hilfe der Quadratsummenzerlegung weitere Indikatoren gebildet, die diesen Einwand berücksichtigen (Lind 1978). In der Forschungspraxis wurde auf den Einsatz dieser komplexeren Meßwerte bislang jedoch weitgehend verzichtet, da sie noch ohne theoretischen Unterbau sind und in empirischen Vergleichen weitgehend dieselben Ergebnisse erbracht haben wie das einfachere Maß Determination Stufe (C-Wert).
Aufbau und Auswertung des MUT werden an anderen Stellen eingehend beschrieben (Lind 1984, 1985a, 2002/im Druck; Lind & Wakenhut 1985). Die Standardversion des MUT in Deutsch und in anderen Sprachen sowie Algorithmen zur Berechnung des C-Wertes können beim Autor angefordert werden.

4.5 Resümee

Der Moralisches Urteil-Test (MUT) erfüllt die eingangs genannten Anforderungen an ein Meßinstrument, das eine adäquate Prüfung der zentralen Hypothesen dieser Untersuchung erlaubt. Der MUT ist, wie wir in den vorangegangenen Abschnitten gezeigt haben, theoriekonform (valide), objektiv (replizierbar) und leistungsabfordernd.
Der MUT erfüllt die Forderung nach theoretischer Validität, da es mit ihm möglich ist, kognitiv-strukturelle Eigenschaften des individuellen Urteilsver-

haltens zu erfassen. Mit dem MUT kann festgestellt werden, ob und in welchem Maß eine Person fähig ist, sich bei der Bewertung von Argumenten auf einen moralischen Standpunkt (*Moral point of view*) zu stellen, d.h. ihr Urteilsverhalten an der moralischen Qualität dieser Argumente auszurichten statt an der Frage, ob diese Argument ihre eigene Meinung stützen oder ihr widersprechen. Die Messung der moralischen Urteilsfähigkeit mit dem MUT ist objektiv; sie kann von anderen Auswertern und in anderen Befragungssituationen in allen Einzelheiten überprüft werden. Die Messung ist nicht auf intuitive, für Verzerrungen anfällige Einschätzungen durch die befragte Person selbst angewiesen. Im MUT braucht die befragte Person nicht anzugeben, wie sie in einer ähnlichen Situation wie dem vorgegebenen Dilemma urteilen würde. Vielmehr wird durch den MUT ihr Urteilsverhalten in situ beobachtet. Da der MUT auch keine subjektive Einschätzung durch den Auswerter und dessen Theoriekenntnisse erfordert, besteht von dieser Seite keine Gefahr für die Neutralität des MUT gegenüber den in unserer Untersuchung zur Debatte stehenden Hypothesen.

Schließlich ist der MUT leistungsabfordernd. Seine Bearbeitung verlangt der befragten Person neben ihren moralischen Einstellungen eine Leistung ab. Sie erhält nur dann einen hohen Testwert auf der Skala moralische Urteilsfähigkeit, wenn sie fähig ist, Argumente konsistent nach ihrer moralischen Qualität zu beurteilen, und zwar auch dann, wenn diese Argumente ihrer eigenen, bekundeten Meinung zur richtigen Lösung eines Dilemma widersprechen. Der experimentelle Beweis für die These, dass die Testwerte des MUT fähigkeitsbedingt sind, ist Teil dieser Untersuchung. Eine Reihe der in den folgenden Kapiteln berichteten Studien befaßt sich direkt mit dieser Frage. So wird einerseits gefragt, ob sich die moralische Urteilsfähigkeit, wie sie im MUT gemessen wird, wie eine Einstellung durch eine einfache Instruktion verbessern läßt, und andererseits, ob fähigkeitsfördernde Programme einen nachweisbaren Effekt auf diesen Meßwert haben. Die leistungsabfordernde Natur des MUT wird in all diesen Studie bestätigt. Anders als der P-Wert des DIT von Rest (1979) läßt sich der C-Wert des MUT nicht durch eine einfache Instruktion verbessern (Kapitel 6). Wie in den Interventionsstudien, die mit dem MJI evaluiert wurden, zeigen auch die MUT-Werte Lerngewinne an. Die MUT-Werte scheinen im Vergleich zu den MJI-Werten sogar sensibler auf Effekte zu reagieren; so ergab im HASMU-Projekt (Oser & Schläfli 1986) die Messung mit dem MJI keinen Effekt, die Messung mit dem MUT jedoch einen

deutlichen Effekt (Kapitel 9). Beim MUT wird die Stufe, auf der das Dilemma diskutiert wird, nicht als Kriterium für hohe Urteilsfähigkeit vorgegeben. Die Probanden sind frei in der Wahl der moralischen Argumentationsebene, da bei der Wahl einer adäquaten Stufe der Dilemmalösung kaum eine objektive und unstrittige Festlegung von "außen"[71] möglich ist. Um einen hohen Testwert im MUT zu erhalten, ist nur Voraussetzung, dass die Befragten sich an der selbst gewählten Moralebene konsistent orientieren. Der Meßwert Determination durch moralische Stufe ist aufgrund dieser Konstruktion logisch unabhängig von den MUT-Meßwerten für die moralischen Einstellungen und daher auch geeignet für die Untersuchung von Fragestellungen über den empirischen Zusammenhang beider Aspekte des moralischen Urteilens (Kapitel 2). Da der MUT auf einem anderen methodologischen Konzept basiert als der MJI, wird bei diesem Meßverfahren keine Einordnung der Probanden in die sechs (oder fünf) Kohlberg-Stufen vorgenommen. Solche Ein-Stufungen mittels MUT-Werten wurden vorgeschlagen (Lind 1985a; Lind & Wakenhut 1985), um eine Vergleichbarkeit mit MJI-Studien herzustellen. Aber es treffen auf sie ähnliche Einwände zu wie auf den MJI. Die für eine Ein-Stufung erforderliche Vermengung von Meßwerten für den affektiven und kognitiven Aspekt des moralischen Urteilens würde viele Fragestellungen ausschließen. Zudem gibt ein direkter Vergleich von MUT- und MJI-Werten, wie Kohlberg (1985) ihn vorschlug, keinen Aufschluß über die Validität des MUT. Dazu müßte die Validität des MJI selbst zweifelsfrei gesichert sein, was aber, wie wir in Kapitel 3 sahen, nicht der Fall ist. Schmied (1981) fand eine mäßig hohe Korrelation zwischen MUT Werten einerseits und MJI-Werten andererseits. Es ist jedoch unklar, welche Schlüsse aus diesem Befund zu ziehen sind. In dieser Studie wurden nur Oberschüler und Studierende untersucht, so dass die Varianz beider Meßwerte und damit auch die Korrelation zwischen ihnen a priori eingeschränkt war. Auch wurde dort die alte Version des MJI benutzt, die selbst nur mäßig hoch (r = .30) mit der neuen Version korreliert (Kohlberg & Candee 1984; s. auch Oser & Althof 1992, S. 75).

Der MUT liegt inzwischen in mehreren Sprachen vor und ist in zahlreichen Untersuchungen in unterschiedlichen Ländern eingesetzt worden (Kapitel 7). Es

71 Diese Wahl spiegelt eher kultur- und aufgabenspezifische Einstellungen wider als ein bestimmtes Niveau der moralischen Urteilsfähigkeit. Siehe Kapitel 7; auch Edwards 1986; Gielen 1986; Locke 1983; Sullivan 1977.

wurden damit Akademiker ebenso wie Hauptschulabsolventen untersucht, ältere Menschen ebenso wie sehr junge. Die jüngsten mit dem MUT untersuchten Personen sind 11-jährige Haupt- und Realschüler. Bei Personengruppen mit geringen Bildungserfahrungen wurde teilweise das Layout (Schriftgröße), die Bewertungsskala (von -2 bis +2) und die Instruktion angepaßt. Wir haben keine Beeinträchtigung der Validität der Meßwerte durch diese Änderungen feststellen können. Gleichwohl muß damit gerechnet werden, dass solche Änderungen, insbesondere aber die Darbietungsform (mit versus ohne Zeitlimits), das generelle Niveau der Testwerte beeinflussen können. Vergleiche über verschiedene Untersuchungen hinweg müssen dies immer berücksichtigen. Mit Ausnahme der EMNID-Untersuchung (Kapitel 5) wurde der MUT bislang ohne Zeitlimit dargeboten.
Zur zusätzlichen, empirischen Sicherung der Validität von veränderten und von in fremde Sprachen übersetzten Versionen haben sich drei Analysen bewährt (Lind 1984; 2002):

1. Die Analyse der Präferenzhierarchie. Gemäß der Theorie sollten die Befragten Stufe-6-Argumente am stärksten akzeptieren, Stufe-5-Argumente etwas weniger usw. und Stufe-1-Argumente am stärksten ablehnen.[72]
2. Die Analyse der Interkorrelationen zwischen den Einstellungen zu den sechs Moralstufen, die nach Kohlberg (195 8) eine Quasi-Simplex-Struktur bilden sollen.
3. Die Analyse der Korrelation zwischen den Einstellungen der Probanden zu den sechs Stufen auf der einen Seite und dem MUT-Meßwert für moralische Urteilsfähigkeit (s. Kapitel 2) auf der anderen. Starke Abweichungen von den prognostizierten Werten waren bislang ausnahmslos durch technische Fehler bei der Datenauswertung oder der Übersetzung bedingt; sie verschwanden, sobald diese Fehler behoben waren.

72 Bei klassisch moralischen Dilemmas wie dem Sterbehilfe-Dilemma bestätigt sich diese Annahme fast ausnahmslos – unabhängig von Merkmalen wie Geschlecht, Alter, Schulbildung, Kulturkreis etc. Bei moralisch weniger anspruchsvollen Dilemmas, wie beim Diebstahl-Dilemma, präferieren die Befragten zumeist eine Argumentation auf Stufe 5 stärker als eine auf Stufe 6. Bezüglich der anderen Stufen bleibt auch hier die Rangheihe weitgehend gewahrt (Lind 1984; 1985a).

5 Moralische Bildung versus moralische Reifung: Das Problem der Regression

Die Bildungstheorie und die Reifungstheorie der moralischen Entwicklung unterschieden sich am stärksten hinsichtlich der Frage, ob sich die moralische Urteilsfähigkeit zurückentwickeln kann. Die Unterschiede zwischen diesen beiden Erklärungsansätze einerseits und der *Sozialisationstheorie* andererseits liegen weniger deutlich auf der Hand, da nach dieser Theorie moralisches Verhalten fast ausschließlich auf moralischen Einstellungen basiert, die in relativ kurzer Zeit verändert werden können, und zwar sowohl in Richtung auf eine größere Akzeptanz bestimmter Moralprinzipien wie auch in Richtung auf eine größere Ablehnung. Die beiden anderen Theorien nehmen an, dass das moralische Verhalten stärker durch seinen kognitiven Aspekt bestimmt wird, der sich weniger schnell ändern läßt.

Die *Reifungstheorie*, wie sie Kohlberg vertritt,[73] sagt voraus, dass das moralische Urteil sich immer von tieferen Stufen zu höheren entwickelt, dass also die Entwicklung der moralischen Urteilsfähigkeit *invariant* ist und es *keine* Regression gibt. Regressionen können, so die Theorie, nur in seltenen Fällen, etwa bei Vorliegen einer starken Beeinträchtigung der geistigen Funktionen auftreten. Die invariante Entwicklungssequenz ist eine Kernhypothese der Reifungstheorie; ihre Widerlegung würde, wie ihre Vertreter betonen, auch die gesamte Theorie in Frage stellen. Die Hypothese einer invarianten Entwicklung des moralischen Urteils besitzt in zweifacher Hinsicht eine hohe Attraktivität – und auch eine große Problematik. Zum einen begründet sie einen großen Optimismus bezüglich der individuellen Entwicklung und der Menschheitsentwicklung insgesamt: Wenn keine Rückentwicklung der Moral eintreten kann, dann ist mit einer stetig anwachsenden Steigerung von Gerechtigkeit und Humanität in der Welt zu rechnen. Offensichtlich widersprechen aber die tatsächliche Menschheitsgeschichte wie auch viele Biographien dieser optimistischen Erwartung. Ungerechtigkeiten und Inhumanität sind, wie jede Nachrichtensendung zeigt, nicht geringer, sondern in letzter Zeit eher wieder größer geworden. Weshalb ist das der Fall? Was wäre, wenn die Theorie nicht stimmt? Dann wäre ein solcher, bedingungsloser Optimismus fatal. Er würde uns auf

73 Siehe Kohlberg 1958; 1984; Colby, Kohlberg et al. 1987; Walker 1986.

dem Gebiet der moralischen Bildung zum Nichtstun verleiten und so nachhaltig zum Abbau der Moral beitragen. Zum andern besitzt die Hypothese einer invarianten Sequenz einen im Vergleich zu vielen anderen psychologischen Theorien sehr hohen Informationsgehalt, d.h., sie trifft eine Vorhersage von großer Präzision[74]. Zwar läßt die Kohlberg-Theorie zu, dass eine Person auch Argumente einer niedrigeren Moralstufe benutzt als die, auf der sie argumentieren *kann*. Aber eine Abnahme der moralischen Urteils*kompetenz*, d.h. eine Regression von einer höheren Stufe zu einer tieferen, gar von Stufe 5 nach Stufe 2, wie sie von Kohlberg und Kramer (1969) berichtet wurde, wird von der Theorie ausgeschlossen.

Nach der *Bildungstheorie* ist dagegen zu erwarten, dass die Entwicklung der moralischen Urteilsfähigkeit nur solange eine stetige Progression aufweist, wie sie durch (institutionalisierte) Bildungsprozesse in Elternhaus, Schule, Hochschule oder Beruf gefördert wird. Bei deren Wegfall ist zu erwarten, dass die Moralentwicklung stagniert oder – bei längerer Bildungsabstinenz – sich zurückentwickelt, also regrediert. Nach dieser Theorie führt – im moralischen Bereich ebenso wie in anderen Fähigkeitsbereichen – das Fehlen von pädagogischen Hilfen und Anwendungsmöglichkeiten zur Rückentwicklung der Fähigkeit, moralische Prinzipien zur Lösung von Konflikten einzusetzen, führt aber nicht notwendigerweise auch zur Aufgabe moralischer Werte. Wie die Sozialisationstheorie sagt auch die Bildungstheorie voraus, dass moralische Ein-

Vorhersage der...	bezüglich des ... ...kognitiven Aspekts	...affektiven Aspekts
Sozialisationstheorie	[keine Aussage]	Veränderung in alle Richtungen möglich
Reifungstheorie	Keine Regression möglich	[keine Aussage]
Bildungstheorie	Regression bei Wegfall von Bildungserfahrungen	Veränderung in alle Richtungen möglich

Abb. 5.1 Forschungsfrage: Ist eine Regression des moralischen Urteils möglich?

74 Zum Konzept des Informationswerts oder *Gehalts* von Theorien s. Popper 1968; zur Bewertung des Gehalts der Kohlberg-Theorie s. Lind 1985a.

stellungen infolge sozialer Einflüsse sich relativ kurzfristig verändern können, und zwar sowohl in positiver wie in negativer Hinsicht. Die Vorhersagen der Bildungstheorie sind jedoch komplexer. Nach ihr ist es denkbar, dass eine Person ein starkes Gerechtigkeitsempfinden entwickelt, während ihre Fähigkeit abnimmt, in konkreten Konflikt-Situationen eine gerechte Lösung zu finden. Diese Konstellation kann etwa dann entstehen, wenn die soziale Umgebung mit Nachdruck moralische Werte vermittelt, den Heranwachsenden aber keine entsprechenden Lernmöglichkeiten offeriert, so dass dieser zwar Ungerechtigkeiten erkennen, aber nicht adäquat dagegen angehen kann. Es bleiben ihm nur Resignation oder Gewalt als Ausweg. Die folgende Tabelle faßt die unterschiedlichen Vorhersagen der drei untersuchten Theorien bezüglich möglicher Regressionen zusammen:

In diesem Kapitel soll das Regressionsphänomen auf der Basis des Zwei-Aspekte-Modells der Moralentwicklung eingehend analysiert werden. Aufgrund der Bildungstheorie versuchen wir die Annahme zu belegen, dass es sich bei den von Kohlberg und Kramer (1969) festgestellten Regressionen um "Pseudo-Regressionen" handelt, die ein Artefakt der unzulänglichen Meßmethode darstellen, und dass eine wirkliche Regression der moralischen Urteilsfähigkeit eintreten kann, wenn die sie stützenden Bildungsprozesse für längere Zeit fortfallen. Wir vermuten daher, dass die vielfach beobachtete Altersabhängigkeit der moralischen Entwicklung kein Indiz für das Wirken einer universellen Entwicklungslogik ist, sondern durch die Tatsache erklärt werden kann, dass in diesen Untersuchungen Alter und Bildung immer *konfundiert* waren, der Effekt des Alters also nicht von dem Effekt der Bildungserfahrungen unterschieden werden konnte. Durch die Bildungstheorie wird keineswegs die Kernidee von der kognitiven Natur der Moralentwicklung in Frage gestellt, wonach moralisches Verhalten nicht nur auf moralischen Einstellungen basiert, sondern auch und vor allem auf einer moralischen Urteilsfähigkeit. Bezweifelt wird aber die Gültigkeit der Emergenztheorie, wonach sich diese Fähigkeit ohne äußere Unterstützung, allein aufgrund altersmäßiger Reifung entwickeln soll. *In Übereinstimmung mit der Bildungstheorie sagen wir eine Stagnation oder gar eine Regression moralisch-demokratischer Kompetenzen bei Menschen voraus, deren sozialer Kontext sie nicht ausreichend stimuliert oder sie gar an den für die Moralentwicklung notwendigen diskursiven Prozessen hindert.*

5.1 Zur Deutung des Regressionsphänomens

Die Kontroverse um das Regressionsphänomen wurde ausgelöst durch (erwartungswidrige) Fälle von Regression bei normalen jungen Erwachsenen, die in der Längsschnittuntersuchung von Kohlberg und Kramer (1969) aufgetreten sind. Sie berichten, dass "zwischen dem Ende der High school und dem zweiten oder dritten Jahr im College bei 20% unserer Untersuchungsgruppe aus der Mittelklasse die Meßwerte für moralische Reife sanken oder regredierten" (S. 109). Dieser Befund löste eine heftige Diskussion unter Anhängern und Kritikern der Kohlberg-Theorie aus, da er offenbar nicht mit der Annahme einer stufenweisen, invarianten kognitiv-moralischen Entwicklung vereinbar ist. Aufgrund der Studie von Kohlberg und Kramer (1969) und anderer Autoren wurde von einigen die Annahme als widerlegt angesehen, dass die Kohlberg-Stufen "harte Stufen" im Sinne von Piaget seien. Es wurden daher verschiedene Revisionen und Alternativen zu Kohlbergs ursprünglichem Stufenschema vorgeschlagen.[75]

Kohlberg hingegen sah in diesen Regressionsfällen zunächst weder einen Widerspruch zur kognitiven Entwicklungstheorie, noch stellte er seine Meßmethode in Frage. Statt dessen versuchten er und seine Mitarbeiter, diese Fälle theoretisch als intervenierende, außermoralische Prozesse zu erklären. Dabei stützten sie sich vor allem auf Perrys (1970) Studien zur ethischen und kognitiven Entwicklung und Eriksons (1966) Theorie der Identitätskrise und Ich-Entwicklung.[76] Nach Ansicht von Kohlberg und Kramer (1969) kommt die moralische Entwicklung im Sinne struktureller Transformationen gegen Ende der Kindheit zum Stillstand, während Ich-Funktionen immer wichtiger werden. Demnach ist die Regression nicht strukturell bedingt, sondern nur eine temporäre, funktionale Phase "im Dienste des Ichs". Nach dieser Sichtweise ist in dieser Lebensphase der Identitätskrise und des "psychosozialen Moratoriums", während der "neue und non-konforme Denk- und Verhaltensmuster ausprobiert werden" (Kohlberg & Kramer 1969, S. 116), eine Regression als normal anzusehen. Dies ist eine interessante Spekulation, aber sie hat den Nachteil, dass

75 Unter anderem von Broughton 1970; Eckensberger & Reinshagen 1980; Gibbs 1977; Kohlberg 1973; Murphy & Gilligan 1980; Rest 1984 und Lind 1985b.

76 Hieran knüpfen auch Döbert und Nunner-Winkler (1975) in ihrer Studie zur Adoleszenzkrise und Identitätsentwicklung an.

eine solche eklektische und etwas willkürliche Fusion von verschiedenen Theorien die Erklärungskraft der kognitiven Entwicklungstheorie als Ganzes gesehen schwächt. Da Identitätskrisen und damit auch Regressionen jederzeit auftreten können, sinkt der Informationswert dieser Erklärung auf null: Nach dieser Theorie kann jedes Ereignis eintreten, keines wird mehr ausgeschlossen.

In der Diskussion um das Regressionsphänomen haben sich zwei andere Argumentationslinien als beständiger erwiesen. Die eine sieht vor, das Kohlbergsche Stufenmodell zu modifizieren; die andere, das Kohlbergsche Schema als gemeinsamen Bezugsrahmen beizubehalten, aber die Meßmethoden zu modifizieren. Zur ersten: Eckensberger[77] argumentiert, dass die Befunde von Kramer und Kohlberg (1969) und anderen nicht nur die Hypothese der invarianten Entwicklungssequenz widerlegen, sondern das gesamte Stufenschema Kohlbergs. Als Konsequenz müßten wir das ursprüngliche Kohlbergsche Stufenmodell zurückweisen und ein neues Modell entwerfen, innerhalb dessen das Regressionsphänomen adäquater erklärt werden kann.[78] Aber u.E. liegt an diesem Punkt kein zwingender Grund vor, um Kohlbergs Stufenmodell, das unser Verständnis der menschlichen kognitiv-moralischen Entwicklung beträchtlich erweitert hat, insgesamt zu verwerfen. Kohlbergs sechs Stufen sind "Idealtypen", die als Grundlage für die Theoriekonstruktion und -messung dienen, aus denen unterschiedliche Hypothesen über den Verlauf der moralischen Entwicklung abgeleitet werden können, die Hypothese einer invarianten Sequenz ebenso wie die Hypothese eines bildungsbedingten Anstieges der Urteilsfähigkeit. Durch Regressionen wäre aber nur die Invarianz-Hypothese widerlegt, nicht aber das Stufen-Modell Kohlbergs selbst.[79] Zudem konnte bislang für keines der alternativen Modelle gezeigt werden, dass es sich besser oder auch nur annähernd so gut bewährt hat wie Kohlbergs Stufen-Modell.

Das bringt uns zur zweiten, bereits erwähnten Argumentationslinie und zu einem neuen Problem, nämlich der Wahl zwischen zwei sehr divergenten Inter-

77 Eckensberger 1983; 1984; Eckensberger & Reinshagen 1980.

78 Für solche alternativen Modelle siehe Gibbs 1977; Haan 1978; Eckensberger & Reinshagen 1980; Murphy & Gilligan 1980.

79 Wir würden auch nicht annehmen, das Konzept der Länge sei falsifiziert, weil wir beobachtet haben, dass der Umfang eines Objektes schrumpft oder dass sich eine bestimmte Distanz nicht in natürlichen Situationen finden läßt. Ob wir das Schrumpfen oder auch die ungleiche Verteilung von Distanzen beobachten, hängt im Gegenteil von einem a priori (nicht-empirischen) Längenbegriff ab.

pretationen der Daten von Kohlberg und Kramer. Sollen wir diese Daten als Indiz für eine echte Regression in der moralischen Entwicklung interpretieren oder aber als Pseudo-Regression, die sich infolge eines systematischen Meßfehlers einstellt, durch den der Verlauf der kognitiv-moralischen Entwicklung umgekehrt wird? Auch hier ist es m.E. ratsam, zunächst nach methodologischen Unvollkommenheiten zu suchen, denen auf der Grundlage einer unveränderten Theorie abgeholfen werden kann, anstatt die Theorie zu ändern oder ihr Hilfsannahmen anzufügen, die ihren *Gehalt* (den Informationswert) reduzieren. Wenn wir Mängel in der Meßmethode als Ursache für Regressionen ansehen, so bedeutet dies nicht, dass nur solche Methoden akzeptabel sind, die zu den erwünschten Ergebnissen führen. Dadurch würde die Theorie immun gegenüber empirischer Kritik, und damit wertlos. Entsprechend problematisch ist Kohlbergs (1976) Vorschlag, die Theorie zu retten, indem man Individuen den Stufen so zuordnet, "dass das Kriterium der sequentiellen Entwicklung erfüllt wird" (S. 47). Vielmehr sollte es darum gehen, die *theoretische Validität* der Erhebungsmethode, d.h. die Übereinstimmung des Konstruktions- und Auswertungsdesigns mit den theoretischen Annahmen zu verbessern. Nur wenn sich zeigen würde, dass eine Verbesserung der theoretischen Validität der Methoden die empirischen Regressionen eliminiert, könnte die Hypothese der invarianten Sequenz als "gerettet" bezeichnet werden.

Anläßlich der Überprüfung seiner früher verwendeten Forschungsmethoden identifizierte Kohlberg verschiedene methodische Ursachen für das Auftreten von Anomalien in den Daten. In einem Fall lag die Ursache im ursprünglichen Auswertungssystem, das seiner Meinung nach nicht klar genug zwischen "Inhalt" und "Struktur" unterscheiden konnte.[80] Kohlberg und seine Kollegen in Harvard revidierten deshalb das Kodierungssystem vollständig (Kapitel 3). Auch wenn die Ergebnisse nun mit der Hypothese der invarianten Sequenz besser übereinstimmen, gibt es noch immer einige Personen, die regredieren.[81] Allein diese Tatsache würde schon genügen, um die Hypothese zu widerlegen, weil, wie Kohlberg (1973) schrieb, "ein einziger Fall der Umkehrung der zeitlichen Abfolge widerlegt die Stufentheorie, wenn es sich dabei nicht um einen manifesten Fall von Meßfehler handelt" (S. 182).

80 Kohlberg 1973, S. 191; 1984.

81 Colby, Kohlberg et al. 1987; Murphy & Gilligan 1980.

Um dem Phänomen der Regression gerecht zu werden, griffen Kohlberg und seine Mitarbeiter das psychometrische Konzept des zufallsbedingten Meßfehlers auf, der auf der Basis einer Testwiederholung innerhalb eines kurzen Zeitraums geschätzt wird (Retest-Reliabilität). Auf dieser Grundlage konnten sie in ihrer Langzeitstudie zeigen, dass die Anzahl von Abwärtsbewegungen in einem Zeitraum von 3 bis 4 Jahren eindeutig geringer ist als die Anzahl der Veränderungen innerhalb eines Monats (Colby et al. 1983). Das ist ein eindrucksvolles Ergebnis, wenn man die Kritik psychometrisch orientierter Psychologen an Kohlbergs Ansatz bedenkt (Kurtines & Greif 1974). Die psychometrische Psychologie beruht jedoch, wie wir in Kapitel 3 zeigten, auf Annahmen, die selbst problematisch sind, da sie der kognitiven Entwicklungstheorie widersprechen. Außerdem läßt sich bezweifeln, ob die neue Auswertungsmethode tatsächlich eindeutiger zwischen Inhalt und Struktur unterscheiden kann.[82]

Bleibt die Frage, ob die beobachtete Regression nicht vielleicht einen bisher nicht identifizierten Meßfehler reflektiert? Die Bedeutung dieser Frage sowohl für die kognitive Entwicklungstheorie als auch für praktische Erziehungsprobleme führt uns noch einmal zu dem Problem der methodologischen Adäquatheit. Möglicherweise haben wir uns so stark auf Probleme der Indexberechnung und Datenanalyse konzentriert, dass wir andere mögliche Meßfehlerquellen übersehen haben. Es gibt Hinweise, dass die "Regressionen" nicht so sehr auf Kodierungsprobleme als vielmehr auf Mängel des Erhebungsverfahrens zurückgehen. Schon 1970 wies Broughton auf Probleme bei der Erstellung von moralischen Interviews hin und identifizierte "einen allgemeinen Mangel an insistierenden Nachfragen" (*probing*). Wenn es, dadurch bedingt, an relevanten Informationen fehlt, kann der Psychologe bei der Klassifizierung von Begründungen gemäß den moralischen Stufen leicht in die Irre geführt werden. Beispielsweise werden die Antworten von Instrumentalisten gewöhnlich der Stufe 2 zugeordnet. Wie aber Döbert und Nunner-Winkler (1975, S. 126-131) gezeigt haben, steht Instrumentalismus nicht eindeutig in Zusammenhang mit Stufe 2. Einige untergeordnete instrumentelle Begründungen lassen sich auch bei Post-Konventionellen (Personen auf Stufe 5 oder 6) finden – wenn die Interviews so durchgeführt werden, dass sie genügend Information enthalten. Zum Beispiel können instrumentelle Begründungen von "zynisch" argumentierenden Post-

82 Siehe Kapitel 3; auch Eckensberger 1983, Lind 1985b, Krebs et al. 1990.

Konventionellen oder von Personen gegeben werden, die ihr moralisches Interesse in bestimmten Situationen wie in der Interviewsituation absichtlich untertreiben.

Colby et al. (1983) fanden bei der Reanalyse der Original-Längsschnittdaten von Kohlberg ebenfalls Hinweise auf "moralischen Zynismus und Relativismus" (S. 72), die ihrer Ansicht nach für die Frage der moralischen Entwicklung irrelevant sind. Während jedoch einige Autoren Tiefstapeln und Zynismus einzig und allein als Indiz für die "metaethische Position" einer Person ansehen, erkennen wir darin auch ein Produkt der Erhebungsmethode. Auch wenn die präsentierten moralischen Dilemmas im MAI rein hypothetisch sind, so nehmen sie doch normalerweise den Befragten gefühlsmäßig in Anspruch, und es ist anzunehmen, dass sie im allgemeinen jemanden dazu veranlassen, auf der höchstmöglichen Stufe des moralischen Urteils zu argumentieren. Möglicherweise ist jedoch in einigen Fällen die Interviewmethode nicht geeignet, dieses Engagement hervorzurufen, mit dem Ergebnis, dass die Konversation künstlich bleibt. In solchen Fällen wird der Befragte, um seine Meinung zu dem Dilemma zu verteidigen, vielleicht lieber "wohltönende" Begründungen vermeiden. Wenn man jedoch diese Interviews liest, fragt man sich, ob der Befragte bei seinen instrumentalistischen Begründungen bleiben würde, wenn man ihn bitten würde, ihre Nicht-Akzeptabilität mit seinen eigenen Begründungen oder denen anderer Personen auf Stufe 2 zu vergleichen. Wenn er nicht wirklich instrumentelle Moralität einem Prinzipienargument vorziehen würde, könnte man die instrumentalistischen Antworten als Hinweis auf nichtmoralische Faktoren (wie z.B. Zynismus oder Untertreibung) oder als Zeichen für eine höchst differenzierte kognitiv-moralische Struktur werten. Es ist auch irreführend, lediglich zu zählen, wie *oft* ein stufentypisches Anliegen geäußert wurde. Wahrscheinlich würden wir entdecken, dass die Häufigkeit, mit der eine Aussage gemacht wird, nicht immer mit ihrer subjektiven Bedeutsamkeit korreliert. Wenn beim Interview oder der Berechnung des individuellen Gesamtwertes diese Ambiguität nicht mit in Betracht gezogen wird, können die nichtmoralischen Äußerungen als moralisches Urteilsverhalten mißverstanden werden und fälschlicherweise den Eindruck von Regression aufkommen lassen. Um die Erklärung, dass die Regression auf einen Meßfehler zurückgeht, zu rechtfertigen, bedarf es daher neuer Untersuchungen.

5.2 Pseudo-Regression: Überprüfung durch eine Längsschnittstudie bei Studierenden

Wenn unsere Erklärung des von Kohlberg und Kramer (1969) berichteten Regressionsphänomens als *Pseudo*regression richtig ist, dann sollten folgende drei Hypothesen empirisch bestätigt werden können: 1. Bezüglich des kognitiven Aspekts des moralischen Urteils tritt *keine Regression* ein. Universitätsstudenten in einem höheren Semester sollten eine höhere moralische Urteilsfähigkeit aufweisen als jüngere Studierende oder gar Schüler. 2. Bezüglich des affektiven Aspekts hingegen können – zumindest bei einigen Studierenden – Veränderungen auftreten, die vom Standpunkt des Kohlberg-Ansatzes aus als Regression gedeutet werden können: Im Laufe des Studiums erhöht sich die Akzeptanz präkonventioneller, instrumenteller Argumente und die Akzeptanz postkonventioneller Argumente wird weniger, allerdings nicht weniger als die Akzeptanz präkonventioneller Argumente. 3. Außerdem beschränken sich solche Fälle rein "affektiver Regression" – wie in der Studie von Kohlberg und Kramer (1969) – auf Studierende mit den höchsten Werten für moralische Urteilsfähigkeit.

Die Überprüfung dieser Hypothesen kann offenbar nicht mit dem MAI durchgeführt werden, da dieses Meßverfahren keine Unterscheidung zwischen kognitiven und affektiven Aspekten erlaubt (Kapitel 3). Daher wurde der *Moralisches Urteil-Test* (MUT) eingesetzt, der eine unabhängige Messung beider Aspekte ermöglicht (Kapitel 4). Der MUT ist zudem neutral gegenüber den drei Theorien, die es zu überprüfen gilt, d.h. sein Aufbau und seine Auswertung bevorzugen oder benachteiligen keine dieser Erklärungsansätze. Prüfdaten stehen uns aus dem Forschungsprojekt Hochschulsozialisation zur Verfügung.[83] Es handelt sich um Längsschnittdaten von 844 männlichen und weiblichen deutschen Studierenden im ersten und fünften Semester aus den Fachbereichen Medizin, Germanistik, Wirtschafts-, Natur-, Ingenieurwissenschaften und Sozialarbeit.

Die Ergebnisse dieser Studie zeigen bei den Studierenden im allgemeinen keine Regression im Hinblick auf die kognitive Struktur des moralischen Urteilsverhaltens. Die moralische Urteilsfähigkeit nimmt in dieser Zeit zu. (Diese progressive Veränderung setzt sich in den weiteren Semestern fort, wie in Ka-

83 Für weitere Angaben zu diesem Projekt siehe Kapitel 1.

pitel 11 gezeigt wird.) Die Anzahl der Studenten, deren Urteilsverhalten ein hohes Maß an Determination durch moralische Überlegungen erkennen ließ (Skalenwerte von 50 bis 100), steigt von 19.4 auf 22.1% an. Das ist kein geringer Anstieg, wenn man den großen Entwicklungsbereich der Kohlberg-Stufen in Betracht zieht.

Bei einem Teil der Studierenden zeigt sich, wie in den Daten von Kohlberg und Kramer (1969), eine "Regression", die sich aber auf den affektiven Aspekt, d.h. auf die moralischen Einstellungen beschränkt. Bei diesen wird die Präferenz für präkonventionelle Begründungen stärker, während die Akzeptanz postkonventioneller Argumente etwas geringer wird. Diese Fälle von "affektiver Regression" beschränken sich, wie in der Untersuchungsgruppe von Kohlberg und Kramer, fast vollständig auf Studierende mit der höchsten moralischen Urteilskompetenz. Das wird deutlich, wenn wir die Gesamtstichprobe in drei Gruppen mit ursprünglich niedrigen (0-9; Stichprobengröße: N= 36), mittleren

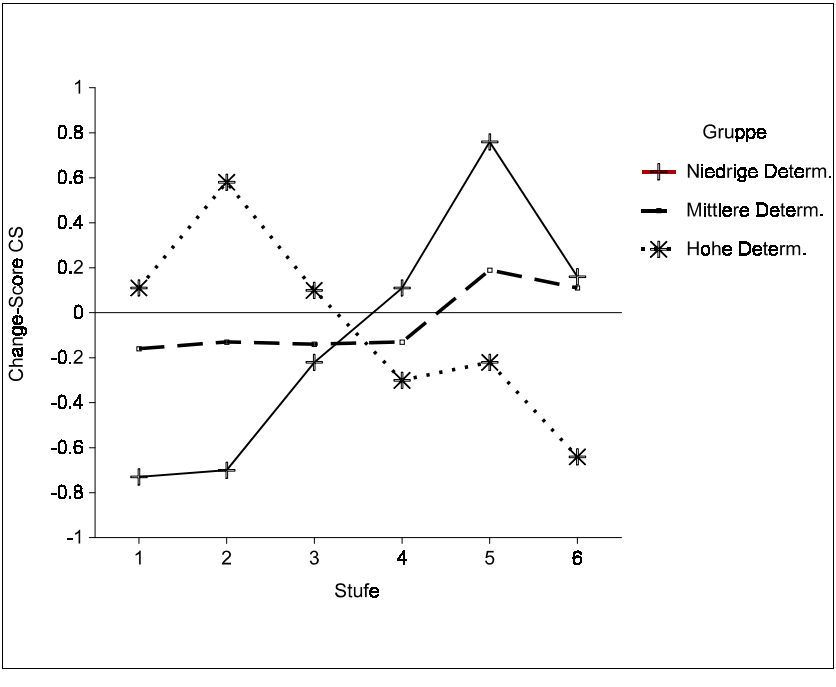


Abb. 5.2 Relative Veränderung der Akzeptanz der sechs Stufen des moralischen Urteils vom ersten zum fünften Semester bei drei Gruppen mit unterschiedlich hoher Urteilsfähigkeit (C-Wert MUT)

(10-49; N= 496) und hohen Skalenwerten (50-100; N= 312) für die moralische Urteilsfähigkeit aufteilen und getrennt analysieren, wie sich die Akzeptabilitätsstruktur der sechs Stufen bei unseren Versuchspersonen verändert. Zu diesem Zweck haben wir für jede der sechs Moralstufen aus der Anzahl der Anstiege (I) und Rückgänge (D) der Akzeptabilität mit folgender Formel einen

Change Score (C) gebildet: $CS = \frac{I - D}{I - D}$

Ein positiver *CS*-Wert bedeutet einen überwiegenden Akzeptabilitätsanstieg, ein negativer Wert einen überwiegenden Rückgang. Wie Abb. 5.2 zeigt, treten nur in der "hohen" Gruppe Regressionen auf. Die "mittlere" und – noch deutlicher – die "niedrige" Gruppe dagegen weisen ein Veränderungsmuster auf, das in Einklang mit der kognitiven Entwicklungstheorie steht. In diesen beiden Gruppen geht die Präferenz für Argumente auf *niedrigen* Stufen zurück,

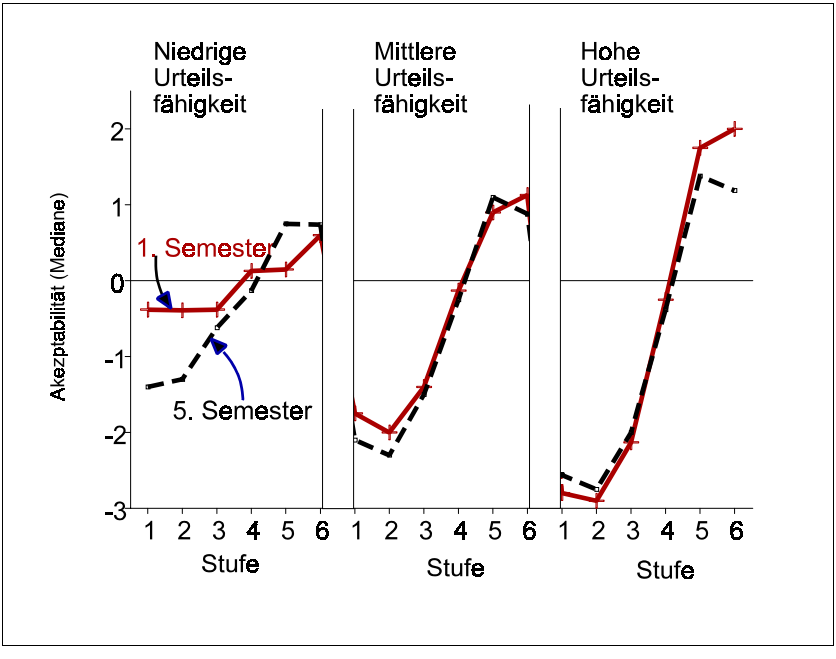


Abb. 5.3 *Absolute* Veränderung der Akzeptanz der sechs Stufen des moralischen Urteils vom ersten zum fünften Semester bei drei Gruppen mit unterschiedlich hoher moralischer Urteilsfähigkeit

während die Akzeptanz postkonventioneller Anliegen zunimmt. Betrachten wir die *absoluten* Veränderungen, dann zeigt sich, dass auch in der Gruppe mit der hohen Urteilsfähigkeit die "Regression" nicht dramatisch ist. Obwohl die Werte aus entwicklungstheoretischer Sicht abfallen, liegen sie in dieser Gruppe bei der zweiten Messung durchweg höher als in den beiden anderen Gruppen: ihre Präferenz für hohe Argumentationsstufen ist größer, und für niedrige geringer als bei den anderen. In der Abbildung sind die Einstellungsprofile für die sechs Stufen der ersten und zweiten Messung getrennt nach den Studierenden mit niedriger, mittlerer und hoher Urteilsfähigkeit (Grad der Determination des Urteilsverhaltens durch moralische Prinzipien; s. Kapitel 4) dargestellt. Da die Veränderungen gruppenspezifisch sind, können sie nicht als Effekt der statistischen Regression gegen den Mittelwert abgetan werden. Um Informationen über die – lineare und nichtlineare – Assoziation zwischen der ursprünglichen kognitiv-moralischen Struktur und der Veränderung der Struktur der Stufenakzeptabilität im Zeitraum vom ersten bis fünften Semester zu erhalten, haben wir die Daten einer multivariaten Varianzanalyse unterzogen. Unsere Hypothesen lauten: 1. Im allgemeinen wird die Kurve, welche die Akzeptabilität der einzelnen Stufen miteinander verbindet, systematisch steiler. 2. In den Gruppen mit ursprünglich niedriger moralischer Urteilsstruktur unterscheidet sich die Änderung im Urteilsprofil – die den Grad der moralisch-kognitiven Entwicklung angibt – in einer vorhersagbaren Art und Weise von der Veränderung bei den Urteilsfähigeren. Um die Ergebnisse mit denen anderer Untersuchungen vergleichbar zu machen, haben wir die Determinationskoeffizienten und die Korrelationskoeffizienten für alle Trends und Effekte errechnet. Die lineare Korrelation zwischen der Gruppenzugehörigkeit und dem Veränderungsmuster erweist sich als hoch bedeutsam ($r = 0{,}39$). Eine zusätzliche, kleinere Korrelation im kubischen Polynom ($r = 0{,}11$), die den S-förmigen Anstieg abbildet, stimmt mit der Hypothese ebenfalls überein. Sie ist aufgrund der beschränkten Größe der Antwortskalen zu erwarten, die gewöhnlich zu Deckeneffekten (Ceiling-Effekten) auf beiden Skalenenden und damit zu solchen S-förmigen Kurven führen.[84]

84 Die Analyse der Polynom-Kontraste erfolgt mit dem Computerprogramm KOSTAS (Nagl et a. 1986; vgl. auch Bock 1975). Beide Werte sind auf dem 5%-Niveau statistisch signifikant. Für die Einschätzung der Bedeutsamkeit ist allerdings die in den Korrelationen zum Ausdruck kommende Effektstärke relevanter (s. Kapitel 9).

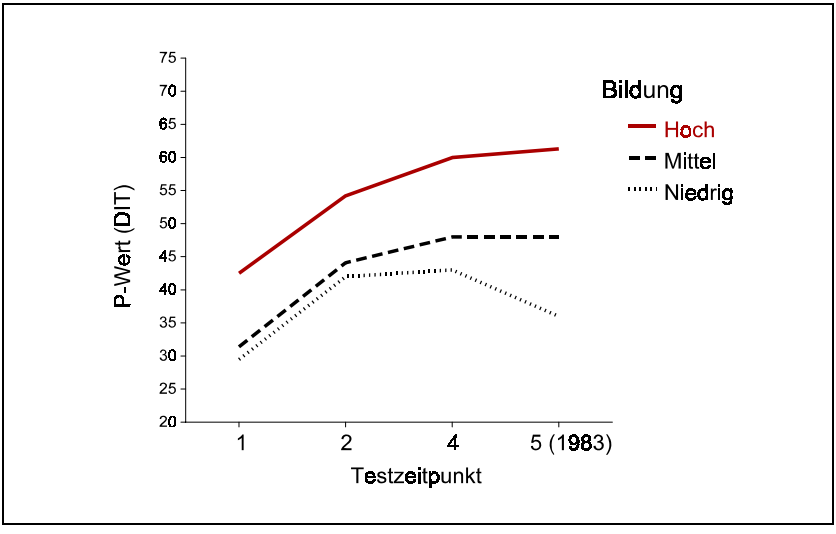


Abb. 5.4 Moralische Entwicklung (DIT) in Abhängigkeit vom Bildungsniveau. Quelle: Längsschnittstudie von Rest 1986, S. 33; N=102

Dieser Befund belegt also die eingangs aufgestellte These, dass es sich bei den Beobachtungen von Kohlberg und Kramer (1969) um eine *Pseudo-Regression* handelt. Der MUT-Befund repliziert die Beobachtungen, die Kohlberg und Kramer mit dem MAI gemacht haben. Aber er schränkt die dort gefundene "Regression" ein auf die affektiven Aspekte des moralischen Urteils, die beim MAI nicht unabhängig von den kognitiven gemessen werden können. Außerdem kann damit gezeigt werden, dass es sich um einen relativ geringen Rückgang handelt. Da sich dieser Rückgang wie im Kohlberg und Kramer-Experiment auf die Gruppe mit der höchsten Urteilsfähigkeit beschränkt, kann es sich hierbei auch um einen methodenbedingten "Deckeneffekt" der Meßinstrumente handeln.

5.3 Echte Regression: Repräsentative Querschnittstudie bei 14-21jährigen

Zwei Umstände haben das Auftreten von echten Regressionen bislang erschwert: zum einen, wie wir gezeigt haben, die Vorkehrungen, die Kohlberg

und seine Kollegen bei den Auswertungsregeln des *MAI* getroffen haben, um invariante Entwicklungsdaten zu erhalten;[85] zum anderen die Tatsache, dass die meisten Studien zu diesem Phänomen an Heranwachsenden durchgeführt wurden, die noch an Bildungsprozessen teilnehmen oder aufgrund einer gehobenen beruflichen Positionen (wie z.B. Management oder Betriebsrat) besondere Bildungserfahrungen genießen. In fast allen Studien wird Wert auf die Feststellung gelegt, dass die Stufe des moralischen Urteils mit wachsendem Alter zunimmt, aber von wenigen Ausnahmen abgesehen hat es kaum ein Autor für wichtig erachtet mitzuteilen, dass in derselben Zeit ebenfalls die Bildungserfahrungen der Befragten zunehmen. So kann es nicht verwundern, dass bis vor kurzem die Autoren nur vage Auskünfte über die Höhe der Korrelation zwischen dem Meßwert im MAI und dem Bildungsstatus gaben. Solche Angaben müssen häufig nachträglich aus den berichteten Daten erschlossen werden.[86] Auf diese Weise ist der *Moral Judgment*-Forschung lange die Tatsache verborgen geblieben, welche Bedeutung das allgemeine Bildungssystem und die in diesem Rahmen möglichen diskursiven Prozesse für die Moralentwicklung haben, und dass bei Wegfall dieses Faktors eine Rückentwicklung dieser Fähigkeit auftreten könnte – eine nicht nur für den Einzelnen, sondern auch für die gesamte Gesellschaft folgenreiche Regression.

Die erste Studie (Rest 1986), in der ein Regressionseffekt nachgewiesen werden konnte, wurde mit dem *Defining Issues Test* (DIT) durchgeführt (Abb. 5.4). In dieser Studie wurde die Veränderung des *P*-Wertes in drei Gruppen untersucht. Die Gruppe, die als "hoch" bezeichnet wird, hat mehr als vier Jahre College-Ausbildung, die "mittel"-Gruppe hatte mindestens zwei Jahre und die "niedrig"-Gruppe weniger als zwei Jahre College-Ausbildung. Zum 1. Meßzeitpunkt waren alle noch in der *High school*. Zum 2. Zeitpunkt waren zwei Jahre seit der ersten Testung vergangen, beim 4. Zeitpunkt vier Jahre und zuletzt (1983) etwa 10 Jahre. Wie Abb. 5.4 zeigt, hat die Gruppe mit der geringsten Bildungserfahrung (weniger als zwei Jahre College) insgesamt den niedrigsten *P*-Wert und nur sie zeigt eine Regression: Die *P*-Werte dieser Gruppe beginnen vier Jahre nach Abschluss der *High school*-Ausbildung abzusinken. Rest (1988) argumentiert zu Recht, dass dieser und andere Befunde eindeutig gegen einen Alterseffekt, also auch gegen die Emergenzthese sprechen. Er postuliert

85 Kohlberg 1973, S. 182, Colby & Kohlberg 1987, S. 38; auch Kapitel 3.

86 Siehe Lind 1985d; auch Kapitel 11.

weiterhin, dass aufgrund dieser Befundlage mittels DIT die Sozialisationsthese zu verwerfen sei und am ehesten solche Thesen gestützt werden, die einen Bildungseinfluß des College in Betracht ziehen. Rest begründet dies mit zwei Hinweisen: a) Er nimmt an, dass die DIT-Werte eine kognitive Fähigkeit ausdrücken und nicht nur unterschiedliche Präferenzen: Sie korrelieren (in anderen Studien) sehr hoch mit Verstehenstests und mit logischem Denkvermögen, IQ und Leistungstests. Zudem können, so Rest, DIT-Werte nicht beliebig, nur auf äußere Veranlassung hin, erhöht werden. b) Die Höhe des Zuwachses an *P*-Werten im DIT hängt stark von der erreichten Stufe logischen Denkvermögens im Sinne von Piaget ab (Rest 1988, S. 188). Diese Befunde stützen die Bildungsthese, aber sie unterstützen sie nur schwach. Der *P*-Wert reflektiert die Präferenz einer Person für *p*ostkonventionelle Argumente (Rest 1979) und stellt daher kein reines Fähigkeitsmaß dar, sondern gleichzeitig auch ein Einstellungsmaß. Veränderungen des *P*-Wertes können daher auch von der Sozialisationstheorie erklärt werden. Da, wie Emler et al. (1983) experimentell bewiesen, Personen relativ einfach durch Instruktionen veranlaßt werden können, einen höheren *P*-Wert zu simulieren (Kapitel 6), kann die Korrelation zwischen dem *P*-Wert und dem Bildungsniveau auch sozialisationstheoretisch erklärt werden. Die Befunde von Rest könnten anzeigen, dass Bildungsinstitutionen wie College und Universität bloß ihr liberales, "postkonventionelles" Werteklima auf die Studierenden übertragen, aber keine moralisch-kognitive Entwicklung fördern.[87] Ein weiterer Schwachpunkt der Rest-Studie ist, dass die Gruppe, in der die Regression beobachtet wurde, relativ klein[88] und ihre Repräsentativität für die Gruppe der Adoleszenten mit weniger als zwei Jahren College unklar ist. Trotz dieser Einschränkungen spricht die Rest-Studie für die These, *dass die moralische Urteilsfähigkeit von Menschen nicht nur zunehmen, sondern auch abnehmen kann, nämlich dann, wenn Bildungserfahrungen ausbleiben.*

87 Diese These wird u.a. von Newcomb (1974) und Dann et al. (1978) vertreten. Für eine kritische Auseinandersetzung damit siehe Lind (1985c).

88 Schätzungsweise 30-40 Personen; in keiner der herangezogenen Veröffentlichungen von Rest (1986a; 1986b; 1988) werden die genauen Meßwerte und die Stichprobengrößen berichtet. Bei der graphischen Darstellung in Abb. 5.4 handelt es sich daher um eine rein graphische Übertragung aus der Darstellung von Rest (1986b, S. 33). Die Darstellung bei Rest (1986a) weicht von den anderen zwar wesentlich ab, aber das Regressionsphänomen ist auch dort sichtbar.

5.3.1 Die EMNID-Studie

Für eine Prüfung der bildungstheoretischen These steht uns durch eine repräsentative Erhebung mit dem MUT bei Jugendlichen in der Bundesrepublik Deutschland (alte Bundesländer) aus dem Jahr 1990 eine breite Datengrundlage zur Verfügung.[89] Die Erhebung zielte auf Heranwachsende und junge Erwachsene im Alter zwischen 14 und 21 Jahren, die eine Haupt- oder Realschule besucht oder abgeschlossen haben. Die untersuchten Personen entstammen also alle dem nicht-gymnasialen, berufsbildenden Sektor. Sie besuchen nach dem Schulabgang in der Regel entweder ein bis zwei Tage pro Woche die Berufsschule oder eine berufsvorbereitende Vollzeitschule. Die Erhebung wurde von Interviewern jeweils in der Wohnung des Befragten durchgeführt. Die Befragungen waren, wie bei Umfrageinstituten üblich, zeitlich begrenzt. Daher sind die berichteten MUT-Werte niedriger und können nicht direkt mit Ergebnissen von Studien verglichen werden, die den Befragten unbegrenzt Zeit lassen. Es handelt sich hier um eine Querschnitt-, und nicht um eine Längsschnittstudie. Es wird oft die Meinung vertreten, dass die Invarianzthese nur durch eine Längsschnittstudie adäquat überprüft werden kann. Dies ist aber nur bedingt richtig. Auch Längsschnittstudien haben Probleme, die ihre Validität einschränken können: z.B. Erinnerungs-, Lern- und Erwartungseffekte (Wohlwill 1977, S. 153 ff.). Da diese Probleme in den Längsschnittstudien von Kohlberg und seinen Kollegen nicht berücksichtigt wurden, läßt sich nicht abschätzen, in welchem Maß solche Effekte zur invarianten Sequenz der MAI-Daten beitragen bzw. Regressionen verhindern. Man kann vermuten, dass ein längeres Dilemma-Interview bei den Befragten stimulierende Effekte hat. Pädagogische Interventionsprogramme zur Förderung der Moralentwicklung haben mit Methoden, die dem Kohlberg-Interview ähneln, einen nachhaltigen Effekt (Kapitel 9). Querschnittstudien sind nicht mit diesen störenden Effekten belastet, aber sie haben andere Probleme, vor allem das “Kohorten-Problem”. Querschnittgruppen repräsentieren nicht nur verschiedene Ausprägungen des Alters oder Bildungsniveaus, sie entstammen immer auch verschiedenen Gene-

89 Die Daten wurden von dem Umfrageinstitut EMNID, Bielefeld, im Rahmen einer Berufsbildungsstudie erhoben und mir für Analysen zur Verfügung gestellt. Der Analyse liegen die Daten von 708 Befragten zugrunde, deren MUT-Daten vollständig waren.

rationen oder Kohorten mit möglicherweise sehr spezifischen Prägungen durch die jeweiligen Umstände der Epoche, in der sie aufgewachsen sind. Die Größe dieser “Störeffekte” hängt aber von zwei Dingen ab: a) von dem Zeitraum, der zwischen den Generationen liegt, und b) von der Art des Phänomens, das wir untersuchen.
Wir können annehmen, dass Kohorten, die kurz aufeinander folgen, auch unter ähnlichen Bedingungen aufgewachsen sind, und dass sich Kohorteneffekte am ehesten auf bestimmte Konventionen und spezifische, zeitbezogene Meinungen auswirken, kaum aber auf die moralische Urteilsfähigkeit. Aufgrund dieser Problemlage scheint eine Querschnittstudie mit zeitlich dicht gestaffelten Alterskohorten eine mindestens ebenso gute Basis für eine Prüfung wie eine Längsschnittstudie zu sein.

5.3.2 Ergebnisse

Die Daten der EMNID-Studie bestätigen die Bildungstheorie: Die moralische Urteilsfähigkeit von Jugendlichen nimmt stetig ab, nachdem sie die Haupt- oder Realschule, und damit das allgemeinbildende Schulsystem verlassen haben. Die Daten zeigen im Einzelnen: a) Das Bildungsniveau hat einen deutlichen Einfluß auf die moralische Urteilsfähigkeit: Die MUT-Werte der Absolventen der Realschule sind in allen Alterskohorten höher als die MUT-Werte der Hauptschulabgänger. b) Der Einfluß der Bildung wird auch durch den charakteristischen Verlauf der “Veränderungskurve” deutlich: Es zeigt sich ein Anstieg der MUT-Strukturwerte bis zum 15. Lebensjahr, wenn noch fast alle Befragten die allgemeinbildende Schule besuchen. Danach nimmt in beiden Gruppen trotz zunehmenden Alters die moralische Urteilsfähigkeit kontinuierlich ab. Die Abnahme vollzieht sich im Altersbereich von 15 bis 19 langsam; das – gegenüber dem Gymnasium – eingeschränkte Bildungsangebot für diese Jugendlichen kann den Rückgang der Urteilsfähigkeit offenbar nur abbremsen, aber nicht aufhalten (nur das kurze "Zwischenhoch" bei den 17jährigen Hauptschulabsolventen paßt zu keiner unserer Thesen). Im Alter zwischen neunzehn

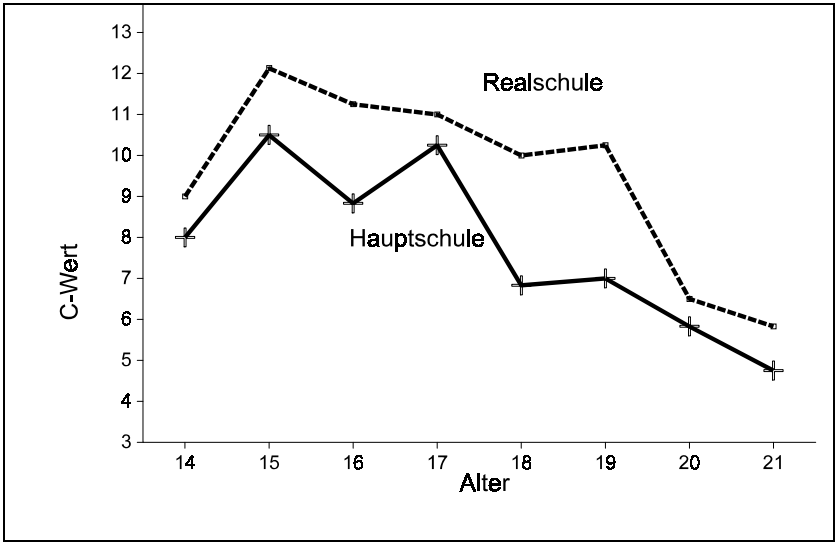


Abb. 5.5 Moralische Urteilsfähigkeit (MUT) in Abhängigkeit vom Alter und von der Schulart bei Jugendlichen von 14 bis 21 Jahren. Quelle: Repräsentativerhebung 1990 von EMNID; N = 780

und zwanzig, wenn die meisten der Befragten voll in das Berufsleben eingetreten sind, sinken die MUT-Werte nochmals scharf ab. Gegen die Interpretation einer Regression ließen sich einige Einwände geltend machen. Man könnte sagen, die älteren Befragten hatten einfach weniger Lust, über moralische Probleme nachzudenken und haben den MUT daher nur nachlässig ausgefüllt. Aber man könnte auch umgekehrt annehmen, dass die jüngeren Befragten weniger Lust zur Bearbeitung der MUT-Dilemmas hatten. Solange für keine der beiden Vermutungen Belege existieren, ist es müßig, darüber zu spekulieren. Bei den Gleichaltrigen, die aufs Gymnasium oder auf die Hochschule wechseln, ist in keiner Altersphase ein “Lusteinbruch” zu verzeichnen. Es könnte weiterhin gesagt werden, dass die älteren Jugendlichen in dieser Studie durch die Anforderungen ihres Berufs und vielleicht auch durch die Gründung einer Familie wenig Zeit haben, sich mit solchen Fragen, wie sie im MUT behandelt werden, zu beschäftigen und ihnen daher eine integrierte und differenzierte Beantwortung schwerfiel, wie sie für einen hohen MUT-Score Voraussetzung ist. Diese Erklärung würde allerdings die Bildungsthese, die wir hier prüfen, voll unterstützen.

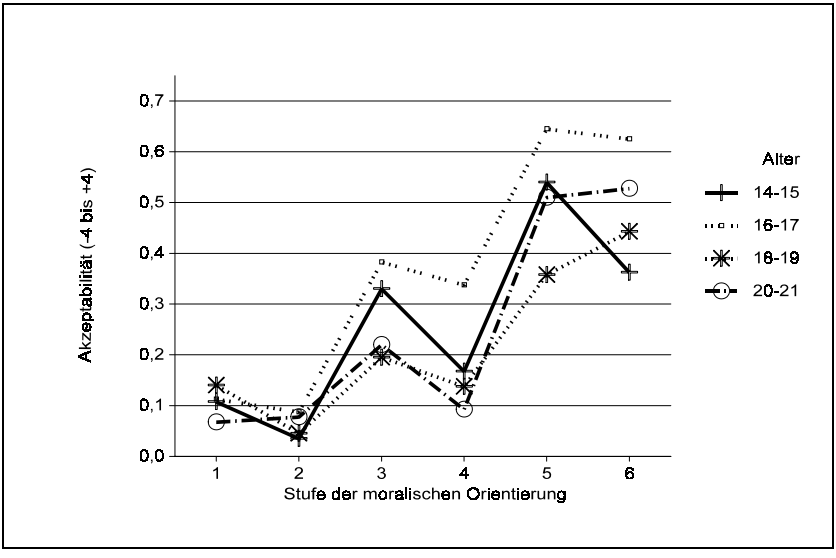


Abb. 5.6 Moralische Orientierungen (MUT) in Abhängigkeit vom Alter. Repräsentativerhebung (alte Bundesländer) von EMNID 1990; N = 780

Wir können auch die umgekehrte Frage stellen: Sind diese Daten mit den beiden anderen konkurrierenden Thesen zur Moralentwicklung in Einklang zu bringen? Die Daten widersprechen eindeutig der Reifungstheorie, die einen Alterseffekt annimmt. Die moralische Urteilsfähigkeit nimmt bei den Haupt- und Realschulabsolventen *nicht* mit dem Alter zu, vielmehr nimmt sie stetig ab. Mit der *Sozialisationsthese* stimmen die Daten auf den ersten Blick überein. Aber sie würden diese These nur dann eindeutig bestätigen, a) wenn sich die moralischen Orientierungen nach einiger Zeit umkehren würden und b) wenn man davon ausgehen könnte, dass die soziale Umwelt der Befragten (Berufsschule, Lehrstätte, Arbeitsplatz) moralisch-kognitive Entwicklung in breiter Front ablehnt oder gar systematisch unterdrückt. Für beide Annahmen gibt es aber kaum Anhaltspunkte. Die moralische Orientierung (hier die Einstellung der Befragten zu den sechs Stufen der Argumentation) verändert sich kaum. Die einzige größere Änderung ist das Anwachsen der Präferenz für Stufe 6, was durch die kognitive Entwicklungstheorie vorhergesagt wird, nicht aber die Sozialisationstheorie. Wir vermuten, dass das Phänomen der Regression moralischer Fähigkeiten eher durch Unterlassung zustandekommt als durch Un-

terbindung. Die Gesellschaft unterläßt es, dem Teil der Jugendlichen, der frühzeitig in die Berufswelt eintritt, ausreichend Gelegenheit zur Allgemeinbildung und zur moralisch-kognitiven Entwicklung zu geben.

5.4 Resümee

Diese Befunde zeigen, dass sich während des Hochschulstudiums *keine* Regression der moralischen Urteilsfähigkeit ereignet, dass eine solche Regression aber zu beobachten ist, wenn man Jugendliche und junge Erwachsene studiert, die in diesem Lebensabschnitt eine berufliche Ausbildung oder eine Arbeit aufnehmen. Sie stützen damit eindeutig die Bildungsthese und widersprechen der Emergenzthese. Insofern widerlegen sie in diesem Punkt den Piaget-Kohlberg-Ansatz, der postuliert, dass bei der moralischen Urteilsfähigkeit keine Umkehrung der Entwicklungssequenz auftritt. Es gibt keine Entwicklungs*logik*, nach der sich die Fähigkeit zum differenzierten und konsistenten Urteilen als Funktion des Alters *entfaltet*. Dazu bedarf es offenbar der Stimulation der Entwicklung durch Bildungsprozesse. Unter Lebens- und Arbeitsbedingungen, die keine solche Stimulation bieten oder den Menschen gar die Möglichkeit der moralischen Reflexion und des ethischen Diskurses entziehen, müssen wir mit einem schleichenden Verlust der moralischen Urteilsfähigkeit rechnen.

6 Bildung versus Sozialisation I: Kann moralische Urteilsfähigkeit wie eine politische Einstellung simuliert werden?

Moral hat mit Kognitionen oder Fähigkeiten nichts zu tun, sagen die Vertreter der Sozialisationstheorie der Moralentwicklung. "Ein Großteil der Varianz des moralischen Urteils bei Erwachsenen, die bislang auf strukturelle Entwicklungsunterschiede zurückgeführt wurde," so Emler und seine Kollegen (1983) in ihrer Kritik der kognitiven Entwicklungstheorie der Moral, "kann adäquater als eine Varianz der politischen Ideologie interpretiert werden" (S. 1079). Veränderungen des moralischen Urteilens und Verhaltens, so Emler (1983), ließen sich besser mit der sogenannten "Selbstdarstellungstheorie" von Johnson und Hogan (1981) erklären: Eine Person übernehme die moralischen Werte und Einstellungen der Gruppe, der sie angehört, um sich als Angehörige dieser Gruppe auszuweisen. Mit anderen Worten, das Konzept der kognitiven Struktur sei weder zur Erklärung noch zur Messung der Moralentwicklung notwendig. Für beides sei das Konzept der (affektiven) Einstellung völlig ausreichend. "Präferenzen benötigen keine Kognitionen (preferences need no inferences)", behauptet Zajonc (1980) im Untertitel seiner programmatischen Kritik an allen "kognitiven" Ansätzen der Psychologie.

Die sozialisationstheoretische Kritik richtet sich primär gegen die kognitive Entwicklungstheorie von Piaget und Kohlberg, die wir im vorigen Kapitel zugunsten der Bildungstheorie der Moral zurückgewiesen haben. Wir müssen uns trotzdem mit dieser Kritik eingehend befassen, da sie sich, wie unschwer zu erkennen ist, auch gegen zentrale Annahmen der Bildungstheorie richtet, die diese mit der Kohlberg-Theorie gemeinsam hat, nämlich gegen die Annahme einer kognitiven Basis des moralischen Verhaltens und die Annahme eines langwierigen Bildungsprozesses als kausaler Faktor der Entwicklung der moralischen Urteilsfähigkeit. Von diesem gemeinsamen *kognitiven* Standpunkt wird vorhergesagt, dass es einer Person *nicht* möglich ist, ihr moralisches Urteil in jede beliebige Richtung zu verändern, wenn ihr dies zum Zweck der Selbstdarstellung geboten erscheint. Aus dieser Sicht kann eine Person zwar eine niedrige moralische Urteilsfähigkeit vortäuschen, nicht aber eine höhere, wenn sie dies möchte oder dazu instruiert wird.

Die kognitiven Theorien gehen sogar so weit zu behaupten, dass nicht nur das moralische Urteil, sondern auch politische Einstellungen oder Ideologien eine strukturelle Basis haben. Demnach sind "linke" und "rechte" Ideologien mehr als bloße Einstellungen; sie reflektieren aus kognitionspsychologischer Sicht auch unterschiedliche Stufen des politischen Verständnisses und der politischen Kompetenz. Vermehrt wendet sich daher das Interesse der *Struktur* politischer Einstellungen zu, wobei auffällige Parallelen zur Struktur des moralischen Urteilens gesehen werden. Diese Sichtweise kann sich bislang vor allem auf empirische Korrelationen zwischen politischen Orientierungen und Verhalten einerseits und moralischem Urteil andererseits stützen.[90] Jedoch reichen Korrelationsdaten nicht aus, um den kognitiven Standpunkt zu untermauern. Eine positive Korrelation läßt sich auch auf der Grundlage der nicht-kognitiven Sozialisationstheorie vorhersagen.

6.1 Das Experiment von Emler, Renwick und Malone (1983)

Emler und seine Kollegen (1983) nehmen an, dass es sich bei der Moral nicht um eine Kompetenz handelt, die sich nur langsam – aufgrund von Reifungs- oder Bildungsprozessen – entwickelt. Aufgrund ihres sozialisationstheoretischen Ansatzes nehmen sie vielmehr an, dass jeder Mensch bereit und fähig ist, sich jeder Stufe des moralischen Urteilens anzupassen, das seiner (politischen) Bezugsgruppe entspricht. Aufgrund dieser Hypothese sagten sie zum Beispiel voraus, dass es gelingen kann, das moralische Urteilsverhalten von Personen allein durch eine entsprechende Instruktion zu verändern, und zwar in jede beliebige Richtung. Die kognitive Entwicklungstheorie kommt zu einer ganz anderen Vorhersage bezüglich des Ausgangs eines solchen Experiments. Sie sagt voraus, dass sich wohl moralische Einstellungen simulieren lassen, nicht aber die moralische Urteilsfähigkeit. Während sich z.B. durch geeignete Instruktionen die Antworten auf einer Einstellungsskala ganz leicht sowohl in negativer als auch in positiver Richtung ändern lassen, dürfte es nicht möglich sein, moralisch-kognitive Kompetenzen auf die bloße Anweisung hin zu erhö-

90 Siehe Binford 1983; Converse 1970; Lind et al. 1985; Rosenberg 1988; Weinreich-Haste 1986.

hen. Dafür sind aus "kognitiver" Sicht ein langwieriger Entwicklungsprozeß bzw. spezielle Erziehungsmethoden vonnöten.

Um ihren Ansatz zu überprüfen, formulierten Emler et al. (1983) zwei Hypothesen: Ihre erste Hypothese war, dass "jene, die sich selbst als rechts beschreiben und konservative Meinungen äußern, relativ hohe Stufe-4-Werte erreichen sollten und relativ niedrige *P*-Werte [im DIT]. Jene, die sich als links bezeichnen und progressive Meinungen äußern, sollten relativ niedrige Stufe-4-Werte haben und relativ hohe *P*-Werte" (S. 1075). Ein hoher Wert auf Stufe 4 weist auf eine positive Einstellung gegenüber dem Erhalt von "Recht und Ordnung" hin; der *P*-Wert repräsentiert den Prozentsatz von Präferenzen für postkonventionelle Argumente. Ihre zweite Hypothese sagt voraus, dass "Personen, die sich selbst als Linke bezeichnen, fähig sind, das Muster moralischer Urteile zu reproduzieren, das für Rechte charakteristisch ist, wenn sie dazu aufgefordert werden, den DIT [s.u.] wie ein Konservativer zu beantworten" (S. 1079). Zudem sagten sie voraus, "dass Rechte fähig sind, das Urteilsmuster eines Linken zu reproduzieren, wenn sie gebeten werden, wie ein Linker zu antworten" (S. 1076). Wie die Autoren zu Recht bemerken, sind die Leistungen der rechtslastigen Testpersonen für die Validität der Theorie entscheidend. Aus Sicht der kognitiven Entwicklungstheorie müßten linksgerichtete Testpersonen die Antworten einer rechtsgerichteten Person reproduzieren können, da letztere auf einer niedrigeren Entwicklungsstufe argumentiert; rechtsgerichtete Testpersonen sollten dagegen nicht in der Lage sein, die für Linke angeblich typische, weiter fortentwickelte Argumentation zu reproduzieren (S. 1076). Emler und seine Mitarbeiter überprüften ihre Vorhersage in einem Experiment mit Universitätsstudenten. Zur Messung der politischen Ideologie benutzten sie eine Selbsteinschätzungsskala und Teilskalen der *New Left*-Skala von Richard Christie und seinen Mitarbeitern. Zur Messung der moralischen Urteilskompetenz entschieden sie sich für die Anwendung des *Defining Issues Test* (DIT) von Rest (1979). In dem DIT sahen die Autoren ein geeignetes Maß "für die strukturell-entwicklungsmäßigen Veränderungen, die dem moralischen Urteil zugrunde liegen" (S. 1075).

Die erste – für die Entscheidung zwischen kognitiven und nicht-kognitiven Ansätzen allerdings unwesentliche – Hypothese wird durch die Befunde von Emler et al. eindeutig bestätigt. Die Korrelation des *P*-Wertes mit der "Philosophie der Neuen Linken" (*New Left*-Skala) beträgt $r = .39$, mit der selbstdefinierten politischen Orientierung (links vs. rechts) $r = .48$; Linke haben einen

höheren *P*-Wert als Rechte.[91] Die Korrelationen der Einstellung der Befragten zu Stufe-4-Argumenten (konventionelle Moral) und der Philosophie der Neuen Linken sowie der selbstdefinierten politischen Einstellung waren hoch negativ (r = -.52 bzw. -.56); Linke lehnen konventionell-moralische Argumente ab, Rechte stimmen ihnen eher zu. Die zweite für unsere Fragestellung sehr wichtige Hypothese wurde ebenfalls bestätigt. Sowohl links- als auch rechtsgerichtete Testpersonen konnten die Werte der Stufe 4 und die P-Werte beim DIT für die politisch Andersdenkenden simulieren. Dieses Ergebnis stimmt mit den Resultaten eines ähnlichen Experiments von Meehan et al. (1979) überein, die den *Survey of Ethical Attitudes* (SEA) von Hogan zur Messung moralischer Einstellungen verwendeten. Sie hatten die Probanden gebeten, den SEA so zu beantworten wie ein linker bzw. ein rechter Bewerber um eine feste Anstellung bei einer politischen Arbeitsgruppe. Auch diese Probanden konnten das moralische Urteilsverhalten von beiden Gruppen gut simulieren.

Emler et al. (1983) zogen aus dem Ergebnis ihres Experiments zwei Schlüsse: Erstens, dass der DIT keine kognitiv-strukturelle Entwicklung mißt, zweitens dass das moralische Urteil keine Fähigkeitsbasis besitzt. Die Autoren übersehen dabei, dass, wenn ihre erste Schlußfolgerung zutrifft, ihre zweite durch das Experiment nicht begründet ist. Wenn der DIT kein valides Fähigkeitsmaß ist, dann sind die Daten von Emler et al. nicht dafür geeignet, die Frage zu entscheiden, ob das moralische Urteil eine kognitiv-strukturelle Basis hat. Tatsächlich ist der DIT seiner Konstruktion nach ein Einstellungstest, bei dem kognitive Aspekte des moralischen Urteils zwar berücksichtigt werden, aber keine dominante Rolle spielen. Der mit dem DIT ermittelte *P*-Wert gibt Auskunft darüber, wie stark Befragte postkonventionelle Moralprinzipien präferieren, aber er sagt nur wenig über die Fähigkeit des Menschen aus, seine Urteile im Licht dieser moralischen Prinzipien kritisch zu überprüfen oder sogar zu modifizieren. Die Befragten werden im DIT z.B. nicht mit Gegenargumenten konfrontiert. Sie sollen nur ankreuzen, wie wichtig ihnen vorgegebene moralische Überlegungen ("Issues") sind (Rest 1979).

Die Tatsache, dass der DIT mit Kohlbergs Maß für moralische Reife r = .60 korreliert, ist kein Beweis dafür, dass er kognitive Aspekt des moralischen Urteils valide mißt. Auch das MAI mißt in hohem Maß den affektiven Aspekt

91 Die *r*-Werte wurden von mir anhand der bei Emler et al. (1983) berichten *F*-Werten berechnet, um die Ergebnisse vergleichbar zu machen (s. Kapitel 9).

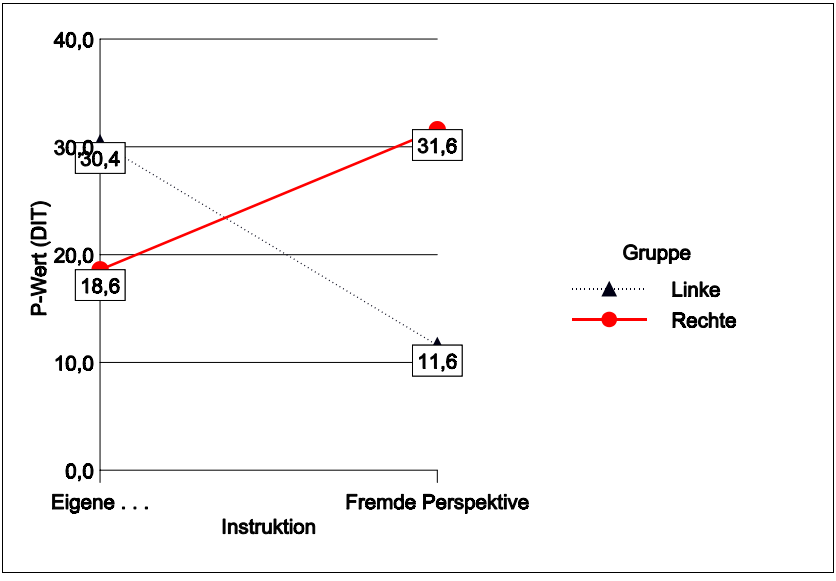


Abb. 6.1 Effekt der Instruktion, im *Defining-Issues-Test* (DIT) die Antworten von politisch sich als "links" bzw. "rechts" bezeichneten Personen zu simulieren. Quelle: Emler et al., 1983.

(s. Kapitel 3), so dass nicht auszuschließen ist, dass die Korrelation durch diesen Aspekt bedingt ist, nicht aber durch den kognitiven Aspekt. Die Ergebnisse des Emler-Experiments legen daher eine andere Interpretation nahe als die von den Autoren gewählte: Sie zeigen dass moralische *Einstellungen* mit politischen Einstellungen korrelieren und dass diese leicht simuliert werden können, aber nicht, dass moralische *Kognitionen* mit politischen Einstellungen korrelieren und dass diese simuliert werden können. Die von Emler und seinen Kollegen aufgeworfene Frage, ob die kognitive Dimension zur Erklärung der moralischen Entwicklung notwendig ist oder ob einfache, sozialisationstheoretische Erklärungen ausreichen, konnte durch ihr Experiment also nicht geklärt werden.

Auf die andere Erklärung von Emler et al. (1983), dass der von ihnen eingesetzte DIT kein Instrument zur Erfassung von kognitiv-strukturellen Aspekten des moralischen Urteilsverhaltens darstellt, gingen die Autoren leider nicht weiter ein. Auch der Autor des DIT, J.R. Rest, blendet diese Frage ganz aus. Rest et al. (1999) nahmen das Emler-Experiment gar zum Anlaß, Simulationsstudien ganz aus ihrem Kriterien-Katalog für die Validität des DIT zu nehmen (S. 115).

Gruppe: N =	Linke 18	Mitte 17	Rechte 18	*r*
Einstellung zur Stufe 4 (MUT) (Affektiver Aspekt)	-0,3	0,2	0,4	0,31
Moralische Urteilsfähigkeit (MUT) (Kognitiver Aspekt)	37,9	28,8	25,3	-0,42

Anmerkung: Die Skala "Einstellung zu Stufe 4" (konventionelle Moral) reicht von -4 bis +4; die Skala "moralische Urteilsfähigkeit" von 0 bis 100. Die Korrelation bezieht sich nur auf die Gruppen "Linke" und "Rechte".

Abb. 6.2 Moralischen Einstellungen und Urteilsfähigkeit (C-Wert, MUT) bei Personen, die sich als politisch links, in der Mitte bzw. rechts einschätzen. Mittelwerte und Korrelationen

Für die Frage nach der kognitiv-strukturellen des moralischen Urteilverhaltens und für die Validität des MUT ist das Emlersche Simulations-Experiment jedoch von großer Bedeutung. Es gibt kaum eine bessere Möglichkeit festzustel len, ob überhaupt eine Fähigkeit (und nicht bloß eine Einstellung) vorliegt und ob das eingesetzte Meßinstrument diese Fähigkeit zu messen vermag.

6.2 Ein modifiziertes Experiment zur Überprüfung der Hypothese von Emler et al. (1983)

Wir haben das Experiment von Emler et al. (1983) in zwei wichtigen Punkten verändert: Erstens statt des DIT, der ein moralischer Einstellungstest ist, wurde der *Moralisches Urteil-Test* (MUT) verwendet, der im Unterschied zum DIT beansprucht, den Fähigkeitsaspekt des moralischen Urteils zu erfassen, und zwar unabhängig von seinem affektiven Aspekt (s. Kapitel 4). Zweitens, im Unterschied zu Emler et al. (1983) haben wir nicht nur Universitätsstudenten ($N = 32$) in die Untersuchung mit aufgenommen, sondern auch Erwachsene mit höherer Bildung ($N = 21$). Von den 53 Versuchspersonen waren 32 männlich

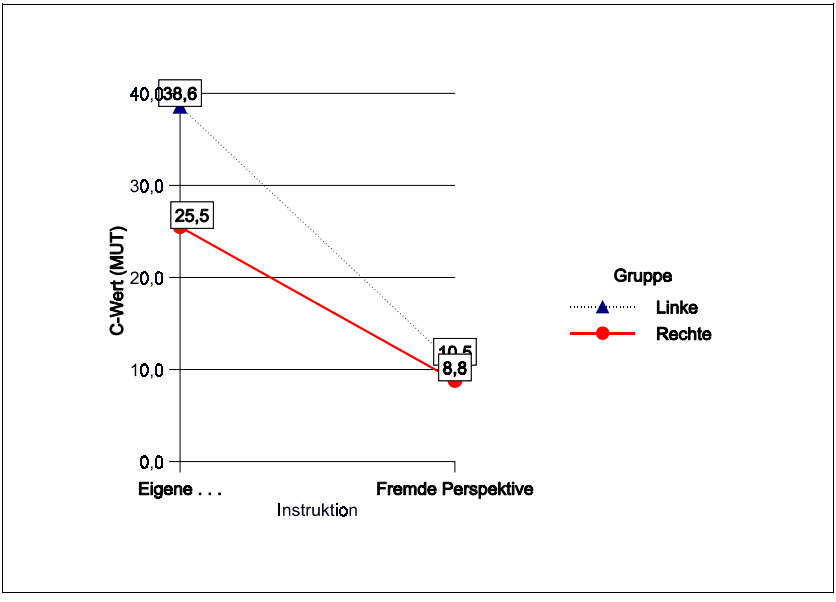


Abb. 6.3 Effekt der Instruktion, im MUT die Antworten von politisch sich als „links“ bzw. „rechts“ bezeichnenden Personen zu simulieren (Lind-Experiment)

und 21 weiblich, im Alter zwischen 19 und 65 Jahren. Wie in Emlers Experiment wurden nach der Selbsteinschätzung der politischen Einstellung drei gleich große Versuchspersonengruppen gebildet: 18 linksgerichtete, 18 rechtsgerichtete, und 17 Personen, die sich als gemäßigt bezeichneten. Diese Gruppen werden hier auch einfach als “links”, “rechts” und “gemäßigt” bezeichnet. Wie im vorigen Experiment wurde jede Gruppe nach Zufall in zwei Untergruppen aufgeteilt. Nachdem jeder den MUT beantwortet hatte, wurden die Versuchspersonen gebeten, den MUT ein zweites Mal auszufüllen. Die eine Experimentalgruppe erhielt die Anweisung, linksgerichtete, die andere rechtsgerichtete Urteile zu reproduzieren.

Zur Klärung der Standpunkte von nicht-kognitiven und kognitiven Theorien wurden im Rahmen dieses Experiments sechs Hypothesen untersucht. Sie werden im folgenden zusammen mit den Befunden dargestellt. 1. Hypothese: Wie wir gesehen haben, besteht zwischen kognitiven und nicht-kognitiven Theorien eine Gemeinsamkeit. Beide sagen voraus, dass die politische Ideologie systematisch mit moralischen Einstellungen korreliert ist. Befund: Rechte politische

	Instruktion:	
Effekt der Instruktion auf die simulierte...	Rechte sollen Linke simulieren	Linke sollen Rechte simulieren
Moralische *Einstellung* - DIT-Studie von Meehan et al. (1979)	+ 0,68	+ 0,81
- DIT-Studie von Emler et al. (1983)	+ 0,67	+ 0,82
Moralische *Urteilsfähigkeit* - MUT-Studie von Lind	- 0,53	+ 0,73

Abb. 6.4 Effektstärke (Korrelationskoeffizient *r*) der Instruktion auf die simulierte moralische Einstellung und Urteilsfähigkeit (MUT). Korrelationskoeffizienten

Ideologie korreliert positiv mit einer hohen Akzeptabilität konventioneller moralischer Urteile ($r = 0.31$). Rechtsgerichtete Personen neigten dazu, Urteile auf Stufe 4 zu akzeptieren, linksgerichtete tendierten zum Widerspruch. Diese Beziehung kann jedoch nicht auf die postkonventionelle Stufe des moralischen Urteils generalisiert werden. Die Einstellung gegenüber Stufe-5-Argumenten korreliert hingegen kaum mit linker Ideologie ($r = 0.04$).[92]

Bei den weiteren Hypothesen folgen aus kognitiven Theorien andere Vorhersagen als aus nicht-kognitiven Theorien.

2. Hypothese: Nicht allein moralische Einstellungen korrelieren mit politischen Ideologien, sondern auch moralische Kognitionen bzw. Fähigkeiten. Befund: Im vorliegenden Experiment war die Beziehung zwischen politischer Ideologie und moralischer *Kognition* hoch, sogar absolut gesehen höher ($|r| = 0.42$) als die obige Korrelation zwischen politischer Ideologie und moralischen *Einstellungen* ($r = 0.31$). Die moralische Urteilsfähigkeit von linken war höher als die von rechten Probanden. Andere Studien bestätigen diesen Befund. Lind et al. (1985) berichten, dass bei Abiturienten die moralische Urteilsfähigkeit

92 Die geringe Korrelation, die sich auch in einer Studie von Weinreich-Haste et al. (1986) fand, erklärt sich möglicherweise aus der Tatsache, dass ausgeprägt postkonventionelle Urteile selten sind (Kohlberg 1984; Lind 1985a), die Varianz bezüglich dieser Stufe also gering ist und so auch Korrelationen mit Stufe-5-Werten nur sehr gering ausfallen können.

(MUT) positiv mit der Akzeptanz demokratischer Gleichheitsgedanken korreliert. Seitz (1991) zeigt in ihrer Studie bei Gymnasiasten (8.-13. Klasse), dass die moralische Urteilsfähigkeit auch positiv mit Demokratie-Verständnis (r = .24), Ambiguitätstoleranz (r = .35) und Anti-Autoritarismus (r = .33) sowie mit einer Abneigung gegen Gewalt (r = .20) korreliert sind. All diese Befunde sind nicht durch die Sozialisationstheorie zu erklären, da diese über die Entwicklung von Fähigkeiten keine Aussagen trifft, aber stehen in Einklang mit der "kognitiven" Sichtweise.

3. Hypothese: Aufgrund der kognitiven Natur des moralischen Urteils wird vorhergesagt, dass alle Versuchspersonen die moralische Urteilsfähigkeit anderer Menschen nur ungenügend reproduzieren können. Ihre eigenen Fähigkeitswerte sollten höher liegen als die Werte, die sie erhalten, wenn sie aufgefordert werden, das Urteilsmuster anderer (linker wie rechter) Personen zu simulieren. Befund: Unter allen Experimentalbedingungen lagen die simulierten Werte für die moralische Urteilsfähigkeit der Versuchspersonen weit unter ihren eigenen Werten.

4. Hypothese: Die Fähigkeit, die moralische Urteilskompetenz anderer zu simulieren, hängt in beträchtlichem Maße von der eigenen kognitiv-moralischen Fähigkeit ab. Zudem sollte das Simulieren linksgerichteter moralischer Urteile stärker von der eigenen Urteilskompetenz abhängen als die Reproduktion rechtsgerichteter Urteile, da erste "komplexer" oder "schwieriger" sind.
Befund: Beide Teilhypothesen werden bestätigt. Die Korrelation zwischen der persönlichen Kompetenz und der simulierten Kompetenz betrug r = .28. Die Korrelation für die Linken, die instruiert wurden, im MUT wie Rechte zu argumentieren, ist r = 0.40, und für die Rechten, die wie Linke argumentieren sollten, ist r = 0.23.

5. Hypothese: Wenn Moral eher eine Frage der Kompetenz als der Einstellung ist, müssen wir erwarten, dass Versuchspersonen zwar im Schnitt die moralische Kognition anderer Menschen nicht korrekt reproduzieren können, aber dennoch in der Lage sind, den *Unterschied* in der moralischen Urteilskompetenz zwischen links- und rechtsgerichteten Personen wahrzunehmen und zu reproduzieren. D.h., wenn Versuchspersonen gebeten werden, die moralischen Urteile Linker zu simulieren, sollten sie eine höhere moralische Kompetenz einbringen, als wenn es in der Interaktion um rechtsgerichtete moralische Urteile geht. Genauer gesagt, die Fähigkeit zur differenzierten Reproduktion der moralischen Kompetenz anderer Menschen sollte von der eigenen kognitiv-

moralischen Entwicklung abhängen, die, wie wir gesehen haben, in gewissem Maße mit der eigenen politischen Einstellung zusammenhängt.

Befund: Die Hypothese wird bestätigt. In Form von Korrelationen ausgedrückt erwies sich die Instruktion, moralische Kognition zu reproduzieren, in der linksgerichteten Gruppe als wirksamer als in der gemäßigten oder rechtsgerichteten Gruppe. Die entsprechenden Korrelationen sind $r = 0.28$ bzw. $r = 0.05$.[93]

6. Zusammenfassend läßt sich die Hypothese aufstellen: Die Instruktion, das moralische Urteilsverhalten bestimmter Personengruppen (mit linken bzw. rechten politischen Einstellungen) zu simulieren, hat ganz verschiedene Effekte, je nachdem, ob man (wie in den Experimenten von Emler et al.1983) die Auswirkungen auf den *affektiven* Aspekt des Urteilsverhaltens betrachtet oder (wie in unserem Experiment) auf den *kognitiven* Aspekt.

Befund: Die Gegenüberstellung der Befunde dieser drei Experimente auf der Basis von Effektstärkemaßen bestätigt diese Hypothese sehr gut. Die in Abb. 6.4 dargestellten Resultate zeigen, dass in den Experimenten von Meehan et al. (1979) und Emler et al. (1983) die Instruktion, moralische Einstellungen zu simulieren sehr effektiv ist, und zwar weitgehend unabhängig davon, um wessen moralische Einstellungen es sich handelt. Die positiven Korrelationen für beide Instruktionen zeigen an, dass die Effekte symmetrisch verliefen. Unser Experiment mit dem MUT belegt hingegen, dass die Instruktion, die moralische *Urteilsfähigkeit* anderer Personen zu simulieren stark davon abhängt, wer wessen Urteilsfähigkeit simulieren soll. Daher sind in unserem Experiment die Korrelationen asymmetrisch. Die Instruktion war lediglich bei Linken effektiv, die die moralische Kompetenz von Rechten reproduzieren sollten, erwies sich aber bei der Gruppe, welche die moralische Kompetenz von Linken produzieren sollte, als nicht befolgbar. Statt höherer Urteilsfähigkeit, wie nach Emlers Theorie zu erwarten gewesen wäre, simulierten Personen mit einer rechten Ideologie eine viel niedrigere, obwohl sie realisiert hatten, dass Linke postkonventionelle Ein-

93 Obwohl die Zuordnung zu den beiden Experimentalbedingungen zufällig erfolgte, waren die MUT-Werte der beiden "gemäßigten" Gruppen sehr verschieden. Die Versuchspersonen, die "rechtslastige" Urteile reproduzieren sollten, waren signifikant häufiger linksgerichtet und moralisch kompetent als die Versuchspersonen mit der "linksgerichteten" Instruktion. Um die Gruppen vergleichen zu können, mußten die Reproduktionswerte für diese Gruppen mittels der persönlichen MUT-Werte vergleichbar gemacht werden.

stellungen haben. Mit anderen Worten, Rechte sind, anders als Linke, offenbar weniger in der Lage, die moralische Urteilsfähigkeit ihrer politischen Gegner korrekt zu reproduzieren. Dieser Befund bestätigt nicht nur, wie die vorhergehenden Befunde, die kognitiven Theorien des moralischen Urteils, sondern legt auch eine "kognitive" Interpretation von politischen Ideologien nahe.

6.3 Die Simulation von realen Personen mit niedriger bzw. hoher moralischer Urteilsfähigkeit (Wasel-Experiment)

Wolfgang Wasel (1994) hat das obige Experiment in abgewandelter Form wiederholt. Um sicher zu stellen, dass die Zielpersonen, die simuliert werden sollen, tatsächlich einen „hohen" oder „niedrigen" Wert hatten, hat er reale Personen mit niedrigen und hohen Testwerten im MUT gebeten, die Antworten von anderen, ihnen gut vertrauten Personen auf die Fragen des MUT zu simulieren. Die Versuchspersonen sollten unter den anderen jeweils eine wählen, mit der sie gern zusammenarbeiten wollten. Wasel hat zwei Gruppen von Versuchspersonen gebildet, eine, die selbst hohe C-Werte im MUT erreichten, aber als Arbeitspartner Personen mit niedrigen C-Werten wählten, und eine, die selbst niedrige C-Werte hatten, sich aber Partner mit hoher moralischer Urteilsfähigkeit aussuchten. Beide Gruppen hat er instruiert, die Antworten ihres jeweiligen Partners zu simulieren.

Wasel hat in seinem Experiment noch deutlichere Belege erhalten für die Annahme, dass der MUT eine Fähigkeit ist. Während die Versuchspersonen mit einer hohen moralischen Urteilsfähigkeit keine Schwierigkeit hatten, das Antwortverhalten von ihren Partnern mit niedrigeren C-Werte zu simulieren, konnten die Versuchspersonen mit niedrigen eigenen C-Werten das Urteilsverhalten ihrer Partner, die eine deutlich höherer Urteilsfähigkeit hatten als sie selbst, *nicht* simulieren. Wie die Abb. 6.5 zeigt, bleiben die simulierten C-Werte in der Gruppe, die einen höheren als den eigenen C-Wert simulieren sollte, noch unter den eigenen. Die Gruppe, die einen niedrigeren Wert simulieren soll hingegen, simuliert tatsächlich einen tieferen Wert, auch wenn sie nicht so tief simuliert, wie die tatsächlichen C-Werte der Zielpersonen sind. Zudem konnte Wasel (1994) zeigen, dass die Genauigkeit, mit der das Antwortverhalten anderer Personen im MUT simuliert werden konnte, mit der eigenen moralischen

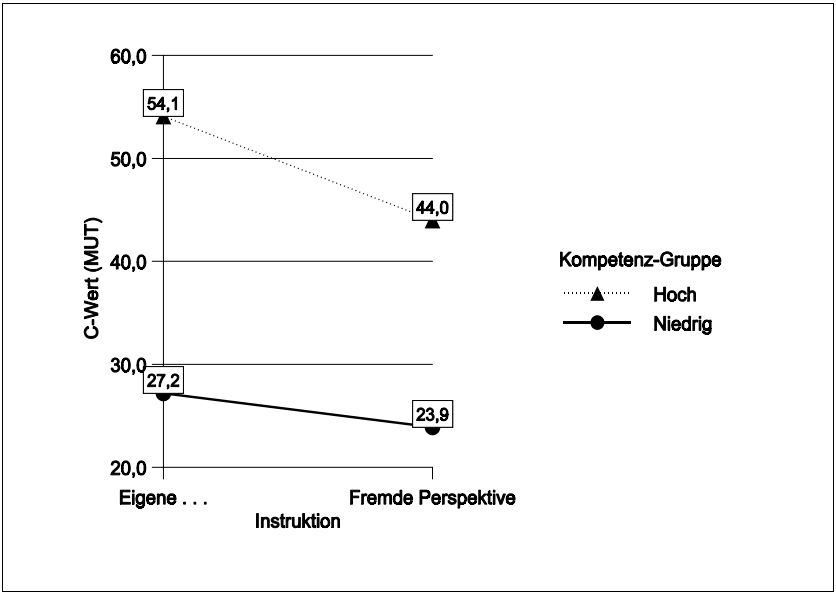


Abb. 6.5 Effekt der Instruktion, im MUT das moralische Urteilsverhalten von Personen zu simulieren, die eine niedrigere bzw. eine höhere moralische Urteilsfähigkeit haben als man selbst. Quelle: Wasel, 1994.

Urteilsfähigkeit korreliert (r = .30). Auch Wasels Experiment bestätigt also die Hypothesen, a) dass das moralische Urteilsverhalten einen starken Fähigkeitsaspekt hat, und b) dass der MUT geeignet ist, diesen Fähigkeitsaspekt der Moral zu messen. Offenbar stellt der MUT für viele Versuchspersonen eine schwierige moralische Aufgabe dar, so dass der C-Wert des MUT tatsächlich eine moralische *Fähigkeit* abbildet und nicht bloß eine moralische *Einstellung*.

6.4 Resümee

Ausgangspunkt der Analyse in diesem Kapitel war die Annahme von Emler et al. (1983), "dass individuelle Unterschiede im moralischen Urteil bei Erwachsenen als Variation auf der Dimension politisch-moralischer Ideologie interpretiert werden können und nicht als eine Variation auf der kognitiv-entwicklungsmäßigen Dimension" (S. 1075). Zur Untermauerung ihrer Annahme verweisen sie auf ihr Experiment, in dem es ihnen mittels einer einfachen

Instruktion gelang, Versuchspersonen mal zur Produktion eines niedrigeren moralischen Urteilswertes als des eigenen zu veranlassen, mal zur Produktion eines höheren. Andere Untersuchungen scheinen diesen Befund zu bekräftigen. So gelang es, Versuchspersonen erfolgreich zu instruieren, wie Politiker oder Philosophen moralisch zu argumentieren (Yussen 1977) oder moralische Urteilsprinzipien vorzutäuschen (Markoulis 1989). Die Instruktion, ein "möglichst ausgereiftes" moralisches Urteil zu reproduzieren, hat hingegen keinen Erfolg gehabt (McGeorge 1975); vielleicht deshalb, weil die befragte Person dabei zu sehr im Unklaren gelassen wird, was mit "ausgereift" gemeint sein könnte. All diesen Experimenten ist gemeinsam, dass sie als Maß für die "instruierte" Moralentwicklung den *Defining Issues Test* (DIT) von Rest (1979) verwendeten, dessen Eignung als Maß für den kognitiven Aspekt durch das Experiment von Emler et al. (1983) in Zweifel gezogen wurde. Rest et al. (1999) haben inzwischen solche Simulationsstudien aus ihrer Liste der Validitätskriterien für den DIT gestrichen und damit indirekt der Anspruch aufgegeben, den Fähigkeitsaspekt der Moral zu messen.

Um die Fähigkeitshypothese Kohlbergs zu überprüfen haben wir daher das Experiment von Emler et mit dem *Moralisches Urteil-Test* (MUT) wiederholt, der eine valide Messung des Fähigkeitsaspekts des moralischen Urteilverhaltens beansprucht (Kapitel 4). In dieser modifizierten Replikationsstudie sowie in dem Experiment von Wasel (1993; siehe oben) konnte nachgewiesen werden, a) dass sich nur moralische *Einstellungen* ähnlich wie politische oder andere Einstellungen leicht simulieren lassen, b) dass aber die moralische *Urteilsfähigkeit* sich *nicht* simulieren läßt. Dieser Befund belegt, dass die kognitive Dimension für das moralische Urteilen eine sehr wichtige Rolle spielt, und dass sozialer Druck oder der Wunsch, einer sozialen Gruppe anzugehören, allein nicht ausreicht, wie die Sozialisationstheorie annimmt, den Prozeß der moralischen Entwicklung zu erklären. Er widerlegt damit eine zentrale Annahme der Sozialisationstheorie der Moral und stützt die beiden hier diskutierten kognitiven Theorien, die Reifungs- und die Bildungstheorie.

Auch in unserem Experiment haben wir den Zusammenhang zwischen moralischem Urteil und politischer Ideologie bestätigt gefunden. Aber dies beweist nicht, wie Emler annimmt, dass das moralische Urteilsverhalten *affektiver* Natur ist. Eher deutet er darauf hin, dass auch politische Ideologien eine *kognitive* Basis haben (siehe auch Rosenberg, 1988; Rest et al., 1999). Vermutlich durchläuft das politische Denken auch eine Reihe von Entwicklungsphasen,

z.B. die Phase der Imitation von politischen Schlagworten (Fishkin et al. 1973), die Phase des Schwarz-Weiß-Denkens sowie der Ambiguitätsintoleranz (Budner 1962) und die Phase des autonomen politischen Urteils, analog den bei Piaget und Kohlberg beschriebenen Phasen der kognitiv-moralischen Entwicklung. Nur wenn wir bei der Untersuchung des *homo politicus* über Dichotomien wie links und rechts, Radikaler und Traditionalist hinausgehen, sind wir laut Robert Holt (1986) in der Lage, das volle politische Ideologienspektrum in seiner affektiven und kognitiven Dimension zu untersuchen.

7 Bildung versus Sozialisation II: Kulturelle Einflüsse auf die moralische Urteilsfähigkeit

Im vorigen Kapitel wurde die These untersucht (und zurückgewiesen), dass das politische Milieu einer Gruppe die moralische Urteilsfähigkeit von Menschen unmittelbar bestimmt.[94] Wir wollen hier untersuchen, ob, wie die Sozialisationstheorie annimmt, das moralische Urteilen durch den weiteren Kontext der *Kultur* geprägt wird oder, wie sowohl die Reifungstheorie als auch die Bildungstheorie der Moral annehmen, durch kulturunabhängige, universelle Entwicklungsbedingungen. Im Hintergrund steht dabei auch wieder die Frage, ob sich moralisch-demokratische Kompetenzen durch allgemeinbildende Maßnahmen fördern lassen, und zwar unabhängig vom besonderen Werte- und Normensystem einer bestimmten Kultur oder Gesellschaft.

Bislang hatten kulturvergleichende Fragestellungen im Bereich der *Moral Judgment*-Forschung die Funktion, die zentrale Invarianz-Annahme der kognitiven Entwicklungstheorie der Moral zu überprüfen. Diese Annahme besagt, so Kohlberg (1971), "fast alle Individuen in allen Kulturen durchlaufen dieselbe Ordnung oder Abfolge von breiten Stufen der Moralentwicklung, auch wenn sie sich hinsichtlich dem Umfang und dem Endpunkt der Entwicklung unterscheiden" (S. 176). Diese Annahme unterstellt also nicht, dass die Moral sich bei allen Menschen in allen Kulturen gleich schnell entwickelt oder dass in jeder Kultur der höchste Stand der Entwicklung nach Kohlbergs 6-Stufen-Modell erreicht wird.[95] Es wird vielmehr nur postuliert, dass die *Reihenfolge* der Entwicklungsstufen überall gleich ist und keine Regressionen auftreten (Kapitel 5). Tatsächlich ist diese Hypothese in verschiedenen Untersuchungen ausreichend bestätigt worden. Zumindest mit dem *Moral Judgment Interview* (MJI)

94 Wir wollen damit nicht ausschließen, daß politische Ideologien *indirekt* großen Einfluß auf die Moralentwicklung nehmen können, zum Beispiel indem sie unterschiedlich stark diskursive Prozesse ermöglichen oder verhindern, institutionalisierte Bildung fördern oder vermindern.

95 Wir gehen in dieser Untersuchung wie Kohlberg in seinen frühen Untersuchungen und dann wieder in späten Arbeiten von sechs Stufen der Moralentwicklung aus, ungeachtet der Tatsache, daß im *Moral Judgment Interview* nur fünf Stufen erfaßt werden (Colby, Kohlberg et al. 1987; Kohlberg 1984).

von Colby, Kohlberg et al. (1987) sind in allen untersuchten Kulturen (neben den USA auch Israel, die Türkei und Taiwan) die von der Theorie vorhergesagten Entwicklungssequenzen gefunden worden.[96] Damit wird zumindest indirekt auch die These bestärkt, dass Moralentwicklung im Kohlbergschen Sinn eher durch universelle Strukturmerkmale der sozialen Umwelt als durch das spezifische Werteklima einer Kultur gefördert oder gehemmt wird. Gleichwohl gibt es gegen eine solche Interpretation Einwände, die sich auf verschiedene Aspekte der Fragestellung beziehen, um die es hier geht:

1. Bisherige Untersuchungen waren selten direkt auf die Frage nach den Entwicklungs- oder Veränderungsbedingungen der menschlichen Moralität gerichtet. Die zum Teil großen Unterschiede bezüglich des Entwicklungsniveaus im moralischen Urteil zwischen verschiedenen Kulturen wurden nur am Rande vermerkt.

2. In kaum einer Untersuchung wurden andere bekannte Bedingungen der Moralentwicklung (wie zum Beispiel das Bildungsniveau) kontrolliert oder so erfaßt, dass eine nachträgliche Kontrolle ihres Einflusses möglich wurde.

3. Die Untersuchungen mit dem MJI waren wegen des damit verbundenen Aufwandes auf sehr kleine Personenzahlen beschränkt, weswegen die Befunde unter dem Vorbehalt stehen, dass sie eine sehr selektive Stichprobe aus der jeweiligen Kultur darstellen und möglicherweise auch nur wenig repräsentativ für eine bestimmte Untergruppe dieser Kultur sind. Dies gilt teilweise auch für DIT-Studien, die oft an kleinen und dazu noch bildungsheterogenen Stichproben durchgeführt wurden.[97]

4. Sie sind aus denselben Gründen auf relativ wenige Kulturen beschränkt geblieben. Zu Recht wurde von mehreren Autoren als nachteilig vermerkt, dass es lange Zeit keine Untersuchungen in (heute zum Teil ehemals) sozialistischen Ländern gab. Dort ließe sich wie kaum in einer anderen Kultureinheit die These eines sozialisatorischen Einflusses von gesellschaftlichen Ideologien prüfen.[98]

96 Hierbei handelt es sich um Längsschnittstudien. Darüber hinaus gibt es zahlreiche stützende Querschnittstudien. Für Übersichten vgl. Edwards 1981, 1986; Kohlberg 1984; Moon 1986.

97 Die Stichproben der DIT-Studien umfaßten gewöhnlich 10 bis 30 Probanden; nur in einem Fall waren es knapp über 100 Probanden (Moon 1986, S. 160).

98 Auf diese Lücke wurde mehrfach hingewiesen (Edwards 1986; Snarey 1985; Moon 1986).

5. Sprachliche Probleme können zu erheblichen Verzerrungen geführt haben. Bei Studien mit dem MJI von Kohlberg wurden in manchen Kulturen, wie zum Beispiel in der Türkei und Taiwan, die Interviews nicht vom Forscher selbst geführt, sondern von einem Dolmetscher übersetzt. Es ist unbekannt, in welchem Maße dadurch eine Verzerrung der Stufenwerte in die eine oder die andere Richtung erfolgt ist. Bei Studien mit dem *Defining Issues Test* (DIT) von Rest wurde nur in wenigen Fällen bekannt, ob die Übersetzung in die Landessprache sorgfältig genug durchgeführt wurde, um von einer äquivalenten Messung zu sprechen.[99]
6. Schließlich wurden in allen kulturvergleichenden Studien zum moralischen Urteil bisher Meßverfahren eingesetzt, in denen affektive und kognitive Aspekte des moralischen Urteilsverhaltens in einen einzigen Skalenwert zusammen gefaßt werden. Eine getrennte Betrachtung kulturbedingter Effekte ist somit nicht möglich. Es kann daher gegen die gefundenen Differenzen eingewendet werden, dass diese weniger einen Unterschied an *Urteilsfähigkeit* widerspiegelten als bestimmte kulturspezifische Werthaltungen und Einstellungen.[100] Damit ist aber jede Schlußfolgerung bezüglich der einen oder der anderen Hypothese ohne klare Grundlage.
Im folgenden sollen die Grundlinien des kognitiv-entwicklungstheoretischen Standpunktes zum Thema Moral und Kultur und dessen Problematik eingehender dargelegt werden, um dann zwei Studien vorzustellen, mit denen wir versucht haben, die aufgeworfenen Fragen zu den kulturellen Bedingungen der Moralentwicklung zu testen, und zwar unter Berücksichtigung der obigen Einwände.

7.1 Zur Problematik kulturvergleichender Moralforschung

Viele psychologische Theorien reklamieren stillschweigend, dass sie immer und überall, also auch in anderen Ländern mit anderen kulturellen Traditionen, anderen Wirtschafts- und Staatsformen und anderen Ideologien gelten. Jedoch viele der Fragen und Aufgaben, die in der psychologischen Diagnostik gestellt

99 Vgl. Rest 1986b, S. 95, 103-104; Moon 1986.

100 Zur (ideologie-) kritischen Diskussion des Kohlberg-Ansatzes vgl. Simpson 1974; Gielen 1984; Snarey 1985; Vine 1984.

werden, um aus den Antworten und Lösungen der Probanden Skalenwerte zu bilden, sind so stark an unsere spezifische Kulturform gebunden, dass entweder die Vertreter anderer Kulturformen diese nicht richtig verstehen können oder aber ihre Antworten und Lösungen von uns nicht richtig verstanden und daher falsch bewertet werden. Hier ein Beispiel: In einem weit verbreiteten Verfahren zur Bestimmung individueller kognitiver Fähigkeiten (des sogenannten "Intelligenzquotienten" oder "IQ") wird die Testperson gefragt, was sie tun würde, wenn sie auf der Straße einen frankierten und adressierte Briefumschlag finden würde. Die "richtige" Antwort, für die man Punkte bekommt, ist, man sollte ihn in den nächsten Briefkasten werfen. Sie bekommt keinen Punkt, wenn sie zum Bespiel sagt, sie würde den Brief der Polizei geben, obwohl das eine richtige Antwort ist, wenn die Person in einem Dorf in Afrika lebt, in dem es keine Briefkästen gibt und der Dorfpolizist die Aufgabe hat, die Post einzusammeln und zu verteilen.[101] Dadurch, dass die Person an den Kriterien eines bestimmten Kulturkreis gemessen wird, wird ihre kognitive Fähigkeit also falsch eingeschätzt.

Kann die gleiche Argumentation auch für moralische Einstellungen und Fähigkeiten gelten? Sind diese so stark durch kulturspezifische Konventionen und Normen geprägt, dass wir die Antworten und Lösungen, die wir erhalten, meist nur innerhalb eines bestimmten kulturellen Kontextes verstehen und beurteilen können? Auf der einen Seite müssen wir feststellen, dass schon innerhalb einer Kultur die Meinungen darüber auseinander gehen können, ob eine bestimmte Situation für ihre Mitglieder ein Handlungsdilemma darstellt oder nicht.[102] Durch dieselbe Situation können bei Personen sehr unterschiedliche Werte und Normen aktiviert werden. So mag es für die eine Person keine moralische, sondern nur eine Frage der Neigung sein, ob sie Fleisch oder Fisch, Wein oder Wasser zum Essen bestellt. Für einen anderen Menschen, dessen religiöse oder ethische Überzeugungen Fleisch oder Alkohol vom Speiseplan ausschließen,

101 Dieses Beispiel bezieht sich auf den weitverbreiteten *Wechsler Intelligenztest für Erwachsene*. Nach dem deutschen Manual des Tests (Wechsler 1961) erhält die Antwort "den Brief zur Polizei bringen" 0 Punkte, die Antwort "den Brief in den Briefkasten werfen" die höchstmögliche Punktzahl von 2.

102 Vgl. dazu den lebensweltlichen Ansatz von Alfred Schütz und seine Anwendung auf die moralische Urteil-Forschung bei Krämer-Badoni & Wakenhut (1985) und Senger (1985). Lind & Wakenhut (1985) diskutieren das Problem in bezug auf die moralische Definition einer Situation.

kann diese Frage zu einem moralischen Problem werden. Ganz sicher wird aber seine moralische Entscheidung stark von seinen Überzeugungen beeinflußt ebenso wie die Moral-Stufe, auf der er das Problem abwägen und diskutieren möchte.

Das Problem läßt sich also auf die Frage zuspitzen: Kann es unabhängig von kulturspezifischen Wert- und Normfragen etwas geben, was eine für alle Menschen (und damit auch für alle Kulturen) gemeinsame Basis abgeben kann für den Diskurs über richtiges und falsches Handeln? Oder ist uns überhaupt keine Beurteilung von Handeln möglich? Es ist offensichtlich, dass eine Welt, in der es keine gemeinsame moralische Basis des Zusammenlebens gibt, ebenso unerträglich und unvorstellbar wäre wie eine Welt, in der jeder Mensch nach völlig uniformen Normen, Werten und Überzeugungen leben müßte. Es ist *eine* Sache, die religiösen Überzeugungen von Menschen zu respektieren und sie nicht an der Ausübung ihrer religiösen Handlungen zu hindern; es ist aber eine andere Sache, bestimmte, religiös begründete Handlungen unter dem Gesichtspunkt allgemeiner ethischer Prinzipien wie denen der Gerechtigkeit, Menschenwürde und wechselseitigen Achtung zu beurteilen und gegebenenfalls auf dieser Basis zu kritisieren. Die Achtung religiöser Bräuche anderer Kulturen kann uns nicht dazu veranlassen, einen relativistischen Ethikstandpunkt einzunehmen und so zum Beispiel die Tötung eines Menschen wegen Gotteslästerung als rechtens zu empfinden. Es entspricht also den Tatsachen, dass von einem universalistischen Standpunkt der Moral aus, wie er von Kohlberg, Habermas und Apel vertreten wird,[103] die Ausrottung von Andersgläubigen nicht akzeptiert werden kann; aber dies bedeutet kein allgemeines Werturteil über eine Kultur oder eine Religion, in der solches praktiziert oder gepredigt wird. Ebensowenig ist beispielsweise die Kritik an der Produktion und am Export von Massenvernichtungswaffen durch deutsche Firmen gleichzusetzen mit einer Bewertung der ganzen deutschen Kultur, denn die Komplexität einer Kultur und der sie konstituierenden Normen und Werte verbietet eine solche verallgemeinernde Beurteilung. Viele Normen und Werte ("Lebensformen") sind Antworten auf bestimmte naturhafte oder soziale Anpassungsnotwendigkeiten, die sich einer moralischen Bewertung entziehen. Als Zwischenergebnis dieser Überlegungen kann man festhalten: Ein Vergleich oder eine Messung der moralischen Entwicklung anderer Menschen und anderer Kulturen kann auf dieselbe

103 Vgl. Kohlberg 1971, 1984; Apel 1976, 1990; Habermas 1983.

Weise begründet werden wie die moralische Beurteilung des Verhaltens anderer Menschen schlechthin, nämlich nur auf der Grundlage von Kriterien, die den ethischen Grundprinzipien des *Kategorischen Imperativs* von Kant oder des Universalisierungsprinzips von Habermas und Apel genügen können.
Die kritische Frage für die ethische Bewertung von Meßmethoden könnte also lauten: *Kann ich wollen, dass die Kriterien der moralischen Einstufung zur Grundlage einer allgemeinen Bewertung von Verhaltens- und Urteilsweisen gemacht werden, also zum Beispiel auch auf mein eigenes Verhalten Anwendung finden*?[104] Ist es zum Beispiel gerechtfertigt zu sagen, dass jene Person oder Kultur mehr wert, besser oder gar moralischer ist, die in Kohlbergs *Moral Judgment Interview* einen höheren *Moral Maturity Score* (MMS) erzielt als die andere? Es ist offenbar so, dass, auch wenn wir als Forscher eine solche Verbindung von Messung und Wertung ablehnen, Ergebnisse der Moralforschung unmittelbar einen wertenden Charakter annehmen, sobald sie zur Grundlage praktischer Politik werden. Diese Wertung kann in den Augen verschiedener Beobachter verschieden ausfallen. Der Befund, dass in Entwicklungsländern eine geringere moralische Urteilsfähigkeit angetroffen wird als in industrialisierten Ländern, kann negative Konsequenzen haben. Er verstärkt dadurch Vorurteile über die "moralische Minderwertigkeit" dieser Kulturen. Er kann aber auch positiv gewendet werden, indem er reichere Länder dazu veranlaßt, dem Aufbau des Bildungssystems dieser Länder mehr Priorität einzuräumen als dem Aufbau militärischer Institutionen, um so die moralischen und demokratischen Fähigkeiten in diesen Ländern zu stärken. Es wäre daher falsch zu sagen, mit der Einstufung der moralischen Urteilsfähigkeit wäre *kein* wertendes Urteil über eine Kultur verbunden. Andererseits muß betont werden, dass dieselben Prinzipien, die eine bestimmte Form des Vergleichs (oder Messung) moralischer Kompetenzen erlauben oder fordern, auch implizieren, dass von solchen Meßwerten nicht der *moralische* Wert einer Person oder einer Kultur abhängig gemacht werden darf. Nach diesen ethischen Prinzipien hat jeder Mensch das

104 Einige der gängigen Methoden zur Messung der Moral können diesem Kriterium offenbar nicht genügen. So haben zum Beispiel Hartshorne und May (1928) zum Mittel der Täuschung gegriffen um festzustellen, ob Kinder in bestimmten Situationen mogeln würden. Diese Autoren haben, wie Kohlberg einmal bemerkte, den Kategorischen Imperativ nicht auf sich selbst angewendet, dessen Einhaltung sie von den Kindern aber erwarteten.

gleiche Recht auf Achtung seiner Menschenwürde, und zwar ohne Ansehen der Person, der Rasse, des Geschlechts oder der moralischen Urteilsstufe.
Das Methodenproblem der Moralforschung kann demnach auf folgende Weise präzisiert werden: *Wie kann die moralische Urteilsfähigkeit von Personen so gemessen werden, dass die Kriterien dieser Messung selbst den Maximen genügen, die sie zur Begründung ihrer Notwendigkeit anführt*? Oder: Wenn wir sagen, dass bestimmte Stufen oder Formen des moralischen Urteils adäquatere Lösungen eines Problems darstellen als andere, wie können wir festlegen, was die jeweils adäquateren sind, ohne das Selbstbestimmungsrecht individueller Problemlösungsstrategien eines jeden einzelnen (und auch jeder Kulturgruppe) zu verletzen? Die frühe moralpsychologische Forschung hat sich diesem Problem kaum gestellt, obwohl ihr dieses teilweise bewusst war. Im Falle von Intelligenztests wurden und werden moralische Einstufungen oft vorgenommen, ohne dass die moralische Natur dieser Bewertungen erkannt wird. Antworten und Lösungen der Versuchspersonen gelten dort allein als Ausdruck ihrer allgemeinen Intelligenz, ihres allgemeinen Verständnisses oder, in extremen Fällen, ihrer sozial-pathologischen Persönlichkeitsstruktur. In explizit moralischen Tests wird dem Problem oft dadurch zu begegnen versucht, dass man es im Prinzip anerkennt, aber seine Lösung unter pragmatischen Gesichtspunkten offen läßt. So hat die schon erwähnte Forschergruppe um Hartshorne und May, die Ende der zwanziger Jahre ein großangelegtes Forschungsprojekt zum moralischen Verhalten von Kindern durchführte, in ihrer theoretischen Abhandlung über die Natur moralischen Verhaltens festgestellt, dass die Einordnung eines bestimmten Verhaltens bei Kindern erst durch moralische und soziale Kriterien zureichend bestimmt werden kann. In bezug auf die Übersetzung ihrer Fragestellung in empirische Forschungsmethoden argumentierten die Autoren aber genau entgegengesetzt, nämlich, es könne keinen Fortschritt geben, wenn diese Kriterien nicht aus der Forschung herausgehalten wurden. Es war ihnen dadurch möglich, bestimmtes Verhalten als moralisch oder unmoralisch einzustufen, ohne dabei die Intentionen der untersuchten Kinder zu berücksichtigen. Wie Pittel und Mendelsohn (1966) feststellten, wurde das Bewertungsproblem in der Moralforschung bis auf wenige Ausnahmen “pragmatisch”, also vom Standpunkt des jeweiligen Forschers und seiner Kultur aus zu lösen versucht. Die Frage, ob dieser Standpunkt auch von den Untersuchungspersonen und ihrem kulturellen Hintergrund geteilt werden konnte, ob also der Standpunkt universalisierbar war, wurde kaum gestellt.

Kohlberg hat dieses Problem auf der Basis einer universalistischen Moralphilosophie und kulturvergleichender Empirie zu lösen versucht. Zur Unterstützung des universalistischen Anspruchs seiner Methode der Messung der moralischen Entwicklung kann er auf eine Reihe weithin anerkannter religiöser und moralphilosophischer Ansichten verweisen, die in Übereinstimmung mit seinem hierarchischen Modell von sechs Moral-Stufen stehen. So billigen ihm Philosophen wie Jürgen Habermas und Karl-Otto Apel zu, dass die Stufen 5 und 6 der moralischen Argumentation, die Stufen der postkonventionellen, prinzipienorientierten Moral, höhere moralische Adäquatheit besitzen als die unteren Stufen der konventionellen oder präkonventionellen Moral.[105] Dadurch hat Kohlberg immerhin sichergestellt, dass das Kriterium für die Messung der Moral nicht, wie das in anderen Ansätzen der Fall ist, bloß das Produkt einer zufälligen Entscheidung oder einer Konvention unter Psychologen und Methodologen ist, sondern auf einer viele Kulturen umfassenden, moralgeschichtlichen Tradition gründet. In den Fällen, in denen empirische Studien (geringe) Abweichungen der Moralentwicklung von der von ihm postulierten Invarianz ergaben, hat er sie als "Kinderkrankheiten" seiner Meßmethode angesehen und sie durch Modifikationen des Auswertungsverfahrens auszugleichen versucht.[106]

Gegen Kohlbergs Ansatz der Begründung der Moralmessung ist vom kulturvergleichenden Standpunkt jedoch einzuwenden, dass es zweifelhaft ist, ob es auf dem induktiven Weg der Anpassung der Meßmethode an empirische Befunde je gelingen kann, das Universalisierungsproblem zu lösen. Es kann nicht in Abrede gestellt werden, dass Kohlbergs Strategie einen bislang einmaligen und, wie ich meine, erfolgreichen Versuch darstellt, moralphilosophisch-abstrakte Forderungen nach Universalisierbarkeit, Reversibilität und Diskursivität auch forschungspraktisch einzulösen. Aber ich glaube nicht, dass dieser Teil des Kohlbergschen Forschungsprogramms aus epistemologischen und ethischen Gründen sein selbst gestecktes Ziel erreichen kann: a) Es gibt keine *logische*

105 Habermas 1985, Apel 1990. Beide gehen sogar soweit zu sagen, daß die Kohlberg-Stufen eine angemessene Beschreibung für die geistesgeschichtliche Entwicklung der Menschheit darstellen. Zu Versuchen der Rekonstruktion der Philogenese auf der Grundlage der Kohlberg-Piaget-Theorie vgl. auch Fetz 1982, Fetz & Oser 1986, Hartmann 1985.

106 Vgl. Colby & Kohlberg 1987; aber auch Kapitel 3.

Möglichkeit, um aus der Tatsache, dass sich die Adäquatheit der Kohlberg-Skala mit ihren Dilemma-Aufgaben in einer beachtlichen, aber kleinen Zahl von Kulturen bestätigen ließ, *mit Sicherheit* zu schließen, dass der Anspruch der Universalisierbarkeit des Invarianzmodells auch mit anderen Dilemma-Aufgaben und in anderen Kulturen oder Subkulturen sich bestätigen lassen wird.[107] b) Die Tatsache, dass Kohlbergs Auffassung von moralischer Adäquatheit bestimmter Dilemmalösungen mit einer Vielzahl philosophischer Meinungen kompatibel ist, entkräftet nicht den Einwand, es handle sich hierbei immer um ein in der untersuchten Kultur festgesetztes, *äußerliches* moralisches Kriterium. Auch wenn wir bestimmte philosophische Einwände gegen die transpersonale und transkulturelle Adäquatheit des Kohlbergschen Stufen-Modells nicht teilen, müssen wir jedoch anerkennen, dass diese Einwände bestehen und damit die Annahme eines universellen Konsenses widerlegen.[108] Indem Kohlberg von seinen Versuchspersonen verlangt, jedes der ihnen vorgelegten Handlungsdilemmas auf der *höchsten* Stufe *autonomer, selbstakzeptierter* Prinzipien zu lösen, um einen hohen Wert auf seiner Moralskala zu erhalten, begeht er, um mit Apel zu sprechen, einen *performativen Selbstwiderspruch.*

7.2 Ein diskursethischer Lösungsansatz des Methodenproblems der kulturvergleichenden Moralforschung

Wir haben mit dem *Moralisches Urteil Test* (MUT) einen Ansatz zur Messung der moralischen Urteilsfähigkeit gewählt, der einerseits den erreichten Stand moralwissenschaftlicher Forschung zu bewahren versucht, andererseits aber von einem internalen und damit einem nicht-konventionellen, kulturneutralen Kriterium für die Angemessenheit eines bestimmten Lösungsverhaltens ausgeht. Mit dem MUT verbindet sich nicht nur der Anspruch, einige der eingangs vorgetragenen Einwände gegen die bisherige kulturvergleichende Forschung zu berücksichtigen. Der MUT ist, da er die Form eines vergleichsweise kurzen, geschlossen Fragebogens hat, geeignet, die Begrenzung von Studien auf wenige Kulturen und wenige Personen zu überwinden. Der MUT ist vor allem mit dem

107 Zum Problem induktiver Verallgemeinerung allgemein vgl. Gadenne 1976, 1990.

108 Vgl. Gibbs 1979; Shweder 1982.

Ziel entwickelt worden, die moralische Urteilsfähigkeit von Personen *logisch unabhängig* von speziellen normativen Orientierungsstufen zu messen, ohne dass dabei der dingliche Zusammenhang zwischen affektiven und kognitivem Aspekt des moralischen Verhaltens getrennt wird. In diesem Ansatz wird *moralische Urteilskompetenz definiert als das Ausmaß, mit welchem selbstakzeptierte moralische Prinzipien – oder moralische Belange – kognitiv organisiert sind, d.h. das Ausmaß, mit dem eigene Moralprinzipien, Normen oder Regeln integriert und differenziert im Urteilsverhalten eines Individuums angewandt werden.*

Diese im Sinne von Locke (1983) "neutrale" Definition der Moral liefert, bei aller Verbesserungsbedürftigkeit, eine Basis für den Vergleich des moralischen Urteilens zwischen verschiedenen Kulturen, die besser universalisierbar und ethisch rechtfertigbar ist als bisherige Meßansätze. Mit dem MUT ist es möglich, die moralische Urteilsfähigkeit unabhängig von den gewählten moralischen Argumentationsstufen zu messen. Es ist damit möglich, bei *jeder* befragten Person zu prüfen, in welchem Ausmaß sie sich bei ihrem Urteil an *moralischen* Merkmalen der Dilemma-Situation orientiert, statt an anderen Kriterien. Um nicht nur prüfen zu können, ob und in welchem Ausmaß Argumente wegen ihres moralischen Gehalts akzeptiert oder zurückgewiesen werden, sondern auch testen zu können, ob die betreffende Person fähig ist, sich an moralischen Kategorien zu orientieren, wenn, wie in einer Konfliktsituation, die zu beurteilenden Argumente ihrer eigenen Meinung zu dem Konflikt *entgegenstehen*, werden im MUT nicht bloß PRO-Argumente zur Beurteilung vorgelegt, sondern auch CONTRA-Argumente. Im Grunde prüft der MUT dadurch nicht nur moralische Urteilsfähigkeit, sondern auch den Kern einer diskursiven, demokratischen Konfliktlösungskompetenz.[109] Der *affektive* oder *Inhaltsaspekt* des moralischen Urteilsverhaltens wird im MUT genauso operationalisiert wie in klassischen Einstellungsfragebögen, nämlich als Summen- bzw. Mittelwert

109 Diese Fähigkeit wird im MUT durch den C-Wert ausgedrückt. Der C-Wert ist der relative Anteil der Urteilsvarianz einer Person, der durch die moralische Qualität der vorgegebenen Argumente bestimmt ist, im Vergleich zu dem Anteil, der durch die Meinungskonformität der Argumente oder den Dilemma-Kontext oder eine Kombination dieser Faktoren bestimmt ist. Das Ergebnis der Berechnung (via intraindividueller Varianzkomponentenzerlegung) wird in Prozentanteilen ausgedrückt und als C-Wert (früher: "Determination Stufe") bezeichnet. Dieser Meßwert ist immer gemeint, wenn von moralischer Urteilsfähigkeit oder dem kognitiven Aspekt des moralischen Urteils die Rede ist (s. Kapitel 4).

über alle Beurteilungen, die eine Person zu den Argumenten einer der sechs Stufen abgibt. Für den affektiven Aspekt werden also in der Regel sechs Meßwerte berichtet; manchmal auch nur einer, wenn es genügt, die am stärksten präferierte Argumentationsstufe anzugeben.

7.3 Drei empirische Studien[110]

In den drei folgenden Studien sollen anhand von MUT-Daten drei Implikationen unserer Konzeptualisierung des Zusammenhangs zwischen Kultur und moralischen Urteilsfähigkeit empirisch überprüft werden. Erstens wird die These geprüft, ob beide Aspekte des moralischen Urteilsverhaltens, der affektive und der kognitive Aspekt, logisch unabhängig voneinander, aber am selben Verhaltenskomplex gemessen werden können, und ob diese Aspekte theoretisch vorhersagbar sind. Zweitens wird die These geprüft, dass sich bedeutsame kulturelle Unterschiede nur hinsichtlich der moralischen Urteils*fähigkeit* (= kognitiver Aspekt) finden lassen, aber nicht hinsichtlich moralischer Orientierungen (= affektiver Aspekt). Drittens wird die These untersucht, inwieweit Bildungseinflüsse und nicht kulturell-sozialisatorische Bedingungen ausschlaggebend für diese Unterschiede der Urteilsfähigkeit sind. Die erste Studie bildet die Grundlage für die beiden folgenden Analysen, in denen Personen mit einem unterschiedlichen kulturellen (genauer: gesellschaftlichen und ideologischen) Hintergrund gegenübergestellt werden. In ihr werden zunächst die zentralen Annahmen des kognitiv-entwicklungstheoretischen Modells der Moral, wie wir es hier rekonstruiert haben, an einer westdeutschen Erhebung bei Studierenden im ersten Semester überprüft. Um den Einfluß kultureller Unterschiede unter Konstanthaltung des Bildungsniveaus abschätzen zu können, werden diese Befunde in der zweiten Studie mit dem moralischen Urteilsverhalten von Personen verglichen, die weitgehend denselben Bildungsstatus (Universitätsstudium, 1. Semester) aufweisen und dieselbe Fachrichtung studieren, aber einen anderen kulturellen Hintergrund haben. In der dritten Studie wird dann anhand einer Studie bei Schülern mit unterschiedlichem Bil-

110 Diese folgenden Abschnitte stützen sich teilweise auf eine frühere Veröffentlichung (Lind 1986b).

dungsstatus abzuschätzen versucht, welchen Einfluß das Niveau der Bildungserfahrung im Vergleich zu kulturell-ideologischen Einflüssen hat.

7.3.1 Inhalt und Struktur des moralischen Urteilsverhaltens bei Studierenden

Die Untersuchungsgruppe der ersten Studie umfaßt 1.469 Universitätsstudenten und -studentinnen im ersten Semester mit den Studienfächern Germanistik, Medizin, Wirtschaftswissenschaften, Naturwissenschaften und Ingenieurwissenschaften. Die Befragten füllten den MUT (Standardversion) als Teil eines ausführlicheren Fragebogens über Studium, Wissenschaft, Beruf, Geschlechterrolle, Gesellschaft und politische Zielsetzungen aus.[111] Die Untersuchung fand im Wintersemester 1977/78 statt.

Die Resultate dieser Studie bestätigen die meisten Postulate der Kohlberg-Theorie und der von uns vorgenommenen Modifikationen: 1. In unserer westdeutschen Untersuchungsgruppe finden wir die Präferenzhierarchie für die sechs Moralstufen nach Kohlberg, die uns als Indikatoren für den affektiven Aspekt dienen, wie sie aufgrund der Kohlberg-Theorie zu erwarten ist: Argumentationen auf Stufe 6 (universelle Prinzipien der Gerechtigkeit) werden von den Studierenden fast immer am stärksten akzeptiert; die unteren Stufen werden mit abfallender Stufennummerierung geringer akzeptiert bzw. stärker abgelehnt.[112] Wenngleich kaum jemand, der von Kohlberg und anderen in offenen Interviews zu moralischen Dilemmas befragt wurde, Argumente der Stufe 6 artikuliert hat, sehen wir hier wie in anderen Studien, dass fast alle Personen postkonventionelle moralische Argumente gegenüber konventionellen oder präkonventionellen bevorzugen. Die Einstellungen zu verschiedenen moralischen

111 Die Größen des Samples und der Subsamples sind unten in Abb. 7.1 wiedergegeben. Die Untersuchung erfolgte 1978 im Rahmen des Teilprojektes Hochschulsozialisation des Sonderforschungsbereichs 23 Bildungsforschung an der Universität Konstanz (s. Kapitel 1).

112 Siehe Lind 1985a, S. 143. Bei den Daten in Abb. 7.2 handelt es sich um Mittelwerte. Sie zeigen aus Vergleichbarkeitsgründen für alle Länder nur die Antworten zum Sterbehilfe-Dilemma an. Beim deutschen Sample wurde bei dieser Darstellung also das Arbeiter-Dilemma weggelassen. Außerdem wurden die Antwortskalen in der Darstellung so transformiert, daß die deutschen Skalen (von -4 bis +4) mit den anderen Skalen, die nur von -3 bis +3 gingen, direkt vergleichbar wurden.

Orientierungen sind hier im großen und ganzen dieselben wie in den USA, wo bisher die einzigen Erhebungen zur Präferenzrangreihe der Stufen vorliegen (Rest 1973).

2. Viele der Studierenden, die wir untersucht haben, zeigen sich in einem vergleichsweise hohen Maße in ihrem Urteilsverhalten durch moralische Überlegungen determiniert. Ihre Bewertungen von Argumenten orientieren sich selbst dann weitgehend konsistent an moralischen Prinzipien, wenn die zu bewertenden Argumente ihrer eigenen Meinung entgegenstanden. Diese Fähigkeit ist, wie wir dieser und anderen Studien entnehmen können, bei Studierenden im ersten Semester noch nicht voll ausgebildet, aber sie ist deutlich höher als bei

	Geistes-wiss.	Natur-wiss.	Wirtschafts-wiss.	Medizin	Ingenieur-wiss.	Summe
Österreich	94	107	186	205	164	756
BRD	159	457	353	211	289	1,469
Niederlande	148	129	123	-	-	400
Polen	181	164	180	156	201	882
Jugoslawien	106	113	152	102	75	548
Summe	688	970	994	674	729	4,055

Abb. 7.1 Sample-Plan der international vergleichenden Studie des FORM-Projekts: Studierende im ersten Semester nach Fach und Land

Personen mit geringerer Allgemeinbildung.[113]

3. Bezüglich des affektiven Aspekts, der Präferenzhierarchie, gibt es keine Bildungs- und Geschlechtsunterschiede. Ein Vergleich der vorliegenden Befunde

113 Vgl. Rest 1973; Keasey 1974; Lind 1985a; s. auch Kapitel 11.

mit dem moralischen Urteilsverhalten von Erwachsenen mit sehr unterschiedlichem Bildungshintergrund zeigt, dass die gezeigte Präferenzrangreihe nicht auf Studierende beschränkt ist, sondern offenbar auch bei Menschen mit weniger Bildungserfahrung in derselben Weise anzutreffen ist.[114] Es scheint also gerechtfertigt anzunehmen, dass der affektive Aspekt des moralischen Urteilens weitgehend unabhängig vom Bildungsniveau ist. (Anders hingegen der kognitive Aspekt, der deutlich mit dem Bildungsniveau korreliert; s. Kapitel 11). Frauen und Männer besitzen eine fast identische Präferenzordnung moralischer

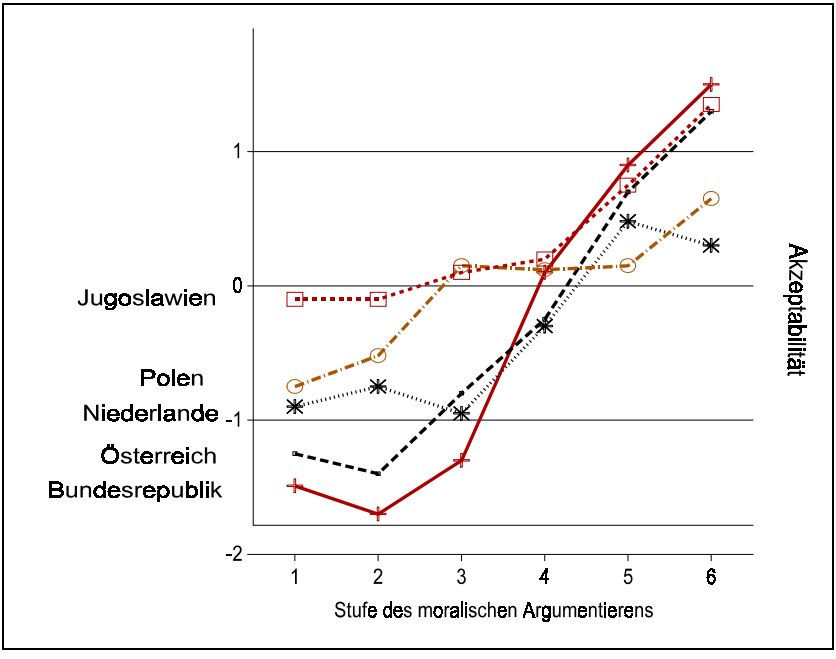


Abb. 7.2 Die Akzeptabilität der sechs Stufen des moralischen Argumentierens nach Kohlberg im MUT. Sample: Studierende verschiedener Ländern im ersten Semester. Quelle: Forschungsprojekt Hochschulsozialisation

114 Für weitere, empirische Bestätigungen der These einer für alle Bevölkerungsgruppen gleichen Präferenzhierarchie der sechs Kohlberg-Stufen vgl. Krämer-Badoni & Wakenhut 1985, S. 216 ff.; Heidbrink 1985, S. 267; Lind 1985a, S. 140 ff.; Oser 1981a, S. 404; Senger 1985, S. 235.

Belange in ihrem Urteilsverhalten. Außerdem ist die Korrelation zwischen Geschlecht und dem kognitiven Aspekt der moralischen Urteilskompetenz sehr gering und trotz der Größe des Samples statistisch nicht signifikant ($r = 0{,}028$). Frauen sind also weder mehr noch weniger moralisch urteilsfähig als Männer (vgl. Lind et al. 1987).

4. Wie im Rahmen der kognitiven Entwicklungstheorie angenommen wird, bilden die Korrelationen zwischen den Stufenpräferenzen in unserer Studentenuntersuchung eine Quasi-Simplex-Struktur (Kohlberg 1963); Das bedeutet, dass die Korrelationen zwischen den Präferenzen für zwei benachbarte Stufen des moralischen Urteils (z.B. Stufe 2 und 3) die höchsten sind, die gefunden wurden, und dass die gefundenen Korrelationen umso geringer werden, je weiter die Stufen von einander entfernt sind (z.B. 2 und 4, Stufe 2 und 5, etc.). Die moralischen Präferenzen der Universitätsstudenten stützen diese Annahme von Kohlberg. Jeder Versuch die Stufen anders anzuordnen erbrachte, dass keine andere Stufenordnung sich besser einem Quasi-Simplex annähert als die mit der theoretischen Präferenzordnung der Stufen 1 bis 6.[115] Die Daten dieser Studie ergeben übrigens einen eindeutigeren Quasi-Simplex als die Interviewdaten Kohlbergs (1963, S. 17).

5. Der kognitive Aspekt der moralischen Urteile der Studierenden, der in der bisherigen *Moral Judgment*-Forschung nicht direkt gemessen werden konnte, zeigt Entwicklungsfortschritte, wie von der Theorie vorhergesagt wurde. Die mit dem MUT erfaßte moralische Urteilskompetenz nimmt mit wachsenden Bildungserfahrungen zu (Kapitel 11). In der Zeit zwischen dem Abschluss des Gymnasiums und dem ersten Jahr an der Universität erweitert sich bei den meisten Studierenden die Kompetenz zur konsistenten Anwendung moralischer Prinzipien bei der Bewertung von Argumenten.

115 Vgl. Lind 1985a, S. 153-157. Wir sind hierbei trotz der Kritik von Kurtines und Greif (1974) immer noch weitgehend auf die einfache Inspektion der Korrelationsmatrix angewiesen. Um ein etwas schärferes Kriterium zu gewinnen, haben wir die Anpassungsgüte der Quasi-Simplex-Struktur mit dem diagonalen Ordnungsverfahren des Statistikprogramms KOSTAS (Nagl et al. 1986) überprüft. Der Vorschlag von Boom und Molenaar (1989), diese Annahme mittels LISREL (Jöreskog 1970) zu testen, erscheint nicht angemessen, da, wie eigene LISREL-Analysen unserer Studentendaten zeigen, schon bei geringen Abweichungen von einem idealen Simplex-Modell die Strukturgleichungen nicht gelöst und daher keine Meßwerte für die Passungsgüte berechnet werden können.

6. Das moralische Urteilsverhalten der Studierenden stimmt mit der Annahme einer kognitiv-affektiven Parallelität überein. Die moralische Urteilskompetenz der Studierenden steht in einem engen Zusammenhang mit dem Ausmaß und der Richtung ihrer Präferenzen für moralische Stufen, d.h., je konsistenter die Studierenden die moralischen Prinzipien in ihrem Urteil benutzen, desto positiver sind jeweils ihre Einstellungen zu höheren Entwicklungsstufen und desto negativer zu niedrigen Moralstufen. Mit anderen Worten im Vergleich zu Studierenden, die mit weniger Konsistenz urteilen -sind Studierende, die eine größere Kompetenz bei der Beurteilung von Argumenten aufweisen, auch diejenigen, die eine stärkere Präferenz für höhere als für niedrigere Moralstufen angeben.

Zusammenfassend kann man sagen, dass das Urteilsverhalten der deutschen Studierenden, die wir hier untersucht haben, weitgehend den theoretischen Erwartungen entspricht, die wir auf der Grundlage der Bildungstheorie der Moralentwicklung formuliert haben. Soweit ein Vergleich mit empirischen Befunden amerikanischer Studien möglich war, hat sich darüber hinaus gezeigt, dass die kulturellen Unterschiede zwischen diesen beiden Ländern, die trotz vieler Ähnlichkeiten bestehen, sich *nicht* in Differenzen bezüglich der moralischen Orientierungen niederschlagen. Ein direkter Vergleich der Urteilsfähigkeit zwischen beiden Ländern ist nicht möglich, da die verwendeten Meßinstrumente (MUT, MJI, DIT oder anderes) keine direkt vergleichbaren Meßwerte produzieren. Eine näherungsweise Einstufung im Sinne von Kohlberg auf der Basis von MUT-Daten erbrachte, dass deutsche Abiturienten überwiegend der Stufe 3 zuzuordnen sind und die Stufe 4 erst am Ende oder nach dem Studium konsolidiert ist (Lind 1985a, S. 159). Dies bestätigt Kohlbergs Befund, der in seiner amerikanischen Längsschnittuntersuchung feststellte, am Ende der *High school* gäbe es nur wenig Stufe-4-Denken und keinen einzigen Fall, der als Prinzipien-Moral (Stufe 5 oder 6) bezeichnet werden konnte. “Stufe 5”, so Kohlberg (1973, S. 192), “wird erst im Erwachsenenalter erreicht und nicht in der Adoleszenz”.

7.3.2 Kultur-Vergleich: Moralisches Urteilsverhalten bei Studierenden in fünf europäischen Ländern

Die Vergleichsuntersuchungen zur deutschen Studentenstudie wurden im Rahmen des *FORM-Projekts* durchgeführt.[116] Anders als in bisherigen Studien wurden in diesem Projekt die untersuchten Gruppen von Personen weitgehend nach Bildungsniveau (alle befanden sich im ersten Semester an der Universität) sowie nach Geschlecht und Studienfachrichtung parallelisiert. Zudem wurde die Untersuchung nicht auf demokratisch verfaßte, marktwirtschaftlich organisierte Länder beschränkt, sondern es wurden auch (damals noch) "kommunistische Gesellschaften" in den Vergleich einbezogen. Die Länder Polen und Jugoslawien (Slowenien) stellten damals sicherlich keine ganz typischen Vertreter des "kommunistischen Blocks" dar. Beide zeigen viele historische, politische und sozio-strukturelle Besonderheiten, die als alternative Erklärungen für mögliche Unterschiede im moralischen Urteilsverhalten ihrer Bürger herangezogen werden können. Aber solche Besonderheiten bestanden und bestehen auch bei den anderen osteuropäischen Ländern und ebenso innerhalb des "westlichen Lagers", also zwischen der Bundesrepublik, den Niederlanden und Österreich, wobei die beiden letztgenannten Länder sicher die größten Ähnlichkeiten aufweisen dürften. Der auffälligste Unterschied zwischen Ost und West war und ist neben der unterschiedlichen Gesellschaftsform die höhere Industrialisierung im Westen und der hohe Anteil ländlicher Lebensformen im Osten sowie, damit zusammenhängend, der deutliche Entwicklungsunterschied des Bildungswesens in Ost und West.

Die Studie wurde in allen fünf Ländern gleichzeitig (Winter 1977/78) durchgeführt. Die Rücklaufquote unterschied sich von Land zu Land, jedoch reichte sie im Durchschnitt über 50%. Mit einigen Ausnahmen erhielten alle Personen den Fragebogen postalisch (In Polen wurden die Fragebogen während der Vorlesung ausgefüllt). Der Fragebogen umfaßte, wie in der deutschen Befragung, neben moralischen Problemen noch andere Fragebereiche. Da in der internatio-

116 Die Abkürzung FORM steht für die französische Übersetzung des Titels der vom Europäischen Koordinationszentrum für Sozialwissenschaftliche Forschung in Wien koordinierten, international vergleichenden Längsschnittstudie "Diplomés de l'Université: leur *Form*ation et leur Conception de la Vie" unter der Leitung von Hansgert Peisert und Władisław Markiewicz. Für weitere Informationen vgl. Sandberger et al. 1982; Framhein & Langer 1984; Dippelhofer-Stiem & Lind 1987.

Stufe:	2	3	4	5	6
Stufe 1	.59	.30	.25	.11	.00
Stufe 2		.29	.30	.08	.03
Stufe 3			.28	.13	.06
Stufe 4				.13	.06
Stufe 5					.25

Abb. 7.3 Produkt-Moment-Korrelationen zwischen den Einstellungen zu den sechs Kohlberg-Stufen, mit der Variable "Land" auspartialisiert. Sample: Studierende im ersten Semester aus fünf Ländern. Quelle: Forschungsprojekt Hochschulsozialisation

nalen Studie die Anzahl der Fragen für jeden Bereich erheblich reduziert werden mußte, wurde hier nur das Euthanasie-Dilemma verwendet.

In bezug auf die Fragestellung der ideologischen Verzerrung liefert diese Untersuchung ähnliche Befunde wie die bundesrepublikanische Studie. Obwohl die Ideologien und Kulturen in Polen und Jugoslawien von denen in den U.S.A. oder Westeuropa zum Zeitpunkt der Erhebung stark abweichen, weisen die Studierenden dieser Länder die gleiche Präferenzordnungen für die sechs moralischen Stufen auf, wie die Studierenden in den "westlichen Demokratien". Es bestehen kaum Variationen bei dem affektiven Aspekt des moralischen Urteilsverhaltens, die auf einen kulturellen Unterschied hinsichtlich der moralischen Orientierungen hinweisen könnten. Die differentielle Akzeptanz der Stufe 1- bis 6-Argumente zeigen für alle fünf Länder die gleiche Präferenzordnung. Abb. 7.2 stellt die mittleren Präferenzen der Studierenden für jede der sechs Stufen der moralischen Urteilskompetenz dar. Den linearen Zusammenhang zwischen der Höhe der Akzeptabilität einer Argumentationsstufe und dem theoretisch vorhergesagten Präferenzrang läßt sich mittels einer multivariaten

Studierende, 1. Semester, aus...	N	Median
BRD	1398	62,07
Österreich	749	46,25
Niederlande	298	43,25
Jugoslawien (Slowenien)	518	33,17
Polen	836	22,50

Abb. 7.4 Moralische Urteilsfähigkeit (C-Werte im MUT). Quelle: Forschungsgruppe Hochschulsozialisation. (Anm.: Nur Euthanasie-Dilemma; die C-Werte sind daher generell höher als bei der Standard-Version)

Varianzanalyse genauer berechnen; er beträgt, ausgedrückt als linearer Korrelationskoeffizient, $r = 0{,}83$ ($F_{1.4011} = 8{,}896.81$; $p < 0{,}001$).[117]
Wie Abb. 7.2 auch zeigt, nehmen in allen fünf Ländern die durchschnittlichen Akzeptanzeinschätzungen kontinuierlich mit den moralischen Stufen zu; allerdings ergeben sich einige geringe Abweichungen von diesem Muster, die möglicherweise zum Teil durch Übersetzungsprobleme bedingt sein können. Studierenden aus Polen haben beispielsweise Stufe 3-Argumente als etwas akzeptabler als Stufe 4-Argumente eingeschätzt. Dies wird durch die sehr hohe Akzeptabilität des vorgegebenen Contra-Arguments auf Stufe 3 bewirkt.[118] Außer diesen geringfügigen Abweichungen stehen die Präferenzordnungen der Studierenden in allen fünf Ländern in Einklang mit der kognitiven Entwicklungstheorie. Die Hypothese einer universell akzeptierten Ordnung der moralischen

117 Für die Berechnung der Korrelation *r* aus dem *F*-Wert siehe Kapitel 9.

118 In der deutschen Version des MUT lautet dass Argument: "Der Arzt handelte falsch (als er die Frau tötete, die ihn bat, ihr Leiden zu beenden,) da er im Gegensatz zu den Überzeugungen seiner Kollegen handelte. Wenn sie gegen Tod auf Verlangen (Euthanasie) sind, dann sollte der Arzt dies nicht tun." Die polnische Übersetzung lautet: "weil er im Gegensatz zu den Prinzipien der Berufsethik handelte" – dieses ist mehr ein Argument der Stufe 4.

Stufen wird auch durch die Interkorrelation der Stufenpräferenzen der Studierenden gestützt, die wie in der deutschen Studie auch in allen vier anderen Ländern einen Quasi-Simplex bilden. Alle benachbarten Stufen sind höher miteinander korreliert, als Stufen, die der Theorie nach weiter voneinander entfernt liegen. Dieser Befund wird besonders gut dokumentiert durch die kombinierten Korrelationen über die fünf Länder hinweg, bei denen die nationalen Mittelwerte auspartialisiert wurden

Im Gegensatz zu dieser Übereinstimmung über moralische Ideale, weicht die moralische Urteilsfähigkeit der Studierenden aus den fünf Ländern sehr voneinander ab. Die Studierenden aus Deutschland und Österreich urteilten am konsistentesten bezüglich der moralischen Qualität der vorgegebenen Argumente. Polnische und jugoslawische Studierende zeigten dagegen eine deutlich geringere moralische Konsistenz in ihrer Beurteilung moralischer Argumente. Dieser Befund stützt sich zwar nur auf die Argumente der Studierenden zu einem Dilemma (Euthanasie); jedoch scheint er eine generelle Tendenz zu reflektieren, da ähnliche Tendenzen auch bei den Argumenten der Studierenden über andere, z.B. politische Belange, aufgetreten sind. Das Ausmaß der moralischen Urteilskompetenz in jeder nationalen Gruppe wird in Abb. 7.2 durch die Steilheit der Verbindungslinie zwischen den einzelnen Akzeptanzwerten zu jeder Stufe dargestellt. Der Zusammenhang zwischen der Kultur-Variable und der Höhe der moralischen Urteilsfähigkeit, wie sie mit dem MUT gemessen wurde, kann auch mittels einer multivariaten Varianzanalyse genauer berechnet werden; er beträgt $r = 0{,}27$ ($F_{4,5865} = 473{,}62$; $p < 0{,}001$). Da diese globale Berechnung für die gesamte Untersuchungsgruppe erfolgt, muß sie unterstellen, dass die Personen strukturell weitgehend ähnlich geurteilt haben. Um diese fragwürdige Annahme obsolet zu machen, kann man die individuelle Urteilsstruktur jedes einzelnen in die Analyse miteinbeziehen (s. Kapitel 4). Diese Berechnung ergibt einen noch größeren “Kultureffekt”. Der Zusammenhang zwischen Land und Moralkompetenz beträgt nun $r = 0.49$ ($F_{1,4} = 246.59$; $p < 0{,}001$).

Wir haben in der zweiten Studie den Blick ganz auf kulturbedingte Effekte gerichtet, da, wie wir durch die Versuchsanlage sicherzustellen versuchten, das Bildungsniveau aller Untersuchungsteilnehmer gleich war. Aber es wäre sicher vorschnell, die gefundene Korrelation allein als Ausdruck unterschiedlicher gesellschaftlicher Systeme oder gar nur als Effekt divergierender Ideologien zu

sehen. Aufgrund unserer Befunde können wir nicht bezweifeln, dass auch ideologische, religiöse oder, wie im Fall von Polen, beide Einflüsse sich im moralischen Urteilsverhalten wiederspiegeln. In der Tat müssen wir einige Aspekte dieses Verhaltens ganz eindeutig solchen Einflüssen zuschreiben, wie zum Beispiel die Antworten zur Dilemmalösung selbst. In Polen, das trotz kommunistischer Herrschaft stärker durch die katholische Religion geprägt ist als irgendeines der anderen Länder, haben die Studierenden zu vier Fünftel die Sterbehilfe abgelehnt; in den anderen Ländern dagegen marschieren die Studenten an der Spitze der Befürworter von Sterbehilfe. Aber diese Einflüsse haben nicht ausgereicht, die Rangfolge der Präferenzen der sechs moralischen Orientierungen in einem der Länder grundlegend zu verändern. Und es ist auch fraglich, ob entgegen des ersten Augenscheins, die kulturellen Besonderheiten für die recht großen Unterschiede bezüglich der moralischen Urteilskompetenz verantwortlich sind. Aufgrund der gegenwärtigen Datenlage ist nicht auszuschließen, dass, wie wir vermuten, die Unterschiede von dem unterschiedlichen Stand der Bildungssysteme verursacht werden. Zwar haben alle Studierenden formal den gleichen Sekundarschulabschluß. Aber in den noch überwiegend agrarwirtschaftlich geprägten Ländern Polen und Jugoslawien befindet sich, wie uns immer wieder berichtet wurde, das öffentliche Schulsystem in einer prekären Lage (schlecht ausgebildete und überlastete Lehrer, Zweischicht-Unterricht, schlechtes Lehrmaterial, geringe Unterstützung durch die – meist selbst bildungsfernen – Eltern).

7.3.3 Kultur versus Bildung als Faktor der Moralentwicklung

Um diese alternativen Möglichkeiten weiter aufklären zu können, sollen nun die Effekte der beiden Hauptvariablen “Kultur” und “Bildung” in einem direkten Vergleich gegenübergestellt werden. Die Daten für diesen Vergleich sind nicht eigens für diesen Zweck erhoben worden, sondern stammen aus der Untersuchung von Clarissa Seitz (1991) über Demokratiefähigkeit und moralisches Urteilen bei einheimischen und Aussiedlerkindern.[119] Sie eignen sich dennoch für unsere Fragestellung, da in dieser Untersuchung der kulturelle Hintergrund ebenso wie der Bildungshintergrund als Sampling-Variable systematisch

119 In die Analyse wurden 171 der insgesamt 218 befragten Personen einbezogen.

	Mittelstufe	Abitur
Einheimische	18,0 (N=101)	53,5 (N= 18)
Aussiedler	19,7 (N= 34)	34,5 (N= 18)

Abb. 7.5 Moralische Urteilsfähigkeit (C-Werte, MUT) bei Aussiedler- und einheimischen Schülern mit verschiedenen Bildungserfahrungen. Quelle: Seitz (1991)

kombiniert wurden. Die untersuchten nicht-einheimischen Jugendlichen haben überwiegend einen osteuropäischen kulturellen Hintergrund. Die meisten von ihnen sind in Polen, Rumänien und der (ehemaligen) Sowietunion geboren und erzogen worden und erst kürzlich in die Bundesrepublik gekommen. Einheimische und Nicht-Einheimische wurden sorgfältig nach verschiedenen Bildungsniveaus parallelisiert (Abiturklassen vs. Mittelstufe des Gymnasiums). Durch den Umstand, dass alle Jugendlichen deutschstämmig und hoch motiviert waren, die deutsche Sprache zu lernen, konnte bei allen Personen die Erfassung der moralischen Urteilsfähigkeit mit der deutschen Version des MUT erfolgen (nur wenige Schüler konnten wegen Sprachschwierigkeiten an der Untersuchung nicht teilnehmen). (Übersetzungsprobleme waren ausgeschlossen.) Zudem konnte die Standardversion mit zwei Dilemmas verwendet werden, was die Stabilität der Befunde erhöht. Die Befragung erfolgte im Klassenzimmer. Für die Bearbeitung des MUT und der anderen Fragen war für alle ausreichend Zeit gegeben. Da einige Untergruppen recht klein sind (N = 18 bei den Abiturienten), können die nachfolgenden Befunde nur mit Vorbehalt verallgemeinert werden.

Die Ergebnisse stärken die These, dass der kognitive Aspekt des moralischen Urteilsverhaltens nur wenig durch kulturell-sozialisatorische Einflüsse bedingt ist, sondern mehr durch Bildungserfahrungen. Bei den einheimischen wie bei den Aussiedlerkindern liegen die Werte der Abiturienten für die moralische Urteilsfähigkeit weit über denen der Mittelstufler. Bei den einheimischen Abiturienten liegen sie dazu nochmal um einiges höher als bei den Aussiedler-Abiturienten. Die multiple Varianzanalyse mit “Kultur” (einheimische vs. Aussiedlerkinder) und “Bildung” (Abitur vs. Mittelstufe) als unabhängige Variable und der moralischen Urteilsfähigkeit (MUT) als abhängige Variable erbringt

einen sehr deutlichen Befund zugunsten der Bildungsvariable. Selbst wenn, um Scheinkorrelationen auszuschalten, der Varianzbeitrag der Bildung um den Varianzbeitrag der Kultur bereinigt wird, verbleibt ein Zusammenhang von $r = 0.52$ zwischen Bildungsniveau und Urteilsfähigkeit. Die Korrelation zwischen kultureller Herkunft und Urteilsfähigkeit ist dagegen nur $r = 0.18$.[120]

7.4 Resümee

Die hier verfolgte zentrale Fragestellung war, ob die moralische Urteilskompetenz kognitive Fundamente aufweist, die universell und unabhängig von kulturspezifischen Werten oder Ideologien sind – wie die von Kohlberg entwickelte kognitive Entwicklungstheorie annimmt. Gemäß Kohlberg implizieren moralische Prinzipien eine universalistische Grundlage für Entscheidungen: "Neben Regelhaftigkeit oder Konsistenz bei der Verwendung einer Begründung für eine Entscheidung, impliziert", so Kohlberg (1958), "ein Prinzip die Universalität und Idealität von solch einer Begründung. Die Grundlage von Entscheidungen sind solche, die für alle wünschenswert wären" (S. 288). Ausgehend von einer relativistischen Perspektive haben Psychologen den sozialisatorischen Einfluß kulturell geprägter Werte und Überzeugungen hervorgehoben. Sie argumentieren, die Theorie gehe von einem bestimmten kultur-historischen Hintergrund aus, und man sollte deshalb erwarten, dass fundamentale moralische Prinzipien sich substantiell von Kultur zu Kultur unterscheiden. Dementsprechend könnten die Annahmen von Kohlberg über das Wesen und den Verlauf der kognitiven Moralentwicklung nur für Mitglieder der industrialisierten Kulturen westlicher Prägung gelten; jegliche Forschung über die kognitiven Entwicklungsaspekte der moralischen Urteilskompetenz könnte ideologisch oder funktional verzerrt sein. Obwohl eine wachsende Zahl von vergleichenden Studien außerhalb der U.S.A. in den vergangenen Jahren unternommen wurden – auch in nicht industrialisierten und "nicht-westlichen" Ländern – konnte die Kontroverse über die potentielle Verzerrung in Kohlbergs Theorie nur teilweise beigelegt werden. Die Untersuchungen weisen interkulturelle Unterschiede bei

120 Beide Korrelationen wurden aus den Varianzanteilen berechnet: $r = \sqrt{r^2}$.

den durchschnittlichen moralischen Entwicklungsstufen auf (gemessen mit dem MJI). Diese Resultate könnten als Bestätigung für relativistische Annahmen gedeutet werden, allerdings falsifizieren sie nicht entscheidend die Annahmen der kognitiven Entwicklungstheorie. Die erhaltenen Unterschiede könnten Hinweise auf Kulturunterschiede darstellen, sowohl in bezug auf die grundlegende moralische Ordnung als auch auf das Tempo der kognitiven Moralentwicklung (oder in bezug auf beide Aspekte). In Israel und in der Türkei durchgeführte Längsschnittstudien stützen die Annahme einer universellen Entwicklungssequenz (Snarey 1985), aber sie ermöglichen keine vollständige Lösung des Problems.

Drei Einwände können und sind gegen die Universalitätsthese geäußert worden. Erstens, die Ähnlichkeit der Entwicklungsordnung könnte durch eine zunehmende Verwestlichung der untersuchten Kulturen bedingt sein. Zweitens, diese Studien haben typischerweise Übergangsstufen von 3 nach 4 sowie Binnenveränderungen innerhalb von Stufen festgestellt, die der Annahme der "strukturellen Ganzheit", wie Kohlberg sie versteht, widersprechen. Drittens, die Universalität der moralischen Prinzipien wurde nicht direkt eingeschätzt, sondern wurde von der moralischen Urteilskompetenz der Individuen abgeleitet, obgleich die Messung dieser moralischen Urteilskompetenz genau das ist, was durch die kulturelle Spezifitätshypothese in Frage gestellt wurde.

Diese Einwände sind aufgrund unserer Untersuchung neu zu gewichten. Sollte es nicht gelingen, das Ausmaß der Verwestlichung präziser zu bestimmen, läuft der erste Einwand Gefahr tautologisch zu sein. Da Vergleiche mit Ländern stattgefunden haben, die vorab als "nicht-westlich" eingestuft wurden, kann der Befund, dass die gleiche Entwicklungssequenz wie in den U.S.A. vorliegt, nicht als ein Indikator für die Verwestlichung dieser Kulturen herhalten. In unserer Studie haben wir auch zwei (zum Zeitpunkt der Untersuchung) sozialistische Länder untersucht. Gemäß ihrer Ideologie, liegt die Last der Verantwortung für eine politische oder moralische Entscheidung auf den Schultern der Kollektive, im Gegensatz zu den westlichen Demokratien, wo die Verantwortlichkeit dem Individuum zugeordnet wird. In der vorliegenden Untersuchung konnten wir zeigen, dass die sechs Typen der Moralität, die in Kohlbergs Schemata beschrieben werden, nicht kulturspezifisch sind, sondern universell akzeptiert werden. Befragte aus ganz unterschiedlichen Kulturen weisen dieselbe, von der Theorie vorhergesagte Präferenzordnung auf. Auch außerhalb der U.S.A., in industrialisierten Kulturen westlicher Prägung (wie der Bundesrepublik

Deutschland, Österreich, den Niederlanden) wie auch in osteuropäischen, sozialistischen Nationen (wie Polen und Jugoslawien) besteht ein erstaunlich universeller Konsens über die moralischen Prinzipien, an denen sich die Lösung von moralischen Dilemmas orientieren soll. Gleichzeitig gibt es zwischen den fünf Ländern beachtliche Unterschiede bezüglich der moralischen Urteilskompetenz der Studierenden, wobei diese Fähigkeit in bezug auf die *selbstakzeptierte* moralische Orientierung der Befragten gemessen wurde. Während beispielsweise Geschlecht und Studienfach keinen bzw. kaum einen Effekt auf die moralische Urteilskompetenz ausübten, scheint die Zugehörigkeit zu einem bestimmten Land (und damit auch zu einem bestimmten Bildungssystem) für die moralische Entwicklung einen großen Unterschied auszumachen. Die Studie über Aussiedlerkinder belegt, dass dafür die nationalen, kulturellen Unterschiede nur vordergründig verantwortlich sind. Es scheint, dass sich dahinter in hohem Maß die Wirkung unterschiedlich weit entwickelter Bildungssysteme verbirgt. Unterschiede in der Qualität der Ausbildung zwischen den Ländern scheinen eine wesentliche Rolle für die Differenzen in der Geschwindigkeit der Moralentwicklung zu spielen.

Insgesamt haben unsere drei kulturvergleichenden Studien eine deutliche Bestätigung der Bildungstheorie der moralischen Entwicklung erbracht und die Sozialisationstheorie geschwächt. Es sind nicht die moralischen Einstellungen, nach denen sich die Kulturen unterscheiden. Das deutlichste Unterscheidungsmerkmal ist vielmehr die moralische Urteilsfähigkeit. Somit kann es sich bei der Moralentwicklung nicht primär um die Anpassung des Individuums an einen bestimmten kulturellen Kontext handeln, sondern um den Erwerb neuer kognitiver Strukturen, die durch die gesellschaftlich bereit gestellten Bildungsprozesse gefördert werden. Die Bildungstheorie stimmt hier mit der kognitiven Entwicklungstheorie überein, die postuliert, dass Gelegenheiten zur Teilnahme an öffentlichen oder privaten Diskussionen über politische Belange wesentlich für die moralische Entwicklung sind. Ein Mangel an solchen Gelegenheiten kann die moralisch-kognitive Entwicklung einer Person verlangsamen oder sie ganz verhindern. In Kulturen, wo die Verantwortung für politische Entscheidungen beim sozialen Kollektiv statt beim Individuum liegt, fühlt dieses sich kaum ermutigt, an moralischen Diskussionen teilzunehmen, und erhält so auch keine Möglichkeit, seine moralische Urteilsfähigkeit weiterzubilden.

8 Förderung der moralischen Urteilsfähigkeit durch Bildungsprozesse: Ein Interventionsexperiment

Die Bildungstheorie macht zwei Vorhersagen, mit denen sie sich von der Sozialisationstheorie wie auch von anderen Theorien abgrenzen läßt. Sie sagt voraus, a) dass das moralische Urteilen einen starken Fähigkeitsaspekt hat, und b) dass sich dieser Aspekt durch langwierige Bildungsprozesse fördern, aber weder von allein entwickeln noch durch bloßen sozialen Anpassungsdruck verändern läßt. Wir haben den Nachweis für die Richtigkeit dieser Vorhersagen in den vorigen Kapiteln hauptsächlich *indirekt* geführt, nämlich dadurch, dass wir die Vorhersagen der beiden anderen Theorien widerlegt haben. Aus den beiden Tatsachen, dass sich eine höhere moralische Urteilsfähigkeit nicht beliebig, durch bloße Instruktion erreichen läßt, und der Tatsache, dass sich scheinbar kulturbedingte Unterschiede in der moralischen Urteilsfähigkeit weitgehend auf Bildungsunterschiede zurückführen lassen, haben wir gefolgert, dass die *Sozialisationstheorie* keine adäquate Erklärung der Moralentwicklung bietet. Aus der Tatsache, dass sich die moralische Urteilsfähigkeit zu-rückentwickeln kann, wenn Bildungsprozesse längere Zeit wegfallen, haben wir den Schluß gezogen, dass auch eine *Reifungstheorie* der Moralentwicklung nicht haltbar ist. Die Bildungstheorie selbst ist aber bislang nicht auf den Prüfstand gestellt worden. Dies soll hier und in den folgenden Kapiteln geschehen.

Hier soll zunächst die bildungstheoretische Kernthese untersucht werden, wonach sich die moralische Urteilsfähigkeit durch die Bereitstellung von moralisch relevanten Bildungserfahrungen fördern läßt. Es soll insbesondere geprüft werden, ob und in welchem Ausmaß eine Kombination aus gezielten didaktischen Methoden (wie der *Dilemma-Diskussions-Methode*) und einem *Just community*-Programm zur Verbesserung des moralisch-demokratischen Klimas der Schule und zur Steigerung der moralischen Urteilsfähigkeit beiträgt. Auch nach der *Reifungstheorie* sollten solche pädagogische Programme einen positiven Effekt auf die Moralentwicklung haben, wenngleich diese Theorie von einer unspezifischen “Stimulation” der Entwicklung ausgeht, statt, wie die Bildungstheorie, von einer gezielten “Förderung”. Eine andere Vorhersage macht die *Sozialisationstheorie*. Danach müßten solche Programme allein die moralischen *Einstellungen* der SchülerInnen verändern, indem sich deren Wertvor-

stellungen den pädagogischen Zielen des Versuchsprogramms der Schule anpassen. Bezüglich des kognitiven Aspekts schließt diese Theorie eine Änderung in der beabsichtigten Richtung nicht aus, sagt sie aber auch nicht explizit voraus, da sie diesen Aspekt ignoriert (Kapitel 6).

8.1 Sozialisationstheoretische Forschung

Aus sozialisationstheoretischer Sicht ist eine Veränderung des moralischen Verhaltens auch kurzfristig, durch die Veränderung des Werteklimas einer Gruppe möglich. Die klassischen Experimente hierzu wurden von der Forschergruppe um Kurt Lewin durchgeführt. Um herauszufinden, wie sich das Gruppenklima auf das Verhalten ihrer Mitglieder auswirkt, bildeten sie kleine Schülergruppen, die unter Anleitung eines erwachsenen Leiters bestimmte Aufgaben erledigen sollten. Die Leiter waren zumeist die Wissenschaftler selbst, die eine von drei Leitungsstilen einnahmen: autokratisch, *Laissez-faire* und demokratisch. Beobachtet wurden die (kurzfristigen) Änderungen des Schülerverhaltens bezüglich Aggressivität und Produktivität. Bekanntermaßen zeigte sich, dass in autokratisch geführten Schülergruppen sehr viel mehr wechselseitige Aggressionen (40 mal soviel) auftraten als in demokratischen Gruppen, dass nur in den autokratischen Gruppen das "Sündenbock-Syndrom" auftrat (einer wurde zur bevorzugten Zielscheibe aggressiven Verhaltens), und dass das Aggressionsniveau besonders anstieg, wenn der autokratische Leiter kurzzeitig abwesend war. Ein weiterer, kurzzeitiger Anstieg von Aggressionshandlungen unter den Schülern war zu verzeichnen, wenn der Lehrer seinen Führungsstil von autokratisch auf *Laissez-faire* änderte, also nach einer strengen Leitung der Gruppen zu einer völligen Freizügigkeit überging – ein Phänomen, das wir gegenwärtig im Verhalten der Völker bestätigt sehen können, die erst kürzlich von einem autokratischen Regime befreit wurden. Befragt, wen sie lieber mochten, nannten 19 von 20 Schülern den demokratischen Leiter und nur ein Schüler den autokratischen (Lewin et al. 1939, S. 279-284). In einer späteren Untersuchung zeigte sich zudem, dass, wenn Gruppen auf sich allein gestellt waren, die autokratisch geführten Gruppen keine eigenen Arbeiten begannen, die *Laissez-faire*-Gruppen zwar *aktiv* aber nicht *produktiv* waren und nur die demokratisch geführten Gruppen auch produktiv waren (Lippitt & White 1973/1947, S. 336).

Obwohl diese Befunde auch heute noch als sehr aufschlußreich für die Rolle des Gruppenleiters gelten können, ist ihr Aussagewert für unsere Fragestellung sehr begrenzt. Gegen diese Befunde kann eingewendet werden, a) dass an den Experimenten nur sehr kleine Schülergruppen beteiligt waren, b) dass diese Gruppen nicht gewachsen, sondern *ad hoc* gebildet wurden und c) dass die "Leiter", die das jeweilige Moralklima erzeugt haben, keine 'normalen' Lehrer waren, sondern Wissenschaftler, die mit einem besonderen Engagement und Interesse an die Sache herangingen. Wir können daher nicht sicher sein, ob sich die positiven Effekte der Demokratisierung, wie sie von Lewin et al. berichtet werden, auch in einer großen, gewachsenen Gruppe, in einer durch Traditionen geprägten Institution (z.B. der Schule) einstellen würden. Zudem muß bezweifelt werden, dass es möglich war, in der kurzen Zeit, die in den Experimenten zur Verfügung stand, außer den moralischen Einstellungen der SchülerInnen auch die *Fähigkeiten* zu verändern, die für moralisch-demokratische Interaktionen in der Gruppe notwendig sind. Vom Standpunkt der Bildungstheorie der Moralentwicklung ist vielmehr zu vermuten, dass der "demokratische" Gruppenleiter die bereits vorhandenen demokratischen Einstellungen und Kompetenzen der Jugendlichen ansprach, während die *Laissez-faire*-Leiter und die autokratischen Leiter die Gruppenmitglieder dazu brachten, auf einer tieferen Stufe zu handeln als es ihren tatsächlichen Fähigkeiten entsprach.

8.2 Der *Just community*-Ansatz: Effekte der moralischen Atmosphäre

Die Bedeutung von Entwicklungs- und Bildungsprozessen (gegenüber reinen Veränderungen des Werteklimas) haben schon Jean Piaget (1973a) und John Dewey (1964/1915) hervorgehoben. Beide gingen von der Annahme aus, dass die Förderung demokratischer Fähigkeiten beim einzelnen und die Demokratisierung des Interaktionsklimas einander wechselseitig bedingen. Aber weder Dewey noch Piaget haben eigene empirische Untersuchungen zu dieser Frage beigesteuert. Erst Kohlberg (z.B. 1987) und seine Kollegen (Power et al. 1987) haben die moralisch-demokratische Atmosphäre zum Gegenstand entwicklungspädagogischer Experimente gemacht. Sie haben dafür langfristige Versuche gewählt, die mindestens auf mehrere Stunden, oft aber auf mehrere Wochen (siehe Kapitel 9) oder, wie im Fall ihrer *Just community* (JC)-Schulprojekte,

mehrere Jahre angelegt waren. Diese Experimente erfüllen damit eine wichtige Voraussetzung für die valide Überprüfung der Bildungstheorie der Moralentwicklung, da nach dieser Theorie die Entwicklung der moralischen Urteilsfähigkeit auf ausreichend Zeit angewiesen ist.

Im Mittelpunkt dieses Kapitels steht der *JC*-Ansatz, oft als *Gerechte Gemeinschaft*-Ansatz übersetzt. Kohlberg und seine Kollegen Clark Power und Ann Higgins berichten von deutlichen Gewinnen bei der Moralentwicklung (Power et al. 1989). Diese Effekte bestätigen, für sich genommen, die beiden "kognitiven" Erklärungsansätze. Der Gewinn an moralischer Urteilsfähigkeit war aber in diesen sehr langfristigen und intensiven Projekten relativ gering,[121] geringer zum Beispiel als der Gewinn, der in relativ kurzer Zeit durch die mehr spezifische Methode der *Dilemma-Diskussion* im Unterricht erzielt wurde. (Kapitel 9). Würde sich durch weitere Untersuchungen der Befund erhärten lassen, dass, wie Enright et al. (1983) vermuten, allgemeine Verbesserungen der "moralischen Atmosphäre" der Schule geringere Wirkungen haben als spezifische pädagogische Curricula, würde die Bildungstheorie der Moralentwicklung gegenüber der Reifungstheorie deutlich gestärkt.

Die bisherigen Untersuchungen stellen jedoch keinen gesicherten Beleg für die eine oder andere Theorie dar, da sie methodologische Schwächen aufweisen. Aufgrund des Designs dieser Experimente ist der Einwand nicht zu entkräften, dass es sich bei dem Zuwachs an moralischer Urteilsfähigkeit in den *JC*-Projekten um einen Effekt der Teilnehmerselektion und/oder der Erhebungsmethode handelt. Die *JC*-Projekte von Kohlberg und seinen Kollegen wurden fast immer mit hochselektiven Schülergruppen durchgeführt. In einem Teil der Schulen bestanden die Experimentalgruppen aus Schülern, die ein besonderes Interesse für das Projekt hatten, das ihnen als ein freiwilliger Zusatzkurs angeboten wurde. In anderen Schulen nahmen nur sozial schwache Schüler teil, die von ihren Lehrern aus sozialpädagogischen Gründen in solche Kurse geschickt wurden. Auch kann nicht ausgeschlossen werden, dass die durchweg mit dem *Moral Judgment Interview* (MJI) gewonnenen Daten mehr eine veränderte moralische Einstellung der SchülerInnen anzeigen als eine gesteigerte mo-

121 Über alle Untersuchungen gerechnet betrugen die Gewinne 5 MMS-Punkte (von 500 möglichen Punkten), wie Power et al. (1989, S. 280) berichten. Dies entspricht einem Gewinn von nur 1% gemessen an der absoluten MMS-Skala (MMS = *Moral Maturity Score*) und beträgt 1/3 des angenommenen Meßfehlers von 15 MMS-Punkten.

ralische Urteilsfähigkeit (Kapitel 3), dass sie also in einem unbekannten Ausmaß einen "Halo"-Effekt wiederspiegeln[122].

Zudem liegen zur Frage, wie die moralische Atmosphäre der *JC*-Schulen verändert wurde, nur punktuelle, zuweilen auch anekdotisch Informationen vor, aber nur wenige systematische Daten. Zwischen der Bedeutung, die verschiedene Aspekte der moralischen Atmosphäre in der Schule auf die Schüler haben, und unserem Wissen darüber besteht, wie Oser und Althof (1992) vermerken, eine Lücke. Es fehlen weitgehend Studien, die sich außer auf das moralische Urteil auch auf andere, handlungsrelevante Einstellungen und Fähigkeiten bei Schülern und auf Aspekte des Schullebens als sozialer Organisation richten, die für die Entwicklung der Demokratie in der Schule bedeutsam sind. Power et al. (1989) haben erste Daten dazu vorgelegt. Sie konnten zeigen, dass sich der pädagogische Umgestaltungsprozeß in *JC*-Schulen auch positiv auf die Struktur der individuellen "Normbegründung" und auf das Gemeinschaftsbewußtsein auswirkt. Diese Studien weisen jedoch auch bezüglich dieser Fragestellung einige Unzulänglichkeiten auf: a) Sie sind weitgehend auf den einzelnen gerichtet und kaum auf die soziale Gruppe (hier: die Gemeinschaft der Schüler und Lehrer). b) Die Anzahl der beobachteten bzw. befragten Schüler ist oft sehr gering, da die Versuche auf Teilschulen (School-within-School) begrenzt waren und zumeist nur sehr zeitaufwendige Erhebungsinstrumente zur Verfügung standen. c) Es wurde bislang weitgehend darauf verzichtet, bei den Schülerurteilen über ihre schulische Umwelt zwischen der subjektiven Relevanz bestimmter Urteilsdimensionen und der Beurteilung selbst zu unterscheiden.[123] Die Überwindung dieser Beschränkungen ist Voraussetzung für eine adäquate Interpretation der Beobachtungen im Hinblick auf Empfehlungen für die schulische Praxis.

Eine spezifische Lücke in unserem Wissen betrifft die moralische Atmosphäre an *deutschen* Schulen und die Entwicklung der moralischen Urteilsfähigkeit von Schülern hier zu Lande. Es liegen hierzu nur punktuelle Informationen vor.

122 Hofstätter (1973) bezeichnet als „Halo-Effekt" die häufig beobachtete Tatsache, daß bei relativ freien Persönlichkeitsbeurteilungen "einzelne Eigenschaften die Beurteilung anderer Eigenschaften in einem durch objektiv feststellbare Korrelationen nicht rechtfertigbaren Maß in Mitleidenschaft ziehen" (S. 191 f.).

123 Auf die Notwendigkeit, zwischen Wahrnehmung und Bewertung zu unterscheiden, hat Barbara Dippelhofer-Stiem (1983) am Beispiel der Hochschulforschung hingewiesen.

So findet sich in den sehr breit angelegten und relativ neuen Studien über die psycho-soziale Entwicklung von 12– bis 16jährigen von Helmut Fend (1990) und seinen Mitarbeitern der Hinweis, dass in diesem Alter die anfänglich vorhandene, kritiklose Akzeptanz der Schule als Institution schwindet: Ältere Schüler fühlen sich in der Schule weit weniger wohl als jüngere (S. 101). Ganz allgemein sinkt "das Vertrauen in geordnete und gerechte Verhältnisse [. . .] von Schuljahr zu Schuljahr" (S. 117). "Politische Analysefähigkeit" hingegen nimmt, wie Fend und seine Kollegen zeigen, im Laufe der Schulzeit sehr deutlich zu. Leistungsanforderungen werden mit wachsendem Alter kritischer betrachtet. Auf dieser Basis kann man erwarten, dass Schüler die moralische Atmosphäre ihrer Schule auch im Laufe des Projekts kritischer beurteilen.
Auch wissen wir wenig darüber, wie sich die Beziehungen zwischen den Schülern untereinander und zwischen Schülern und Lehrern gestalten, wie sich die Identifikation der Jugendlichen mit ihrer Schule vollzieht und wie sich ihre Einstellungen zu Regelverstößen bilden. Jugendliche im Alter zwischen 10 und 16 fallen häufig durch Normverletzungen auf (Klockhaus & Habermann-Morbey 1986). Inwieweit erkennen – und anerkennen – sie Normen und sind sich dessen bewusst, dass sie oder ihre Gleichaltrigen eine Norm verletzen? Inwieweit fühlen sich Jugendliche in diesem Alter bereits (mit)verantwortlich für die Aufstellung von Regeln und die Überwachung ihrer Einhaltung? Ist es ihnen in diesem Alter bereits wichtig, an Entscheidungsprozessen in der Schule teilzuhaben?

8.3 Fragestellung

Die Hypothesen, die sich aus unserer Gegenüberstellung der drei Theorien der Moralentwicklung ergeben haben, sowie die offen Fragen, die durch die kritische Sichtung der bisherigen Forschung zum *Just community*-Ansatz aufgeworfen wurden, sollen im folgenden anhand der Ergebnisse des schulischen Erprobungsprojekts "Demokratie und Erziehung in der Schule" (DES) zu klären versucht werden (Lind & Raschert 1987). Ziele und Anlage des DES-Projekts orientierten sich weitgehend am *JC*-Ansatz. Das wichtigste Gestaltungsprinzip des *JC*-Ansatzes ist, dass alle Beteiligten sich wechselseitig achten und soziale und normative Konflikte auf der Grundlage allgemeiner Prinzipien, im Diskurs, zu lösen versuchen, statt sie zu ignorieren oder mit Gewalt zu entscheiden. Dies

setzt die Bereitschaft und Fähigkeit aller, Schüler wie Lehrer, voraus, sich in sozialen Prozessen lernend zu verändern (Entwicklung als Ziel der Erziehung).[124] Die Initiatoren des DES-Projekts sind von der Annahme ausgegangen, dass Jugendliche moralisch-demokratische Einstellungen und Fähigkeiten in der Schule nur lernen können, wenn das Schulleben demokratisch gestaltet ist und den Heranwachsenden Gelegenheit bietet, solche Fähigkeiten zu üben. Es hat sich gezeigt, dass Konflikte und Probleme in der Schule nicht vermieden werden können, indem sie ignoriert oder vom Lehrer im Alleingang, ohne Mitwirkung der Schüler, gelöst werden. Es bestand die Überzeugung, dass demokratisches Lernen nur angeregt und gefördert wird, wenn Konflikte und Probleme im Unterricht und im Leben der Schule selbst durch gewaltfreien Meinungsstreit, auf der Basis allgemein geteilter moralischer Prinzipien bearbeitet werden. Mit dem DES-Projekt sollen hier drei Fragen näher untersucht werden: 1. Inwieweit sind die beteiligten Schüler bereit für demokratische Veränderungen, und inwieweit halten sie zentrale Werte und Prinzipien demokratischen Zusammenlebens (wie Gerechtigkeit, Beteiligung an Entscheidungen, wechselseitige Achtung usw.) für wichtig? 2. Verändert sich die moralisch-demokratische Atmosphäre in den Schulen infolge des Interventionsprogramms des DES-Projekts? 3. Hat sich die (moralisch-demokratische) Urteilsfähigkeit der teilnehmenden Schüler durch das DES-Projekt fördern lassen, und zwar über das Maß hinaus, in dem es vom normalen, allgemeinbildenden Unterricht bereits gefördert wird?

8.4 Anlage und Durchführung der Untersuchung

Das DES-Projekt weist gegenüber den bisherigen *JC*-Programmen einige Besonderheiten auf, die für die Beurteilung der Ergebnisse von Belang sind: An dem *Just community*-Programm zur Förderung der Moralentwicklung nahmen

124 Für Darstellungen verschiedener *Gerechte Gemeinschaft*-Schulen in den U.S.A. vgl. u.a. Higgins 1987; Lind 1987; Power et al. 1989; Oser & Althof 1992; Schreiner 1989. Zu den philosophisch-erziehungswissenschaftlichen Hintergründen dieses Ansatzes vgl. Kohlberg 1987. Zur Aufnahme und Diskussion im deutschen Sprachraum vgl. u.a. Aufenanger et al. 1981; Dobbelstein-Osthoff 1990; Edelstein 1986, 1987; Gilgenmann 1984; Hagemann et al. 1982; Heidbrink 1992; Reinhardt 1980; Oser & Althof 1992; Schreiner 1981, 1992; Schirp 1990.

Schule	Schuljahr	1. Befragung	2. Befragung	Beide Befragungen	Gesamtpersonen
Schule A	6./8.	32	44	24	100
Schule B	5./7.	28	33	28	89
	5./6.	21	20	44	85
Schule C	6./7.	11	7	41	59
	8./9.	17	17	28	62
Gesamt		109	121	165	395

Abb. 8.1 DES-Projekt: Auswertungsplan

10 – bis 15jährige Kinder (bisher zumeist 15 – bis 18jährige) teil. Es wurden *alle Kinder einer Jahrgangsstufe* in das Programm einbezogen. Die Förderung wurde nicht auf eine bestimmte Gruppe begrenzt, wie dies in anderen Projekten normalerweise geschah. Bei den Schülern und Schülerinnen in NRW handelte es sich insgesamt also weder um besonders gute Schüler großbürgerlicher Herkunft, die sich, wie etwa die Schüler der Scarsdale-Schule in den USA, den "Luxus" demokratischer Teilhabe am Schulleben leisten konnten, noch um besonders "erziehungsbedürftige" Jugendliche, wie die in der Roosevelt High School in der Süd-Bronx. Schließlich wurden die Fördermaßnahmen ausschließlich von *regulären Lehrern* und nicht von speziell ausgebildeten Pädagogen oder Psychologen durchgeführt. Die Lehrer erhielten eine kurze Einführung in den *JC*-Ansatz und begleitende Beratung über die Projektlaufzeit. Insbesondere im ersten Projektjahr waren kaum projektspezifische Effekte zu erwarten, da fast alle Lehrer diese Zeit benötigten, um den Ansatz neben ihrer normalen Arbeit kennen- und einsetzen zu lernen. Als besonders schwierig erwies sich für viele Lehrer die Unterscheidung zwischen genuin *moralischen* Dilemmas, bei denen es keine "technische" Lösung gibt, und allgemeinen Problemgeschichten, die sich mit Geschick und Wissen lösen lassen.

Das Design der Untersuchung (zu der auch eine Vorstudie bei 200 Schülern gehört) bietet verschiedene Möglichkeiten, um die Frage nach der Wirkung des DES-Projekts und der Größe der Effekte näherungsweise zu beantworten. In

einer Schule (Schule „B“)[125] wurden hintereinander zwei 5. Klassen in den Versuch einbezogen. Dies ermöglicht die Abschätzung der Veränderung der Schulatmosphäre während des DES–Projekts unabhängig von altersbedingten Veränderungen in der Schulwahrnehmung von einer generell affirmativen zu einer generell kritischen Sichtweise. In einer Schule „C“ nahmen zwei verschiedene Jahrgänge (die 6. und 8. Klasse) zur selben Zeit am Projekt teil. Aus ihrem Vergleich lassen sich Anhaltspunkte für das Ausmaß von “normalen”, schulbedingten Veränderungen gewinnen, ebenso wie aus dem Vergleich mit einer Vorstudie, an der Schüler aus den Klassen 8 bis 12 teilgenommen hatten.[126] Die DES-Schulen repräsentieren drei unterschiedliche Schulformen: eine *Hauptschule* (Schule A; drei 6. Klassen), eine *Realschule* (Schule B; zweimal drei 5. Klassen) und ein *Gymnasium* (Schule C; jeweils zwei 6. und 8. Klassen). Dadurch kann die Frage beantwortet werden, ob die Wirkung des *JC*-Ansatzes auf bestimmte Schulformen oder Schülergruppen begrenzt ist. Da jeweils nur eine Schule pro Schulform teilnahm, lassen sich mögliche Unterschiede natürlich nicht verallgemeinern.

An der Untersuchung nahmen 395 Schüler und Schülerinnen teil, von denen die meisten zweimal befragt werden konnten. Insgesamt wurden 580 Interviews durchgeführt. Die *Untersuchungsgruppe* bestand aus 61% Jungen und 38% Mädchen. Von etwa knapp der Hälfte (45%) liegen zuordenbare Daten aus zwei Befragungen vor, und zwar Daten aus einer Befragung *vor* der pädagogischen Intervention und aus einer *danach* (Vorher-Nachher-Design). Die restlichen Schüler haben entweder wegen Klassen- oder Schulwechsel tatsächlich nur an einer der beiden Befragungen teilgenommen oder die beiden Datensätze lassen sich einander nicht zuordnen. Wir bezeichnen die zuzuordnenden Fälle in dieser Studie als “Kerngruppe” oder einfach als “Kern”.[127] Da sich in dem

125 Aus Datenschutzgründen werden die drei Schulen hier mit den Buchstaben A, B und C bezeichnet.

126 Es fand keine Zufallsaufteilung der Schüler statt. Der Wert von zufällig gebildeten Kontrollgruppen wird oft überschätzt. Eine Zufallsaufteilung der Schüler erlaubt keine vollständige Kontrolle über externe Einflüsse, vielmehr verändert sie die Realität des Experiments so stark, daß Aussagen über Ursachen und Effekte noch schwieriger werden.

127 Aus Datenschutzgründen erfolgte die Zuordnung über einen vierstelligen “Geheim-Code”, der sich aus dem eigenen Geburtsmonat, den Anfangsbuchstaben der Eltern (Fortsetzung...)

Vergleich der Ergebnisse von Gesamt– und Kerngruppe keine nennenswerten Unterschiede ergaben, wurden mit einer Ausnahme alle nachfolgenden Analysen mit der Gesamtgruppe ausgeführt. Die multivariate Varianzanalyse mit Meßwiederholung zur Abschätzung der Effektstärke bezieht sich nur auf die Kerngruppe (siehe den *Auswertungsplan* in der Tabelle unten).

Die Befragungen wurden klassenweise während des Unterrichts durchgeführt.[128] Außer den im Unterricht fehlenden Schülern und Schülerinnen gab es keine "Ausfälle". Teilnahmeverweigerungen sind nicht vorgekommen. Die meisten Jugendlichen haben den Fragebogen gern beantwortet. In einzelnen Fällen haben Schüler und Lehrer Themen aus dem Fragebogen zum Gegenstand von Diskussionen in der Klasse gemacht. Der Anteil unbeantworteter Fragen oder nicht auswertbarer Antworten (z.B. Mehrfachankreuzungen) ist vernachlässigbar gering, obwohl die Fragen für diese Altersgruppe zum Teil recht schwierig waren. Da die Anzahl der Nichtbeantwortungen von Frage zu Frage schwankt, ergeben sich bei den einzelnen Auswertungen zum Teil leicht variierende Gruppengrößen.

8.4.1 Meßansatz und Meßinstrumente

Im DES–Projekt wurden gut erprobte Fragebogen und Tests eingesetzt, zu denen bereits zahlreiche Daten aus US-amerikanischen und deutschen Schulversuchen vorlagen. Zur Erfassung moralisch-demokratischer Aspekte des Schullebens aus Schülersicht wurden der *Moralische Atmosphäre Fragebogen* (MAF) und eine deutsche Version des *Sociomoral Reflection Measure* (SRM) von John Gibbs et al. (1982) eingesetzt. Beim SRM muß die befragte Person, ähnlich wie beim MJI, selbst Argumente zu moralischen Dilemmas produzieren. Jede Person erhält auf der Basis ihrer Antworten einen Stufenwert von 1

127(...Fortsetzung)
und dem eigenen Geschlecht zusammensetzte. Dieses Verfahren ist allerdings nicht empfehlenswert für ähnliche Studien, da die Schüler damit entweder überfordert sind oder es zu Mogeleien reizt.

128 Die Durchführung der Befragung oblag dem Landesinstitut für Schule und Weiterbildung, Soest. Koordination: Heinz Schirp und Peter Dobbelstein-Osthoff.

bis 5, je nachdem, welche der Stufen der moralischen Reflexion im Sinne von Kohlberg (1984) diese Antworten repräsentieren.[129]

Der MAF basiert weitgehend auf dem von Power, Higgins und Kohlberg ausgearbeiteten *Moral Atmosphere Questionnaire* zur Erfassung moralisch-demokratischer Strukturen im Schulleben. Lind und Link (1986) haben diesen Fragebogen unter strukturpsychologischen Gesichtspunkten weiterentwickelt und einer Validierungsstudie unterzogen. Für das DES–Projekt wurde der MAF nochmals leicht gekürzt und, wo es sich als notwendig erwies, an die Sprachgewohnheiten junger Schüler angepaßt.[130] Der Teil des MAF, der zum sogenannten "Verantwortungsurteil" wurde, ähnlich dem MUT, wurde nach den Prinzipien des Experimentellen Fragebogens (Kapitel 4) zu einem Fähigkeitstest ausgestaltet. Wir kommen darauf zurück.

Der MAF besteht aus 14 Teilen, die im Original mit *A* bis *N* gekennzeichnet sind (siehe *Anhang*). Die Auswertung des Teils *B* ("Wichtigkeit/moralische Bedürfnisse") erfolgt auf Einzelfragen-(Item)Ebene. Alle anderen Auswertungen werden auf Skalenebene durchgeführt. Die 24 Skalen werden hier fortlaufend durchnummeriert. Die Skalen 12 bis 19 und die Skalen 22 bis 24 spielen für die Fragestellung unserer Untersuchung keine zentrale Rolle; auf ihre Analyse wird daher im Ergebnisteil verzichtet.

Die Skala 1 *Gute Beziehungen zu Mitschülern* umfaßt: Freundlichkeit, Hilfsbereitschaft und Verträglichkeit. Das Item "Gegen andere zusammenhalten" paßt nicht sehr gut in diese Skala, da es für Schüler sehr unterschiedliche Bedeutungen haben kann. Für manche drückt sich darin eine gute Beziehung zu Mitschülern aus, für andere eine schlechte (Die Item-Test-Korrelation, die Trennschärfe, für diese Frage beträgt entsprechend nur $r = 0.05$). Die Skala 2 *Gute Schüler-Lehrer-Beziehung* besteht nur aus zwei Items: "Respekt und Achtung" sowie "Offenheit und Gesprächsbereitschaft". Das Item "Die Lehrer wollen die Schulordnung immer durchsetzen" korreliert gering negativ mit dieser Skala; d.h. in den Augen der Schüler ist es der Gerechtigkeit und Offenheit teilweise entgegengesetzt. Es wurde daher nicht in die Skala aufgenommen. Die Skala 3

129 Die Schlußauswertung der SRM-Daten war bei der Endredaktion dieser Arbeit noch nicht vor. Eine Teilauswertung findet sich bei Lind & Althof (1992).

130 Die Überarbeitung des MAF für das DES-Projekt wurde von Georg Lind und Peter Dobbelstein-Osthoff unter Mitarbeit von Sibylle Reinhardt und Fritz Oser durchgeführt.

Gute Meinung über die Schule umfaßt die Frage-Items gerechte Schulordnung, stolz auf die Schule sein, gute Gemeinschaft und gute Ausbildung. Alle vier Beurteilungsdimensionen sind in dieser Schülergruppe positiv korreliert, wie auch der hohe α-Wert anzeigt. Die Skala 4 *Wenig Regelverstöße* bilanziert Aussagen der Schüler zu Regelverletzungen im Schulalltag: Konkurrenzstreben, Schwänzen, Mogeln, Schlägereien, Diebstähle, Vandalismus. Die Skala 5 *Partizipationsmöglichkeiten* faßt drei Antworten der Schüler zusammen, die sich auf Teilnahme an Entscheidungen in der Schule beziehen: Bereitschaft der Lehrer zum Gespräch, Mitbestimmung und Stimmrecht. Das erste Item ist auch in der Skala 2 enthalten. Die Skala 6 *Partizipationswunsch* stützt sich auf fünf Fragen: zwei Fragen zur Wichtigkeit von Mitbestimmung und Stimmrecht der Schüler und drei Fragen darüber, ob Schüler es richtig finden, dass sie an der Aufstellung von Regeln, an der Festlegung des Strafmaßes bzw. an der Bestrafung beteiligt werden. Die Skalen 7 bis 11 betreffen *Inhalt, Partner* und *Ausmaß der Kommunikation* der Schüler.

Mit der Skala 12 *Mogeln: Selbst abschreiben* wird beschrieben, ob die Schüler (unter bestimmten Umständen) in einer Klassenarbeit abschreiben oder abschreiben lassen würden. Mit der Skala 13 *Mogeln: Andere schreiben ab bzw. lassen abschreiben* wird beschrieben, wie sie das Verhalten ihrer Mitschüler wahrnehmen. Die Skala 14 *Helfen: Sollen und tun* umfaßt drei Frage-Items zur Hilfestellung gegenüber Mitschülern. Die Skalen 15 bis 17 *Helfen: Stufe der Normbegründung* geben das Normverständnis wieder, das dem helfenden Verhalten der Schüler zugrunde liegt oder das sie bei ihren Mitschülern und Lehrern vermuten. Die Skalenwerte (von 7 bis 14) bilden eine Guttman-Skala. Sie können aber nicht direkt mit den Stufen verglichen werden, die auf Interviewdaten basieren. Die Skalen 18 *Gute Lehrer* und 19 *Guter Unterricht* fassen differenzierte Urteile der Schüler über ihre Lehrer und den Unterricht zusammen. Die Items stammen aus einem Fragebogen von Schwarz (Jerusalem).

Die Skala 20 *Verantwortungsurteil-Test* (VUT) wird unten näher erläutert.

Die Skalen 22 bis 24 *Gemeinschaft* zeigen drei verschiedene moralische Einstellungen der Schüler an: Präkonventionalität, Konventionalität und Postkonventionalität im Sinne von Kohlberg (1984).

Beschreibung der *Modi der Verantwortung*	Modus
Verantwortung und Verpflichtung werden von der Person als *dasselbe* angesehen.	1
Verantwortung und Verpflichtung werden unterschieden. Die Person fühlt sich aber nur sich selbst gegenüber verantwortlich.	2
Es wird anerkannt und gefordert, daß jeder für sich und sein Wohlergehen verantwortlich ist.	2/3
Verantwortung heißt, das "Gute" tun, unabhängig vom Urteil anderer, aber sie ist auf persönliche Bekannte beschränkt.	3
Verantwortung bedeutet vor allem die Achtung der Gefühle anderer. Wer keine Rücksicht auf die Gefühle nimmt, verdient keine Achtung.	3/4
Verantwortung als ein gegenseitiges Gefüge von Gefühlen und Übereinkünften zwischen Menschen in einer Partnerschaft oder Gemeinschaft	4

Abb. 8.2 Modi der Verantwortung nach Power et al. (1989).

8.4.2 Verantwortungsurteil-Test (VUT)

Das Konzept des Verantwortungsurteils bildet nach Power, Higgins und Kohlberg (1989) eine Grunddimension der moralischen Atmosphäre einer Gemeinschaft. Hiernach sind Urteile über Art und Umfang personaler Verantwortung wichtig für das Wohlergehen anderer. Sie postulieren, dass Verantwortungsurteile die subjektiven Handlungsentscheidungen leiten und über den Umgang miteinander die Organisation gemeinschaftlicher Lebensformen beeinflussen. Die durch Verantwortungsübernahme realisierte Gemeinschaftlichkeit bestimmt (zusammen mit dem Bemühen um Gerechtigkeit) nicht nur die in einer Gruppe bestehende moralische Atmosphäre. Sie beeinflußt auch, so die Autoren, die individuelle sozial-moralische Entwicklung durch die Erfahrung konkreter interaktiver Verhaltensweisen und des Raumes, der dem einzelnen zur aktiven Mitgestaltung und Veränderung gewährt wird. Power, Higgins und Kohlberg

(1989) beschreiben sechs qualitativ unterschiedliche *Modi des Verantwortungsurteils*, die in einer hierarchischen Ordnung zueinander stehen sollen (siehe Tabelle unten). Die Entwicklung der Verantwortung ist gekennzeichnet durch 1. den Übergang von einer extrinsischen zu einer intrinsischen Handlungsmotivation; 2. die Ausdehnung des Personenkreises, für dessen Wohlergehen man sich verantwortlich fühlt (z.B. nur die Freunde oder alle Mitglieder der Gruppe); 3. einer zunehmenden Anerkennung der Gemeinschaft als "Wert an sich".

Wie beim MUT können auch beim VUT kognitive und affektive Aspekte logisch unabhängig voneinander gemessen werden. Der VUT umfaßt die Fragen H und K1 bis K6 und L1 bis L6 des MAF (siehe *Anhang*). Bezogen auf ein vorgegebenes Dilemma (Frage H), 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. Moralisch-demokratische Urteilsfähigkeit erweist sich, wie in Kapitel 4 aufgezeigt wird, dadurch, daß der oder die Befragte in der Lage ist, sich mit moralischen Argumenten anderer diskursiv auseinanderzusetzen, insbesondere dann, wenn diese Argumente von der "Gegenseite" kommen. Diese Fähigkeit wird im VUT wie im MUT durch die Vorgabe von Contra-Argumenten geprüft. Der oder die Befragte soll Argumente beurteilen, die seiner oder ihrer Meinung entgegenstehen. Für die Ermittlung eines Meßwertes (C-Wert) der Urteilsfähigkeit werden die Antworten jeder befragten Person einer intraindividuellen Varianzkomponentenzerlegung unterzogen und die Determinationskoeffizienten für die beiden realisierten Designfaktoren berechnet:

(1) Richtung der Argumente (für oder gegen Helfen) und
(2) Modi der Verantwortung

Der Koeffizient für den zweiten Faktor stellt das Maß für die "*moralische Urteilsfähigkeit (VUT)*" dar. Er gibt an, wie stark sich die Person bei der Beurteilung der vorgegebenen Argumente am Verantwortungsmodus (statt an der Richtung der Argumente) orientiert. Wenn die getestete Person sich bei ihrer Bewertung von Argumenten gar nicht am Verantwortungsmodus, sondern zum Beispiel vollständig an der Übereinstimmung der Argumente mit ihrer eigenen Meinung orientiert, erhält sie den Skalenwert 0; wenn sie sich hingegen bei der Beurteilung der Argumente zu 100% am Modus orientiert, erhält sie den

Skalenwert 100. Die Befragten können also Werte zwischen 0 und 100 erhalten.[131] Bezüglich der affektiven Aspekte wird erfaßt, welche Einstellung die Befragten zu jedem der sechs Modi des Verantwortungsurteils haben. Die Skalenwerte können hier von 1 (Ablehnung) bis 5 (Zustimmung) reichen.

8.4.3 Statistische Auswertungen

Die meisten Skalen wurden durch Summierung der entsprechenden Items und Division durch die Anzahl der Items gebildet (Mittelwertskalen). Der Wertebereich entspricht so der Breite der Antwortskalen jedes Items. Eine Ausnahme davon bildet die Skala 15 ("Helfen: Stufe der Normbegründung"), die sich von 7 bis 14 spannt, und die Skala 20 ("Verantwortungsurteil: VUT-C-Wert"). Fehlende Angaben wurden bei den Skalen 1 bis 19 durch den individuellen Mittelwert ersetzt. Fehlten mehr als 50% der Angaben, wurde die gesamte Skala als nicht beantwortet betrachtet. Die üblichen Kennwerte der Skalenkonsistenz (*Cronbachs α* und der *Reproduzierbarkeitskoeffizient* für Guttman-Skalen) liefern zufriedenstellende Werte (siehe Abb. 8.3).

Anmerkungen (zur nachstehenden Tabelle)

a. Die Nummern beziehen sich auf den *Moralische Atmosphäre Fragebogen* (MAF), der in dieser Untersuchung als "Fragebogen B" vorgegeben wurde (siehe Anhang).
b. Umgepolte Items werden durch Klammern "()" gekennzeichnet.
c. Reproduzierbarkeitskoeffizient; der minimale Repr.Koeff. beträgt hier: .78.
d. Minimaler Repr.Koeff.: .69.
e. Die Frage O13 ("Sie zeigen, daß sie von jedem/r Schüler/in eine feste Meinung haben") wurde wegen geringer 'Augenscheinvalidität' nicht in die Skala „Guter Lehrer" aufgenommen.
f. Bei der Skala „Verantwortungsurteilsfähigkeit" wurde kein α–Koeffizient berechnet, da hier die Konsistenz als Indikator für die kognitive Strukturiertheit des *individuellen* Urteilsverhaltens definiert wird und nicht als Merkmal des Tests.

131 Die Berechnung der Strukturwerte zum Verantwortungsurteil wird bei Link & Lind (1988) dargestellt.

Name der Skala	Item–Nr.	Breite	α
1. Gute Beziehung zu Mitschülern	A1, A2, A3, (A4)	1 – 5	468
2. Gute Beziehung zu Lehrern	A5, A6	1 – 5	687
3. Gute Meinung von der Schule	A8, A11, A12, A13	1 – 5	652
4. Wenig Regelverstöße	(A14), A15, -A16, A17, A18, A19	1 – 5	550
5. Partizipationsmöglichkeiten	A6, A9, A10	1 – 5	510
6. Partizipationswunsch	B9, B10, D1, D2, D3	1 – 5	745
7. Kommunikation über schulische Dinge	E3, E4, E5, E6	1 – 5	531
8. Komm. über persönliche Dinge	E7, E8, E9, E10	1 – 5	531
9. Komm. mit Mitschülern	E3, E4, E7, E8	1 – 5	646
10. Komm. mit Lehrern/ Eltern	E5, E6, E9, E10	1 – 5	575
11. Komm. insgesamt	E3 – E10	1 – 5	683
12. Mogeln: Selbst abschreiben	F1, F3, F5, F7	1 – 3	827
13. Mogeln: Andere schreiben ab/ lassen abschreiben	F2, F4, F6, F8	1 – 3	716
14. Helfen: Sollen und Tun	H1, H2, H4	1 – 3	587
15. Helfen: Eigene Stufe der Normbegründung	(I1), (I2), I3, I4, I5, I6, I7	7 – 14	0,87[c] 0,82[d]
16. Helfen: Wahrgenommene Normbegründung bei Schülern	(J1), (J2), J3, J4	1 – 4	628
17.und bei Lehrern	(J5), J6, J7	1 – 4	482
18. Gute Lehrer (Wahrnehmung)	O1, (O2), O3–O7, (O8), (O9), O10, (O11), O12, O14–O16, (O17), (O18), O19	1 – 5	87
19. Guter Unterricht (Wahrnehmung)	P1, P2, P3, P5	1 – 5	722
20. Verantwortungsurteil (VUT-C-Wert)	H, K1–K6, L1–L6	0 – 100	–

Abb. 8.3 MAF: Unterskalen, Items, Skalenbreite und Skalenkonsistenz (Cronbachs α)

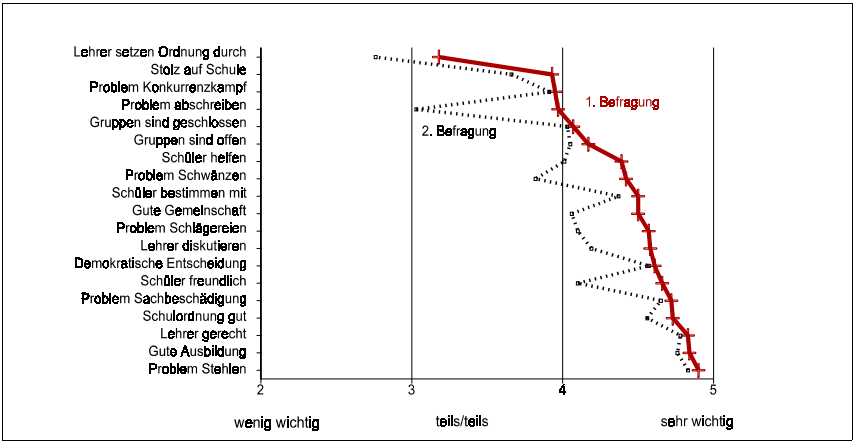


Abb. 8.4 Wichtigkeit einer moralischen Atmosphäre in der Schule aus der Sicht von Schülern und Schülerinnen. (Fragen B1 bis B19 des MAF, siehe *Anhang*). Veränderung in ca. zwei Jahren

8.5 Wichtigkeit und Wahrnehmung der moralischen Atmosphäre in der Schule

Die Schüler und Schülerinnen wurden gebeten, anhand von 19 vorgegebenen Aussagen anzugeben, wie wichtig für sie verschiedene Dinge in der Schule sind und in welchem Ausmaß diese Dinge ihre Schule beschreiben. Bei der Frage nach der Wichtigkeit können wir die Antworten der Schüler direkt interpretieren. Sie repräsentieren die Bedeutung, die ihrer Meinung nach diesen Dingen zukommt. Wenn sie sagen, etwas sei ihnen wichtig oder unwichtig, dann können wir diese Antworten zunächst einfach so nehmen wie sie sind.

Bei der Frage nach der Wahrnehmung sollten die Schüler auf einer Skala von 1 bis 5 ankreuzen, ob der fragliche Sachverhalt bezogen auf ihre Schule *gar nicht*, *teilweise* oder *voll und ganz stimmt*. Wenn wir uns die "Wahrnehmung" der Schüler näher ansehen, müssen wir bedenken: a) Während ihnen bestimmte Dinge *völlig* unwichtig oder aber *extrem* wichtig sein können, werden bei der Wahrnehmung extreme Kategorien wie "trifft nie zu" oder "trifft völlig zu" von den Schülern vermutlich nur selten gewählt werden. b) Veränderungen, die sich langsam und graduell vollziehen, werden von Menschen, die direkt daran beteiligt sind, oft kaum oder gar nicht bewusst registriert, auch wenn sie einem gelegentlich zu Besuch kommenden Beobachter sofort auffallen. c) Jugendliche

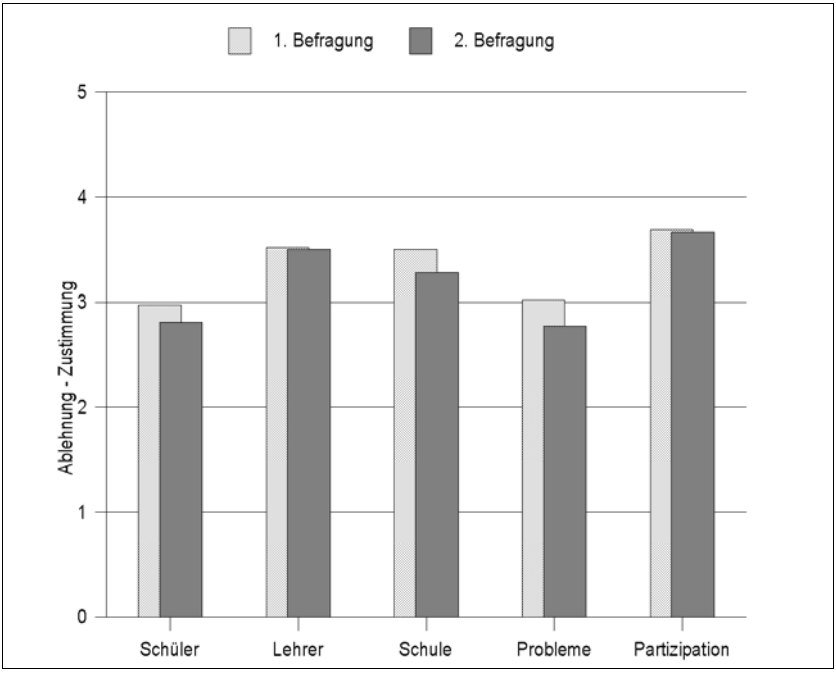


Abb. 8.5 Die Veränderung der Wahrnehmung der moralischen Atmosphäre der Schule durch die Schüler und Schülerinnen in Bezug auf Mitschüler, Lehrer, Schule, Probleme und Partizipation (Skalen 1 bis 5; Frageblock A des MAF, siehe Anhang)

entwickeln mit wachsenden kognitiven Fähigkeiten strengere Maßstäbe für die Wahrnehmung ihrer Umwelt, und dieser Prozeß wird durch entwicklungsfördernde Maßnahmen, wie sie das DES-Projekt durchgeführt hat, noch verstärkt. Nicht jede – in den Augen der Schüler – negative Veränderung muß daher auch eine tatsächliche Verschlechterung des Schulklimas darstellen. Die folgenden Befunde beziehen sich jeweils auf die Gesamtgruppe (N= 395).

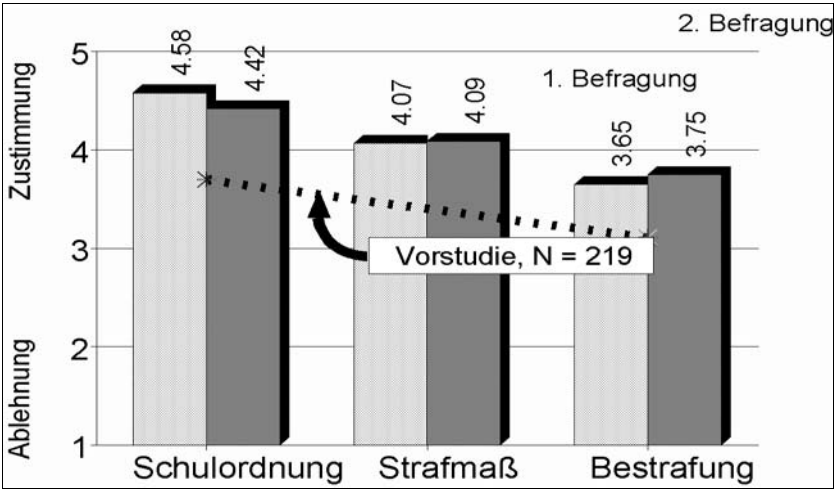


Abb. 8.6 Beurteilung der Mitbestimmung bei Fragen der Schulordnung, der Strafbemessung und der Bestrafung durch die SchülerInnen. (Fragen D1 - D3)

8.5.1 Die Wichtigkeit der moralischen Atmosphäre in der Schule

Die befragten Schüler zeigen sich sehr engagiert für Gerechtigkeit und Gemeinschaftlichkeit in der Schule, für mehr Freundlichkeit und Hilfsbereitschaft unter den Schülern, für gerechte und gesprächsbereite Lehrer, für gute Lernmöglichkeiten, gegen Regelverstöße in der Schule – und für die Mitwirkung der Schüler und Schülerinnen an Entscheidungsprozessen in der Schule. Die größte Bedeutung messen sie dem Problem des Stehlens in der Schule bei. Aber fast genauso wichtig ist für sie eine gute Ausbildung und dass die Lehrer gerecht zu ihren Schülern sind. Aber auch demokratische Entscheidungen und die Mitbestimmung der Schüler sind für sie immer noch sehr wichtig bzw. mehr als nur wichtig. Am unteren Ende der Wichtigkeitsskala befindet sich die Aussage, Lehrer sollen die Schulordnung durchsetzen. Dies finden die Schüler am wenigsten wichtig ("Teils-Teils"). Dieser Befund steht in starkem Kontrast zu früheren Untersuchungen, aus denen sich ergab, dass "alle Kinder der Meinung sind, dass der Lehrer für Ordnung sorgen soll" (Hofer et al. 1986, S. 236). Zu den Dingen, die wichtig, wenn auch nicht sehr wichtig sind, gehören stolz auf die Schule zu sein, die Probleme Konkurrenzkampf und Mogeln zu verringern,

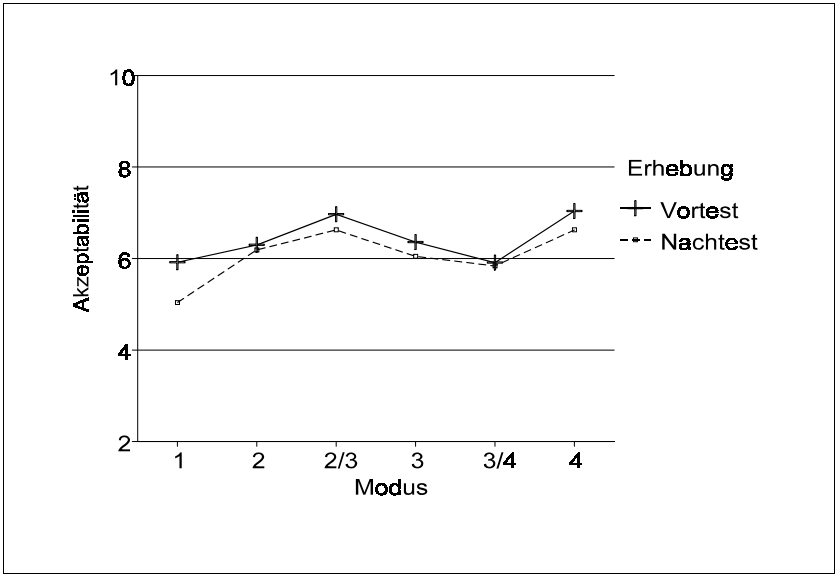


Abb. 8.7 Veränderung der Akzeptabilität verschiedener Modi des moralischen Argumentierens bei SchülerInnen (VUT). Quelle: DES-Projekt; N = 395. Fragen: K1 - K6, L1 - L6; Mediane

und ob es offene Gruppen bzw. ob es geschlossene Gruppen gibt, die gegen andere zusammenhalten.

Dieser Befund mag für manche Erzieher – Eltern wie Lehrer – überraschend sein. Wie in Kapitel 10 gezeigt wird, unterschätzen viele Lehrer stark das Engagement der Schüler für eine gute Schule. Die Bedeutung einer guten Schule für die Schüler veränderte sich im Verlauf des DES-Projekts kaum.

8.5.2 Die Beurteilung der moralischen Atmosphäre in der Schule durch die Schüler

In ihrer Wahrnehmung des Lebens an ihrer Schule sind die Schüler eher zurückhaltend. Die am meisten bevorzugte Antwortkategorie war “stimmt teilweise”. Die Schüler können also nicht rückhaltlos den Aussagen zustimmen, dass ihre Mitschüler hilfsbereit und freundlich sind, ihre Lehrer gerecht und respektvoll mit Schülern umgehen, die Schule eine gute Gemeinschaft ist und es nur wenig Regelverstöße gibt. Wir können diese zurückhaltende Beurteilung zum

Teil als Ausdruck einer allgemeinen Urteilstendenz der Jugendlichen betrachten.

Damit wäre aber die Aussage, die darin zum Ausdruck kommt, nicht hinreichend gewürdigt. Denn den Schülern ist, wie wir gesehen haben, ein gutes Schulklima alles andere als unwichtig und sie urteilen durchaus differenziert. Wie die zusammenfassende Analyse der Wahrnehmungen der Schüler auf Skalenebene zeigt, beurteilen sie die *Partizipation*smöglichkeiten und die *Schul*atmosphäre als vergleichsweise gut; mit ihren *Lehrern* scheinen sie zufriedener zu sein als mit ihren *Mitschülern*; der Aussage, dass es wenig Regelverstöße (*Probleme*) gibt, können sie relativ selten zustimmen. In manchen Punkten wird das Urteil der Schüler kritischer, in manchen bleibt es konstant. Kritischer wird ihr Urteil in Bezug auf ihre Mitschüler und die Schulgemeinschaft; auch Probleme und Regelverstöße scheinen in Folge der offenen Diskussionen, die in den Gemeinschaftssitzungen dazu stattfinden, mehr ins Bewusstsein zu treten. Unverändert bleibt die positive Beurteilung ihrer Lehrer und der Mitwirkungsmöglichkeiten für Schüler. Dies zeigt, dass solche Aussagen nicht allein Ausdruck einer Urteilstendenz sind, und dass Jugendliche im Laufe ihrer Entwicklung nicht generell kritisch-distanzierter werden.

8.6 Partizipation

Wie wir sahen, besteht unter Schülern ein hohes Interesse an eigener demokratischer Mitwirkung. Damit ist nicht gesagt, dass die Schüler eine demokratische Mitbestimmung an Entscheidungen auch für richtig halten. Die anderen Schüler, so sagte einer der Jugendlichen einmal während eines Interviews, seien dafür noch nicht reif; es wäre daher besser, der Lehrer treffe alle Entscheidungen allein. Denken alle Schüler so? Wir haben diese Frage in verschiedenen Variationen gestellt, um verschiedene Aspekte berücksichtigen zu können, und zwar Fragen zur Mitwirkung a) bei der Aufstellung von Schulregeln (was erlaubt und verboten ist), b) bei der Festlegung des Strafkatalogs (was bei Verstößen gegen Regeln zu folgen hat) und c) bei der Bestrafung von Schülern, die eine Regel übertreten haben (Fragen D1 bis D3 des MAF; siehe *Anhang*).

Wie die Antworten zeigen, hält die überwiegende Zahl der Schüler und Schülerinnen die demokratische Mitwirkung an allen drei Punkten für *richtig*. Selbst die Forderung, dass Schüler daran beteiligt werden sollen, Regelverletzer zu

bestrafen, findet überwiegend Zustimmung, auch wenn hier einiger Widerspruch zu verzeichnen ist. An der Aufstellung von Regeln sollten Schüler mitwirken, darin waren sich die Befragten ziemlich einig. Sie antworten hier noch entschiedener als die (älteren) Schüler, die in der Vorstudie befragt wurden. Die Meinung der Schüler zur Mitwirkung bleibt über den Projektzeitraum weitgehend stabil.

8.7 Verantwortungsmodus und Urteilsfähigkeit

Bezüglich des *Inhalts des Verantwortungsurteils* haben wir herausgefunden, dass die Schüler im großen und ganzen die höheren Modi gegenüber den tieferen bevorzugen. Sie präferierten vor allem den Modus 4. Sie präferierten aber auch den tieferen Modus 2/3 (lies: zwei Strich drei), der die Verantwortung für das eigene Wohlergehen, Eigentum und die eigenen Ziele in den Vordergrund

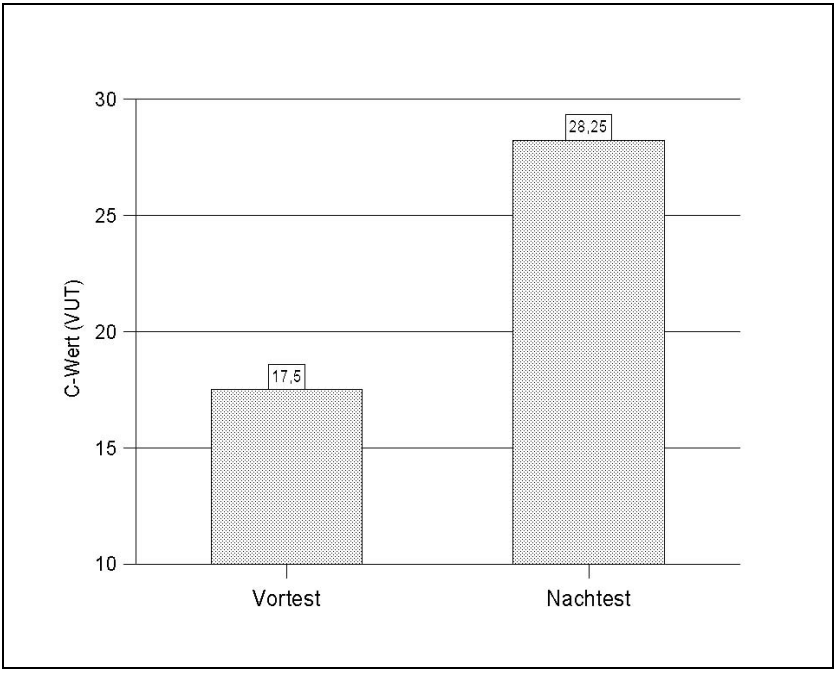


Abb. 8.8 Moralische Urteilsfähigkeit (VUT; C-Wert; Skala 20) vor und nach der DES-Intervention. Mediane

Schule (Vortestjahr)	Interventionsdauer (Monate)	Vortest (Median)	Nachtest (Median)	Median-Differenz	Effekt pro Jahr
Schule A (1988)	29	22,00 n=44	41,00 n=66	19.00	7.86
Schule B (1987)	24	14,88 n=47	32,50 n=56	17.62	8.81
Schule B (1988)	14	13,50 n=54	17,25 n=61	3.75	3.21
Schule C (1988)	20	20,67 n=89	28,00 n=88	7.33	4.40
Mittlerer DES–Effekt:	22	17.76	29.69	11.93	6.51
Schätzungen des Bildungseffekts:					
Vergleich 6. vs. 8. Kl. (Schule C)	24	16	26.5	10.50	5.25
Vergleich 8. vs. 11. Kl. (N=219; Vorstudie)	36	30.75	41.5	10.75	3.58

Abb. 8.9 Veränderung der moralischen Urteilsfähigkeit (C-Werte/VUT)

stellt. Dieses Ergebnis wird durch die Vorstudie deutlich bestätigt, die bei älteren Schülern eine sehr ähnliche, zweigipflige Verteilung der Präferenzen der Modi ergab. (Diese Abweichung von der idealen, monoton ansteigenden Präferenzlinie macht möglicherweise eine Neuformulierung der Theorie des Verantwortungsurteils notwendig, wie es von Kohlberg und seinen Kollegen vorgeschlagen wurde; vgl. Link & Lind 1988). Die Schülerurteile verändern sich von der ersten zur zweiten Befragung in einer Weise, wie wir es aufgrund der kognitiven Entwicklungstheorie und der Befunde der Moralisches-Urteil-Forschung zu erwarten haben. Es zeigt sich, dass sich die Präferenzreihenfolge (Inhalt) nicht oder nur wenig verändert, dass aber, als Ausdruck gewachsener Urteilsfähigkeit, *alle* Merkmale bei der zweiten Erhebung etwas weniger stark gewichtet werden.

Direkte Hinweise auf entwicklungsbedingte Veränderungen der Urteilsfähigkeit der Schüler sollten sich in der *Struktur* der individuellen Verantwortungsurteile zeigen. Hier lautet die Frage: Haben die Schüler überhaupt erkannt, dass es sich

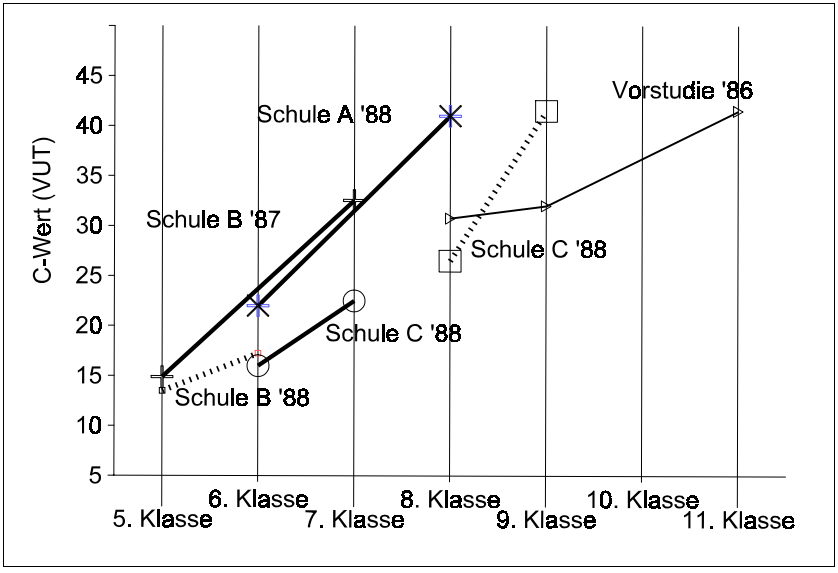


Abb. 8.10 Absolute Effekte der Intervention im DES-Projekt; Veränderungen der C-Werte (VUT) in verschiedenen Schulen und Klassen.

um qualitativ verschiedene, unterschiedlich gut rechtfertigbare Argumente handelt? Waren sie fähig, die Argumente konsistent, nach ihrer Zugehörigkeit zu einem bestimmten Modus zu beurteilen? Und: wie verändert sich diese Fähigkeit im Verlauf des DES-Projekts? Die Struktur des Verantwortungsurteils verändert sich eindeutig in Richtung einer größeren Urteilskompetenz. Die Schüler haben sich in den ein bis zwei Jahren, von der ersten bis zur zweiten Befragung, soweit entwickelt, dass sie die qualitativen Unterschiede in den Argumenten besser zu erkennen und die Argumente in der Chris-Geschichte (Frage H) konsistenter zu beurteilen vermochten.

Dieser Effekt zeigt sich am Grad, in dem sich die Schüler bei der Beurteilung der vorgegebenen Argumente an dem Modus der Verantwortung (statt an ihrer Meinungskonformität) orientierten, oder anders gesagt, am Grad, mit dem der Modus das Urteilsverhalten der Schüler determiniert. Die Stärke des Effekts läßt sich quantitativ ausdrücken durch die Korrelation zwischen dem Testzeitpunkt und der Höhe der Urteilsfähigkeit; sie beträgt $r = 0.41$. Dem entspricht ein interventionsbedingter Varianzanteil von 17,8%, ermittelt durch eine multivariate Varianzanalyse (MANOVA) mit wiederholter Messung.

8.8 Veränderung der moralischen Atmosphäre der Schule durch die pädagogische Intervention

Wir haben oben festgestellt, welche Veränderungen sich in dem Zeitraum ergeben haben, in dem das DES-Projekt an den beteiligten Schulen gelaufen ist. Wie eingangs erwähnt, können wir diese Veränderungen aber nicht einfach den Wirkungen des Projekts zurechnen. Manche Veränderungen hätten auch eintreten können, wenn die Schulen ihr normales Unterrichtsprogramm durchgeführt hätten. Andere Veränderungen gehen vielleicht stärker auf die individuelle Entwicklung der Schüler als auf Veränderungen des Schulklimas zurück. Inwieweit können wir diese Effekte also darauf zurückführen, dass in den DES-Schulen Dilemma-Diskussions-Stunden und *Just community*-Treffen durchgeführt wurden? Wir haben oben gesehen, dass es nicht leicht ist, diese Frage anhand der Angaben der Schüler zu beantworten. Deren Wahrnehmungen sind immer eine Funktion sowohl der Umwelt, die sie wahrnehmen, als auch ihrer eigenen, kognitiv-moralischen Entwicklung. Wir müssen immer damit rechnen, dass Schüler im Laufe ihrer Entwicklung "strengere" Maßstäbe an ihre Schule anlegen und sich ihre Beurteilungen schon aus diesem Grund verändern, nämlich zu einer weniger positiven Sichtweise. Dieses Problem kann man nur lösen, indem man versucht, den Einfluß der entwicklungsbedingten Veränderung der Wahrnehmungen auszuschalten. Wir können dies im vorliegenden Fall dadurch erreichen, dass wir für die Beurteilung der moralischen Atmosphäre zwei *gleichaltrige* Schülergruppen heranziehen, die ihre Schule zu zwei verschiedenen Zeitpunkten beurteilt haben. Solche Beurteilungen liegen uns aus einer Schule vor, und zwar von zwei 5. Klassen, die 1987 und 1988 eingeschult wurden. Die Klasse von 1987 hat die Schule im ersten Jahr des DES-Projekts erlebt und beurteilt; die Klasse von 1988 im zweiten Jahr. Da wir unterstellen dürfen, dass sich beide Klassen zum Zeitpunkt der Erhebung im großen und ganzen auf dem gleichen Stand der moralisch-kognitiven Entwicklung befunden haben, können Differenzen in der Beurteilung zwischen beiden Gruppen als Ausdruck der Veränderungen des Schulklimas in dieser Zeit gewertet werden. Der Vergleich der Angaben dieser beiden Schülergruppen zeigt in bezug auf fast alle Fragen eine Verbesserung der schulischen Atmosphäre (Abb. 8.11):

- Es ist den Lehrern und Lehrerinnen in der Realschule offenbar gelungen, die Beziehung zwischen Schülern, Lehrern und ihrer Schule deutlich zu ver-

bessern. Die Lehrer gehen nach dem ersten Projektjahr in der Wahrnehmung der Schüler gerechter mit ihnen um und zeigen sich offener und gesprächsbereiter gegenüber Problemen und Fragen von Schülern. Dieses positive Bild wird durch die Angaben der Schüler zur Skala "Guter Lehrer" vollauf bestätigt. Wie Antworten zu einzelnen Fragen zeigen, haben sich die Lehrer im ersten Projektjahr in folgenden Punkten besonders deutlich verändert: Die Lehrer versuchen, die Gefühle der Schüler zu verstehen, hören sich die Meinung der Schüler an, sind für die Schüler da, wenn diese sie brauchen, gehen auf die Schüler ein und machen ihnen Mut. Außerdem, so kann man aus dem Vergleich der Schülerantworten von 1987 und 1988 schließen, machen sich die Lehrer seltener über Schüler lustig, sie bevorzugen seltener die guten Schüler, und sind eher bereit, Fehler zuzugeben.

- Der Unterricht von 1988 wird von den Schülern deutlich besser beurteilt als der von 1987, gleiches gilt für die Qualität der Ausbildung an ihrer Schule.
- Nach einem Projektjahr stellt sich die Schule den Schülern als eine Gemeinschaft dar, in der Verbotsregeln stärker eingehalten und Hilfegebote mehr befolgt werden. Bei den Angaben über sich selbst und über ihre Mitschüler mogeln sie weniger als die Fünftklässler von 1987; sie finden es schlecht für sich und andere. Beim Thema Helfen finden Argumente im Jahr 1988, die dagegen sprechen, einem Mitschüler zu helfen, oder die Gleichgültigkeit ausdrücken, noch weniger Zustimmung als bei den Schülern ein Jahr davor. Die Veränderung konnte hier nicht sehr groß sein, da die Werte für die Stufe der Normbegründung (Skala 15) bereits 1987 sehr hoch waren. Bemerkenswert ist die stärker ausgeprägte Wahrnehmung der Schüler 1988 bezüglich der Hilfsbereitschaft ihrer *Mitschüler* (Skala 16). Es scheint, dass es den Lehrern und Lehrerinnen mit den Veranstaltungen im Rahmen des DES-Projekts gelungen ist, die wechselseitige Fehleinschätzung unter den Schülern bezüglich ihrer Hilfsbereitschaft zu verringern. Veränderungen im normativen Verhalten der Schüler werden auch durch die unterschiedliche Wahrnehmung von Regelverstößen angezeigt (Skala 4). Die Schüler berichten nicht nur, dass weniger ab-geschrieben wird, sondern auch über weniger Vorkommnisse von Diebstählen und Sachbeschädigungen.
- Die Fünftklässler von 1988 sehen mehr *Mitwirkungsmöglichkeiten in der Schule* als die von 1987.
- Die einzige negative Veränderung ist bezüglich der Kommunikationsstruktur zu verzeichnen. 1988 ist die Häufigkeit von Gesprächen mit Mitschü-

Nr.	Skala	1987 (N=89)	1988 (N=85)	Veränderung
1.	Gute Beziehung zu den Mitschülern	3	3	0.0
2.	Gute Beziehung zu den Lehrern	3.5	4	0.5
3.	Gute Schule	3.5	3.8	0.3
4.	Wenig Regelverstöße	3	3.2	0.2
5.	Partizipationsmöglichkeit	3.7	4	0.3
7.	Kommunikation über schulische Dinge	2.8	2.8	0.0
8.	Kommunikation über persönliche Dinge	2.5	2.5	0.0
9.	Kommunikation mit Mitschülern	2.3	2	-0.3
10.	Kommunikation mit Lehrern und Eltern	3	3	0.0
12.	Mogeln: Selbst abschreiben (lassen)	1.3	1	-0.3
13.	Mogeln: Andere schreiben ab	1.7	1.5	-0.2
14.	Helfen: Sollen und selbst tun	2.7	2.7	0.0
15.	Helfen: Eigene Normbegründung	13	13.1	0.1
16.	Helfen: Normbegründung bei Mitschülern	2.5	2.8	0.3
17.	Helfen: Normbegründung bei Lehrern	3	3	0.0
18	Gute Lehrer	3.5	3.8	0.3
19	Guter Unterricht	3.8	4	0.2

Abb. 8.11 Die Wahrnehmung der moralischen Atmosphären der Schule *B* durch die Fünftklässler von 1987 und von 1988 (Mediane)

lern geringer als 1987. Diese negative Veränderung kann ein unbeabsichtigter Nebeneffekt der Bemühungen der Lehrer sein, mit den Schülern in einen intensiveren Kontakt zu treten. Die weitere Analyse der einzelnen Antworten anstelle der Skalen ergab, dass sich diese negative Veränderung auf den Kontakt zu Mitschülern außerhalb der eigenen Klasse beschränkt und dieses Kommunikationsdefizit zum Zeitpunkt der 2. Befragung wieder ausgeglichen wird. Die Fünftklässler von 1988 steigern ihre Kommunikation mit den Schülern außerhalb der eigenen Klasse stärker als ihre Vorgänger von 1987.

Zusammenfassend kann man also sagen, dass in der Realschule (der einzigen Schule, aus der geeignete Vergleichsdaten vorliegen) nach dem ersten Projektjahr die Schüler eine deutliche Verbesserung des Schulklimas sehen. Auch wenn dieser Befund für sich genommen noch kein Beweis dafür ist, dass sich die moralische Atmosphäre der Schule durch das DES-Projekt entscheidend verbessert hat, so stützt er diese Interpretation doch erheblich.

8.9 Bewirkt die pädagogische Intervention über den Effekt der Schule hinaus eine Steigerung der Urteilsfähigkeit (VUT)?

Bewirkt die pädagogische Intervention über den Effekt der Schule hinaus eine Steigerung der Urteilsfähigkeit (VUT)?Wir wollen diese Frage hier nur in bezug auf die Veränderung bei der Urteilsfähigkeit untersuchen. Um diese Frage beantworten zu können, wird eine Schätzbasis für die Größe der Veränderungen im Untersuchungszeitraum benötigt und eine für den Anteil des Effekts, der allein auf die Veranstaltungen im Rahmen des DES-Projekts zurückgeführt werden kann. Angesichts der schmalen Datenbasis können wir nicht mehr als eine vorläufige Schätzung abgeben. (Soweit ich weiß, ist dies die erste Schätzung des absoluten Effekts im Bereich der Moralerziehung.)
Der globale Vergleich der Strukturwerte für das Verantwortungsurteil zeigt, dass sich die Urteilsfähigkeit der Schüler in allen Schulen nach oben entwickelt hat (siehe die Tabelle unten). Die Entwicklungsfortschritte betragen zwischen 3 und 19 Prozentpunkten (bei einer rein theoretisch möglichen Punktzahl von ca. 80 bis 85 Punkten).
Bei der Beurteilung dieser Entwicklung ist zunächst zu berücksichtigen, dass sie sich über einen Zeitraum von 14 bis 29 Kalendermonaten vollzog. Um die Ergebnisse aus den verschiedenen Erprobungen besser vergleichen zu können, beziehen wir die Entwicklungsfortschritte auf ein Kalenderjahr. Diese betragen zwischen 3,2 und 8,8 Prozentpunkten. Einen ersten Anhaltspunkt für den Effekt des normalen Unterrichts während eines Schuljahres ohne besondere Förderprogramme erhält man aus dem Vergleich der zwei verschiedenen Jahrgangsklassen, die in der Schule C am Projekt beteiligt waren. Der mittlere C-Wert - des Verantwortungsurteils beträgt *vor Beginn* des Projekts in der 6. Klasse (Median=) 16.0 und in der 8. Klasse 26.50. Die Differenz beträgt 10,5. Einen Indikator für die pädagogische Effizienz *eines* Schuljahres erhält man, wenn

man diese Differenz durch die Schuljahrdifferenz (=2) dividiert: Der jährliche Zuwachs beträgt demnach *5.25* C-Punkte. Dies entspricht einer Effektstärke von r = 0,23. Ermittelt man den Ein-Jahres-Effekt der pädagogischen Intervention für das Verantwortungsurteil, so erhält man einen Wert von *6,51* Prozentpunkten (r = 0,26). Dieser Wert liegt noch über dem Effekt des normalen Schulunterrichts. Die Differenz beider Werte (1,26 Punkte) ist möglicherweise eine zu niedrige Schätzung der spezifischen Effektstärke des DES-Projekts. Der obige Schätzwert für den normalen Unterricht ist relativ hoch. Von der Vorstudie liegen uns Daten von 117 Gymnasiasten in den Klassen 8, 9 und 11 vor, die eine niedrigere Schätzung des Effekts des normalen Unterrichts ergeben, nämlich einen Wert von *3,58* für den Ein-Jahres-Effekt. Gemessen hieran erbrachte das DES-Projekt also fast drei C-Punkte mehr Förderung der Urteilsfähigkeit als der normale Schulunterricht. Demnach wäre der Effekt des DES-Projekts fast doppelt so hoch wie der normale Schuleffekt. Abb. 8.11 stellt diese Befunde graphisch dar: Die Effekte sind in fast allen Fällen größer als die Effekte des normalen Unterrichts. Jeder C-Punkt gibt an, wie hoch die mittlere Urteilsfähigkeit (Median) in einer der beteiligten Gruppen ist. Diese Fähigkeit wird hier gemessen als der Grad, mit dem das individuelle Antwortverhalten durch den Verantwortungsmodus der Argumente determiniert wird (C-Wert), - d.h. als ihre Konsistenz des Urteils in bezug auf die Verantwortungsmodi. Je steiler die Verbindungslinie zwischen zwei Meßpunkten, desto größer ist der Effekt.

Zudem haben wir den DES-Effekt durch eine globale Mittelwertbildung über alle beteiligten Schulen und Jahrgänge hinweg geschätzt. Diese Schätzung berücksichtigt nicht, dass das DES-Projekt an den Schulen mit sehr unterschiedlicher Intensität durchgeführt wurde. So könnte man vermuten, dass zum Beispiel in der Realschule die zweite Schülergeneration (mit den geringeren Zuwachsraten) nicht mehr mit gleicher Intensität am DES-Projekt teilhatte wie die erste. In der Tat berichtete der Schulleiter, dass im zweiten Jahr eine leichte "Motivationskrise" zu verzeichnen war, die seines Erachtens u.a. damit zu tun hatten, dass die Lehrer kaum Anhaltspunkte über den Erfolg ihrer Bemühungen hatten. Auch vermuten wir, dass die 6. Klasse in der Schule C, die nur einen geringen Anstieg der Urteilskompetenz aufweist, weniger Gelegenheit zu Dilemma-Diskussionen und demokratischen Versammlungen hatte als die 8. Klasse, die während dieser Zeit einen Landschulaufenthalt durchführte, bei dem das DES-Programm im Mittelpunkt stand.

Es sollte jedoch nochmals betont werden, dass es sich hier um erste Effektschätzungen auf einer relativ schmalen Datenbasis handelt. Sie allein genügen nicht, um darauf weitreichende pädagogische oder schulpolitische Entscheidungen zu gründen. Aber zusammen mit vielen anderen Studien, die konsistent positive Effekte berichten (siehe Kapitel 9), können durchaus als Bestätigung der Erwartung gelten, dass mit dem Kohlberg-Oser-Ansatz eine Förderung der Urteilsfähigkeit der Schüler möglich ist, und zwar über das hinaus, was die Schule immer schon leistet (siehe Kapitel 11).

8.10 Resümee

Die vorliegende Untersuchung zeigt, dass es an deutschen Schulen unterschiedlicher Schulformen möglich ist, durch gezielte pädagogische Maßnahmen die moralische-demokratische Atmosphäre der Schule zu verändern und die Urteilsfähigkeit von Schülern und Schülerinnen zu fördern. Die wichtigsten Ergebnisse sind:

- Die zentralen pädagogischen Ziele des *Just community*-Ansatzes (Gerechtigkeit, wechselseitige Achtung, Einhaltung von Regeln, demokratische Partizipation u.ä.m.) haben in den moralischen Bedürfnissen und Sorgen der Schüler und Schülerinnen eine direkte Entsprechung. Sie halten eine gerechte Gemeinschaft von Schülern und Lehrern für wichtig, teilweise sogar für sehr wichtig; sie legen mehr Wert auf die Mitwirkung an Entscheidungen in der Schule, als ihnen von vielen Erziehern zugetraut wird. Die Schüler sprechen sich sehr für die Einhaltung von Regeln und Normen in der Schule aus, auch wenn sie diese nach eigenen Angaben nicht immer einhalten.
- Die Schüler paßten sich nicht, wie die Sozialisationstheorie vorhersagt, den Wertvorgaben des *Just community*-Projekts an. Im Gegenteil, nach Abschluss des Projekts wiesen sie ein differenziertes und kritischeres Werte- und Wahrnehmungsmuster auf als zu Beginn.
- Die moralische Atmosphäre in der Schule, insbesondere die Beziehung zwischen Schülern und Lehrern, wurde, wie sich aufgrund der beiden kognitiven Theorien der Moralentwicklung prognostizieren läßt, mit dem Kohlberg-Oser-Ansatz der Moralerziehung deutlich verbessert.

- Die Entwicklung moralisch-demokratischer Fähigkeiten (Verantwortungsurteil) der Jugendlichen wurde mit diesem Ansatz merklich gefördert, und zwar über das Maß der jeweiligen Förderung hinaus, die der normale Schulunterricht zu leisten scheint. Der (zusätzliche) Effekt des DES-Projekts betrug 6,51 C-Punkte gegenüber einem geschätzten Effekt des normalen Schulunterrichts von 5,25 bzw. 3,58 C-Punkten. Ein Vergleich mit den Effektstärken anderer Interventionsstudien wird durch die Unterschiedlichkeit der verwendeten Meßskalen (hier: VUT) erschwert. Wir hatten eingangs erwähnt, dass der mittlere Effekt der Schulprogramme, die Power et al. (1989, S. 208) analysiert haben, 5 *MMS*-Punkte betrug, das sind 1% der MMS-Gesamtskala (von 100 bis 500).

Dies kann als Hinweis gelten, dass das DES-Projekt vergleichsweise große Effekte erzielt hat. Aufgrund der Tatsache, dass an diesem Interventionsexperiment nur drei Schulen beteiligt waren, verbietet es sich, Aussagen über die Wirkung von Kontextvariablen wie geographische Region, Geschlechterverteilung, Ausländeranteil, sozio-ökonomischer Hintergrund etc. zu treffen. Auch kann die Tatsache, dass drei verschiedene Schultypen (Haupt-, Realschule und Gymnasium) im Versuch vertreten waren, nicht dazu benutzt werden, Aussagen über diese Schultypen zu machen. Wir können aber daraus schließen, dass die Wirksamkeit des Kohlberg-Oser-Ansatzes der Moralerziehung offenbar nicht auf einen bestimmten Schultyp oder bestimmte Schüler begrenzt ist.
Der aufgrund der Lewin-Studie erwartbare "Chaos-Effekt" infolge der Demokratisierung des Schullebens ist ausgeblieben. Die beteiligten Lehrer berichten von keiner gesteigerten Aggressionshäufigkeit weder in den speziellen Veranstaltungen dieses Interventionsprogramms noch in anderen Schulstunden. Im Gegenteil, nach Einschätzung des Autors, der in größeren Abständen diese Schulen besucht hat, verliefen die großen *Just community*-Versammlungen, an denen 100 bis 150 Schüler teilnahmen, mit wachsender Dauer des Projekts zunehmend geordneter, obwohl das Engagement der Teilnehmer für die anstehenden Sachfragen eher größer zu werden schien. Dies bestätigt unser Annahme, dass die Demokratisierung des Interaktionsklimas nicht einseitig, allein als institutionelle Maßnahme, vollzogen werden kann, dass vielmehr gleichzeitig die moralisch-demokratischen Fähigkeiten der Teilnehmer gefördert werden müssen.

Dieses Kapitel war ganz der Veränderung der Urteilsfähigkeit auf Seiten der Schüler gewidmet. Dies bedeutet aber nicht, dass sich die Lehrer als Interaktionspartner in diesem Prozeß nicht auch entwickelt haben. In der Studie von Lewin et al. (1939) haben sie die Rolle eines Funktionsinhabers (Gruppenleiter) gespielt. Diese Rolle war begrenzt auf Verhaltensweisen, die Lehrer oder andere Erwachsene in ihrem normalen Verhaltensrepertoire haben. Nur diese konnten durch die Instruktion des Versuchsleiters für die kurzen Experimente aktiviert werden, um das Gruppenklima zu beeinflussen. Im vorliegenden Versuch ging es jedoch um die Beeinflussung durch ein grundlegend verändertes Interaktionsklima. Dafür war es notwendig, dass sich auch die moralisch-demokratischen Kompetenzen der Lehrer entwickelten. Die “moralischen” Wahrnehmungen von Lehrern werden in Kapitel 11 untersucht.

9 Bedingungen für den Effekt von Interventionen zur Förderung der moralischen Urteilsfähigkeit: Eine Meta-Analyse

Die bisherigen Analysen belegen, dass der kognitive Aspekt tatsächlich eine herausragende Rolle im moralischen Verhalten spielt. Sie legen den Schluß nahe, dass die moralische Urteilsfähigkeit des einzelnen kaum wirkungsvoller und adäquater geändert (wir sagen besser: gefördert) werden kann als durch eine pädagogische Stimulation in diskursiven Prozessen. Dieser Befund, so scheint mir, bestärkt auch die Vorbehalte gegen Theorien, die behaupten, dass das Moralverhalten allein über den sozialisatorischen Anpassungsdruck an bestehende Normen veränderbar sei, wie auch gegen Theorien, die die beobachteten Veränderungen allein auf das Wirken genetisch gesteuerter Reifungsprozesse zurückführen. Zwar bestimmen beide Einflußfaktoren, in welchem Rahmen Moralentwicklung stattfinden kann: Jedoch setzen soziale Normen und Institutionen ebenso wie genetische Faktoren Grenzen. Sie stellen per se keine fördernden Bedingungen für die Moralentwicklung dar.

Die Basis unserer bisherigen Analysen ist jedoch noch zu schmal, um die beobachteten Änderungen in der moralischen Urteilsfähigkeit eindeutig auf die Wirkung des Bildungssystems oder kognitiv-entwicklungspsychologischer Interventionen zurückzuführen oder um sagen zu können, welche Bedingungen deren Effektivität begünstigen oder hemmen. Um den Einfluß solcher Bedingungen wie das besondere Engagement und pädagogische Geschick der Versuchsleiter, günstige Schülerauswahl, Schulform, sozio-kultureller Hintergrund, Alter der Schüler usw. untersuchen zu können, bedürfte es Experimenten von einer Größenordnung, die kaum durchführbar erscheinen. Zumal es sich hier um Hypothesen handelt, die nicht im Labor, sondern nur im “Feld”, also unter nur teilweise kontrollierbaren Bedingungen, überprüft werden können.

Die Frage nach den begünstigenden und hemmenden Bedingungen läßt sich m.E. aber auch genauso gut durch eine zusammenfassende Analyse von bisher durchgeführten Experimenten zum Kohlberg-Ansatz beantworten. Dies gelingt umso besser, je größer die “natürliche” Variation der Bedingungen ist, die mit den pädagogischen Interventionen erzeugt wurde. Es soll dadurch auch untersucht werden, inwieweit sich die Effektivität des Kohlbergschen Interven-

tionsansatzes zur Förderung der moralischen Urteilsfähigkeit verallgemeinern läßt. Als Methode bietet sich die *Meta-Analyse* an (Glass et al. 1978).

9.1 Systematische Wirksamkeitsstudien als Basis für pädagogische Reformen

Wie Mauermann (1988) feststellte, sind wir im sozio-moralischen Bereich der Bildung größtenteils auf subjektive Erfahrungsberichte einzelner Pädagogen angewiesen. Nur selten liegen systematische Wirksamkeitsstudien vor. Viele Ansätze sind so vage beschrieben, dass fast jede pädagogische Handlung hierunter eingeordnet und fast jedes Ergebnis als Erfolg gewertet werden kann. Ganz selten liegen objektive Beobachtungen der Umsetzung pädagogischer Empfehlungen in konkrete pädagogische Methoden vor. Zudem waren (und sind) die Methoden zur Feststellung von Wirkungen oft unzulänglich. Oft wird nur das Wissen von Menschen über moralische Begriffe oder individuelle Regelverstöße erfaßt, aber nicht die Fähigkeit des einzelnen, allgemeine Moralregeln in konkretes Urteilsverhalten umzusetzen. Wo wissenschaftliche Untersuchungen über die Wirkung von traditionellen Moralerziehungsprogrammen vorliegen, sind die Befunde bislang eher enttäuschend.[132] Neue pädagogische Programme haben daher nur eine Chance von der Schule akzeptiert zu werden, wenn nachgewiesen werden kann, dass sie wirksam sind.

Für die Einschätzung der Wirksamkeit eines pädagogischen Ansatzes ist die wissenschaftliche Forschung unerläßlich. Zwar ist auch der persönliche Augenschein für die Entscheidung wichtig, ob ein bestimmter Erziehungsansatz übernommen werden sollte. Das eigene Urteil aufgrund persönlicher Kenntnis kann durch Experimente oder Statistiken nur schwer ersetzt werden. Für Eltern, Lehrer, Schulleiter und Schulaufsicht ist die eigene Anschauung der konkreten Umsetzung in der Praxis eine wichtige Voraussetzung für das Vertrauen in die Tauglichkeit eines neuen Innovationskonzepts. Aber den Betroffenen ist es selten möglich, einen pädagogischen Ansatz so umfassend zu evaluieren, dass

132 So konnte für den moralpädagogischen Werteklärungsansatz nach Raths et al. (1978) trotz seiner großen Verbreitung bislang nicht der Nachweis für seine Wirksamkeit erbracht werden (siehe Lockwood 1978, S. 344; Leming 1981, S. 150; Oser & Althof 1989, S. 48 ff.; Mauermann 1988).

eine verläßliche Urteilsbildung möglich ist. Die eigene Erfahrung aus der Teilnahme an wenigen Interventionen reicht nicht aus, um zuverlässige Schlüsse über die Tauglichkeit dieser Interventionen zu ziehen. Ein bestimmter Ansatz kann fälschlicherweise abgelehnt werden, wenn er ein einziges Mal getestet wurde und genau in dieser Situation versagt; er kann aber auch unberechtigterweise akzeptiert werden, nur weil die Umstände ein günstiges Ergebnis zustande gebracht haben, das dann der Ansatz für sich reklamiert.
Erfolg wie Mißerfolg einer pädagogischen Intervention haben in der Regel viele Ursachen, die sich durch wenige Erprobungen nicht unterscheiden lassen. Will man diese anderen Möglichkeiten der Erklärung eines "Erfolgs" ausschließen und damit belegen, dass überwiegend der pädagogische Ansatz für die Wirksamkeit verantwortlich gemacht werden kann, muß eine *systematische* Überprüfung erfolgen, d.h., der Ansatz muß mehrmals unter verschiedenen Bedingungen wiederholt werden, und zwar so, dass vermutete andere Ursachen für die festgestellten Wirkungen ausgeschlossen werden können.
Über diesen Grundsatz der systematischen empirischen Validierung pädagogischer Interventionen besteht heute weitgehend Einigkeit, wenn auch umstritten ist, wie er im einzelnen realisiert werden kann. Auf der einen Seite gibt es den Standpunkt, dass grundsätzlich *alle* fremden Einflußfaktoren auszuschließen sind, und zwar durch folgende Maßnahmen bei der Versuchsanordnung (nach Campbell & Stanley 1966):

a) Vorher/Nachher-Messungen
b) Experimentalgruppen (Personen, die an dem Programm teilnehmen) und Kontrollgruppen (Personen, die nicht daran teilnehmen)
c) Eine rein zufällige Auswahl der Versuchspersonen (hier der Schüler)
d) Die rein zufällige Aufteilung auf Experimental- und Kontrollgruppen
e) Eine repräsentative Auswahl der Umgebungsbedingungen (Zeit, Gesellschaft, geographischer Region, und vieles andere mehr)

Auf der anderen Seite wird argumentiert, dass es unmöglich und daher sinnlos ist, *alle* denkbaren Alternativerklärungen für den Erfolg eines pädagogischen Ansatzes überprüfen zu wollen. Der oben genannte Maßnahmekanon ist nicht nur schwer zu realisieren, weil er sehr zeit-, personal- und geldaufwendig ist; seiner Umsetzung stehen im Bereich der pädagogischen Forschung oft auch unüberwindbare politische, rechtliche und ethische Gründe entgegen.

Das Programm der Ausschaltung *aller* Alternativursachen ist, so wird aus wissenschaftstheoretischer Sicht gesagt, auch prinzipiell nicht zu realisieren, da der Katalog möglicher Alternativursachen und daher auch die Liste der experimentellen Maßnahmen, die man zur Kontrolle oder Ausschaltung dieser Ursachen zu beachten hätte, immer unvollständig ist, wie umfangreich man sie auch konzipiert. Es ist prinzipiell unmöglich, so viele Variationen in ein Experiment einzubauen, dass alle denkbaren Einflußgrößen kontrolliert oder ausgeschaltet werden können. Damit bleibt es gewöhnlich dem Belieben des Forschers anheimgestellt, welche Bedingungen einer Intervention "systematisch" variiert oder durch Kontrollgruppen ausgeschaltet werden und welche nicht.[133]

Nehmen wir beispielsweise an, der Erfolg einer moralpädagogischen Intervention ist durch die Zusammensetzung der Schülerschaft in einer Klasse bedingt. Gemäß dem Maßnahmekatalog zur Kontrolle von "Gefahren" für die Validität von Versuchsanordnungen von Campbell und Stanley (1966) müßte dieser Einfluß durch eine Zufallsaufteilung der Schüler einer Schule auf Experimental- und Kontrollgruppen neutralisiert werden. Aber damit ist keineswegs jede Alternativerklärung ausgeschlossen.

Zudem erweist sich diese Maßnahme selbst als ernsthafte Gefährdung der Validität des Experiments. Denn da diese Aufteilung wiederum nur von einer bestimmten, verfügbaren Gruppe von Schülern ausgehen kann, ließe sich durch eine Zufallsaufteilung nur die Erklärung ausschließen, dass das ganz spezielle Klima einer Klasse den Interventionsausgang bedingt hat, nicht aber die Vermutung, dass er nur in dieser Schule oder in diesem Jahrgang zu erzielen ist. Andererseits würde die zufällige Aufteilung der Probanden bedeuten, soziale Strukturen aufzulösen, die für ein sozialpsychologisches Experiment von eminenter Bedeutung sein können. Der Ausgang des Experiments wäre gerade durch diejenige Maßnahme stark verzerrt, die dazu gedacht war, Verzerrungen im Hinblick auf den Untersuchungszweck auszuschalten. Da die Förderung des moralischen Urteils in intakten Klassengruppen stattfinden soll, würde die zufällige Aufteilung der Schüler die Validität der Interventionsstudie ver-

133 Ich folge hier den Grundzügen der Argumentation von Gadenne 1976 und Westmeyer et al. 1987; vgl. auch Lind 1985a.

mindern. Dieses Beispiel zeigt, dass eine rein "methodologische" Experimentalistik ihren eigenen Ansprüchen nicht genügen kann.[134]

Das tiefere Problem der rein methodologischen Planung von Experimenten liegt darin begründet, dass die Beweisführung dort einseitig den Befürwortern (oder den Gegnern) einer Hypothese aufgebürdet wird. Durch diese Einseitigkeit wird die Beweisführung zumeist unmöglich. Akzeptabler erscheint ein "konstruktivistisches" Beweisprinzip als Grundlage für die Planung von Experimenten, das man analog zum argumentativen Verfahren der Rede und Gegenrede des Wissenschaftstheoretikers Paul Lorenzen so formulieren könnte: Die Beweislast für und gegen die zu prüfende Hypothese liegt auf beiden Seiten und ist daher auf Proponenten und Opponenten fair zu verteilen. Von den Befürwortern eines Ansatzes muß verlangt werden, dass sie empirische Belege für seine Wirksamkeit erbringen. Sie sollten darüber hinaus auch zeigen können, dass jene Ansätze, die sie ersetzen wollen, weniger geeignet sind, die gewünschte Wirkung zu erzielen. Natürlich kann sich dieser Nachweis nicht auf eine theoretische Argumentation beschränken. Die Proponenten eines Ansatzes müssen diesen der "Gefahr" des Scheiterns an der Wirklichkeit aussetzen, bevor sie ihren Ansatz als valide ausgeben.

Aber es kann nicht verlangt werden, dass die Proponenten jeden nur denkbaren Einwand antizipieren und durch experimentelle Prüfung auszuschalten versuchen. Wurde ein Ansatz in diesem Sinn validiert, dann ist es an den Kritikern, ihre Einwände zu formulieren und ebenfalls der Gefahr des Scheiterns an der Wirklichkeit auszusetzen. Von der Kritik ist fairerweise derselbe Grad an Validität zu verlangen wie von der Behauptung, gegen die sie sich wendet. Ganz allgemein wird man feststellen, dass bei Beachtung solch fairer "Spielregeln" sich überschaubare und damit erfüllbare Forderungen an die Versuchsanordnung ergeben. Die zu wählende Versuchsanlage hängt hier nicht von allen Einflußgrößen ab, an die man denken könnte, sondern nur von jenen, die man als begründet annehmen muß.

Es sollte deutlich geworden sein, dass wir bei der Beurteilung der Wirksamkeit pädagogischer Interventionen nicht vor der Wahl zwischen absolut sicheren Aussagen und subjektiver Anschauung stehen. Begründete Behauptungen lassen sich durchaus wissenschaftlich streng überprüfen. Wahrheiten gelten so lan-

134 Es kann jedoch Gründe geben, Kontrollgruppen einzuführen, wenn z.B. begründete Alternativerklärungen ausgeschlossen werden sollen.

ge, wie sie nicht als trivial erweisen oder empirisch widerlegt wurden. Die pädagogische Forschung muß dafür sorgen, dass Aussagen über die Wirksamkeit pädagogischer Ansätze einer systematischen Prüfung unterzogen werden, wobei begründete, alternative Erklärungen berücksichtigt werden müssen. Dies ist ein gemeinschaftliches Unternehmen, in dem nicht ein einzelner endgültige Beweise für oder gegen eine Wirksamkeitsthese erbringen kann, sondern in dem Einzelstudien Mosaiksteine in einem Puzzle sind, zu dem einzelne Forscher oder Forschungsgruppen ihren Teil beitragen.

9.2 Interventionsmethoden nach Kohlberg

Der Kohlberg-Ansatz der Moralerziehung beruht im wesentlichen auf zwei Methoden, zum einen der "Dilemma-Diskussion" und zum anderen der "Just community".[135] Die Dilemma-Diskussion besteht im Kern in der Diskussion von kurzen (fiktiven) Geschichten, in denen zwei etwa gleich wichtige Normen und Werte aufeinander treffen, so dass der Handelnde, ganz gleich wie er sich entscheidet, immer eine Norm verletzen muß. Die Dilemma-Geschichten können aus dem Kontext des jeweiligen Fachunterrichts, aus dem Alltag der Schüler oder aus der einschlägigen Moral-Literatur entnommen werden. Die Diskussion solcher Geschichten soll die Schüler dazu befähigen,

a) moralisch übergeordnete Argumente und Perspektiven zu finden, die eine Lösung des Dilemmas ermöglichen oder wahrscheinlich machen,
b) sich diskursiv (statt mit Gewalt) mit Argumenten für und gegen die gewählte Lösung auseinanderzusetzen.[136]

Bei der Methode der "Gerechten Gemeinschaft" werden nicht mehr nur "fachlich" relevante Dilemmas zur Diskussion gestellt, sondern auch Probleme, die Schüler und Lehrer in ihrem Zusammenleben direkt be-

135 Eine gute Übersicht über den Kohlberg-Ansatz der Moralerziehung geben Kohlberg (1987), Higgins (1987, 1989) und Power et al. (1989). Zum Ansatz selbst und seinen verschiedenen Formen gibt es inzwischen umfangreiche Literatur. Vgl. vor allem Edelstein 1986, 1987; Oser 1987b; Oser & Althof 1992; Schreiner 1989; auch Kapitel 8.

136 Für weitere Hinweise zur Methode der Dilemma-Diskussion siehe Oser & Althof 1992, Power et al. 1989.

treffen. Bei den hier untersuchten Interventionsstudien spielt diese Methode jedoch nur eine untergeordnete Rolle

9.3 Datengrundlage: Berücksichtigte Publikationen

Als Grundlage für die Meta-Analyse dienen alle erreichbaren Veröffentlichungen über Interventionsstudien, die sich auf den Kohlberg-Ansatz stützen und die zwischen 1970 und 1984 veröffentlicht wurden. Neuere Interventionsstudien werden für diese Analyse nicht herangezogen, da sie zumeist mit der neuen Version des MJI evaluiert wurden, was die Vergleichbarkeit der Befunde verringert. (Die Aufgabe, alle erreichbaren Studien nach dem neuen Auswertungshandbuch von Colby, Kohlberg et al. 1987 zu reanalysieren, würde ein größeres Forschungsprojekt erfordern.) Der Meta-Analyse liegen insgesamt 48 Publikationen (Zeitschriftenartikel und unveröffentlichte Dissertationen) zugrunde, in denen über 141 – zumeist experimentell angelegte – Studien berichtet wird.
Grundsätzliche Gemeinsamkeit herrscht bezüglich des verwendeten Interventionsansatzes und der Meßmethoden. Die vorliegenden Studien unterscheiden sich jedoch in mehrerer Hinsicht, so nach

- Alter der Teilnehmer (alle Altersgruppen ab 6 Jahren)
  Umfang der untersuchten Schülergruppen (13 bis 193 Schüler)
  sozio-demographischer und ethnischer Hintergrund der Versuchsteilnehmer (überwiegend US-Amerikaner)
- Design der Studie (1 bis 270 Tage zwischen Vor- und Nachtest; mit und ohne zweitem Nachtest; mit 7 Tagen bis 1 Jahr Zeitabstand zwischen erstem und zweitem Nachtest; mit und ohne Kontrollgruppen)
- Qualifikation des Versuchsleiters (Psychologe, Lehrer, *peers*)
- Art der pädagogischen Intervention (Dilemma-Diskussion, Gerechte Gemeinschaft und Mischformen unter Einschluß von Ethikunterricht, Rollenspiel)
- Dauer der pädagogischen Intervention (zwischen 1 Tag und 2 Jahren)
- Intensität der pädagogischen Intervention (zwischen 20 Minuten und 12 Stunden pro Woche)

- Art und Inhalt der eingesetzten Meßinstrumente (Kurz- und Langform des MJI, verschiedene Anzahl von Untertests)

Diese zahlreichen Variationen zeigen sowohl eine Schwäche als auch eine Stärke der Berichtslage: eine Schwäche, weil keine einfache Vergleichbarkeit gegeben ist und daher höchst widersprüchliche Ergebnisse und Interpretationen möglich sind. Die Variationsbreite der Studien bedeutet aber auch eine große Stärke, weil das Ziel dieser Meta-Analyse auf die Frage zugespitzt werden kann: Zeigt sich ein deutlicher Effekt der Kohlberg-Methode(n) auf die moralisch-kognitive Entwicklung von Schülern und Erwachsenen *unabhängig* von diversen methodischen und forschungstechnischen Variationen?[137]

9.4 Analyse-Methode

Um hinsichtlich der Effekte eine meßtechnische Vergleichbarkeit herzustellen, wurden alle Untersuchungen von uns reanalysiert und ihre Effektstärkemaße vereinheitlicht. Als Basis der Vereinheitlichung standen folgende Maße zur Auswahl:

- *Signifikanzmaße*: Maße wie *t, F* oder χ^2 scheiden als Maße für die Effektstärke aus, da sie stark von der Größe der untersuchten Stichprobe und der Wahl des Auswertungsdesigns abhängen und Interventionsstudien mit unterschiedlichen Designs damit nicht direkt verglichen werden können.[138] In der Beschränkung auf Maße der statistischen „Signifikanz" ist ein gravierender Mangel der Meta-Analysen von Lockwood (1978), Higgins (1980), Leming (1981) und Enright et al. (1983).
- *Nachtestwerte:* Die Differenzen der Nachtestwerte zwischen Experimental- und Kontrollgruppe hängt sehr von den Anfangswerten, also dem Entwicklungsstand der Schüler *vor* Beginn der pädagogischen Intervention ab, und

137 Die Quellen für die Untersuchungen, auf die sich die nachfolgenden Ergebnisse der Meta-Analyse beziehen finden sie im Anschluss an das Literatur- und Quellenverzeichnis. Eine tabellarische Übersicht über die Ergebnisse im einzelnen kann vom Autor angefordert werden.

138 [Vgl. u.a. Hays (1963), Carver (1978; 1993)].

sagen daher – für sich genommen – nichts über die Wirksamkeit von pädagogischen Interventionen aus.

- *Differenzwerte zwischen Vor- und Nachtest:* Der Vergleich von Nach- und Vortests hebt die Nachteile der Nachtestwerte auf. Die Differenzwerte sie sind jedoch nicht über verschiedene Untersuchungen hinweg vergleichbar, da verschiedene Meßinstrumente und Meßskalen verwendet wurden.
- *Effektstärke-Maß d*: Dieses Maß drückt eine Veränderung zwischen Vor- und Nachtest als Anteil an der Standardabweichung aus, womit verschiedene Studien vergleichbar werden. Das *d*-Maß wird z.B. von Schläfli et al. (1985) in ihrer Meta-Analyse von Interventionsstudien verwendet, die mit dem DIT von Rest evaluiert wurden. Nachteilig ist aber, a) dass das *d*-Maß unanschaulich, weil nicht nach oben begrenzt ist, und b) dass damit nur Studien verglichen werden können, die Mittelwerte und Standardabweichungen berichtet haben. Damit werden viele wichtige Studien, die mit Vierfelderanalysen (χ^2) oder Varianzanalysen arbeiteten, von der zusammenfassenden Analyse ausgeschlossen (Glass et al. 1978).
- *Korrelationen:* Beide Nachteile werden bei Verwendung des Korrelationskoeffizienten r_{xy} als Maß für Effektstärken vermieden. Er stellt einen guten Weg dar, Untersuchungen über Effekte pädagogischer Methoden miteinander vergleichbar zu machen, auch wenn die Ergebnisse in ganz unterschiedlicher Form berichtet werden. Der Koeffizient kann aus Mittelwertdifferenzen, aus *t*-, *F*- und χ^2-Werten und aus Varianzanteilen berechnet werden. Er ist unabhängig von der Größe der Untersuchungsgruppe und er ist nach unten und oben begrenzt (Grenzen: -1 und +1).

Leider wird in Interventionsstudien in den seltensten Fällen der Wert r_{xy} berechnet, so dass dieser Wert für fast alle Studien nachträglich aus den berichteten Signifikanz-Kennwerten (χ^2, *t*, *F*) oder, wo diese fehlen, aus Mittelwertdifferenzen und Standardabweichungen berechnet werden mußte.[139] Die *r*-Werte werden direkt, ohne Umrechnung in Fischers *z* verglichen und zusammengefaßt, da beide Verfahren praktisch zu den gleichen Ergebnissen kommen (Glass et al. 1978, S. 147). Als Zwischenstufe werden hier die quadrierten Korre-

139 Bei der Umrechnung von *t*-Werten in *r* wurde statt der Produkt-Moment-Korrelation die punktbiseriale Korrelation berechnet. Die Werte lassen sich vergleichen, auch wenn sie nicht exakt übereinstimmen.

(1)	$r^2 = Chi^2 / n$	für Vierfeldertafeln via Chi^2-Test mit n =Stichprobengröße
(2)	$r^2 = t^2 / (t^2 - 2 + n)$	für Mittelwertvergleiche via *t*-Tests
(3)	$r^2 = F / (F - 2 + n)$	für Varianzanalysen via *F*-Tests
(4)	$r^2 = QS_b / (QS_b + QS_w)$	für Varianz- und Regressionsanalysen, mit QS_w = Quadratsumme *within groups* und QS_b = Quadratsumme *between groups*
(5)	$r^2 = d^2 / (d^2 + 4)$	für Umwandlung der Effektgröße *d*
(6)	$r^2 = Z^2 / n$	mit *Z* aus der Tabelle für die exakte Irrtumswahrscheinlichkeit *p*

Abb. 9.1 Formeln für die Umrechnung von Statistiken in Effektstärke r^2

lationskoeffizienten, die sogenannten Determinationskoeffizienten r^2 berechnet.[140] Die Umrechnungsformeln für r_{xy} bzw. r^2 sind in Abb. 9.1 wiedergegeben (vgl. Bredenkamp 1970; Glass et al. 1978, S. 148 ff.; Cohen 1988).

Bei dem Koeffizient für die Effektstärke r_{xy} steht *X* für die Variable "Intervention" (Ausprägungen: vorhanden vs. nicht vorhanden) und *Y* für die gemessene Veränderung an moralischer Urteilsfähigkeit, wobei der Zusammenhang kausal gedeutet werden darf, weil es sich hier durchweg um experimentelle Interventionsstudien handelt.

Die in Zeitschriften veröffentlichten Berichte enthielten in vielen Fällen nicht genügend Angaben, um das Effektstärkemaß r_{xy} zu berechnen. Meistens müssen die notwendigen Angaben schwer zugänglichen Forschungsberichten oder Dissertationen entnommen werden. Aber selbst bei Heranziehung dieser "grauen" Literatur konnten in unserem Fall nur bei 74 von 141 Studien alle Informationen ermittelt werden, die für die Berechnung eines Effektstärkemaßes

140 Determinationskoeffizienten sind wenig geläufig und werden daher in dieser Analyse nicht benutzt. Gleichwohl stellen sie sogenannte PRE-Maße mit vorteilhaften Eigenschaften – wie Addierbarkeit – dar (vgl. Nagl et al. 1986, Kapitel X, Seite 32 ff.).

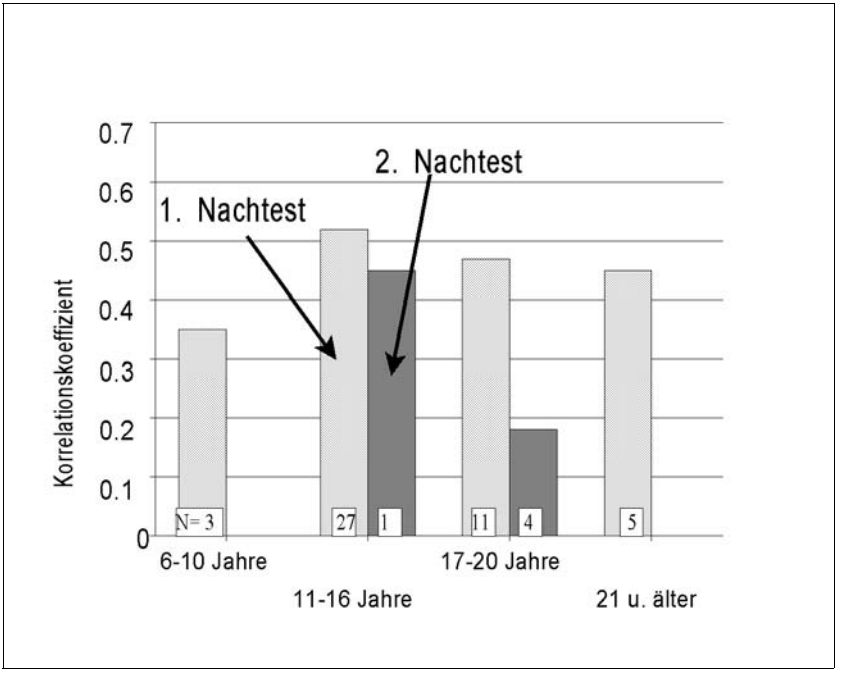


Abb. 9.2 Effekt der Intervention in verschiedenen Altersgruppen (Effektstärke als Korrelation zwischen Testwertänderung und Interventionsgruppe)

notwendig sind. Bei den anderen Studien fehlten einige oder alle wichtigen Informationen (n, χ^2, t oder F).[141] Bei der Umrechnung der berichteten Kennwerte in die Effektstärke r_{xy} mußte auf das jeweilige Design der Studien und die Art der Kennwerte-Berechnung durch die Autoren Rücksicht genommen werden. Viele Auswertungen waren vor allem deshalb ungenügend, weil individuelle Unterschiede beim Beginn der Intervention nicht durch entsprechende Kennwerte für wiederholte Messungen berücksichtigt wurden.

141 Hier muß einmal mehr beklagt werden, dass es nicht zur Pflicht gehört, diese Informationen in Publikationen oder Dissertationen mitzuteilen. Es ist mehr als bedauerlich, dass in fast jeder zweiten Forschungsarbeit so unvollständig berichtet wird, dass aus ihr keine Schlüsse auf die getestete Hypothese gezogen werden können. Die Mitteilung, ein Ergebnis sei "signifikant" gewesen oder ein p von soundso sei unterboten worden, ist für sich genommen, ohne Kenntnis der Stichprobengröße oder der Mittelwertdifferenzen, wertlos.

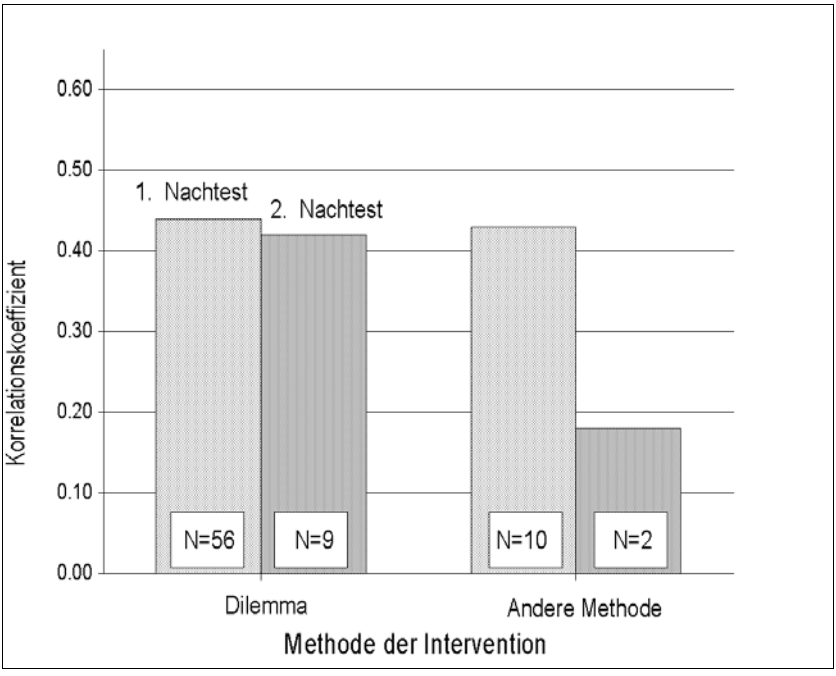


Abb. 9.3 Interventionsmethode: Effektstärke von Dilemma-Methode versus andere Methoden

Da das Effektstärkemaß r_{xy} auf die Variation der Meßwerte in einer Unter suchungsgruppe standardisiert wird, führt diese Nachlässigkeit meistens zu einer *Unterschätzung* der Effektstärke.

Da, wo es möglich war, wurden Kennwerte für *wiederholte* Messungen gewählt oder von uns nachträglich berechnet. Am häufigsten wurden *t*-Werte aus dem Vergleich der mittleren moralischen Urteilsfähigkeit vor und nach der Intervention berichtet. Diese wurden von uns nach Formel (2) in r_{xy} umgerechnet. Waren Kontrollgruppen vorhanden, dann haben die Autoren gewöhnlich das Ergebnis einer Varianzanalyse berichtet. In diesem Fall haben wir den *F*-Wert nach Formel (3) oder die Varianzanteile nach Formel (4) in r_{xy} umgerechnet. Auch hier wurde, wenn möglich, die Tatsache berücksichtigt, dass es sich um Meßwiederholungen handelt. Es wurde aber darauf verzichtet, die berichteten Zahlenwerte durch Zuhilfenahme irgendwelcher weiteren Annahmen von uns aus zu korrigieren, wenn dafür keine feste Grundlage vorhanden war. Entsprechend unserer Fragestellung nehmen wir hier (anders als von manchen vor-

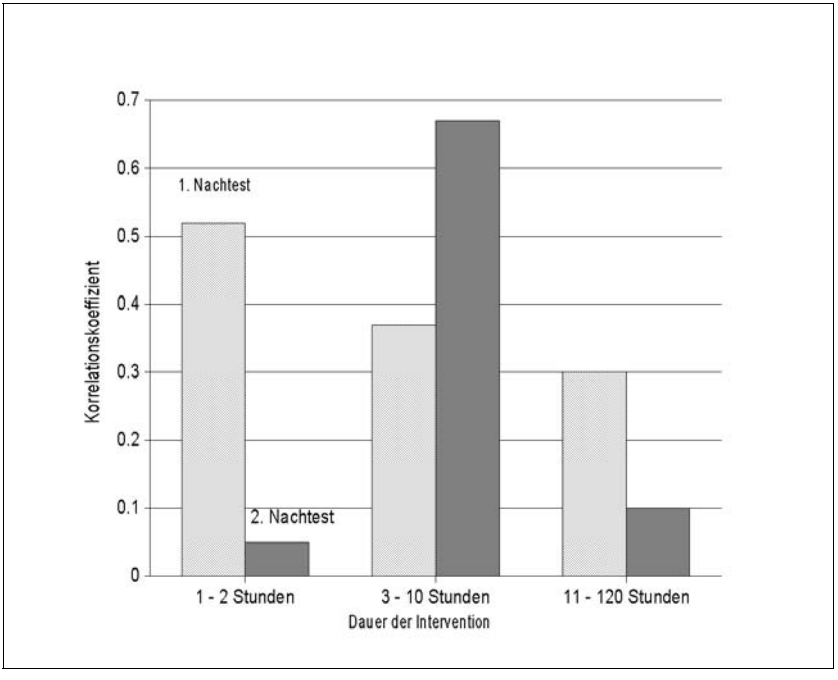


Abb. 9.4 Einfluß der *Interventionsdauer* auf die Veränderung der moralischen Urteilsfähigkeit (C-Wert, MUT)

geschlagen) keine Bewertungen der Güte der Untersuchungen vor. Es kommt uns nur darauf an zu sehen, ob der Kohlberg-Ansatz *unabhängig* von methodischen Schwächen und Stärken einer pädagogischen Intervention einen substantiellen Effekt hat. Dass wir die Schwächen einzelner Studien nicht berücksichtigen, stellt für die Prüfung der Kohlberg-Methode eine zusätzliche Bürde dar. Geringe oder negative Effekte können auf methodische Mängel der Untersuchungsdesigns statt auf die Kohlberg-Theorie zurückgehen. Das nehmen wir in Kauf, da wir auch bei der praktischen Anwendung des Kohlberg-Ansatzes von sehr unterschiedlichen Bedingungen ausgehen müssen.

9.5 Ergebnisse

Das Hauptergebnis unserer Meta-Analyse ist: *Die auf Kohlberg gründende Dilemma-Diskussions-Methode zur Förderung der moralisch-kognitiven Entwicklung hat einen deutlichen und nachhaltigen Effekt bei Schülern, und zwar weitgehend unabhängig von methodischen Variationen, von Dauer und Intensität der Intervention und von den forschungstechnischen Methoden des Effektnachweises.* Besonders hervorzuheben ist, dass in keiner einzigen Untersuchung ein negativer Effekt auftrat und nur selten geringfügige Korrelationen (als Maß für die Effektstärke) zu verzeichnen waren, dass aber in fast einem Drittel der Untersuchungen Effektstärken von mehr als $r = .50$ gefunden wurden. Mehr als 50% der Untersuchungen berichten von Zusammenhängen zwischen Intervention und Veränderung der moralischen Urteilsfähigkeit, die über $r = .40$ liegen. Die Werte liegen über vielen Effektstärken, die in anderen Bereichen der psychologischen und pädagogischen Behandlung gefunden wurden (Lipsey & Wilson, 1993).

Obwohl dieser Befund im großen und ganzen für alle Untersuchungsgruppen und -bedingungen gilt, ergaben sich Unterschiede bezüglich der Effektstärken, die für den praktischen Einsatz der Kohlberg-Methoden von Bedeutung sind: Das *Alter der Probanden* spielte nur insofern eine Rolle, als die deutlichsten Erfolge der Kohlberg-Methode in der Gruppe der 11-16jährigen Schüler gefunden wurden. Eine Studie erbrachte, dass in dieser Gruppe auch mit der größten Nachhaltigkeit des Interventionseffekts zu rechnen ist (Abb. 9.2). Dieser Befund steht in Kontrast zu der weit verbreiteten Meinung (z.B. Schläfli et al., 1985), dass der Effekt in diesem Alter geringer sei als im frühen Erwachsenenalter.

Bei der *Dauer der Intervention* scheint das Optimum bei 3 bis 10 Stunden zu liegen. Kürzere Interventionen zeigen zwar auch eine beträchtliche Wirkung ($r = .52$; Basis: 17 Studien), der Effekt scheint jedoch von geringerer Dauer zu sein. Studien, die einen zweiten Nachtest durchführten, zeigen, dass der Effekt der Intervention schon nach kurzer Zeit nicht mehr nachzuweisen ist ($r = .05$; 2 Studien). Moralpädagogische Interventionen von längerer Dauer haben etwa

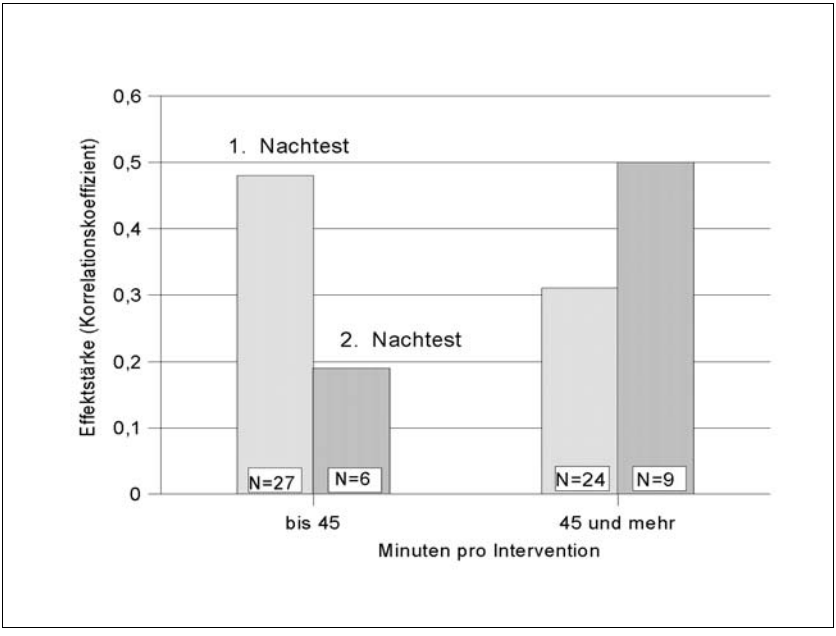


Abb. 9.5 Einfluß der Intensität der Intervention auf die Veränderung der moralischen Urteilsfähigkeit (C-Wert, MUT)

den gleichen Effekt ($r = 0.30$; Basis 16 Studien) wie mittellange, aber auch hier ist der Effekt nicht so nachhaltig ($r = 0.10$; 5 Studien;).[142]

Betrachtet man die *Intensität der Interventionen* und verfolgt, wie sie sich im durchschnittlichen Umfang der einzelnen Veranstaltungen pro Woche ausdrückt, dann zeigt sich, dass auch wenig intensive Dilemma-Diskussionen (eine Schulstunde bzw. 45 Minuten pro Woche) einen starken Einfluß auf die Entwicklung der moralischen Urteilsfähigkeit haben, dass sich intensivere Interventionen (von mehr als einer Schulstunde pro Woche) jedoch als die längerfristig wirksameren herausstellten. Auf den ersten Blick erstaunlich ist die Tatsache, dass Ethik-Unterricht und Interventionen, die mit allgemeinen, moralbezogenen Geschichten oder mit Rollenspiel-Aufgaben die Diskussion in den

142 Der Begriff "nachhaltige Wirkung" bezieht sich hier immer auf den Vergleich der Vortest-Werte mit den Messungen in einem *zweiten* Nachtest. Da nur wenige Studien einen solchen zweiten Nachtest durchgeführt haben, ist die Datengrundlage entsprechend schmal.

Schülergruppen in Gang brachten, fast ebenso effektvoll sind wie klassische Dilemma-Diskussionen. Langfristig erwies sich die Dilemma-Methode aber als die wirkungsvollere Methode.
Schließlich läßt sich anhand dieser Interventionsstudien sehen, welchen Einfluß die *Qualifikation des Interventionsleiters* auf die Wirksamkeit der Intervention hat. Auch sie scheint für die Höhe der unmittelbaren Effekte weniger bedeutsam als für die Nachhaltigkeit der Wirkung. Eine langfristig wirksame Förderung wurde am häufigsten festgestellt, wenn ein (in der Kohlbergschen Interventionsmethode ausgebildeter) Lehrer die Leitung der Intervention innehatte. Beim zweiten Nachtest wurde ein weiteres Anwachsen der moralischen Urteilsfähigkeit festgestellt (Basis: 3 Studien). Waren (Entwicklungs-) Psychologen oder ausgebildete Gleichaltrige die Versuchsleiter, so kam es zu einer Stabilisierung bzw. zu einem leichten Abfall der Werte. Natürlich liegt hier die Erklärung auf der Hand. Die Lehrer blieben in den meisten Fällen auch noch nach dem Interventionsexperiment mit den Schülern in Kontakt und dieser kann für ein weiteres Ansteigen der Urteilsfähigkeit verantwortlich gemacht werden. Diese Annahme müßte in weiteren Experimenten genauer untersucht werden. Die Tatsache der Nachhaltigkeit der pädagogischen Intervention durch Lehrer zeigt jedoch, dass es sich bei dem Kohlberg-Ansatz um eine von Lehrern praktizierbare Pädagogik handelt.

9.6 Resümee

Die Analyse ergibt also, *dass Interventions-Methoden nach Kohlberg weitgehend unabhängig von methodischen Variationen, von Dauer und Intensität der Intervention und von den forschungstechnischen Methoden des Effektnachweises einen substantiellen Effekt auf die moralisch-kognitive Entwicklung von Schülern haben.* Trotz heterogener Interventionsbedingungen war in keiner Studie ein negativer Effekt zu verzeichnen. Die kompetenzfördernde Wirkung solcher Interventionen konnte noch nach Wochen und Monaten nachgewiesen werden.

Das Ergebnis unserer Meta-Analyse steht weitgehend in Einklang mit den Meta-Analysen anderer Autoren.[143] Unsere detaillierten Analysen der Studien ergaben aber weiterführende Befunde, die zum Teil gängigen Annahmen widersprechen: a) Die vielzitierte These, wonach Erwachsene stärker von moralisch-kognitiven Interventionen profitieren als Kinder, hat sich nicht bestätigen lassen; es zeigte sich, dass 11- bis 16-jährige am meisten von Kohlberg-Interventionen profitieren.

b) Kurze Interventionen von 1-2 Stunden Dauer sind u.U. wirksamer als lange.

c) Die "klassischen" Kohlberg-Methoden der Dilemma-Diskussion und der Gerechte Gemeinschaft-Schule sind wirksamer als Methoden wie Ethikunterricht oder Rollenspiel.

Der auffälligste Unterschied ergab sich hinsichtlich der Interventionsstudien, die mit dem *Defining Issues Test* (DIT) von Rest durchgeführt wurden, und zwar bezüglich der Rolle, die das Alter der Probanden spielt (Schläfli et al. 1985). Auch die DIT-Studien zeigen durchgängig, dass der Kohlberg-Ansatz eine fördernde Wirkung hat. Gleichzeitig legen sie jedoch nahe, dass Fördermaßnahmen bei Erwachsenen eine viel größere Wirkung haben als bei Kindern und Jugendlichen. Diesem Befund wird von einigen Autoren große Bedeutung zugemessen, da er wichtig für die Frage ist, in welchem Alter Moralerziehung einsetzen sollte (Edelstein (1987, S. 188; Oser & Althof 1992, S. 155). Edelstein sieht darin eine Bestätigung für die Vermutung, dass Erwachsene anders als Jugendliche und Kinder einen "Nachholbedarf" an moralischer Entwicklung haben. Insofern scheint der Befund aus den Interventionsstudien, die mit dem DIT evaluiert wurden, die *Reifungstheorie* der Moralentwicklung zu bestätigen. Von Nachholbedarf kann aber nur gesprochen werden, wenn davon ausgegangen werden kann, dass es so etwas wie eine "natürliche" Reifung der Moral gibt. Die Projekte, die mit Kohlbergs MJI evaluiert wurden, kommen hingegen, wie wir gesehen haben, zu einem ganz anderen Ergebnis: Sie bestätigen, wie bereits Lockwood (1978) feststellte, dass Kinder auf einem niedrigen Entwicklungsniveau mehr von Interventionsstudien nach dem Kohlberg-Ansatz profitierten als andere. Sie zeigen die größte Wirksamkeit bei Jugendlichen zwischen 11 und 16 Jahren, also der jüngsten untersuchten Gruppe, an. Dieser Befund paßt besser zu der von uns vertretenen *Bildungstheorie* der

143 Vgl. Rest 1974; Lockwood 1978; Stuhr 1978, Lawrence 1980; Higgins 1980; Leming 1981; Enright et al. 1983; Schläfli et al. 1985.

Moralentwicklung, die davon ausgeht, dass die mit Abstand am häufigsten evaluierte Dilemma-Diskussions-Methode für das Entwicklungsniveau dieser Altersgruppe am angemessensten ist.

Wie kommt diese Diskrepanz zustande und welche Aussage ist nun die richtige? Die Vermutung liegt nahe, dass die Diskrepanz vor allem durch die unterschiedlichen Meßmethoden bedingt sind. Für die Frage, welcher Aussage mehr Vertrauen geschenkt werden kann, sind zwei Studien aufschlußreich, die mit dem Moralisches Urteil-Test (MUT; Kapitel 4) und dem Verantwortungsurteil-Test (VUT; Kapitel 8) evaluiert wurden. In beiden Studien zeigte sich generell, dass durch die moralpädagogische Intervention tatsächlich nicht, wie die *Sozialisationstheorie* impliziert, die Einstellungen zu bestimmten Moralstufen verändert werden (es zeigten sich keine neuen Prioritäten in der Präferenz für bestimmte Argumentationsstufen), sondern allein der kognitive Aspekt der Struktur des moralischen Urteils gefördert wurde. In beiden Studien zeigte sich auch, dass relativ junge Menschen deutlich von moralpädagogischen Interventionen profitieren. In der ersten Studie, dem Interventionsprojekt von Oser und Schläfli bei Schweizer Banklehrlingen im Alter zwischen 15 und 21 Jahren, zeigte sich im Strukturwert des MUT ein deutlicher Zuwachs. Die Korrelation zwischen Maßnahme und Effekt betrug $r = 0{,}26$.[144] Auch im zweiten Interventions-Projekt, "Demokratie und Erziehung in der Schule", über das bereits im Kapitel 8 berichtet wurde, zeigte sich ein deutlicher Zugewinn bei der Urteilsfähigkeit (VUT), obwohl die teilnehmenden SchülerInnen erst 10-13 Jahre alt waren. Beide Studien lassen uns zu der Überzeugung kommen, dass die reifungstheoretisch inspirierte These vom spezifischen Nachholbedarf an moralischer Entwicklung im Erwachsenenalter nicht zutrifft, und dass die

144 Median des C-Wertes vorher: 24,3, nachher: 31,3; Kontrollgruppe: 25,3; eigene Berechnungen; bei der Kontrollgruppe liegt kein Nachtestwert für den MUT vor. Die Größe der Experimentalgruppe war $N = 41$; Kontrollgruppe $N = 21$. Diese Werte werden hier erstmals publiziert. Für eine Darstellung der Gesamtstudie siehe Oser & Schläfli (1986). Dagegen ließ sich, wie Oser und Althof (1992, S. 106) berichten, mittels des parallel verwendeten Socio-Moral Reflection-Fragebogen von Gibbs *keine* Veränderung des moralischen Urteils feststellen. Der fehlende Effekt deutet auf eine im Vergleich zum MUT geringere Änderungssensitivität des Interviewverfahrens hin, und nicht, wie Oser und Althof (1992) ohne Kenntnis der obigen MUT-Daten vermuteten, darauf, dass "eine Intervention von einer Woche zu kurz" ist (S. 162). Interventionen von dieser Dauer und Intensität haben, wie wir oben gesehen haben, durchweg deutliche Effekte erbracht.

mit Kohlbergs *MJI* ermittelten Effekte valider sind als die mit dem *DIT* evaluierten Ergebnisse.

Zusammengefaßt stützen alle – mit drei sehr verschiedenartigen Meßverfahren – evaluierten Interventionsexperimente die bildungstheoretische Sicht von Moralentwicklung und Moralpädagogik: a) Bei den Effekten kognitiv orientierter Moralerziehung handelt es sich nicht um eine Anpassung oder Sozialisation in Hinblick auf bestimmte moralische Normen, sondern um die Förderung der moralischen Urteilsfähigkeit. b) Die moralische Urteilsfähigkeit läßt sich bei Kindern und Jugendlichen ebenso gut oder vielleicht sogar noch besser fördern als bei Erwachsenen. c) Rein theoretische Einführungen in Moralphilosophie und Moralpsychologie haben kaum eine Wirkung im Gegensatz zu handlungsorientierten Dilemma-Diskussionen oder Just community-Programmen. Anders lautende Befunde zu a) bis c) sind auf Studien begrenzt, die mit dem DIT von Rest (1986) evaluiert wurden und müssen den methodologischen Besonderheiten dieses Meßverfahrens zugerechnet werden. Der DIT mißt nicht, wie der MUT – und in gewissem Maß auch das MJI, die *kognitive* Seite des moralischen Urteilsverhaltens, sondern allein die *affektive* Seite des moralischen Urteils, nämlich moralische Einstellungen (Lind, 1996c). Effekte in DIT-Studien spiegeln daher eher moralische Einstellungsänderungen wieder, wie sie beispielsweise bei Erwachsenen durch theoretische Schulung erreicht werden können, als eine Entwicklung der moralischen Urteils*fähigkeit*. DIT-Studien sind daher nur von bedingtem Wert als Grundlage für moralpädagogische Empfehlungen.

Schließlich konnte mit der Meta-Analyse auch gezeigt werden, d) dass, wie bereits Enright et al. (1983) festgestellt haben, auch relativ kurze, aber intensive Interventionen eine nachhaltige Wirkung hinsichtlich der Förderung der moralischen Urteilsfähigkeit haben. Dies ist ein ermutigendes Ergebnis. Moralische Urteilsfähigkeit läßt sich zwar nicht durch bloße Indoktrination und Wertevermittlung fördern, aber der notwendige Aufwand für eine erfolgreiche Moralerziehung scheint durchaus überschaubar und, angesichts der großen Bedeutung, dem moralische Fähigkeiten für das Individuum und der Gesellschaft in einer Demokratie zukommen, auch geboten.

10 Moralische Stereotype und pluralistische Ignoranz: Die Wahrnehmung moralischer Bedürfnisse von Schülern durch Lehrer

Aus bildungstheoretischer Sicht erfordert der Prozeß der moralisch-kognitiven Entwicklung offenbar einen anderer Typ des Erziehers als aus reifungs- oder sozialisationstheoretischer Sicht. Erzieher bzw. Lehrer stellen für die Heranwachsenden hier wichtige Interaktionspartner dar, die sich nicht abwartend verhalten oder versuchen, Werte und Einstellungen normativ zu verändern, sondern auf die moralischen Bedürfnisse der Jugendlichen eingehen. Der Beitrag der Lehrer zur Förderung der moralischen Urteilsfähigkeit ist nicht immer leicht auszumachen, aber er ist kaum in Abrede zu stellen (Kapitel 8). Es ist zwar richtig, dass, wie Piaget annimmt, die Gleichaltrigengruppe (*peers*) eine eminente Bedeutung für die Entwicklung von sozialen Bindungen (Freundschaften und Gruppen) und sozialer Kooperation hat (Keller & Edelstein 1986; Stolz 1987). Aber es ist fraglich, ob es Kindern, auf sich allein gestellt, gelingen kann, jene Gerechtigkeitsstrukturen zu entwickeln, die für die Lösung von Dilemmas und Konflikten unabdingbar sind.

Bereits Foerster (1918) wies auf den Unterschied zwischen sozialem Lernen einerseits und, wie er es nannte, staatsbürgerlichem Lernen andererseits hin. Während es beim sozialen Lernen nur um eine einheitliche Meinungsbildung oder Gesinnung geht, wie sie sich in der Cliquen- und Bandenbildung im frühen Jugendalter zeigt, geht es bei den staatsbürgerlichen Fähigkeiten darum, widerstreitende Interessen und Ansichten zu harmonisieren. Auch Garz (1989) unterstreicht die Notwendigkeit, zwischen Freundschaftsentwicklung und Moralentwicklung zu differenzieren. Die "Verbindung zwischen Versprechen, die einem Freund gegenüber gegeben wurden, und dem Kohlbergschen Ansatz [ist] eher als Grenzfall denn als Testfall für eine Theorie der Gerechtigkeit anzusehen [. . .] Für Freundschaftsbeziehungen [ist] gerade kein Austarieren von Rechten und Pflichten konstitutiv, sondern [. . .] im Gegenteil eine 'Verpflichtung zum Ungleichgewicht'." (S. 123, Fußnote).

Wie schwer sich Kindergruppen zum Beispiel bei der Lösung eines Aufteilungsproblems tun (wie teilen Kinder zwei Kinokarten auf, wenn mehr als zwei Kinder sie haben wollen?), zeigen die Gruppenexperimente von Oser (1981a, S. 58). Wenn sie allein sind oder nur allgemeine Ratschläge erhalten, was alles bei

der Lösung des Verteilungskonflikts zu berücksichtigen ist, kommt die Diskussion der Kinder kaum von der Stelle. Aber sobald der – erwachsene – Versuchsleiter eine "Gerechtigkeitsregel" vorschlug, wurde die Interaktionsstruktur (Oser) deutlich besser. Wie dramatisch der Änderungsgewinn ist, zeigt die Höhe des Interventionseffekts an: $r = 0.74$ (Oser, 1981a, S. 173).[145]

Die Rolle, die Lehrer nicht nur potentiell, sondern tatsächlich bei der Förderung moralisch-demokratischer Fähigkeiten bei Heranwachsenden spielen können, hängt natürlich von mehreren Bedingungen ab. Zum Beispiel: Wissen Lehrer von ihren Möglichkeiten und nutzen sie diese auch? Lassen ihnen der Lehrplan und andere Verpflichtungen genügend Raum im Unterricht, um Schülern die Bedingungen der Möglichkeit für moralische Diskurse einzuräumen? Lehrer werden zwar über das Funktionieren von Personen im allgemeinen unterrichtet, aber was wissen sie über die besonderen moralischen Bedürfnisse und Sorgen ihrer Schüler?

Das Unterrichten im Klassenzimmer ist auch auf das Wissen um diese Antworten angewiesen. Es ist ein interaktiver Vorgang, in dem der Lehrer sich *sowohl* auf generelles Wissen über didaktische Regeln *als auch* auf spezifisches Wissen über die Situation und die beteiligten Personen verlassen muß. Wenn beispielsweise ein Schüler den Unterricht stört, muß der Lehrer zwischen verschiedenen Handlungsentwürfen wählen, je nachdem *wie er den spezifischen Fall bewertet*. Wenn die Störung als ein Zufallsprodukt erscheint, wird der Lehrer sie möglicherweise übersehen; wenn dieser Vorfall durch maliziöse Intentionen des Schülers bedingt ist, mag der Lehrer eine Strafe als adäquate Reaktion erachten; sollte jedoch die Störung einen Hinweis auf eine Enttäuschung des Schülers über eine vorangegangene Lehrerentscheidung darstellen, mag eine integrative Handlung die angemessenste Art und Weise sein, mit der Situation umzugehen. Im Idealfall wird der Lehrer die Situation eingehend erforschen oder notfalls – bevor er seine Schlüsse zieht – eine Moraldiskussion über das Thema führen. In der Praxis sind allerdings schnelle Entscheidungen gefragt; dies impliziert, dass die Perzeption des Lehrers von konkreten

145 Berechnung des Verfassers auf der Basis der Formeln in Kapitel 9. Eine "hundertprozentige" Wirkung der Intervention in der beabsichtigten Richtung würde durch einen Wert von $r = 1.0$ angezeigt, keine Wirkung durch $r = 0.0$ und eine vollkommen negative Wirkung durch $r = -1.0$.

Situationen, größtenteils durch *generalisiertes Wissen über Schüler* determiniert wird.

Diese Art von Wissen wird oft als *Schema* oder *Stereotyp* bezeichnet, da es kaum auf einer tatsächlichen Erforschung der Situation und den beteiligten Personen durch den Lehrer basiert, sondern größtenteils auf *indirekten* Quellen, wie vergangenen Erfahrungen als Lehrer oder Schüler, auf didaktischen Autoritäten oder auf Fachliteratur (Dann 1989a,b; 1990; Kramis 1990). Stereotype sind nicht *a priori* falsch. Ganz im Gegenteil: In vielen Fällen stellen sie das genaueste Erziehungswissen dar, das zur Verfügung steht. Ein erfolgreiches Klassenzimmermanagement scheint tatsächlich oft von solchen generalisierten Wissensbeständen über Schüler abzuhängen (Dann et al. 1987, Fraser et al. 1987, Shulman 1987). In den Bereichen der Lehrer-Schüler-Interaktion, in denen aus verschiedenen Gründen dem Lehrer kein direkter Zugang zur Situation möglich ist, stellt indirekt erworbenes Wissen sogar die einzige Informationsquelle dar.

Stereotype sind jedoch immer Wissen aus zweiter Hand. Um präzise zu sein, bedürfen sie einer ständigen Korrektur durch Rückmeldung von den Schülern. Wenn sie zu vage oder gar falsch sind, führen sie zu einer Störung oder gar einem Stillstand in der Schüler-Lehrer-Interaktion. Daher ist in Bereichen, in denen Lehrer einen persönlichen Zugang zu den Bedürfnissen und Sorgen der Schüler haben, der direkte Wissenserwerb vorzuziehen – wenn nicht gar aus ethischen Gründen geboten. Dies gilt besonders für schulbezogene moralische Sichtweisen der Schüler.

Diese Studie über das Wissen der Lehrer über Schüler sollte als eine erste Erkundung betrachtet werden. Sie orientiert sich an dem sozialpsychologischen Konzept der *pluralistischen Ignoranz*[146] und den kognitiven Theorien der Moralentwicklung und Moralerziehung[147]. Diese Studie wurde entworfen, um in bezug auf das Wissen der Erziehenden über Schüler zwei Fragen zu beantworten: 1. In welchem Ausmaß besitzen Lehrer ein Schema über *den* Schüler? 2.

146 Das Konzept der *Pluralistischen Ignoranz* geht auf Floyd Allport 1924, zurück. Siehe auch Krech & Crutchfield 1948, Breed & Ktsanes 1961. Neuere Forschung zu diesem Konzept scheint es, wie die Recherche in *PsycLit* ergab, in der Psychologie nicht zu geben.

147 Siehe u.a. Kohlberg 1984, 1987; Lind 1985a; Oser 1986; Power et al. 1989; und Kapitel 2.

Wie genau ist dieses Wissen? Um diese Forschungsfragen zu beantworten, sollen hier Schülerantworten zu Themen des Schullebens mit den Perzeptionen der Lehrer über diese Schülerantworten verglichen werden. Die Vergleichsdaten entstammen einer Reihe von Studien über die moralische Atmosphäre von Schulen, an denen ich beteiligt war. Diese Studien ermittelten u.a. die "moralischen Bedürfnisse" von Schülern von Klasse fünf bis zwölf. Im einzelnen wurde ermittelt, wie wichtig ihnen zum Beispiel folgende Dinge sind:

- Gegenseitige Hilfeleistungen der Schüler, auch wenn sie nicht der gleichen Clique angehören
- Zugehörigkeit zu einer Clique, die Außenstehenden gegenüber zusammenhält
- Gerechte und respektvolle Behandlung durch die Lehrer
- Eine Schule, in der sich Lehrer und Schüler umeinander kümmern
- Eine Schule, die eine gute Ausbildung gewährleistet
- Schüler, die wenig schummeln
- Eine demokratische Form der Beschlußfassung in der Schule, d.h. jeweils eine Stimme für Lehrer und Schüler

Die meisten Schüler zeigten großes Interesse daran, den Fragebogen zu beantworten. Ein Schüler bemerkte: "Diese Themen werden nie von uns in der Klasse diskutiert; dies sollte öfters passieren". In einigen Fällen wurden Schüler und Lehrer durch den Fragebogen dazu angeregt, die Themen "Gemeinschaft" und "Gerechtigkeit" in der Klasse zu besprechen.
Als ich die Ergebnisse einem Lehrerpublikum präsentierte, fragte ich mich, ob sie Überraschung hervorrufen würden, oder ob die Lehrer die Bedürfnisse und Sorgen der Schüler bereits vor meinem Vortrag kannten. Ich bat daher das Publikum, mir ihre Sichtweisen über das, "was ein Durchschnittsschüler der Klassen 5 bis 12 zu jedem dieser Themen sagen würde", darzustellen. Selbstverständlich kann diese Aufforderung Irritationen erzeugen. Die befragten Lehrer wurden aus diesem Grund ermutigt, die Produktion einer Klischee-Antwort zu unterlassen, falls es ihnen unangebracht erschien. Ich muß noch anmerken, dass die Lehrer weder über ihre Sichtweisen des moralischen Klimas der Schule befragt wurden, noch ob diese mit den Schüleransichten übereinstimmten. Statt dessen richtete sich mein Interesse darauf herauszufinden, ob Lehrer ein Wissen

über die moralischen Bedürfnisse von Schülern haben und wie genau dieses Wissen ist.

Plausibilitätsüberlegungen führen zu zwei völlig konträren Vorhersagen der Lehrerantworten. Auf der einen Seite sollten Lehrer Schüler sehr gut kennen. Sie interagieren täglich mit Schülern, und sie haben Zugang zu der Fachliteratur, die auf einen großen Forschungsfundus über Einstellungen, Intentionen und Verhalten von Schülern fußt. Zusätzlich ergab eine neuere Studie, dass "ohne Ausnahme alle Lehrer dem Vertrautsein mit den moralischen Belangen und Einstellungen der Schüler eine große Relevanz zuordnen" (Brugman & Dohle 1990, S. 13). Auf der anderen Seite wird die Lehrer-Schüler-Kommunikation oft durch den offiziellen Lehrplan eingeschränkt. Dementsprechend erhalten vom Curriculum nicht vorgesehene Schülerfragen oft den Status einer Störung, die lediglich von den Lernzielen ablenken. Dieses Problem verdient besondere Berücksichtigung, weil die Intentionen, Bedürfnisse und Sorgen der Schüler in Lehrbüchern für Lehrer selten enthalten sind. Es wurden bislang kaum Studien über das Wissen der Lehrer über Schülereinstellungen durchgeführt.[148] In Dunkin und Biddles (1974) Modell zur Erforschung des Unterrichts im Klassenzimmer, das "eine enorme Auswirkung auf dieses Gebiet hatte" (Shulman 1987, S. 5), wird das Erlangen der Lehrkompetenz als ein einsamer Prozeß konzipiert: Obwohl das Modell äußerst komplex ist, wird der Diskurs mit Schülern nicht in Betracht gezogen.

Größtenteils bezieht sich die Forschung auf das Regelwissen der Lehrer, d.h. auf Wissen, das in Sätzen wie dem folgenden ausgedrückt werden kann: *Wenn* eine Störung vorkommt, *dann* tue dies oder jenes. Um Regelwissen erfolgreich anzuwenden, bedarf es neben der Kunst des Lehrens gerade auch des Wissens über die Personen und Umstände der alltäglichen Klassenzimmerinteraktion. Effektives und verantwortungsvolles Unterrichten erfordert beide Arten des Wissens (Oser 1986, 1990; Brugman & Dohle 1990).

148 Zu den wenigen Studien gehört jene von Morton-Williams (zit. bei Raven 1973), die gezeigt hat, daß Lehrer die Wichtigkeit beruflicher und familiärer Werte bei Schülern *unter*schätzen. Den Hinweis verdanke ich Hanns-Dietrich Dann.

10.1 Untersuchungsanlage

Es wurden zwei Gruppen von Schülerdaten verwendet und mit den Lehrerperzeptionen verglichen. Der erste Datensatz stammt von Schülern unterer Klassen, die an der laufenden Forschung des Projekts "Demokratie und Erziehung in der Schule" (DES) in Nordrhein-Westfalen teilnahmen (s. Kapitel 8). Von 1987 bis 1989 wurden 580 Fragebogeninterviews durchgeführt. Aus dieser Befragung wurden die Daten der Schüler der fünften bis siebten Klassen herausgezogen, die eine Personenzahl von $N = 405$ ergaben. Der Datensatz von den älteren Schülern stammt aus einer Studie über die Entwicklung der sozialen Verantwortung (Link & Lind 1988). Für die beabsichtigte Analyse werden hiervon nur die Daten der Schüler der Vollzeitschule verwendet (Daten der Teilzeitschüler der Berufsschulen wurden ausgeschlossen). Folglich bleiben $N = 143$ Befragte der Klassen acht bis zwölf übrig. Somit besteht die komplette Schülerstichprobe aus $N = 548$ Sekundarschülern.

Die Lehrerdaten wurden bei den verschiedensten Gelegenheiten in den Jahren 1987 bis 1989 vom Autor selbst erhoben. Es beteiligten sich $N = 129$ Lehrer der verschiedenen Schultypen in der Altersklasse von 21 bis 63 Jahren und einer Lehrerfahrung von 2 bis 33 Jahren; 44% waren männlich und 56% weiblich. Unser Ziel bestand darin, das generelle Wissen der Lehrer über die Sichtweisen *der* Schüler in Erfahrung zu bringen. Die Schüler und Lehrer in dieser Untersuchung sind daher einander nicht zugeordnet; sie stammen von verschiedenen Schulen und Schultypen.

Den Schülern wurden insgesamt 22 Fragen zum Thema "moralische Bedürfnisse und Sorgen der Schüler" gestellt. Diese Fragen sind Teil des *Moralische Atmosphäre Fragebogens* (s. Kapitel 8). Eine vollständige Liste der hier verwendeten Fragen ist im Anhang enthalten (Fragen B1 bis B19 und D1 bis D3). Diese Fragen berühren die fünf Hauptbereiche des Schulalltags:

1) Schüler-Schüler-Interaktion (4 Aussagen)
2) Schüler-Lehrer-Interaktion (3 Aussagen)
3) Beziehung der Schüler zu ihrer Schule im ganzen (6 Aussagen)
4) Einhaltung von Regeln in der Schule (6 Aussagen)
5) Schülerbeteiligung bei Entscheidungen (3 Aussagen)

Bei den ersten vier Bereichen wurden Schüler gebeten, die Aussagen gemäß ihrer subjektiven Relevanz einzustufen, und zwar von "1" („nicht wichtig") bis "5" („sehr wichtig"). Bei der fünften Fragedomäne konnten die Schüler die Aussagen von "1" ("Meiner Ansicht nach ist das völlig falsch") bis "5" ("Komplett richtig") einstufen. Da bei einigen Lehrerstudien nur eine dreistufige Skala für die letzte Domäne benutzt wurde, stützt sich die folgende Analyse auf eine transformierte Skala, die von "1" („falsch") bis "3" („richtig") reicht.

Die Schülerfragebogen wurden während des Unterrichts ausgefüllt. Die Lehrerfragebogen wurden bei Seminaren über Schulerziehung ausgeteilt. Vor dem Ausfüllen der Fragebogen wurden keine Informationen über die Schülerdaten gegeben. Die Formulierung der einzelnen Fragen war für alle Versuchspersonen identisch. Die Lehrer erhielten allerdings, wie bereits vermerkt, die Zusatzinstruktion, den Fragebogen so auszufüllen, wie die *meisten* Schüler typischerweise die Fragen beantworten würden. Sie wurden gebeten, sich in einen typischen Schüler hineinzuversetzen, der die 5. bis 12. Klasse irgendeiner Sekundarschule besucht. Sie wurden außerdem gebeten, für jede Frage anzugeben, ob sie nur wenig oder viel Abweichung zwischen den Schülern erwarten würden. Als Maß der Effektgröße wählten wir Cramérs *V* (Cramér 1966).[149] *V* stellt das Ausmaß des Zusammenhangs zwischen der Variablen "Gruppe" (Lehrer und Schüler) auf der einen Seite und den Antworten auf der anderen Seite dar.[150] Ein hoher *V*-Wert bedeutet eine geringere, ein niedriger *V*-Wert eine große Genauigkeit der Stereotype der Lehrer. Statistische Signifikanztests fanden hier keine Anwendung.

Inhalt und Struktur der Stereotype: Sowohl die kognitive Entwicklungstheorie (Kapitel 2) wie auch neuere Studien über Lehrer (Dann et al. 1987) legen eine

149 *Ein hoher V-Wert* indiziert eine geringe Übereinstimmung der Stereotype der Lehrer mit den wirklichen Werthaltungen der Schüler. Der *V* Koeffizient kann von χ^2 abgeleitet werden: $V = \frac{\chi^2}{N * (q - 1)}$

mit *q* als der kleineren Anzahl der Kategorien der zwei Variablen. Daher kann *V* Werte von 0 bis 1 annehmen, auch wenn die Zwei-Weg-Häufigkeitstabelle nicht quadratisch ist. *V* wurde mit dem statistischen Computerprogramm KOSTAS (Nagl et al. 1986) berechnet.

150 In den Abbildungen wird die relative Häufigkeit der höchsten Antwortkategorie ("sehr wichtig") für Schüler (Balken) und Lehrer (Linien) dargestellt. Der Bereich der Zustimmungsraten über 40 % wurde zusätzlich grau schattiert.

Differenzierung zwischen *Inhalt* und *Struktur* der Lehrerstereotype nahe. Während der Inhalt als der Mittelwert aller Antworten eines Befragten definiert wird, wird die Struktur definiert als die Beziehung zwischen ihren Antworten. Mit anderen Worten, die Strukturanalyse zielt auf die *Bedeutung* der individuellen Antworten ab. Die Strukturanalyse widmet sich beispielsweise der Frage, was der "Wunsch nach einer guten Schule" für Schüler und Lehrer bedeutet, d.h., welche anderen Bedürfnisse werden aus der Sichtweise der Schüler oder der Lehrer von diesem einzelnen Bedürfnis impliziert bzw. ausgeschlossen? Da sich Personen solcher Bedeutungsstrukturen mehr oder weniger bewusst sind, beschlossen wir die "implizite" Bedeutungsstruktur zu analysieren, d.h. die Struktur, die sich aus dem Muster der Antworten ergibt. Bei dieser Strukturanalyse werden für jede Stichprobe separate Korrelationskoeffizienten zwischen Paaren von Aussagen berechnet. Um einen Gesamteindruck der Bedeutungsstrukturen der Schüler und der Lehrer zu erhalten, werden diese Korrelationen zusätzlich einer Faktorenanalyse unterzogen.

10.2 Ergebnisse

10.2.1 Lehrer haben ein Stereotyp über "den" Schüler

Die meisten Lehrer scheinen stark streotypisierte Vorstellungen von Schülern zu haben. D.h., Lehrer scheinen auch Schüler definitiv zu kennen, mit denen sie noch nie persönlich Umgang hatten. Dieser Schluß basiert auf drei Befunden. Erstens, 70 bis 100% der Lehrer kommen der Aufgabe der Beschreibung der Antworten eines typischen Schülers nach. Zweitens, nur 5 bis 15% geben an, dass die Schülerantworten in Bezug auf moralische Bedürfnisse und Sorgen etwas oder gar stark variieren. Drittens, es ergeben sich relativ geringe Abweichungen bei den Lehrerantworten über die Sichtweisen der Schüler. Die Spanne ihrer Durchschnittseinschätzungen der Schülerantworten zu den Fragen B1 bis B19 ist größer als die Spanne der durchschnittlichen Schülerantworten. Lehrer scheinen sehr zuversichtlich bezüglich ihres Wissens über Schüler zu sein.

10.2.2 Die Stereotype der Lehrer über Schüler sind oft sehr ungenau

Im großen und ganzen fallen die typischen Lehrerstereotype mit den typischen Schülerantworten zusammen. Die Antworten der Lehrer weisen ein sehr ähnliches Muster auf wie die der Schüler, die über die gleichen Themen interviewt wurden. Gemäß diesem Gesamteindruck scheinen Lehrer die moralischen Bedürfnisse und Sorgen der Schüler sehr gut zu kennen. Man kann allerdings Bedenken gegen diesen Schluß anmelden, da er sich nur auf eine Gesamtanalyse stützt. Die Lehrereinschätzungen sind schließlich nicht auf ein wirkliches Kennen der Schüler gegründet, deren Daten für den Vergleich herangezogen wurden, sondern basieren auf konventionellem Wissen. Außerdem schließt der Befund der großen Ähnlichkeit zwischen den Durchschnittsantwortprofilen der Lehrer und Schüler nicht aus, dass trotzdem Unterschiede bestehen können. *Die eingehende Analyse der Daten zeigte in der Tat auffällige Ungenauigkeiten der Wahrnehmung bei den LehrerInnen sowohl in Bezug auf ihren Inhalt wie auch ihre Struktur.* Bezüglich des *Inhalts* des Lehrerwissens über Schüler ergab sich folgendes Bild:

- *LehrerInnen unterschätzen stark, welche Wichtigkeit eine gute Ausbildung für Schüler hat.* Man könnte meinen, dass diese Unterschätzung durch die Lehrer darauf zurückzuführen ist, dass sie "vorsichtig" antworten, also ihre Antworten zumeist um den Mittelpunkt der vorgegebenen Skala liegen und sich nur wenig darum streuen (technisch gesprochen würde ein solches Antwortmuster die Deutung nahelegen, dass die Lehrerantworten überwiegend ein Ergebnis von Zufallsprozessen, einer "Regression zur Mitte", sind). Dies ist aber nicht der Fall. Die Lehrerstereotype tendieren mehr in Richtung des Nullpunkts ("nicht wichtig") als die tatsächlichen Schülerantworten, also nicht auf den Skalenmittelwert ("3") hin. Zudem ist die Spanne der Lehrerantworten größer als die der Schüler. Dies widerlegt die Annahme, dass Lehrer generell Schülerwichtigkeiten "vorsichtig" einschätzen oder dass ihre Antworten gar als Ausdruck eines Zufallsprozesses zu deuten sind. Was auch immer die richtige Erklärung ist, die Diagnose der Ungenauigkeit von Lehrerwahrnehmungen bleibt bestehen.
- *Die Wahrnehmung der Lehrer ist am genauesten in bezug auf die Schüler-Lehrer Interaktion und die sie betreffenden moralischen Bedürfnisse und Sorgen.* Alle drei Fragen, welche die Schüler-Lehrer-Interaktion betreffen, befinden sich unter den vier am genauesten beantworteten Fragen. Lehrer schei-

nen sehr wohl wahrzunehmen, dass die Schüler von ihnen gerecht und respektvoll behandelt werden wollen (70 – 80% der Schüler weisen dies als "sehr wichtig" aus), sich eine Offenheit der Lehrer für Diskussionen wünschen (ungefähr 60%) und kaum Wert darauf legen, dass die Lehrer die Schulregeln durchsetzen (etwa 5-15%).

- *Lehrer sind nur mäßig mit den moralischen Bedürfnissen der Schüler in bezug auf Schüler-Schüler Interaktion und Schüler-Schule-Beziehungen vertraut.* Obwohl generell eine gute Übereinstimmung zwischen den Durchschnittsantworten der Schüler und Lehrer besteht, zeigen sich signifikante Unterschiede, wenn man die "sehr wichtig"-Antworten untersucht. Lehrer unterschätzen durchschnittlich das Ausmaß der Wichtigkeit, das Schüler sowohl dem freundschaftlichen und hilfsbereiten Umgang miteinander zuordnen wie auch dem Bedürfnis nach Cliquen, die gegenüber Fremden zusammenhalten. Lehrer unterschätzen auch den Wunsch der Schüler nach einer guten Gemeinschaft ("Schüler und Lehrer sorgen sich umeinander"). Viele Lehrer zeigten sich überrascht, als sie hörten, dass eine große Mehrheit der Schüler einer guten Ausbildung in der Schule (60-80%) einen großen Wert beimaß. Der *V*-Koeffizient für das letzte Item beträgt 0.41, was bedeutet, dass die Lehrerperzeptionen markant von den realen Bedürfnissen der Schüler abwichen
- *Viele Lehrer unterschätzen das Ausmaß der Sorgen, die sich Schüler über Regelverletzungen in der Schule machen.* Die Einhaltung der Regeln wird von den Schülern als wichtiger erachtet als viele Lehrer glauben. Besonders jüngere Schüler der Klassen 5 bis 7 sorgen sich beachtlich um Vandalismus, Diebstahl und Kämpfe zwischen den Mitschülern. Schule schwänzen und das Abschreiben bei Prüfungen scheint für diese jungen Schüler auch von Belang zu sein, wenngleich in einem geringeren Ausmaß. Ältere Schüler kümmern sich weniger um diese Dinge. Sie sorgen sich mehr um starke Konkurrenz (Konkurrenzkampf wird aus der Lehrersicht meist nicht als Regelverletzung betrachtet, kann jedoch aus Schülersicht als Verletzung von Solidaritätsgeboten bezeichnet werden.).
- *Viele Lehrer unterschätzen den Schülerwunsch nach demokratischer Partizipation an der Formulierung und Durchsetzung von Schulregeln.* Unabhängig vom Alter weisen die meisten Schüler Wünschen wie "Schülermitspracherecht bei der Entscheidungsfindung in der Schule" und "demokratische Wahlprinzipien" eine große Wichtigkeit zu. Sie sagen außerdem, dass die Schülerbeteiligung bei der Entscheidungsfindung in der Schule richtig ist, obwohl ältere

Schüler die demokratische Teilhabe an Entscheidungsfindungsprozessen als Schülerrecht zurückhaltender betrachten.

10.2.3 Für Lehrer und Schüler haben moralische Bedürfnisse und Sorgen auch eine verschiedene Bedeutung

Was verstehen Schüler und Lehrer unter dem, was wir sie gefragt haben? Verstehen sie darunter jeweils das gleiche oder etwas anderes? Von den benutzten Begriffen sind drei für den moralischen Erziehungsprozeß besonders kritisch: „gute Gemeinschaft", „Cliquen" und „demokratische Partizipation". Wir haben die Bedeutungen dieser drei Begriffe auf zwei Wegen untersucht:

a) Durch die Analyse der Korrelationen dieser drei Begriffe mit den anderen Fragen über die moralischen Bedürfnisse und Sorgen, die wir den Schülern gestellt haben
b) Durch die Analyse der Muster dieser Korrelationen mittels Faktorenanalyse[151]

Beide Analysen zeigen, dass sich Schüler und Lehrer nicht nur hinsichtlich ihrer Antworten zu den einzelnen Fragen stark unterscheiden, sondern auch was ihr Verständnis zentraler Begriffe anbelangt. Die Stereotype der LehrerInnen über die moralischen Bedürfnisse von SchülerInnen weisen erhebliche *Struktur*unterschiede auf:

- *Aus Sicht der Schüler beinhaltet der Wunsch nach einer guten Gemeinschaft so etwas wie den Wunsch nach Freundschaft*. Freundschaftsorientierung korreliert hoch mit dem Wunsch der SchülerInnen, dass ihre Mitschüler einen freundlichen und hilfsbereiten Umgang miteinander pflegen (auch wenn sie unterschiedlichen Cliquen angehören), dass Schulregeln gerecht und gut sind und dass Lehrer und Schüler Probleme in offener Diskussion Probleme angehen

151 Die Anzahl der Faktoren wurde durch die Verwendung des Eigenwerts 1 als Abbruchkriterium etabliert. Die Faktorenextraktion wurde durch eine schiefe Rotation (Promax-Methode) determiniert. Die Kalkulation erfolgte mit KOSTAS (Nagl et al. 1986).

Korrelationen können durch einen „Faktor" repräsentiert werden, der Kohlbergs Stufe 3 („guter Freund") der Moralperspektive entspricht.

- *Der Wunsch der Schüler nach einer guten Schulgemeinschaft, die auf Fürsorge, Hilfsbereitschaft und Freundschaft gegründet ist, korreliert positiv mit den meisten anderen moralischen Bedürfnisse und Sorgen.* Er korreliert besonders deutlich mit Wünschen bezüglich der Schüler-Lehrer-Interaktion: Viele der Schüler, die eine gute Schulgemeinschaft wollen, wünschen sich, dass Lehrer offene Diskussionen über Probleme und Konflikte fördern und – in geringerem Ausmaß – einen gerechten und respektvollen Umgang mit Schülern pflegen, sowie für die Einhaltung der Schulregeln sorgen. Schüler, die eine "freundliche Schule" für wichtig halten, sind auch der Meinung, dass Schüler ein Mitspracherecht bei Schulentscheidungen haben sollten und dass Regelverletzungen in der Schule reduziert werden müßten.
- *Das Verständnis der LehrerInnen von der "Schule als einer guten Gemeinschaft" paßt recht gut mit dem Verständnis der Schüler zusammen.* Auch bei - Lehrern ist dieses Thema eng mit Freundlichkeit, Hilfsbereitschaft und Fürsorge zwischen den Schülern verbunden. In mancher Beziehung jedoch weicht das Verständnis der LehrerInnen deutlich von dem der SchülerInnen ab. *Lehrer scheinen eine engere Beziehung zwischen dem Wunsch der Schüler nach einer guten Gemeinschaft und ihrem Wunsch nach weniger Regelverletzungen zu sehen.* Alle Korrelationen zwischen "guter Gemeinschaft" und den Fragen zum Bereich der "Regeleinhaltung" fallen in der Lehrergruppe höher als in der Schülergruppe aus . Das gleiche gilt für den Bereich "Schüler-Schule-Beziehung", mit einer wichtigen Ausnahme: *Bei den Schülern korreliert der Wunsch nach einer guten Gemeinschaft mit dem Wunsch nach demokratischer Partizipation, während bei den Lehrern keine Korrelation zwischen den beiden Variablen vorliegt.*
- *Lehrer tendieren dazu, die Bedeutung von Cliquen für Schüler verzerrt einzuschätzen. Cliquen sind für jeden Schüler Teil einer guten Gemeinschaft. Doch besonders für jüngere Lehrer scheinen Cliquen die gegenteilige Bedeutung aufzuweisen.* Viele (vor allem die jungen) Schüler möchten gerne Mitglied einer Clique sein, die Fremden gegenüber zusammenhält. Aus Schülersicht scheint der Wunsch, Mitglied einer Clique zu sein, eine positive Bedeutung zu besitzen. Er korreliert in ihrer Sicht außerdem (moderat) mit den Wünschen nach guten Lehrer-Schüler-Beziehungen, gutem Schulleben (sowohl gerechte Regeln wie auch gute Ausbildung) und weniger Regelverletzungen.

- *Aus der Lehrerperspektive ist der Wunsch der Schüler, Mitglied einer Clique zu sein, mit fast allen Aspekten einer guten Schule negativ korreliert.* Schüler, die eine Cliquenbildung favorisieren, werden so wahrgenommen, als würden sie sich wenig um Freundlichkeit und Hilfsbereitschaft kümmern und weniger als andere Schüler dies tun. Ebenso wie ihnen die Einhaltung der Schulregeln, das Führen offener Diskussionen und die Partizipation an der größeren Gemeinschaft weniger relevant sind. Die weitere Analyse zeigt, dass besonders jüngere Lehrer die Bedeutung der Cliquen für ihre Schüler mißzuverstehen scheinen. (In anderen Bereichen weisen allerdings jüngere Lehrer genaueres Wissen über die Schüler auf als ihre Kollegen mit längerer Berufserfahrung.)
- *Lehrer scheinen mißzuverstehen, was Schüler mit dem Wunsch nach demokratischer Partizipation in der Schule meinen.* Die Analyse zeigt, dass die Bedeutungsstruktur der Lehrer in bezug auf demokratische Partizipation schlecht mit denen der Schüler zusammenpaßt. Die Wertschätzung der Schüler hinsichtlich demokratischer Entscheidungsfindungen korreliert positiv mit allen anderen Bereichen moralischer Bedürfnisse. Schüler, die demokratische Entscheidungsfindungen wollen, halten auch gerechte und gute Regeln für sehr wichtig. Beide Fragen wiesen eine Korrelation von $r = 0.37$ in der Schülerstichprobe auf. Viele Lehrer scheinen sich dieses hoch positiven Zusammenhangs nicht bewusst zu sein und anzunehmen, dass die Befürworter einer demokratischen Verwaltung unter den Schülern sich wenig um gute und gerechte Regeln kümmern würden. In dieser Gruppe ist die Korrelation denn auch deutlich niedriger *(*$r = 0.15$).
- *Besonders die Erzieher mit langer Lehrerfahrung missverstehen die Konnotationen der Schüler.* In dieser Gruppe ist die Korrelation sogar negativ ($r = -0.12$). Ältere Lehrer nehmen offensichtlich auch an, dass ein Schülerwunsch nach demokratischer Verwaltung eine Opposition gegen die Lehrer darstellt, die Regeln durchsetzen ($r = -0.25$), während die Schülerantworten diesen Schluß nicht zulassen. Jüngere Lehrer haben ein genaueres Wissen über die demokratischen Wünsche der Schüler. Mehr Lehrerfahrung scheint in dieser Hinsicht nicht die Genauigkeit der Wahrnehmung zu fördern.
- Die zusammenfassende Analyse der Struktur der moralischen Bedürfnisse und Sorgen in der Gruppe der Schüler ergibt drei Faktoren, die eine Ähnlichkeit zu den Stufen drei bis fünf in Kohlbergs (1984) Modell der Moralentwicklung aufweisen: Der Faktor 3 “Freundschaft-Orientierung” scheint Bedürfnisse oder

Motive zu bündeln, die sich der dritten Stufe (guter Junge – nettes Mädchen) der moralischen Orientierung der individuellen moralisch-kognitiven Entwicklung zuordnen lassen. Der Faktor 1 "Norm-Orientierung" verbindet die moralischen Bedürfnisse, die durch Kohlbergs Stufe 4 repräsentiert werden (Einhaltung von Gesetz und Ordnung). Der Faktor 2 "Diskurs-Orientierung" ähnelt der Stufe 5 (legaler Vertrag und prinzipielle Moralität) in Kohlbergs Skala der Moralentwicklung.[152] Die Faktorenanalyse der Lehrerwahrnehmungen ergibt vier Faktoren. Die Faktorenstruktur der Lehrer ähnelt deutlich der Struktur der Schüler. Die Lehrer scheinen die "Norm-Orientierung" in zwei Faktoren zu differenzieren: Ein Normfaktor ist durch den Wunsch charakterisiert, dass Schüler selbst die Verantwortung für die Einhaltung von Schulnormen ergreifen (Faktor 1), der andere Normfaktor entspricht dem Wunsch, dass Lehrer die Schulregeln durchsetzen.

10.3 Resümee

Die Untersuchung in diesem Kapitel richtete sich auf das mögliche Wissen der Lehrer über die moralischen Bedürfnisse und Sorgen von Schülern. Ausgehend von der Annahme, dass Lehrer normalerweise Entscheidungen unter Handlungsdruck treffen müssen, waren wir besonderes an ihrem generellen Wissen, d.h. an ihren Stereotypen von *dem* Schüler (im Gegensatz zu einem ihnen persönlich bekannten Schüler), interessiert. Die Stereotype der Lehrer wurden bezüglich ihres Inhalts und ihrer Struktur untersucht. Wir verglichen (a) die Durchschnittsantworten der Lehrer und Schüler zu 22 Themen, die sich auf Moralverhalten und Schule bezogen, und (b) die Bedeutung dreier Themen für diese zwei Gruppen.

Die Untersuchung ergab: 1. Die meisten Lehrer besitzen eine ausgeprägte Stereotypvorstellung über *den* Sekundarschüler. Lehrer glauben definitiv, die Schüler zu kennen. 2. Die typischen Lehrerstereotype über Schüler ähneln zwar im großen und ganzen den typischen Schülerantworten, aber im einzelnen weisen sie beachtliche Ungenauigkeiten bei der Wahrnehmung der Schüler auf,

152 Für eine Diskussion der Korrelation zwischen der individuellen moralischen Entwicklung der Schüler und der moralischen Atmosphäre in der Schule siehe Power et al. 1989. Für Studien über die moralische Entwicklung von Lehrern siehe Althof 1990 und Rulon 1990.

und zwar sowohl in bezug auf ihren Inhalt als auch auf ihre Struktur: Was den Inhalt ihrer Stereotype angeht, so scheinen Lehrer den Wert zu unterschätzen, den Schüler einer guten Schule zuschreiben. Die Lehrer nehmen die Schülerantworten zur Schüler-Lehrer-Interaktionen am genauesten wahr. Sie sind allerdings nur mäßig vertraut mit den moralischen Bedürfnissen und Sorgen der Schüler in bezug auf Schüler-Schüler-Interaktion und Schüler-Schule-Beziehungen. Lehrer unterschätzen markant das Ausmaß, mit dem Schüler sich Sorgen über Regelverletzungen in der Schule machen, und ihren Wunsch nach Mitbestimmung bei der Gestaltung und Durchsetzung von Schulregeln. Was die Struktur ihrer Stereotype betrifft, so passen die Vorstellungen, die Lehrer mit "Schule als eine gute Gemeinschaft" verbinden, nur mäßig gut mit denen der Schüler zusammen. Lehrer assoziieren den Wunsch nach einer guten Gemeinschaft mit weniger Regelverletzungen als die Schüler. Bei Schülern korreliert der Wunsch nach einer guten Gemeinschaft mit dem Wunsch nach demokratischer Partizipation; für Lehrer gilt dies nicht. Lehrer scheinen die Bedeutung, die Cliquen für Schüler haben, mißzuverstehen. Aus Schülersicht ist das Vorhandensein einer Gruppe, die gegenüber anderen zusammenhält, integraler Teil einer guten Gemeinschaft. Bei den Lehrern ist der Wunsch der Schüler, Mitglied einer Clique zu sein, weniger positiv (oft negativ) mit fast allen Aspekten einer guten Schule korreliert. Jüngeren Lehrern ist, wie die Daten zeigen, das Bedürfnis der Schüler nach Cliquen noch weniger vertraut als den älteren Lehrern. Für sie hat der Wunsch der Schüler nach demokratischer Partizipation eine wesentlich negativere Konnotation als für die Schüler. Für Schüler ist der Wunsch nach demokratischer Partizipation positiv mit dem Wunsch nach Einhaltung der Schulregeln verbunden, während für Lehrer dieser Wunsch ein Akt der Opposition gegen die Lehrer, welche die Regeln durchsetzen, darstellt. In diesem Fall haben die älteren Lehrer eine ungenauere Wahrnehmung als die jüngeren.

Das Wissen der Lehrer über *den* Schüler ist besser als man erwarten kann, wenn man sich die ihnen gestellte Aufgabe vor Augen führt. In Anbetracht der Tatsache, dass sie generalisiertes Wissen wiedergaben, und der Ungenauigkeit dieses Wissens, müssen wir ihre Vorstellungen dennoch als *Stereotype* bezeichnen. (Zugegeben: Die meisten, wenn nicht sogar alle unsere Entscheidungen stützen sich auf Stereotype, d.h. auf generalisiertes Wissen, das mehr oder minder genau ist). Wenn Stereotype über andere Personen in der eigenen Gruppe, zum Beispiel über Schüler im Klassenzimmer, auffällig ungenau sind, dann

sprechen wir von *pluralistischer Ignoranz*. Auf Grund der Resultate dieser Studie ziehen wir den Schluß, dass, zumindest *in bezug auf einige zentrale Aspekte der moralischen Bedürfnisse und Sorgen der Schüler, viele Lehrer ein beachtliches Ausmaß an pluralistischer Ignoranz demonstrieren*.

Selbstverständlich könnten wir sehr plausible Erklärungen für einige dieser Ungenauigkeiten der Lehrerurteile finden. Vielleicht hatten viele der Lehrer beispielsweise die Frage im Hinterkopf, wie Schüler sich *gewöhnlicherweise* benehmen, und nicht, wie sie sich benehmen *wollen*. Es ist zu beachten, dass wir nicht das moralische Verhalten der Schüler aufgezeichnet haben, sondern ihre Wünsche in bezug auf ein gutes moralisches Klima in der Schule; auch befragten wir die Lehrer nicht nach beobachtbarem Verhalten der Schüler. Das Hauptinteresse der Pädagogik gilt dem Verhalten der Schüler. Es sollte jedoch weder als der einzige Gegenstand des Interesses, noch sollte es als ein einfaches oder eindeutiges Konzept gelten. Verhalten kann nur eindeutig von einer bestimmten Perspektive aus gedeutet werden. Aus der Lehrerperspektive kann beispielsweise ein spezielles Verhalten als aggressives Verhalten betrachtet werden, während aus der Schüler- oder Elternperspektive das gleiche Verhalten als eine defensive Handlung ausgelegt wird. Ohne die zu einer Handlung gehörigen Bedürfnisse des Akteurs zu kennen, scheint es kaum möglich, bzw. unwahrscheinlich, dass die an einer Interaktion beteiligten Personen Übereinstimmung darüber erreichen, welche Bedeutung eine Handlung hat. Eine möglichst genaue Wahrnehmung der moralischen Bedürfnisse und Sorgen der Schüler durch die Lehrer ist, so legen es unsere Befunde nahe, eine notwendige Bedingung für ethischen Diskurs und Konsens im Klassenzimmer.

Diese Befunde werfen natürlich auch Fragen zum Beruf des Lehrers und zur Lehrerausbildung auf. Hat pluralistische Ignoranz einen Einfluß auf den Schulalltag? Welche Ursachen hat die pluralistische Ignoranz? Wie kann ihr begegnet werden? Wir können auf diese Fragen keine abschließende Antwort geben. In bezug auf den Einfluß der pluralistischen Ignoranz, scheint mir, hat sie einen stark hemmenden Effekt auf den moralischen Diskurs zwischen Lehrern und Schülern.[153] Hinweise auf den moral-behindernden Effekt der pluralisti

153 Ich erinnere mich an ein Ereignis meiner eigenen Schulzeit: Eines Tages kam unsere Klasse nach der vierten Stunde von einem Ausflug zurück. Unser Lehrer der fünften Stunde vermutete, dass wir eine Schulnorm gebrochen hatten und war darüber sehr aufgebracht. Er "wußte", ohne sich durch Rückfragen zu vergewissern, dass Schü-
(Fortsetzung...)

schen Ignoranz stammen hauptsächlich aus der Forschung über das Hilfeverhalten. Dort wird gezeigt, dass pluralistische Ignoranz eine der Hauptursachen für das Unterlassen von Hilfeleistungen ist, wenn jemand sie braucht. Wenn beispielsweise eine Person "weiß", dass das Opfer selbst keine hilfsbereite Person ist, oder wenn angenommen wird, dass andere Anwesende kompetentere Hilfe leisten könnten, ist sie weniger geneigt zu helfen (Latané & Darley 1970, Bierhoff 1980). Ähnlich wird ein Lehrer, der von der Annahme ausgeht, dass die Schüler von moralischen Belangen wie Gerechtigkeit, Fürsorge, Hilfsbereitschaft und Verantwortung völlig unberührt bleiben, sie möglicherweise dementsprechend behandeln. Der Lehrer könnte von der Annahme fehlgeleitet sein, der Charakter der Schüler könne leicht aus ihrem Verhalten abgeleitet werden. Dieses Verhalten ist oft mehr durch die Situation (von welcher der Lehrer ein wichtiger Bestandteil ist) determiniert als durch die Dispositionen des Schülers. Die Forschung zeigt, dass Schüler ihre moralischen Wünsche offenbaren, wenn Erzieher ein förderndes Diskursklima etablieren.[154] Ein förderndes Diskursklima stellt sich nur dann ein, wenn der Lehrer sich selbst als wichtige Ursache für das störende Verhalten von Schülern erkennt, und wenn er respektlose Kommentare über Schüler unterläßt und auf unmittelbare Bestrafungen verzichtet.[155]

Über die Ursachen der pluralistischen Ignoranz und die Möglichkeiten einer Verringerung gibt es wenig gesichertes Wissen. Das Phänomen der pluralistischen Ignoranz wurde bereits 1924 von Floyd Allport entdeckt; seitdem wurde dieses Thema allerdings nur wenig erforscht. Krech und Crutchfield (1948) nahmen an, dass pluralistische Ignoranz sich besonders in großen Gruppen entwickelt, in denen "die einzelnen Mitglieder kaum in unmittelbare Kom-

153(...Fortsetzung)
ler, die vor der 5. Stunde die Schule betraten, unerlaubterweise das Schulgrundstück verlassen hatten. Er war sich dessen so sicher, dass er jede Stellungnahme von Schülern unterband: "Ich weiß ganz genau was Ihr getan habt, versucht erst gar nicht, Euch zu entschuldigen!" Dieser Vorfall hatte eine dramatische Auswirkung auf die Atmosphäre im Klassenzimmer. Wir entschieden von da an, jegliche "unnötige" Kommunikation mit diesem Lehrer zu unterlassen; der Klassensprecher trat vom Amt zurück.

154 Oser 1986, 1990; Kohlberg 1987; Power et al. 1989; Perret-Clermont 1990.

155 In einer Studie über aggressives Verhalten im Klassenzimmer zeigten Dann et al. (1987), dass solches Lehrerverhalten hoch mit gutem Lernklima und wenigen Schülerstörungen korreliert.

munikation miteinander treten und die tatsächlich vorkommende Kommunikation oft durch rigide Kanäle geschieht" (S. 389). Nach der Studie von Breed und Ktsanes (1961) über pluralistische Ignoranz in bezug auf Einstellungen zur Rassensegregation in den USA der fünfziger Jahre, hat aber die Gruppengröße allein darauf keinen Einfluß. Dennoch weisen die Befunde darauf hin, dass falsche Stereotype oft mit eingeschränkte Kommunikationsbedingungen korrelieren. In der Schule wird Lehrer-Schüler-Kommunikation oft eingeschränkt durch a) den vom offiziellen Curriculum verursachten Zeitdruck und b) durch die Sorge der Lehrer, dass sie den vielseitigen Erwartungen von Kollegen, Eltern, Schülern und Schulverwaltung, auf die sie in der Ausbildung nicht hinreichend vorbereitet werden, nicht gerecht werden könnten.[156] Wenn Lehrer keine Möglichkeit haben, ihr professionelles Wissen in der Lehrer-Schüler-Kommunikation ständig zu überprüfen, kann dies leicht in pluralistische Ignoranz münden.

Demnach muß die Frage lauten: Wie kann das professionelle Wissen gegen pluralistische Ignoranz geschützt werden? Es scheint, dass dies nur dadurch geschehen kann, dass Lehrer mit Schülern auch über moralische Bedürfnisse und die Lösung moralischer Konflikte regelmäßig reden. Wie dies geschehen könnte, haben Kohlberg (1987), Oser und Althof (1989), Higgins (1987) und andere ausführlich dargestellt. Ich möchte einen solchen moralischen Diskurs hier anhand einer Episode schildern, die aus dem Projekt "Demokratie und Erziehung in der Schule" (DES) stammt (s. Kapitel 8). Schüler wurden immer wieder bei der Verletzung der Norm, das Schulgrundstück während der Pausen nicht zu verlassen, angetroffen. Die Lehrer dieser Schule verließen sich nicht auf ihr "Wissen um die Situation", sondern beteiligten sich an einem Moraldiskurs mit ihren Schülern. Die Schüler hatten dadurch Gelegenheit zu erläutern, warum sie die Regelverletzung begangen hatten. Sie sagten, sie mußten das Schulgrundstück verlassen, um ihr Frühstück einzukaufen. Sie befanden sich also in einem Normenkonflikt. Eine "Gerechte Gemeinschaft"-Sitzung wurde anberaumt, um das Problem zu diskutieren. Es wurde gemeinsam entschieden, auf dem Schulgelände einen Kiosk einzurichten, so dass dieser Konflikt vermieden werden konnte.

156 Junglehrer erfahren den Unterschied zwischen der Ausbildung und derartigen Erwartungshaltungen als Praxisschock (Dann et al. 1978; Fend 1977).

Wie von Higgins et al. (1984) gezeigt wurde, vermindern solche Moraldiskurse bei Schülern die pluralistische Ignoranz. Wir vermuten, dass sie auch bei Lehrern zu einer Verringerung von stereotypisierter Wahrnehmung und einem genaueren Wissen über Schüler beitragen können. Moraldiskurse im Klassenzimmer können über eine Vielzahl von Themen abgehalten werden. Kohlberg und Kollegen schlagen zwei Hauptthemen vor: (a) fiktive moralische Dilemmas, die in verschiedenen Unterrichtsfächern behandelt werden können, zum Beispiel in Gemeinschaftskunde, Sozialwissenschaften, Religion, Geschichte, Sport, Biologie, und (b) reale moralische Konflikte im Schulleben, die mit Fragen der Gerechtigkeit, Normverletzung oder persönlichen Achtung zu tun haben (s. Kapitel 8 und 9). Schirp (1990) hat darüber hinaus vorgeschlagen, Moraldiskurse auf ethische Probleme außerhalb der Schule auszudehnen und im Unterricht auch Themen aus dem Alltagsleben der Schüler, der Gemeinde oder aus anderen Teilen der Gesellschaft zu behandeln.

Eine Reduktion der pluralistischen Ignoranz in der Schule wird nicht leicht zu erreichen sein. Sie setzt einige einschneidende Veränderungen voraus. Lehrer und Schüler müssen lernen, moralische Probleme in einer konstruktiven Art und Weise zu diskutieren. Insbesondere müssen sie jeweils lernen, sich die Gründe der Gegenseite anzuhören und zur Erreichung ihrer Ziele Argumente statt Macht einzusetzen. Lehrer müssen nach Oser (1990) die Fähigkeit erlangen, "vollständige Diskurse" zu initiieren und in Gang zu halten. Wie Sibylle Reinhardt (1990) bemerkt hat, müssen Lehrer für die Schüler sowohl kulturelle Traditionen repräsentieren wie auch die spezifischen Formen des Verstehens bei Schülern achten und akzeptieren. Zur Erreichung dieser enorm wichtigen und schwierigen Aufgaben bedürfen die Lehrer zusätzlicher Ausbildung sowohl während des Studiums wie auch während der Berufsausübung. Das Curriculum muß Zeit für regelmäßige Moraldiskurse im Klassenzimmer und in der Schulgemeinschaft vorsehen. Wir müssen alle die außerordentliche Relevanz erkennen, die der Förderung moralisch-demokratischer Kompetenzen zukommt. Dies schließt ein, dass wir zu schätzen wissen, was die Schule und andere allgemeinbildende Bereiche unserer Gesellschaft jetzt schon dazu beitragen (s. Kapitel 11), aber auch sehen, welche Herausforderung diese Aufgabe angesichts einer sich ständig wandelnden Welt für die Schule bedeutet.

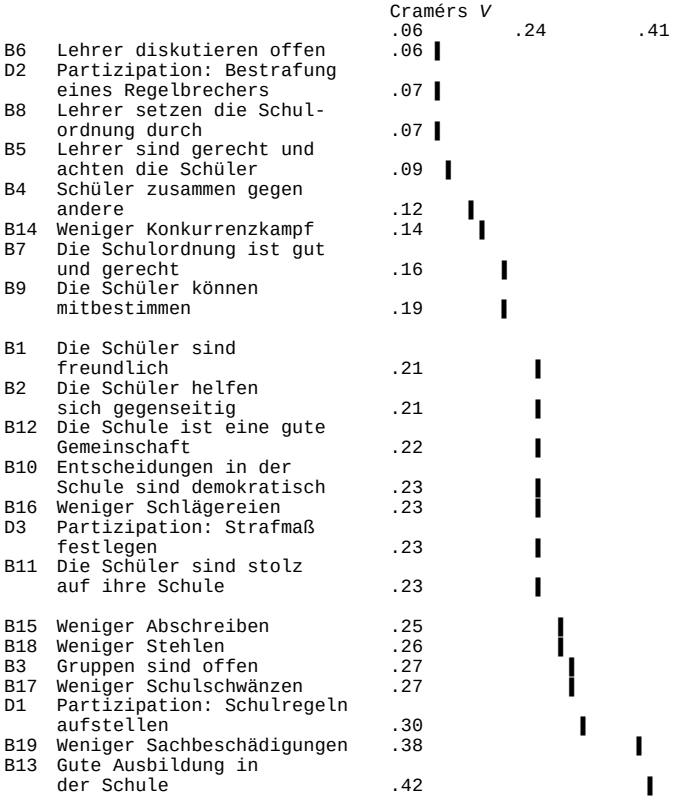


Abb. 10.1 Die Genauigkeit der Wahrnehmung der Schülermeinungen durch die Lehrer. Große V-Werte zeigen eine *geringe* Genauigkeit an.

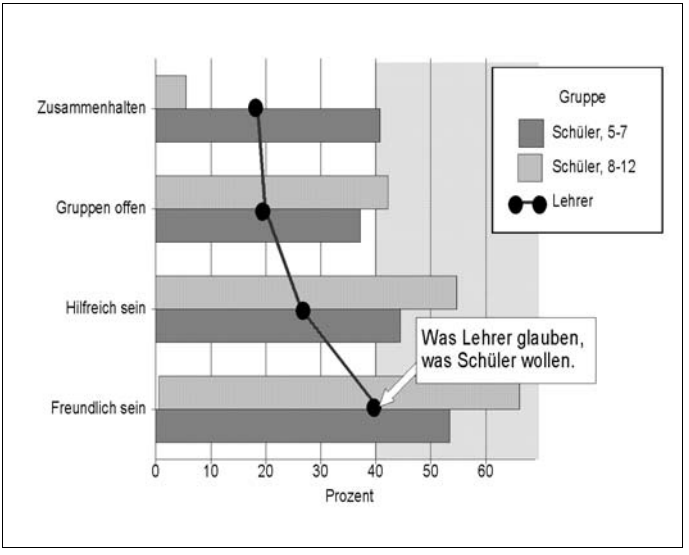


Abb. 10.2 Was Schülern wichtig ist bei ihren Mitschülern (Nennungen „sehr wichtig“)

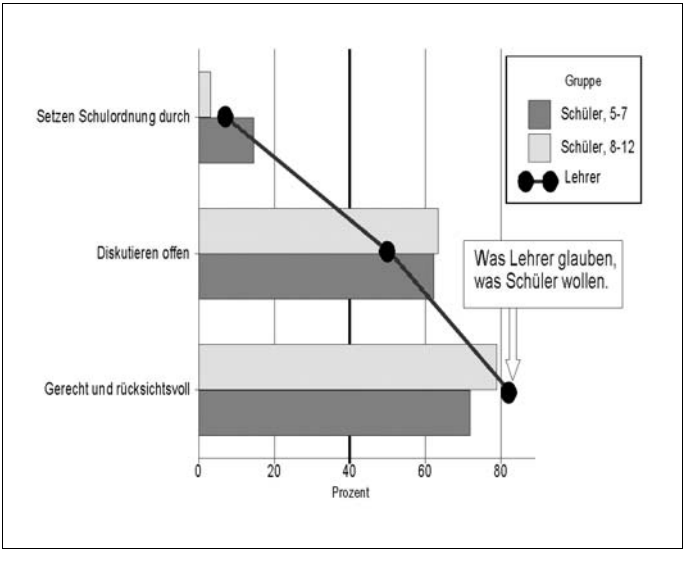


Abb. 10.3 Was Schülern wichtig ist bei ihren Lehrern (Nennungen „sehr wichtig“)

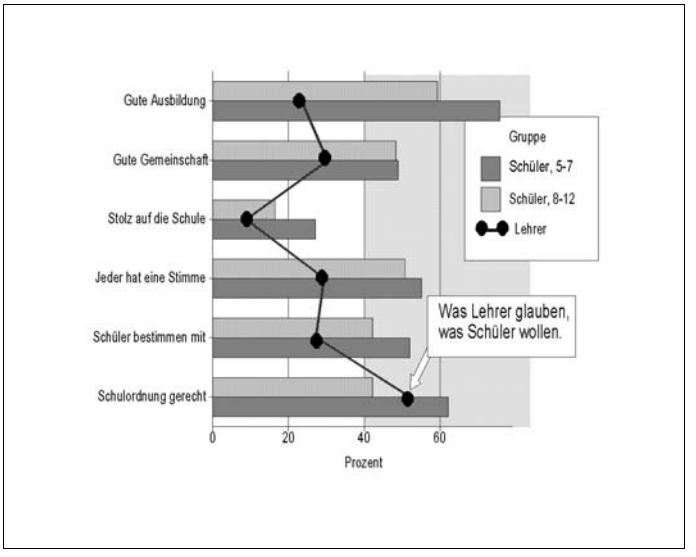


Abb. 10.4 Was Schülern *wichtig* ist an ihrer Schule (Nennungen „sehr wichtig“)

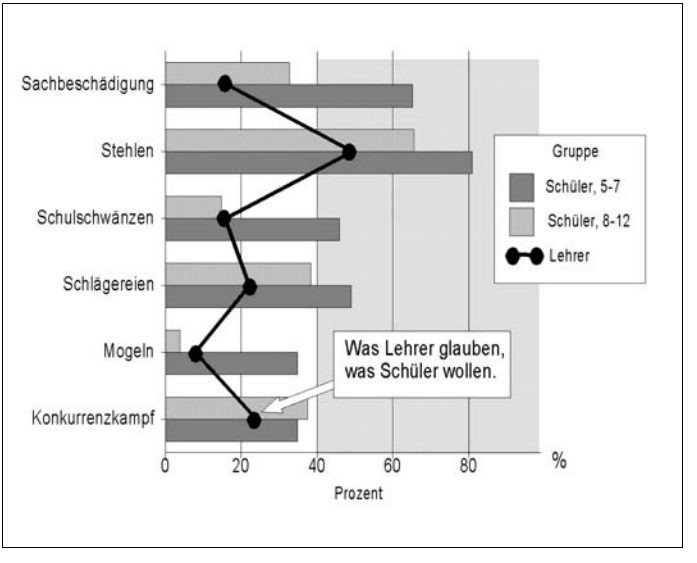


Abb. 10.5 Was Schülern wichtig ist: Weniger Regelverletzungen (Nennungen „sehr wichtig“)

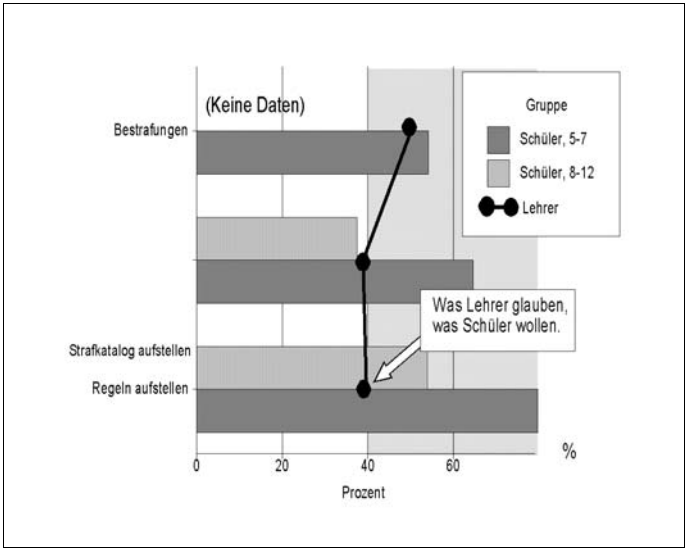


Abb. 10.6 Was Schüler bezüglich mehr Partizipation für richtig halten (Nennungen "ist richtig")

Korrelation zwischen der der Wertschätzung von Zusammenhalt gegen Andere (Item B4)
mit anderen Aspekten des Moralischen Klimas in der Schule
bei Lehrern (Stereotyp) und Schülern

MIT DEN ANTWORTEN...		*Lehrer*	*Schüler*	*Differenz*
B1	Schüler: sollen freundlich miteinander umgehen	-.27	+.11	+++
B2	Schüler: sollen einander helfen	-.32	+.12	+++
B3	Schüler: offene Gruppen	-.43	+.15	+++
B5	Lehrer: sollen gerecht und respektvoll sein	-.08	+.05	+
B6	Lehrer: sollen offen mit Schülern diskutieren	-.25	+.08	+++
B8	Lehrer: sollen Ordnung konsequent durchsetzen	+.01	+.18	+
B7	Schule: Vorschriften sollen gerecht sein	+.01	+.23	++
B11	Schule: Schüler sollen auf sie stolz sein können	-.08	+.34	+++
B12	Schule: Gute Gemeinschaft			
B9	Schule: Schüler sollen mitbestimmen	-.05	+.21	++
B10	Schule: Entscheidungen sollen demokratisch sein	-.04	+.21	++
D1	Mitbestimmung: Schulordnung	+.10	+.16	
D3	Mitbestimmung: Strafmaßnahmen (konkret)	+.07	+.12	
D2	Mitbestimmung: Festlegung von Strafen (allgemein)	+.01	+.14	+
B14	Probleme: Weniger Konkurrenzkampf	-.23	+.04	++
B15	Probleme: Weniger abschreiben und mogeln	-.03	+.27	++
B16	Probleme: Weniger Schlägereien	-.14	+.13	++
B17	Probleme: Weniger Unterricht schwänzen	-.07	+.26	++
B18	Probleme: Weniger stehlen	+.02	+.16	+
B19	Probleme: Weniger Sachbeschädigungen	-.10	+.20	+++

+/-	= pos./neg. Differenz $r > .10$
++/- -	= pos./neg. Differenz $r > .20$
+++/- - -	= pos./neg. Differenz $r > .30$

Abb. 10.7 Das semantische Feld von Gruppenzusammenhalt für Lehrer und Schüler

Korrelation zwischen der der Wertschätzung von demokratischen Entscheidungen (Item B10) mit anderen Aspekten des Moralischen Klimas in der Schule bei Lehrern (Stereotyp) und Schülern

MIT DEN ANTWORTEN...		*Lehrer*	*Schüler*	*Differenz*
B1	Schüler: sollen freundlich miteinander umgehen	-.02	+.19	++
B2	Schüler: sollen einander helfen	-.05	+.21	++
B3	Schüler: offene Gruppen	+.04	+.23	+
B4	Schüler: Gruppe soll gegen andere zusammenhalten	-.04	+.21	++
B5	Lehrer: sollen gerecht und respektvoll sein	+.12	+.23	+
B6	Lehrer: sollen offen mit Schülern diskutieren	+.31	+.29	
B8	Lehrer: sollen Ordnung konsequent durchsetzen	+.04	+.15	+
B7	Schule: Vorschriften sollen gerecht sein	+.15	+.37	++
B11	Schule: Schüler sollen auf sie stolz sein können	+.04	+.25	++
B12	Schule: Gute Gemeinschaft	+.17	+.27	+
B14	Probleme: Weniger Konkurrenzkampf	+.26	+.18	
B15	Probleme: Weniger abschreiben und mogeln	-.11	+.11	++
B16	Probleme: Weniger Schlägereien	+.07	+.23	+
B17	Probleme: Weniger Unterricht schwänzen	-.11	+.21	+++
B18	Probleme: Weniger stehlen	+.03	+.18	+
B19	Probleme: Weniger Sachbeschädigungen	+.02	+.22	++

+/-	= pos./neg. Differenz $r > .10$
++/- -	= pos./neg. Differenz $r > .20$
+++/- - -	= pos./neg. Differenz $r > .30$

Abb. 10.8 Das semantische Feld von demokratischer Mitbestimmung für Lehrer (Stereotyp) und Schüler

Korrelation zwischen der der Wertschätzung von Guter Gemeinschaft (Item B12)
mit anderen Aspekten des Moralischen Klimas in der Schule
bei Lehrern (Stereotyp) und Schülern

		Lehrer	*Schüler*	*Differenz*
B1	Schüler: sollen freundlich miteinander umgehen	+.47	+.40	
B2	Schüler: sollen einander helfen	+.46	+.47	
B3	Schüler: offene Gruppen	+.38	+.34	
B4	Schüler: Gegen andere zusammenhalten	-.28	+.17	+++
B5	Lehrer: sollen gerecht und respektvoll sein	+.14	+.24	+
B6	Lehrer: sollen offen mit Schülern diskutieren	+.27	+.40	+
B8	Lehrer: sollen Ordnung konsequent durchsetzen	+.22	+.21	
B7	Schule: Vorschriften sollen gerecht sein	+.38	+.22	-
B11	Schule: Schüler sollen auf sie stolz sein können	+.54	+.41	-
B13	Schule: Gute Ausbildung	+.47	+.35	-
B9	Schule: Schüler sollen mitbestimmen	+.04	+.23	+
B10	Schule: Entscheidungen sollen demokratisch sein	+.17	+.27	+
D1	Mitbestimmung: Schulordnung	+.07	-.02	
D3	Mitbestimmung: Strafmaßnahmen (konkret)	+.19	+.02	-
D2	Mitbestimmung: Festlegung von Strafen (allgemein)	+.17	+.12	
B14	Probleme: Weniger Konkurrenzkampf	+.48	+.20	-
B15	Probleme: Weniger abschreiben und mogeln	+.37	+.20	-
B16	Probleme: Weniger Schlägereien	+.46	+.29	-
B17	Probleme: Weniger Unterricht schwänzen	+.45	+.26	-
B18	Probleme: Weniger stehlen	+.36	+.25	-
B19	Probleme: Weniger Sachbeschädigungen	+.45	+.29	-

Erklärung:

+/-	= pos./neg. Differenz $r > .10$
++/- -	= pos./neg. Differenz $r > .20$
+++/- - -	= pos./neg. Differenz $r > .30$

Abb. 10.9 Das semantische Feld von “guter Gemeinschaft” bei Lehrern und Schülern

Faktor:			Frage:	*Schüler; n= 547*
[F3] Freundschaft-Orientierung (Typ 3)				
	.29	.00	-.70	B2. Schüler: Gegenseitig helfen
	.44	.25	-.69	B12. Schule: Eine gute Gemeinschaft
	.34	.00	-.64	B1. Schüler: Freundlich sein
	.31	.21	-.63	B3. Schüler: Gruppen sind offen
	.00	.27	-.54	B6. Lehrer: Offen diskutieren
	.43	.33	-.48	B7. Schulordnung gut und gerecht
	.00	.30	-.39	B5. Lehrer: Gerecht und rücksichtsvoll
	.30	.00	-.37	B14. Probleme: Weniger Konkurrenzkampf
[F1] Norm Orientierung (Typ 4)				
	.77	.00	-.33	B17 Probleme: Weniger Schwänzen
	.72	.00	-.40	B19 Probleme: Weniger Sachbeschädigung
	.70	.00	-.23	B15 Probleme: Weniger Mogeln
	.67	.00	-.42	B16 Probleme: Weniger Schlägereien
	.58	.00	-.37	B18 Probleme: Weniger Diebstähle
	.54	.21	-.40	B11 Stolz auf die Schule sein
	.47	.26	-.47	B13 Eine gute Ausbildung
	.42	.00	-.27	B8 Lehrer: Schulordnung durchsetzen
	.40	.36	-.24	B4 Schüler: Gegen andere zusammenhalten
[F2] Diskurs Orientierung (Typ 5)				
	.00	.75	.00	D3 Partizipation: Bestrafung aussprechen
	.00	.66	.00	D2 Partizipation: Strafen festlegen
	.31	.60	-.40	B9 Schule: Die Schüler können mitbestimmen
	.00	.55	.00	D1 Partizipation: Schulregeln aufstellen
	.37	.47	-.49	B10 Schule: Demokratische Mitbestimmung

Faktor Interkorrelation

	1	2	3	Faktor:
1.	1.00			[F1] Norm-Orientierung (Typ 4)
2.	.34	1.00		[F2] Diskurs-Orientierung (Typ 5)
3.	-.59	-.42	1.00	[F3] Freundschafts-Orientierung (Typ 3)

Faktorenanalyse: Schiefwinklige Rotation.
Matrix: Faktor-Item-Korrelationen (Faktorstruktur)

Abb. 10.10 Die Struktur der moralischen Wünsche der Schüler

Faktor:				Frage:	*Lehrer; n= 129*
[F3] Freundschafts-Orientierung (Typ 3)					
.24	.00	-.74	.30	3. Schüler: Gruppen sind offen	
.00	.00	-.72	.39	2. Schüler sind hilfsbereit	
.00	-.21	-.66	.49	1. Schüler sind freundlich	
.00	.00	.58	.00	4. Schüler halten gegen andere zusammen	
.00	.35	-.49	.24	6. Lehrer diskutieren offen	
.00	.00	-.29	.00	5. Lehrer sind gerecht und rücksichtsvoll	
[F1] Norm-Orientierung (Typ 4): Die Verantwortung der Schüler					
.98	.38	.00	.00	D2 Partizipation: Strafen festlegen	
.77	.00	-.28	.42	18. Probleme: Weniger stehlen	
.77	.00	-.45	.45	16. Probleme: Weniger Schlägereien	
.71	.00	-.44	.27	14. Probleme: Weniger Konkurrenzkampf	
.68	-.22	-.47	.62	19. Probleme: Weniger Sachbeschädigung	
.54	.00	-.25	.53	13. Schule: Gute Ausbildung	
[F4] Norm-Orientierung (Typ 4): Die Verantwortung der Lehrer					
.00	.00	.00	.59	8. Lehrer setzen Schulordnung durch	
.54	.00	-.66	.68	12. Schule ist gute Gemeinschaft	
.52	-.39	-.37	.69	17. Probleme: Weniger Schwänzen	
.00	.00	-.42	.66	11. Schüler sind stolz auf Schule	
.27	.00	-.22	.62	7. Schulordnung ist gut und gerecht	
.43	-.43	-.30	.59	15. Probleme: Weniger Mogeln	
[F2] Diskurs-Orientierung (Typ 5)					
.00	.71	.00	.00	10. Schule: Demokratische Beteiligung	
.00	.61	.00	.00	9. Schule: Schüler können mitbestimmen	
.47	.58	.00	.00	D3 Partizipation: Bestrafung aussprechen	
.52	.54	.00	.00	D1 Partizipation: Schulregeln aufstellen	

Faktor Interkorrelation

	1	2	3	Faktor:
1.	1.00			[F1] Norm-Orientierung (Typ 4): Schüler-Verantwortung
2.	-.05	1.00		[F2] Diskurs-Orientierung (Typ 5)
3.	-.34	.10	1.00	[F3] Freundschafts-Orientierung (Typ 3)
4.	.50	-.30	-.52	[F4] Norm-Orientierung (Typ 4): Lehrer-Verantwortung

Abb. 10.11 Die Struktur der Lehrer-Wahrnehmung in Bezug auf die Wünsche der Schüler

Globale Korrelation zwischen den Faktoren in beiden Gruppen (siehe Gebhardt 1968): *r = .643*

Ähnlichkeit zwischen einzelnen Faktoren in beiden Gruppen:*

SCHÜLER (n= 547):

.[F1] Norm-Orientierung (Typ 4)
| .[F2] Diskurs-Orientierung (Typ 5)
| | .[F3] Freundschafts-Orientierung (Typ 3)
| | |
| | | LEHRER (n= 127)

-.46	-.10	.92	[F3] Freundschafts-Orientierung (Typ 3)
.81	.65	-.48	[F1] Norm-Orientierung (Typ 4): Schüler-Verantwortung
.90	.17	-.66	[F4] Norm-Orientierung (Typ 4): Lehrer-Verantwortung
-.33	.71	-.17	[F2] Diskurs-Orientierung (Typ 5)

*Individuelle Korrelationen zwischen den Faktoren (Fischer & Roppert 1964).

Abb. 10.12 Grad der strukturellen Ähnlichkeit zwischen Lehrern und Schülern

11 Bildungserfahrungen als Bedingung für moralische Entwicklung

Mit der kognitiven Entwicklungstheorie haben Piaget und Kohlberg den Blick für die dem moralischen Verhalten zugrundeliegenden, kognitiven Prozesse geschärft. In ihren einschlägigen Arbeiten haben sie auch immer wieder auf die Bedeutung von Moralerziehungsprogrammen für die Moralentwicklung hingewiesen. Aber ihre reifungstheoretische Grundposition, charakterisiert durch die zentralen Annahmen einer "Entwicklungslogik" und einer "invarianten Entwicklungsabfolge", haben ihnen und ihren Nachfolgern lange Zeit den Blick verstellt für die, wie wir heute wissen, herausragende Bedeutung von allgemeinen Bildungsprozessen für die moralisch-kognitive Entwicklung. In keiner der Arbeiten von Kohlberg ist ein entsprechender Hinweis zu finden, obwohl er und seine Kollegen in vielen Untersuchungen Bildungsdaten mit erhoben haben. Bezeichnend ist, dass die Höhe der Korrelation zwischen dem Bildungsniveau und den MJI-Werten oft gar nicht oder nur beiläufig, in Fußnoten, berichtet wurde. Will man sich eine Vorstellung von dieser Korrelation verschaffen, muß man verschiedene Zahlenwerte aus Fußnoten und Anmerkungen zusammenstellen und hieraus die Effektstärke berechnen. Solche Nachberechnungen ergeben deutliche Hinweise für die Richtigkeit der Bildungsthese. So ergibt sich aus den Daten der Dissertation von Kohlberg (1958, S. 102) eine Effektstärke des Bildungsniveaus von $r = 0.65$; aus den Daten der Studie von Candee, Graham und Kohlberg (1978) eine Effektstärke von $r = 0.71$. Aber diese Daten sind als Beleg der Bildungstheorie der Moralentwicklung nicht ausreichend. Zum einen beruhen sie auf dem *Moral Judgment Interview*, dessen Problematik wir in Kapitel 3 ausführlich diskutiert haben, und auf die wir unten nochmals eingehen werden, soweit sie für das Thema dieses Kapitels relevant ist. Zum anderen fehlen in den MJI-Studien Hinweise auf die Effekte von Selektionsprozessen, die bei Korrelationen mit dem Bildungsniveau immer eine Rolle spielen können. Schließlich gibt es bislang keine Studie, die Aussagen darüber zuläßt, ob nicht-institutionalisierte Bildungsprozesse, wie sie etwa bei Erwachsenen am Arbeitsplatz ablaufen können, einen Effekt auf die Moralentwicklung haben.

11.1 Die Ausgangslage

Wir haben in den vorangegangenen Kapiteln die Bildungsthese weitgehend dadurch stützen können, dass wir gezeigt haben, dass die alternativen Erklärungen von Moralentwicklung als Produkt von Reifungs- bzw. Sozialisationsprozessen nicht zutreffen können. In den Kapiteln 8 und 9 konnte darüber hinaus anhand einer Schulreformstudie und einer zusammenfassenden Meta-Analyse gezielter Interventionsstudien gezeigt werden, dass spezielle Fördermaßnahmen, die vor allem moralische Diskussionen beinhalteten, ohne Zweifel zur Förderung der moralischen Urteilsfähigkeit beitragen. Schließlich liegen inzwischen, worauf ich noch eingehen werde, mehrere Studien vor, in denen eine sehr hohe Korrelation zwischen Bildungsniveau und moralischer Urteilsfähigkeit gefunden wurde.

Aber in bezug auf die These, dass Bildungsprozesse eine hinreichende Bedingung für die Moralentwicklung sind, lassen diese Befunde noch einige Fragen offen: a) Die bisherigen Befunde entstammen Studien, die mit Meßinstrumenten durchgeführt wurden, in denen nicht zwischen affektivem und kognitivem Aspekt unterschieden wurde. Dadurch ist nicht auszuschließen, dass sie auch oder sogar überwiegend eine Anpassung an das besondere Werteklima von Bildungsinstitutionen reflektieren, aber kein Wachstum der moralischen Urteils*fähigkeit*. b) Die meisten Befunde stellen eine globale Korrelation zwischen dem Bildungsstatus von Befragten und ihrer Moralentwicklung dar; sie können daher statt Bildungseffekte auch *Selektionseffekte* reflektieren. c) Zur Absicherung von Bildungseffekten gegenüber *Stichprobeneffekten* sind Studien mit größeren Untersuchungsgruppen notwendig. Aufgrund des geringen Umfangs der Untersuchungsgruppen, auf denen die meisten Befunde beruhen, könnten zufällige oder systematische Auswahleffekte für einen Teil der beobachteten Korrelation verantwortlich sein. d) Das Problem der *Kohorteneffekte* ist nur mit *Längsschnittstudien* zu lösen. Dieses Problem spielt in unserem Zusammenhang aber nur eine untergeordnete Rolle, da inzwischen mehrere Bestätigungen für die hohe Korrelation aus unterschiedlichen Befragtengenerationen vorliegen. e) Schließlich gab es bislang kaum Studien, an denen die *Langfristigkeit* von unmittelbaren oder mittelbaren Bildungseffekten geprüft werden konnte.

11.2 Forschungsstand

Bis Mitte der siebziger Jahre konnte kaum gesagt werden, ob Bildungsprozesse überhaupt etwas zur Förderung der moralischen Urteilsfähigkeit beitragen.[157] Das bis dahin dominierende, rein affektive Einstellungs-Paradigma erschwerte – und verhinderte – den Nachweis eines Bildungseffekts in der, wie des heißt, "affektiven Domäne". Die Befundlage dieser Zeit könnte man sehr vereinfachend auf folgenden Nenner bringen: Entweder es zeigte sich überhaupt kein Effekt (das war meist der Fall). Zeigten sich jedoch Effekte, dann waren sie oft untereinander inkonsistent, oder es zeigten sich – manchmal – konsistente Effekt, dann konnten sie als "bloße Einstellungsänderung" mit dem Hinweis auf methodische Artefakte ("soziale Erwünschtheit") in Zweifel gestellt werden. In der Öffentlichkeit hat sich daher der Eindruck verfestigt, dass Schulen und Hochschulen nur "fachliches" Wissen, aber keine sozialen Fähigkeiten vermitteln (Lind 1985c).

Mit dem Aufkommen der kognitiv orientierten Moralpsychologie bahnte sich eine Änderung dieses Verständnisses der Moral an. Immer unübersehbarer wurden die direkten und indirekten Hinweise für die Bedeutung von Bildungsprozessen für die Entwicklung der moralischen Urteilsfähigkeit. Bereits anfangs diese Jahrhunderts konnte Levy-Suhl (1912; siehe auch Heidbrink 1989) zeigen, dass moralisches Verhalten einen kognitiven Anteil hat und mehr eine Fähigkeit als eine Einstellung darstellt. Auch die Studien von Hartshorne und May (1928) zur Natur des Charakters enthalten, obschon nicht dafür ausgelegt, Hinweise auf den kognitiv-entwicklungstheoretischen Aspekt der Moral: Die Konsistenz des Verhaltens war bei älteren Jugendlichen größer als bei jüngeren (Burton 1963). Aber dies waren eher Nebenbefunde, die in der Wissenschaft kaum Resonanz fanden.

Auf den kognitiven Aspekt des moralischen Handelns und die mögliche Bedeutung von Bildungsprozessen für die Entwicklung der moralischen Urteilsfähigkeit hat als einer der ersten Piaget (1973/1932), dann auch vor allem Kohlberg (1984; 1987), aufmerksam gemacht. Aber Kohlberg konnte keine Klarheit über den Einfluß allgemeiner Bildungsprozesse auf die Moralentwicklung

157 In einem Interview gefragt, was das College, außer Wissensvermittlung, auf dem Gebiet der Persönlichkeitsentwicklung bewirke, sagte der Sozialpsychologe und Bildungsforscher Theodore Newcomb (1974), "Frankly, very little that is demonstrable" (S. 73).

schaffen, vor allem weil das von ihm und seinen Kollegen entwickelte Meßinstrument, das *Moral Judgment Interview* (MJI), im Hinblick auf die Prüfung dieser Hypothese zwei *Mängel* aufweist (s. Kapitel 3): 1. Im MJI kann der kognitive Aspekt des moralischen Urteils nicht direkt und unabhängig vom affektiven Aspekt erfaßt werden. Wegen dieser Vermengung stellen MJI-Daten keine geeignete Basis für die Entscheidung dar, ob nun die Bildungsthese oder die Sozialisationsthese zutrifft. Hohe Korrelationen zwischen Bildungsniveau und MJI-Werten können auch als Ausdruck der Wirksamkeit eines Sozialisationsdrucks interpretiert werden. 2. MJI-Daten sind qua Auswertungstechnik zugunsten der Invarianzthese verzerrt. Daher erlauben MJI-Daten keine unvoreingenommene Testung der Bildungs- und Emergenztheorie. Vielmehr begünstigen sie a priori die zweite Theorie.

Rest (1969) berichtet von drei Studien mit dem DIT, die einen hohen Zusammenhang zwischen dem Bildungsniveau und der Moralentwicklung ergaben. Die Effektstärken betrugen $r = 0.62$, $r = 0.67$ und $r = 0.60$ (S. 100). Der *Defining Issues Test* (DIT) von Rest (1979) ist neutral gegenüber der Invarianzthese; auf ihn trifft also der zweite Einwand nicht zu. Bezüglich der Invarianzthese ist mittels DIT ein unverzerrter Test der Bildungsthese möglich. Aber bei dem DIT-Instrument sind wie beim MJI affektive und kognitive Aspekte ununterscheidbar vermengt. Der von Rest vorgeschlagene Indikator für moralische Entwicklung, der *P-Wert*, besagt, wie konsistent (= kognitiver Aspekt) eine Person postkonventionelle moralische Argumente bevorzugt (= affektiver Aspekt), so dass ein hoher P-Wert auch Änderungen von moralischen Einstellungen abbildet. Daher erlaubt der DIT keine klare Unterscheidung zwischen der Bildungs- und der Sozialisationsthese.

Mit dem *Moralisches Urteil Test* (MUT) wurde ein Instrument geschaffen, a) das unverzerrte Daten für die Prüfung der Invarianz- und der Bildungsthese liefert, b) mit dessen kognitiven Meßwerten die These einer bildungsbedingten Fähigkeitsentwicklung gegen die These von sozialisationsbedingten Einstellungsänderungen unverzerrt getestet werden kann und c) das sich aufgrund seiner Kürze und der einfachen Auswertbarkeit auch bei größeren Untersuchungsgruppen einsetzen läßt. Die Validität des MUT wird allein durch das Konstruktions- und Auswertungsdesign sichergestellt und nicht durch empirische Kriterien. Dadurch wird, anders als beim MJI, eine Zirkularität in bezug auf die Prüfung der Hypothesen vermieden (Kapitel 2 und 4). Durch die unabhängige Erfassung der kognitiven und der affektiven Aspekte können Annahmen über

die Natur des Effekts von Bildungsprozessen (Förderung von moralischen Kompetenzen oder Zwang zur Anpassung an soziale Normen) direkt miteinander verglichen werden.

In den vorstehenden Kapiteln konnte bereits eine Klärung in bezug auf die Hauptthese dieser Arbeit vorgenommen werden. Im Kapitel 5 wurde die *Emergenz- oder Reifungstheorie* widerlegt. Die von Rest (1986) gefundenen Regressionen, die unverträglich mit der Annahme einer invarianten Entwicklungssequenz sind, konnten in einer repräsentativ angelegten Studie bei Jugendlichen mit dem MUT deutlich bestätigt werden. Diese Daten bestätigen auch die Vermutung von Rest, dass diese Regressionen ihre Ursache in mangelnden Bildungsgelegenheiten haben. Gegen diese Deutung könnten folgende Einwände bestehen: Vom emergenztheoretischen Standpunkt könnte eingewendet werden, dass es sich beim Regressionsphänomen um "natürliche" Abbauprozesse handelt, die von der Reifungstheorie nicht ausgeschlossen würden. Auf unsere Daten angewandt würde dieser Einwand jedoch bedeuten, dass moralische Entwicklung bereits bei 14 bis 16-jährigen ihren biologischen Höhepunkt erreicht hätte und dann nur noch abnehmen könnte. Alle bisherigen Untersuchungen sprechen jedoch gegen diese Annahme. Entwicklungszunahmen werden inzwischen bis ins hohe Erwachsenenalter nachgewiesen (Kohlberg 1984; Rest 1986b). Ein weiterer Einwand könnte daher lauten, dass es sich bei berufstätigen oder in der Berufsausbildung stehenden Jugendlichen um eine besonders selegierte Bevölkerungsgruppe handelt, bei der die moralische Reifung atypisch verläuft. Dieser Einwand ist auf der Basis der bisher diskutierten Daten nicht ganz von der Hand zu weisen, wenngleich er bereits eine starke Einschränkung der Emergenzthese darstellt. Streng genommen läßt diese Theorie, wie Kohlberg zu Recht betonte, Regressionen nur in ganz wenigen Ausnahmen zu. Die Emergenztheoretiker müßten, um gehaltvolle Vorhersagen zu treffen, die Menschen spezifizieren können, bei denen ein frühzeitiger Abbau der moralischen Urteilsfähigkeit eintritt, und darüber hinaus erklären können, weshalb diese Regressionen gerade bei Menschen auftreten, die ihre allgemeine Bildung bereits abgeschlossen haben. Beides ist bislang nicht geschehen. Schließlich könnte eingewendet werden, dass das Regressionsphänomen auf einen "Deckeneffekt" zurückgeht, wie wir ihn in den MJI-Daten angedeutet fanden (Kapitel 3), dass es sich also um ein methodisches Artefakt handelt. Wir haben bereits in Kapitel 5 gezeigt, dass der MUT selbst noch bei

Universitätsstudenten sensitiv für Entwicklungsänderungen ist, so dass auch dieser Einwand nicht stichhaltig ist.[158]
Wir können daher vorläufig bilanzieren, dass die Emergenzthese durch unsere Untersuchung widerlegt worden ist. Selektions-, Kohorten-, Sampling und Deckeneffekte müssen jedoch als ernstzunehmende Erklärungen für bestimmte Phänomene weiterhin in Betracht gezogen werden, auch wenn diese keine Erklärung bieten, die die Emergenzthese retten könnte. Das diese These widerlegende Regressionsphänomen wird durch die genannten Effekte nicht in Frage gestellt. Wir werden uns daher im empirischen Teil dieses Kapitels stärker auf die Abgrenzung der Bildungsthese von der Sozialisations- und vor allem der Selektionsthese konzentrieren.

11.3 Ergebnisse

Die für unsere Argumentation verfügbaren Befunde werden im folgenden mit Verweisen auf die Quellen zusammenfassend referiert.

11.3.1 Mit dem Abitur ist die Entwicklung nicht abgeschlossen

Wir wollen zunächst klären, ob sich Bildungseffekte mit dem hier hauptsächlich verwendeten Meßinstrument, dem MUT, auch noch im frühen Erwachsenenalter nachweisen lassen. Wir haben diese Frage bereits Ende der siebziger Jahre anhand des Vergleichs von Abiturienten mit Wissenschaftlern verschiedener Fachrichtungen untersucht. Abb. 11.1 zeigt die Befunde hinsichtlich des strukturellen C-Wertes (MUT).[159] Abiturienten orientieren sich in ihrem Urteils-

158 Dieser Einwand ist damit nicht "erledigt". Wir müssen bei allen hier besprochenen Meßmethoden mit einem Deckeneffekt rechnen, da sie sich sämtlich auf Kohlbergs Modell der Moralentwicklung beziehen, welches ursprünglich für die Phase der mittleren und späten Adoleszenz (14 bis 20 Jahre) ersonnen wurde.

159 Beim Vergleich dieser Meßwerte mit anderen Untersuchungen ist zu beachten, daß hier eine ältere Version des MUT benutzt wurde, der wegen einigen inadäquat formulierten Items generell etwas niedrigere C-Werte (früher "Determination Stufe" genannt) erbrachte. In allen anderen Untersuchungen wurde die 2. Version des MUT aus dem Jahr 1977 benutzt, die derzeitig noch gültig ist.

verhalten noch deutlich weniger an moralischen Normen als die wissenschaftlich arbeitenden Hochschulabsolventen (C = 27,1 versus 38,8). In Abb. 11.1 sind auch die Werte für den Wert "Faktor Pro-Contra" angegeben, der besagt, in welchem Maß sich die Befragten bei der Beurteilung der Argumente im MUT an deren *Meinungskonformität* orientierten. Hier sehen wir, dass Abiturienten sich zwar wesentlich weniger an der Meinungskonformität als an moralischen Prinzipien orientieren, dass sie dies aber mehr tun als die Wissenschaftler. Die von Portele (1985) untersuchten Wissenschaftler (vom Assistenten bis zum Lehrstuhlinhaber) lösen sich bei der Beurteilung von moralischen Argumenten fast völlig von ihrer eigenen Meinung. Andere Personengruppen, zum Beispiel die von Wischka (1982) untersuchten Strafgefangenen, orientierten sich bei dieser Aufgabe überwiegend, zum Teil mehr als 50%, an ihrer eigenen Meinung zu dem Problem, über das argumentiert wurde.

11.3.2 Moralentwicklung korreliert systematisch mit Bildungserfahrungen

Ein mehr systematischer Hinweis auf die Rolle von Bildungserfahrungen für die Entwicklung der moralischen Urteilsfähigkeit kann einem Vergleich entnommen werden, in dem mehrere Untersuchungsgruppen recht grob nach dem Grad des Bildungsniveaus geordnet wurden, das sie repräsentieren. Das untere Ende der Bildungsskala bilden die bereits erwähnten Straftäter, die hier nochmals nach psychiatrisierten und nichtpsychiatrisierten Personen getrennt wurden (Wischka 1982). Den mittleren Bereich repräsentieren die Gymnasiasten der Mittel- und der Oberstufe. Wir haben die Lehrlinge aus den Studien von Heidbrink (1985) und von Oser (s. Lind 1989b) zwischen die Mittel- und Oberstufenschüler am Gymnasium eingereiht, da sie dem Alter nach den Oberstufenschülern entsprechen, aber in der Regel weniger allgemeinbildende Schulung erhalten.[160] Abb. 11.2 enthält diese bildungsbezogene Vergleichsanordnung und die dazugehörigen C-Wert des MUT. Sie zeigt, dass eine deutliche Beziehung zwischen dem Ausmaß der Bildungserfahrungen und der gemessenen moralischen Urteilsfähigkeit besteht. Dieses Ergebnis stützt sich auf insgesamt ca. 2000 untersuchte Personen, also auf ein weit größeres Sample als

160 Die Reihenfolge der beiden Lehrlingsgruppen soll keine Bildungsunterschiede andeuten.

bislang mit dem MJI untersucht werden konnte. Gleichwohl stammen die Vergleichswerte aus verschiedenen Studien, die unter unterschiedlichen Bedingungen und zu jeweils anderen Zeiten durchgeführt wurden, so dass der Einwand, es handele sich hier zumindest zum Teil um Kohorten- oder Stichprobeneffekte mit diesen Daten nicht ausgeräumt werden kann. Dafür werden, wie bereits erwähnt, Längsschnittdaten benötigt. Solche Daten wurden bereits im Rahmen der Diskussion des Regressionsphänomens dargestellt (Kapitel 5). Sie zeigen, dass sich die moralische Urteilsfähigkeit auch noch nach dem Abitur, im ersten Studienabschnitt, zwischen dem ersten und fünften Semester deutlich erhöht.

In Abb. 11.3 werden die MUT-Daten aus verschiedenen Quellen (vor allem aus dem gesamten Längsschnitt vom ersten bis zum dreizehnten Semester, also teilweise bis nach dem Studium) zusammen dargestellt und mit den Befunden aus Studien verglichen, die sich anderer Verfahren zur Messung der moralisch-kognitiven Entwicklung bedienten. Wenn man den Studierenden-Längsschnitt, der sich auf die Werte jener Befragten beschränkt, für die über den gesamten Erhebungszeitraum vollständige MUT-Daten vorliegen (wodurch gewährleistet ist, dass keine selektiven Dropout-Effekte das Ergebnis verzerren, dann zeigt sich auch hier eine kontinuierliche Entwicklung der moralischen Urteilsfähigkeit. Interessanterweise ist diese Entwicklung im zweiten Studienabschnitt stärker als im ersten. Der Übergang in den Beruf hat im allgemeinen also keinen hemmenden Einfluß auf die Moralentwicklung, wie man aufgrund der Sozialisationstheorie vielleicht erwarten könnte (das sogenannte “Praxisschock-Phänomen”). Aus diesem Bild fallen, wie neuere Auswertungen zeigten. die Medizin-Studenten heraus. Ihr Studium, so scheint es, trägt nicht zur Entwicklung moralischer Fähigkeiten bei (Lind, 2000a).

11.3.3 Kohorten- und Selektionseffekte bezüglich der moralischen Urteilsfähigkeit

Hinweise auf Kohorten- und Selektionseffekte finden sich nur vereinzelt in der Literatur (siehe Lind 1985 a; 1985 b). Es gab bislang keine systematische Untersuchung zu diesem Thema. Mit der kombinierten Quer- und Längsschnittstudie von Grabow (1990) liegt erstmals eine systematische Datenbasis vor, auch wenn die Fragestellung dieses Autors nicht auf unser Thema zielte.

In einem Teil seiner Studie analysierte Grabow (1990) mit dem MUT das moralische Urteilsverhalten von Fachschülerinnen der Richtung Sozialpädagogik. Die Untersuchungsanlage enthält vier verschiedene "Kohorten" bzw. Generationen, nämlich Teilnehmerinnen des "Vorpraktikums", Schülerinnen der Unter- und der Oberstufe, und Teilnehmerinnen des einjährigen Berufspraktikums (= Querschnitt). Jede dieser Kohorte wurde zweimal im Abstand von einem halben Jahr untersucht (= Längsschnitt). Damit ist es möglich, bildungsbedingte Veränderungen im Urteilsniveau von kohortenbedingten zu unterscheiden. Diese Teilstudie belegt sehr deutlich den Bildungseffekt. Es zeigt sich, dass im Längsschnitteil in allen vier Kohorten eine Entwicklung der moralischen Urteilsfähigkeit zu verzeichnen ist. Nimmt man hingegen den Querschnitteil der Studie, dann zeigt sich ein Kohorten-Effekt: die Berufspraktikantinnen haben etwas niedrigere C-Werte als die Oberstufenschülerinnen. Hätte man die Bildungsthese allein am Querschnitt untersucht, wäre man – fälschlicherweise – zu dem Schluß gelangt, dass das Berufspraktikum einen hemmenden Einfluß auf die moralische Entwicklung ausübt.

Die Studie von Grabow eignet sich auch zur Prüfung von Selektionseffekten. In dieser reinen Querschnittstudie wurden Angehörige verschiedener Ausbildungsgänge im ersten und im zweiten Ausbildungsjahr befragt. Hier können also Selektions-, aber keine Kohorten-Effekte analysiert werden. Wie Abb. 11.4 zeigt, ergeben sich für alle drei von Grabow untersuchten Fachrichtungen Bildungsgewinne. Aber es ergeben sich auch unübersehbare Hinweise für Selektionseffekte. Da uns entsprechende Informationen fehlen, läßt sich nicht weiter untersuchen, ob diese Selektionseffekte ihrerseits auf unterschiedliche Bildungserfahrungen zurückgeführt werden können. Aber wie wir schon in anderen Studien gefunden haben (Lind 1985a), können Selektionsprozesse tatsächlich in Zusammenhang mit der moralischen Urteilsfähigkeit stehen. Die Gründe für diesen Zusammenhang liegen aber noch weitgehend im Dunkeln. Es kann sich dabei um eine Art "Selbstselektion" der Schüler handeln, indem die moralische Urteilsfähigkeit das Berufswahlverhalten beeinflußt, oder um eine "Fremdselektion", indem die Fähigkeit über die Schulnoten implizit mit beurteilt wird.

Auch für diese These gab es bis vor kurzem kaum Belege. Mit der Studie von Niemczynski et al. (1988) bei Erwachsenen verschiedener Altersgruppen von 20 bis 80 in Polen liegen *MJI*-Daten getrennt nach Bildungsniveau und Geschlecht vor. Die Befunde dieser Forscher belegen zunächst noch einmal eindrucksvoll, dass das Alter *per se* keine Moralentwicklung bewirkt. Es ergaben sich keine Hinweise für eine Höherentwicklung im mittleren und späten Erwachsenenalter, was aufgrund der Emergenzthese, aber nicht aufgrund der Bildungsthese zu erwarten gewesen wäre.[161] Es zeigen sich aber selbst im hohen Alter von 70 bis 80 Jahren noch sehr deutlich Bildungsunterschiede. In der Gruppe der Männer weisen die Befragten mit einer Universitätsausbildung eine um 20 bis 30 MMS-Punkte höhere Urteilsfähigkeit auf als die Befragten, die nur einen Gymnasialabschluß haben. Die Tatsache, dass sich bildungsbezogene Unterschiede in den MJI-Werten noch im hohen Alter nachweisen lassen, kann aber schwerlich allein auf die direkte Wirkung der Bildungserfahrung zurückgeführt werden. Wie wir in Kapitel 5 sahen, scheint sich die moralische Urteilsfähigkeit nach dem Ende der institutionalisierten Bildung zurückzubilden. Es ist aber möglich, dass die frühere Bildung einen *indirekten* Effekt hat, nämlich dass sie Einfluß auf berufliche und soziale Positionen einer Person hat, die wiederum durch ihre spezifischen Anforderungen die Urteilsfähigkeit unterschiedlich stark fordern und fördern.

Es könnte sein, dass diese Erklärung auch auf das Phänomen zutrifft, auf das die Abb. 11.7 hinweist: In der Gruppe der Frauen zeigt sich, neben dem besagten Bildungseffekt, auch eine Regression. Die von Niemczynski et al. (1988) untersuchten Frauen weisen in der Altersgruppe von 20 bis 35 dieselbe Höhe der moralischen Urteilsfähigkeit auf wie die Männer – ein Befund, der sich mit der bisherigen Forschungslage deckt (vgl. dazu Lind et al. 1987). Die höheren Altersgruppen weisen aber eine deutlich niedrigere Urteilsfähigkeit auf. Obwohl es sich hier um eine Querschnittstudie handelt, ist dieses Phänomen nur schwer als Kohorten-Effekt zu erklären. Denn dann müßte sich eine solche Abnahme auch bei den Männer zeigen. Aufgrund der Bildungsthese könnte dieses Phänomen dadurch erklärt werden, dass die meisten Frauen im mittleren

161 Die Graphiken wurden auf der Basis der Zahlen erstellt, die von den Autoren berichtet werden.

und höheren Alter ihre berufliche Karriere zugunsten der Familie aufgegeben haben und dadurch weit weniger Möglichkeiten für diskursive Prozesse halten als Männer, die in gehobenen Stellungen oder in politischen Zirkeln mehr Gelegenheit dazu geboten bekamen. Aber dies ist, wie gesagt, eine bloße Schlußfolgerung aus einer theoretischen Prämisse, die noch der empirischen Überprüfung bedarf.

11.4 Resümee

Eine weitverbreitete Auffassung besagt, dass sich Bildung hauptsächlich oder nur auf die Entwicklung von fachlichen Fertigkeiten auswirkt, und dass für die Entwicklung der Moral die Sozialisation im vor- und außerschulischen Bereich verantwortlich ist. Mit den hier vorgelegten empirischen Studien und Experimenten konnte diese allgemeine Auffassung eindeutig *widerlegt* werden. Es konnte gezeigt werden, dass Bildung nicht nur mit der Moralentwicklung korreliert, sondern dass sie dafür eine *ursächliche* Bedingung darstellt. *Institutionalisierte Bildungsprozesse,* so die Schlußfolgerung aus allen in dieser Arbeit vorgelegten Befunden, *sind eine notwendige und hinreichende Bedingung für die Entwicklung der moralischen Urteilsfähigkeit beim einzelnen.*
Diese Schlußfolgerung basiert zum einen auf einer ungewöhnlich hohen Korrelation zwischen der Moralentwicklung und dem Umfang der Bildungserfahrungen, die von allen Studien bestätigt wird, die – wenigstens teilweise – den Fähigkeitsaspekt des moralischen Verhaltens erfaßt haben. Sie basiert aber auch auf der Widerlegung einer Reihe von Alternativerklärungen, die für den korrelativen Befund angeboten worden sind. Aufgrund der in dieser Arbeit vorgelegten Daten können die beiden am häufigsten genannten Alternativen, die Sozialisations- und die Emergenzthese, als eindeutig widerlegt gelten. *Moralische Urteilsfähigkeit wird nicht durch sozialen Anpassungsdruck (Indoktrination oder dergleichen) gefördert. Auch kann aufgrund der umfassenden Befundlage ausgeschlossen werden, dass sich die kognitive Seite der Moral von alleine, rein altersbedingt entwickelt.*

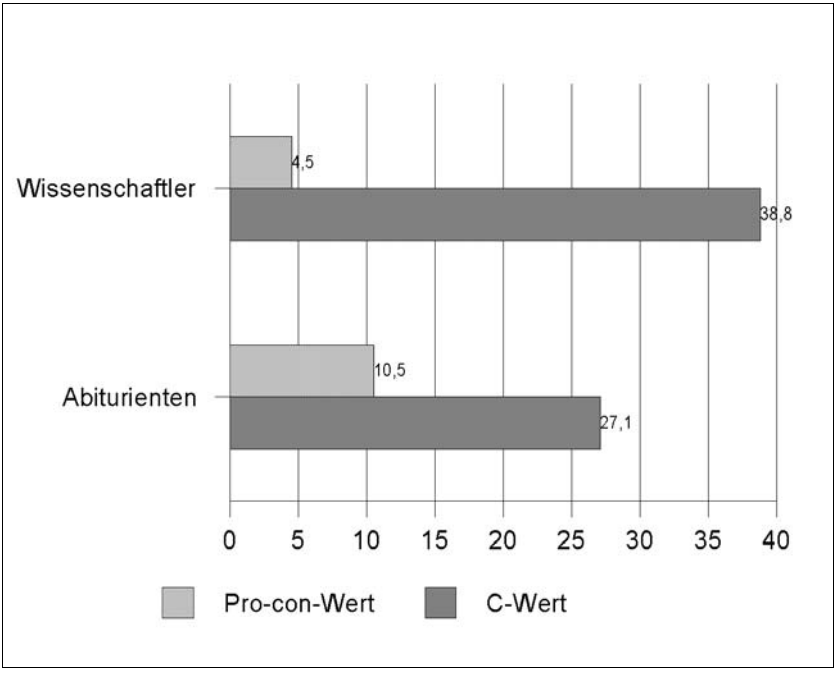


Abb. 11.1 Moralische Urteilsfähigkeit bei Wissenschaftlern und Abiturienten. Beide Erhebungen mit der MUT-Version von 1976. Quellen: Wissenschaftler-Studie von Portele (1983), N=48. Abiturienten-Studie des FORM-Projekts Lind (1985a), N=516

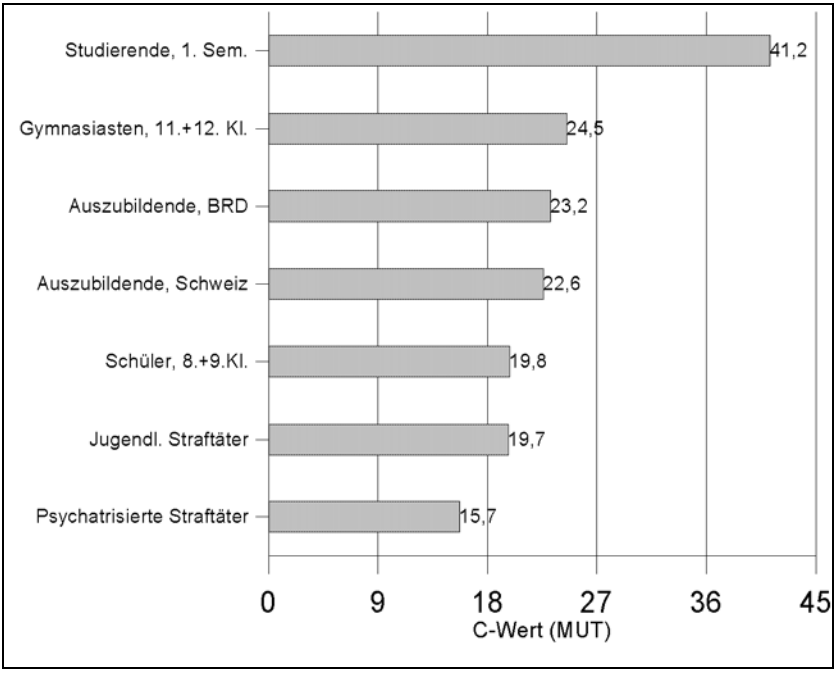


Abb. 11.2 Moralische Urteilsfähigkeit (C-Werte, MUT Version 1977) bei Gruppen mit unterschiedlichen Bildungserfahrungen

Erläuterungen zur Graphik:

Gruppe	Gruppengröße, Erhebungsjahr, Quelle
Studierende, 1.Sem.:	n= 1288, 1978, FORM-Projekt*
Gymnasiasten, 11.+12.Kl.:	n= 42, 1987, Link & Lind (1988)
Auszubildende, Schweiz:	n= 579, 1983, HASMU-Projekt (Lind 1989b)
Auszubildende, BRD:	n= 256, 1978, Heidbrink (1985)*
Gymnasiasten, 8.+9.Kl.:	n= 87, 1987, Link & Lind (1988)
Jugendliche Straftäter:	n= 26, 1981, Wischka (1982)
Psychiatrisierte Straftäter:	n= 15, 1981, Wischka (1982)

Meßinstrument: MUT, Version 2 (1977), 5er Antwortskala, * = 9er Antwortskala.

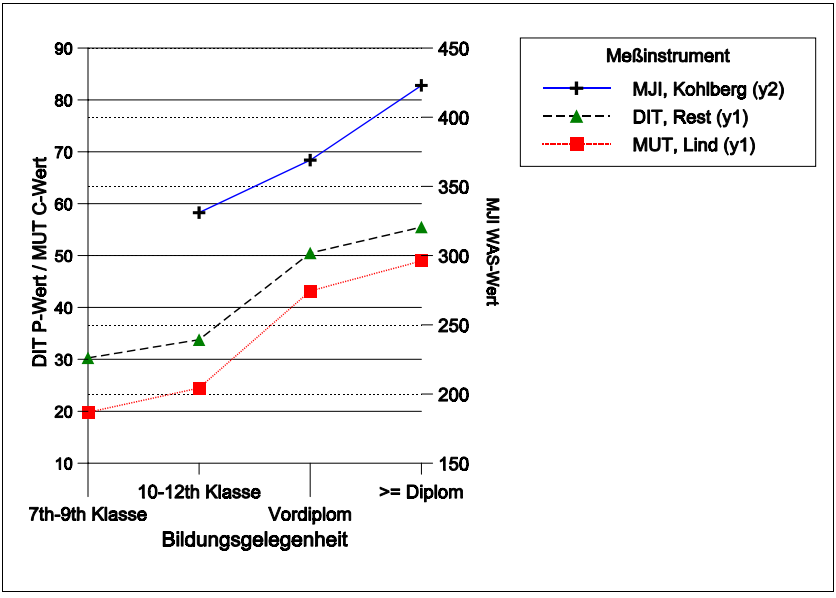


Abb. 11.3 Moralisch-kognitive Entwicklung in Abhängigkeit von Menge und Qualität der Bildungsgelegenheiten mit drei verschiedenen Verfahren gemessen (Quellen für MJI-Studie: Colby & Kohlberg et al., 1987, S. 109; für DIT-Studie: Rest, 1979; für MUT-Studie: Lind, 2002)

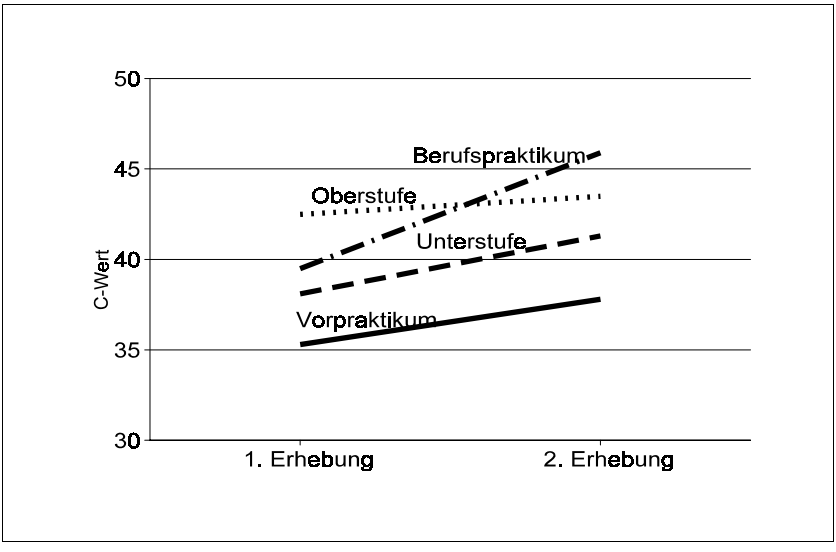


Abb. 11.4 Moralische Urteilsfähigkeit (C-Wert, MUT) bei Fachschülerinnen der Sozialpädagogik mit unterschiedlichem Bildungsniveau. Quelle: Grabow (190).

Erläuterung zur Graphik:

Der Abstand zwischen 1. und 2. Erhebung betrug jeweils ½ Jahr.

Gruppen:

Vorpraktikum	n= 27 (1. Erhebung), n= 25 (2.)
Unterstufe	n= 60 (1.), n= 58 (2.)
Oberstufe	n= 59 (1.), n= 59 (2.)
Berufspraktikum	n= 32 (1. Halbjahr), n= 29 (2. Halbjahr)

Beim „Berufspraktikum" handelt es sich um verschiedene Personen, bei den anderen Gruppen um follow-up-Erhebungen bei denselben Personen.
Meßinstrument: MUT, Version 1977, 5er Antwortskala.

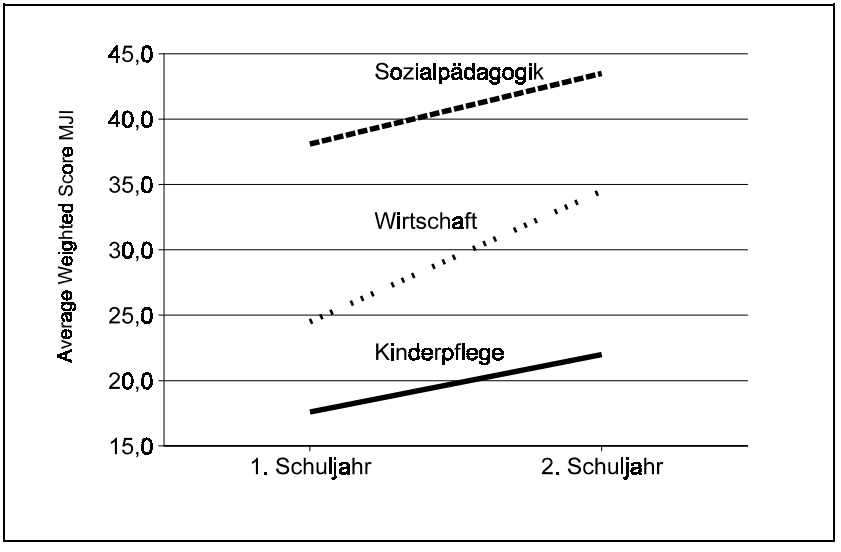


Abb. 11.5 Moralische Urteilsfähigkeit (C-Wert, MUT) in Abhängigkeit von Bildungsdauer (Längsschnitt) und Selektion nach Fachgebiet. Quelle: Grabow (1990)

Erläuterungen zur Graphik:

Gruppen:

FS Sozialpädagogik (Anfang)	n= 33 (Vorpraktikum, Erhebung 0)
FS Sozialpädagogik (Ende)	n= 61 (Berufspraktikum, Erhebung 2)
Berufsfachschule Wirtschaft (Anfang)	n= 27 (Unterstufe)
Berufsfachschule Wirtschaft (Ende)	n= 26 (Oberstufe)
Kinderpflegerinnen (Anfang)	n= 17 (Unterstufe)
Kinderpflegerinnen (Ende)	n= 11 (Oberstufe)

Meßinstrument: MUT, Version 1977, 5er Antwortskala.

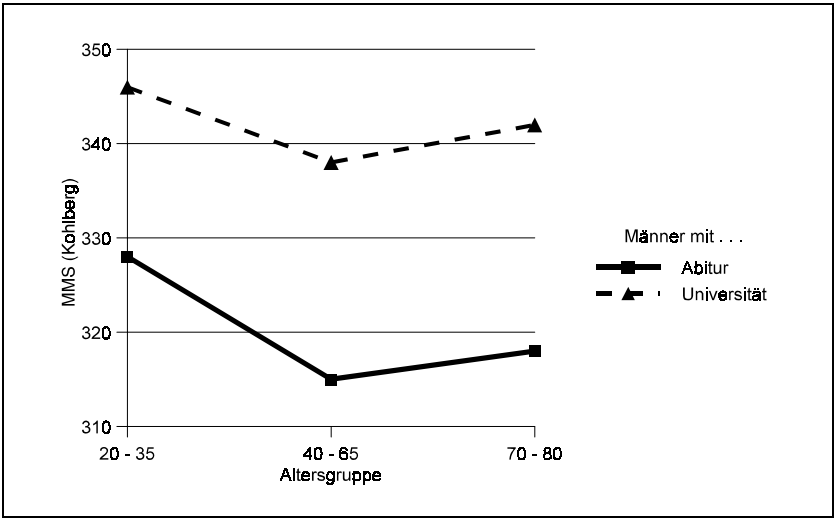


Abb. 11.6 Moralische Urteilsfähigkeit (Kohlbergs MJI) bei *Männern* mit verschiedenen Schulabschlüssen und Alter (n=130, Polen; Quelle; Niemczynski et al., 1988)

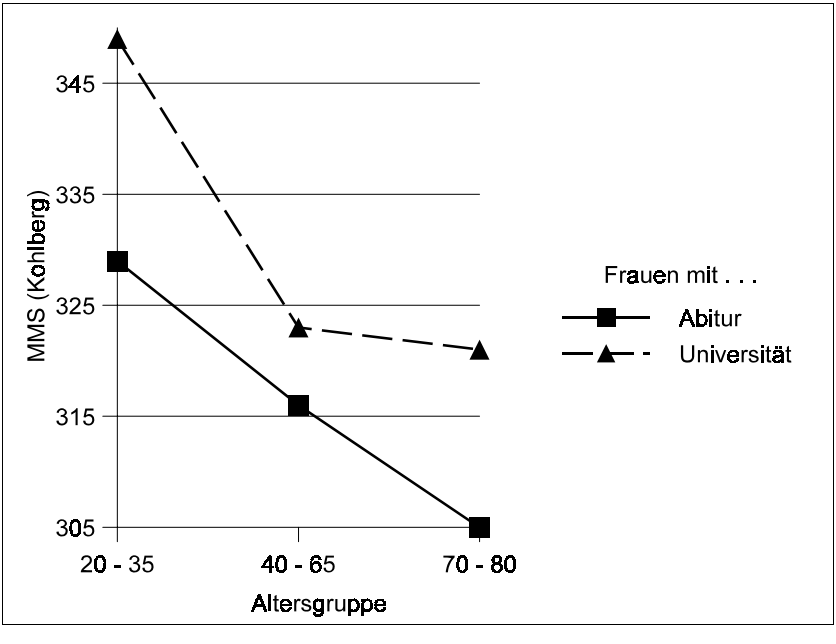


Abb. 11.7 Moralische Urteilsfähigkeit (MJI) bei Frauen in Abhängigkeit von Bildungsniveau und Alter (N = 161 Polen). Quelle: Niemczynski et al. 1988

12 Schlußfolgerungen

Theodor W. Adornos (1973) eindringliches Plädoyer für eine Erziehung, damit "Auschwitz nicht noch einmal sei", stand Pate bei der Entstehung dieser Untersuchung vor fast zwanzig Jahren. Adorno nahm im gleichen Aufsatz *Erziehung nach Auschwitz* auch eine Folgerung aus dieser Untersuchung vorweg: Nur Menschen, die autonom denken und urteilen gelernt haben, deren Mut zur eigenen Vernunft gefördert wurde, können der Barbarei widerstehen.

I.

Noch sehr eindringlich sind mir die Bilder von Auschwitz, Treblinka und anderen Orten des Menschenmordens in Erinnerung, die uns in der Schule gezeigt wurden und die mich, ein Kind der Tätergeneration, seitdem mit der Frage bewegen: Wie hat es dazu kommen können? Was kann man, was können *wir* tun, damit dies nie und nirgendwo mehr passieren möge? Die überall im Land auflodernden Fanale der Unvernunft, des Hasses und der Gewalt erinnern uns jetzt daran, dass es Auschwitz immer wieder geben kann. Der Auftrag der Erziehung ist nie erfüllt, er muß immer auf's Neue definiert und zu erfüllen gesucht werden.

Nach dem Schrecken des Zweiten Weltkrieges, nach dem kläglichen Zusammenbruch der Weimarer Demokratie und nach der Terrorherrschaft der Nazis, haben Philosophen wie Karl Jaspers, Pädagogen wie Heinrich Roth und Politiker aller Parteien den Schluß gezogen, dass auf Dauer nur enorme Anstrengungen in Erziehung und Bildung die Demokratie sichern können, damit nicht nur der Wohlstand erhalten bleibt, sondern auch Friede und soziale Gerechtigkeit zur stabilen Grundlage des Zusammenlebens werden. In einer demokratischen Verfassung geht die Macht vom Volke aus. Regierung und Politik können nur so gut sein, wie die demokratische Kompetenz der Bürger, die diese kontrollieren und personell beschicken sollen.

Die Einsichtigen damals sahen Bildung nicht nur als individuelles Investitionsgut, sondern auch als gesellschaftliches. Bildung war ihnen nicht nur liberales "Bürgerrecht" (Dahrendorf), sondern auch demokratische Bürgerpflicht und sollte, wie schon Tocqueville (1976/1835) aufgrund seiner exzellenten Analyse der jungen *Demokratie in Amerika* erkannte, für die Gesellschaftspolitik höchs-

te Priorität haben: “Die Freiheit, der öffentliche Friede und selbst die soziale Ordnung [können] ohne Bildung nicht mehr bestehen” (S. 613). Das gilt für Industrienationen ebenso wie für die Länder der Dritten Welt. Bildung ist, wie der Philosoph Dieter Böhler (1992) es formulierte, tätige *Zukunftsverantwortung*.

II.

Fast fünfzig Jahre lang hat sich die Struktur der westdeutschen Demokratie als stabil erwiesen, nicht zuletzt deshalb, weil in der Bundesrepublik tatsächlich eine enorme Verbreiterung des Bildungswesens stattgefunden hat. Ohne die Verlängerung der Schulzeit für alle, ohne die Aufhebung der konfessionellen Segregation der Schulen (ich erinnere mich noch gut an den konfessionell geteilten Pausenhof an meiner Grundschule und die oft heftigen Kämpfe zwischen protestantischen und katholischen Jugendbanden), ohne die Öffnung des Gymnasiums für breite Bevölkerungsschichten und ohne die Vervielfachung der Studentenzahlen hätte unsere Demokratie unter der wachsenden Last von Verteilungskonflikten, Umweltzerstörungen, politischer Korruption, Wettrüstung usw. vermutlich kaum standhalten können. Die Wohlfahrt aller hat davon profitiert, dass unsere Schulen und Hochschulen eine immer breitere Schicht von Bürgern und Bürgerinnen dazu befähigt, die sozialen, politischen und moralischen Probleme des Lebens mit Vernunft und ohne Gewalt zu bewältigen. Auch unsere Wirtschaft verdankt ihren Erfolg dem sozialen Frieden, vielleicht mehr als umgekehrt.

Doch einiges ist auch falsch gelaufen in dieser Zeit, sonst würden nicht so viele zweifeln an den Kernsätzen unserer Verfassung, und es gäbe nicht die Hitze der Asyldiskussion und die daraus entstandenen Brände in den Asylantenheimen. In der Öffentlichkeit gilt Bildung heute nicht mehr viel. Es scheint, dass mit der Distanz zur Nazi-Katastrophe die Zahl der Einsichtigen abnimmt und die Versuchung immer größer wird, unser Bildungssystem zu vernachlässigen, es gar gegenüber anderen Bereichen des Staates als Luxus anzusehen, der in Zeiten der ökonomischen Krise Einsparungsmöglichkeiten bietet. Das ist nicht nur bei uns so. In den USA und England wurde der Rückzug aus der bildungspolitischen Verantwortung in den achtziger Jahren zum politischen Programm erhoben. Dieser Rückzug wird verbrämt durch das Schlagwort von “Bildung als individueller Karriere-Investition”. Bildung gilt nur noch als Waffe des Einzel-

nen für seinen Überlebenskampf, wofür er oder sie auch allein zu zahlen hat. Diese Politik ohne Zukunftsbezug zeitigt heute verheerende Folgen für viele Bürger dieser Länder. Wohlstand, Gerechtigkeit und sozialer Frieden, so kritisierte der Nationalökonom John Kenneth Galbraith, sind ohne ein gut funktionierendes Bildungssystem nicht möglich.

Aber auch bei uns sank in den letzten Jahren der Wert von Bildung und Erziehung. Das zeigte sich nicht nur am Schrumpfen der Finanzmittel für Bildung, sondern auch an der Erstarrung der Bildungspolitik, der Fehlleitung von Mitteln in eine einseitige, technikorientierten Bildung und, schließlich, auch an dem hilflosen Rufen nach neuen oder alten Identitäten und Werten. Die *Erstarrung der Bildungspolitik* angesichts großer sozialer Dynamik hat zu teilweise chaotischen Zuständen im Bildungssystem geführt. Ich denke an die in Unordnung geratene Schullandschaft mit ihren großen, geographisch bedingten Disparitäten der Schulversorgung (Bargel & Kuthe 1992), den gegenwärtigen (1993; redaktionelle Anmerkung GL) Fehlbestand von 300.000 Studienplätzen an den Hochschulen und Universitäten und an die Tatsache, dass das Gymnasium vielerorts mit einer Aufnahme von über 50% eines Altersjahrgangs de facto zur Gesamtschule geworden ist, obwohl weder ihr Fächerkanon noch die pädagogikferne Ausbildung ihrer Lehrer dafür taugen.

Es ist aber vor allem die *einseitige Förderung technisch-wirtschaftlicher Bildung*, die unsere Gesellschaft instabil macht, indem sie immer mehr die Kluft verbreitert zwischen innovativen Eliten und dem Rest der Bevölkerung, zwischen hochindustrialisierten Ländern und dem Rest der Welt. Sie heizt den immer schnelleren ökonomisch-technologischen Wandel unserer Gesellschaft an (Rationalisierung, Atomenergie, Kommunikationstechnologien, Computer, Gentechnik) und vergrößert viele Probleme (Überbewaffnung, Bürgerkriege, Umweltbelastung, Arbeitslosigkeit, Ohnmachtsgefühle, Kriminalität), statt sie bewältigbar zu machen. Der geförderte Teil der Bevölkerung entwickelt immer komplexere Techniken, der andere, größere, vernachlässigte Teil hat immer mehr Schwierigkeiten, die Veränderungen wahrzunehmen und zu verstehen, sie zu bewältigen und zu nutzen. Es wäre aber wichtig, diese Kluft zu verringern, indem einerseits Bildung und Erziehung für die gesamte Bevölkerung hier und in den Ländern der Zweiten und Dritten Welt verbessert würden, und andererseits, indem im gleichen Umfang wie die technischen auch die moralisch-demokratischen Kompetenzen gefördert würden. Notwendig wäre eine Art “SBI”, eine strategische Bildungsinitiative, die dafür sorgt, dass Bildung und

Erziehung gleichmäßiger verteilt und qualitativ weiterentwickelt werden. Nur die breite Förderung von Beurteilungs- und Mitbestimmungsfähigkeiten der Bevölkerung stärkt die Fähigkeit einer demokratischen Gesellschaft, neue Technologien zu verantworten. Elitäre Ethikkommissionen sind kein Ersatz für den verfassungsgemäßen politischen Ethikdiskurs mündiger, autonomer Bürger und Bürgerinnen. Wird die Kluft zwischen technischen Eliten und Bürgern unüberbrückbar, bleibt nur dumpfe Resignation (die da oben werden schon wissen, was sie machen) oder ebenso dumpfe Gewaltbereitschaft, hoch empfänglich für bequeme Feindbilder und aggressive Botschaften.

Angesichts wachsender Gewalt rufen viele, Laien und Experten, Politiker und Sozialforscher, nach *alten oder neuen Identitäten und Werten*, die unsere Schulen vermitteln sollten. Aber dieser Ruf geht an dem Problem vorbei. Die Reportagen über Rechtsradikale und Skins zeigen fast alle, dass es diesen offensichtlich nicht an Identifikation mit Werten (Heimat, Nation, Gerechtigkeit, Gemeinschaft usw.) mangelt. Nachfragen lassen aber erkennen, dass diese Werte ihnen seltsam fremd sind, dass es den gewaltgeneigten Jugendlichen an der Fähigkeit mangelt, den Sinn solcher Werte zu verstehen und sie differenziert und integriert anzuwenden. Mit Adornos Worten, es mangelt ihnen an der "Kraft zur eigenen Reflexion, zum unabhängigen Denken, zum Nicht-Mitmachen". Es ist leicht, sich irgendwelche "Identitäten" oder Orientierungen anzueignen, aber schwer, diese im praktischen Leben anzuwenden und die dabei auftretenden, unausweichlichen Konflikte konstruktiv zu lösen. Hierbei, so zeigt die moderne Moralforschung, benötigen die Heranwachsenden gezielte Förderung durch Bildung und Erziehung. Aber auch wir neigen immer mehr zu bequemen Lösungen. Lieber, so scheint es, füttern wir unsere Kinder, je nach politischem Standpunkt, mit alten oder neuen Werten ab, statt unsere Gedanken daran zu geben, wie wir sie bei dem schwierigen Prozeß des Mündigwerdens unterstützen könnten.

III.

In den Siebziger Jahren wurden wir Zeuge einer Renaissance von Erziehung und Bildung, aber auch eines großen Streits über deren Grundlagen und Ziele. Der Streit darüber, ob Bildung wirklich zur Demokratie beiträgt oder nicht, hat nicht wenig zu Resignation und Stillstand in der Bildungspolitik beigetragen.

Zu denen, die ihren Überzeugungen aus vor-demokratischer Zeit treu bleiben, dass Bildung nur für wenige da sei, gesellten sich bald Progressive, die in jedem Bildungssystem eine Gefahr für die individuelle Freiheit und die soziale Gerechtigkeit sahen. Die Schule, so klagten sie, verfestige die Ungleichheit der Menschen, indem sie ungleiche soziale Chancen perpetuiere, und mit ihrem Curriculum offen oder versteckt alte, undemokratische Normen und Werte vermittle. Selbst neue Lerninhalte und neue Schulformen wie die Gesamtschule könnten daran nichts grundsätzlich ändern.

Das Vorverständnis der Siebziger Jahre ließ keinen Platz für die Vorstellung von Moral als einer *Fähigkeit*. Moralisch-demokratische Fähigkeiten wurden den "affektiven Erziehungszielen" zugeschlagen und rein als Frage der richtigen Einstellung definiert. Man glaubte allgemein, es genüge, um "gut" zu handeln, immer das Moralische und die Demokratie zu *wollen*. Konservative und Progressive stritten nur darum, *welche* Werte vermittelt werden sollten. Die Konservativen wollten traditionelle Werte vermittelt sehen ("Mut zur Erziehung"). Sie glaubten, die Nazi-Herrschaft resultierte aus einem Mangel an bürgerlichen Tugenden wie Anstand und Pflichtbewußtsein. Progressive hatten zunächst Schwierigkeiten, eine Alternative zu benennen. Ihre Haltung reichte von der Ablehnung jedweder Moralerziehung ("Anti-Pädagogik") bis zur Erziehung zu neuen Werten ("neue Identität"). Gemeinsam war beiden Positionen, dass sie auf dem moralisch-demokratischen Gebiet nur Einstellungen und Werte walten sahen, auf die allein sie die Erziehung verpflichten wollten. Jede Erziehung aber, die nur auf die Vermittlung von Werten, ob alte oder neue, zielt, verfehlt den Auftrag demokratischer Bildung, jene Fähigkeiten zu fördern, die für einen fairen, gewaltfreien Interessenausgleich und eine gemeinsame, vernünftige Bewältigung sozio-moralischer Probleme notwendig sind; Fähigkeiten, die Bedingung dafür sind, dass die späteren Bürger staatliche Institutionen wie Parlament und Rechtsprechung mit Leben füllen und nutzen können, dass sie über eigene und fremde Interessen im Konfliktfall beraten und befinden können.

IV.

Unglücklicherweise folgte auch die damals auf breiter Front einsetzende empirische Bildungs- und Erziehungsforschung diesem nicht-kognitiven Vorver-

ständnis von Moral und erbrachte daher kaum handfeste Belege für demokratisierende Leistungen des Bildungssystems. Außer fachlichen, berufsbezogenen Fertigkeiten, so schien es, förderten Schulen und Hochschulen allenfalls moralische und demokratische *Einstellungen*. Zu den entsprechenden Kompetenzen konnte die Forschung nichts sagen.

Es bedurfte, um die Bedeutung und Entstehung moralischer Fähigkeiten zu verstehen, jener Revolution der Sichtweise, die eng mit den Namen Piaget und Kohlberg und der von ihnen ausgearbeiteten *kognitiven* Entwicklungstheorie verbunden ist. Dieser Ansatz brach mit der Vorstellung, dass Moralität allein eine Frage der richtigen, "guten" Einstellung ist. Er öffnete neue Wege der Erforschung der *kognitiven* Grundlage und Entwicklung moralischen Verhaltens. Bis dahin war gesicherte Erkenntnis, a) dass, entgegen alltäglicher Vorurteile, fast alle Menschen gute moralische Einstellungen haben (fast alle sind *gegen* Stehlen, Lügen, Morden ebenso wie *für* Gerechtigkeit, Ehrlichkeit, Helfen usw.), und b) dass sie aber oft nicht in Übereinstimmung mit diesen Einstellungen handelten. Piaget und Kohlberg haben gezeigt, dass dies eine natürliche Folge der Tatsache ist, dass zwischen moralischen Orientierungen und moralischem Verhalten komplexe *kognitive* Prozesse geschaltet sind, die dem Handelnden bestimmte Fragen beantworten helfen: Was genau bedeutet eine bestimmte moralische Norm, z.B. die Norm *Du sollst nicht stehlen*? Was ist eine *gerechte* Verteilung von Sachen? Wie löse ich ein moralisches Dilemma, wenn, wie immer ich mich entscheide, ein mir wichtiges Prinzip verletzt wird? Welches Prinzip ist wichtiger? Kann es Ausnahmen von Prinzipien geben, und wenn ja, unter welchen Bedingungen? Wie kann man mit einer anderen Person eine vernünftige Lösung eines Konflikts erreichen? Die Theorie von Piaget und Kohlberg erklärt uns, dass die Antworten auf solche Fragen je nach kognitivem Entwicklungsstand verschieden ausfallen können. Nur wer eine hohe moralische Urteilsfähigkeit besitzt, das heißt, nur wer fähig ist, seine vorgefaßten Meinungen auf der Grundlage moralischer Prinzipien zu reflektieren, viele Aspekte einer Situation zu berücksichtigen und die Meinung und die Argumente von Andersdenkenden zu achten, kann "gute" Entscheidungen treffen.

Der neue Denk- und Forschungsansatz von Piaget und Kohlberg hat uns die Schwachstellen der Lern- und Sozialisationstheorie der Moralentwicklung vor Augen geführt (Kapitel 6 und 7), und damit gezeigt, weshalb der bloße Ruf nach Werten keine erfolgversprechende pädagogische Methode sein kann (vgl. Kohlberg 1987; Oser & Althof 1992). Der Ansatz hat uns letztlich auch den

Nachweis ermöglicht, dass an der Entwicklung moralisch-demokratischer Kompetenzen Bildungsprozesse maßgeblich beteiligt sind. Aber dieser Nachweis war, wie in der hier vorgelegten Untersuchung ausführlich dargelegt wird, erst nach einer gründlichen Revision theoretischer und methodologischer Positionen des Piaget-Kohlberg-Ansatzes möglich (Kapitel 3 und 5).

V.

Der enge Zusammenhang von Moral, Bildung und Demokratie wird schon lange behauptet und mit Beispielen zu belegen versucht. Aber mit ebensolcher Überzeugung wird er auch oft in Abrede gestellt oder sogar ein Grundwiderspruch zwischen ihnen zu beweisen versucht. Erst die modernen Methoden experimenteller, sozialwissenschaftlicher Forschung haben es ermöglicht, diese Kontroverse objektiv und auf einer breiten Datenbasis zu behandeln und den empirischen Beweis für die Behauptung anzutreten, dass Bildungserfahrungen entscheidend für die Moralentwicklung sind.

Der *Moralisches Urteil-Test* (MUT) wurde ursprünglich entwickelt für den Einsatz in der international vergleichenden Längsschnittstudie "Hochschulsozialisation" (Kapitel 1). Als Mitglied der Forschungsgruppe hatte ich 1973/74 die Aufgabe erhalten, ein Instrument zur Messung der Moralentwicklung von Studierenden zu finden oder zu entwickeln, das die Prüfung von zentralen Hypothesen der Kohlberg-Theorie ermöglichen sollte, das vor allem den kognitiven oder Fähigkeitsaspekt zu messen erlaubt und das daneben möglichst einfach anzuwenden und auszuwerten ist. Die letzte Bedingung resultierte aus dem komplexen Design der Studie, das eine sehr umfangreiche, international zusammengesetzte Untersuchungsgruppe implizierte. Wir benötigten jedoch weniger eine neue, einfacher zu handhabende Meßtechnik als ein theoretisch valides Meßinstrument zur Überprüfung von Hypothesen. Daher sollte sich die Neukonstruktion eng an den Vorgaben von Piagets und Kohlbergs Theorie der Moralentwicklung orientieren. Die hierin enthaltene neue Sichtweise von Moral als einer (kognitiven) Fähigkeit schloß alle Meßmethoden aus, die eine Definition von Moral bloß als Einhaltung von sozialen Normen oder bloß als moralische ("gute", "richtige") Einstellungen voraussetzten. Die neue, "kognitive" Sichtweise, so stellte Kohlberg (1958, 1977, 1981) wiederholt fest, verlangt eine völlig neue psychologische Methodologie. Im Zusammenhang mit der Überprü-

fung der These von der kognitiv-affektiven Parallelität mußten Kohlberg und seine Kollegen (Kuhn et al. 1977) aber eingestehen, dass mit dem eigenen Instrument, dem *Moral Judgment Interview* (MJI), diese Forderung nicht befriedigend einlösbar ist. Obwohl sie von der Annahme ausgingen, dass der kognitive Aspekt des moralischen Urteilsverhaltens mit dem MJI gemessen werden kann, operationalisierten sie diesen Aspekt bei der Prüfung der Parallelitätsthese nicht mittels des MJI, sondern mittels eines Kognitionstests, der ohne moralischen Bezug war (Kapitel 2 und 3).

Der MUT wurde für die Beantwortung von Forschungsfragen und für Evaluationszwecke entwickelt. Für die Individualdiagnostik ist er nicht gedacht. Sollte er dafür eingesetzt werden, müßte er eigens dafür angepaßt und neu validiert werden. Unter anderem müßte man dem Problem mehr Beachtung schenken, dass unterschiedliche Dilemma-Typen verschiedene Ebenen der moralischen Reflexion oder Argumentation verlangen. Nicht jede Lösung von Verhaltenskonflikten bedarf einer Behandlung auf den höchsten Moralstufen. Nur moralische Konflikte im engeren Sinne erfordern dies. Wir haben das insofern berücksichtigt, als wir für den MUT zwei Dilemmas ausgewählt haben, von denen wir aufgrund von Vorstudien annehmen konnten, dass ihre Lösung postkonventionelle moralische Überlegungen erforderte. Wie alle Studien mit dem MUT übereinstimmend zeigten, sehen bei beiden Dilemmas die Befragten die postkonventionellen Stufen der Argumentation (Stufe 5 und 6) als die adäquatesten an (Kapitel 4; Lind 1985a). Wir haben den Eindruck, dass die sonst in der Forschung verwendeten Dilemmas nicht immer eine so hohe Reflexionsebene erforderten.

Wir hatten uns entschlossen, mit dem MUT direkt Urteilsverhalten in einer wirklichkeitsnahen Situation zu erfassen, statt von der befragten Person eine Auskunft darüber zu erbitten, wie sie sich in einer ähnlichen Dilemma-Situation, wie sie im Test geschildert wird, entscheiden *würde* und wie sie ihre Entscheidung begründen *würde*. Solche Auskünfte wären hypothetisch; wir müßten davon ausgehen, dass sie, wie wir aus der Persönlichkeitsforschung wissen, durch vielerlei Wahrnehmungseinflüsse der befragten Person (wie Halo-Effekte, Ankerreiz-Effekte, soziale Erwünschtheit usw.) gefiltert und verzerrt sind. Vorversuche mit einem Fragebogen, in dem die Befragten angeben sollten, mit welchen (von mehreren vorgegebenen) Argumente sie am ehesten ihre Entscheidung begründen würden, zeigten, dass die Antwort oft weniger von der moralischen Qualität der Argumente als von (für uns) nebensächlichen stilisti-

schen Aspekten der Formulierung beeinflußt wurde ("So würde ich das nicht sagen"). Im MUT entfallen beide Probleme: Die Befragten müssen nicht über ihr (übliches) Verhalten in ähnlichen Situationen berichten und sie müssen auch nicht sagen, ob sie ihre Begründung so oder anders formulieren würden. Sie sollen nur beurteilen, für wie akzeptabel sie eine Reihe von Argumenten für oder gegen eine bestimmte Dilemma-Lösung halten. Und nur dieses *in situ*-Verhalten wird vom MUT aufgezeichnet und analysiert (Kapitel 4).

Die mit dem MUT verbundenen Intentionen haben sich weitgehend erfüllt. Er stellt neben Einstellungsmaßen auch Werte für den Fähigkeitsaspekt des moralischen Urteilsverhaltens zur Verfügung. Anhand dieser Meßwerte können die Annahmen der kognitiven Theorie und anderer Theorien der Moralentwicklung neutral und objektiv geprüft werden. Der MUT stellt daher eine methodologische Alternative zum MJI dar. Er ist, wie sich gezeigt hat, sensitiv für Meßwertänderungen in Interventionsstudien und seine Meßwerte können eine relativ große Entwicklungsspanne abbilden (von 10jährigen Hauptschülern bis 30jährigen Universitätsabsolventen). Während beim MJI selten andere Werte als Stufe 3 und 4 (bei 5 bzw. 6 Stufen insgesamt) resultieren, finden sich beim MUT individuelle Meßwerte von 0 bis 100 (also aus dem gesamten Skalenbereich) und Gruppenmittelwerte von unter 10 bis 50 Punkte. Schließlich ergeben sich beim MUT kaum fehlende Daten oder Totalausfälle, was auf eine große Akzeptanz des Testformats bei den Befragten hinweist.

Neben und über die genannten Einzelaspekten hinaus halte ich für den größten Vorzug des MUT, dass er, wie diese Untersuchung belegt, einen differenzierten, unverzerrten Nachweis der Effekte von Bildungsprozessen auf die moralische Urteilsfähigkeit von Heranwachsenden ermöglicht hat. Erst mit den MUT-Experimenten konnte gezeigt werden, dass dieser Effekt weder auf – leicht revidierbare – moralische *Einstellungen* beschränkt ist, noch dass er als Effekt von *Alters- oder Reifungsprozessen* wegerklärt werden kann.

Die Begrenzungen des MUT sind angesprochen worden und es kann kein Zweifel bestehen, dass dieses Meßinstrument in dem Maße weiterentwickelt werden muß, wie unser Verständnis des moralischen Urteilsverhaltens voranschreitet. Beides bedingt sich wechselseitig. Eine Einschränkung betrifft die Einzelfalldiagnostik, wofür der MUT nicht konstruiert wurde. Hierfür bietet sich nach wie vor das MJI als das geeignetste Verfahren an. Eine andere Begrenzung ergibt sich aus der Vernachlässigung situativ bedingter Differenzie-

rungsprozesse des moralischen Urteilens. Hieran wird die Weiterentwicklung und Verbesserung dieses Meßverfahrens vor allem ansetzen müssen.

VI.

Die hier vorgelegten Studien mit dem MUT haben ergeben, dass Bildung und Erziehung einen sehr wichtigen Faktor für die Entwicklung der individuellen moralischen Urteilsfähigkeit darstellen. Es handelt sich hierbei weder um Effekte der Anpassung an soziale Normen noch um das Produkt genetisch gesteuerter Reifung. Sicherlich haben soziale Normen und biologische Reifungsprozesse in bestimmten Phasen der Moralentwicklung auch wichtige Funktionen, aber ihr Einfluß wurde, wie diese Untersuchung zeigt, bislang weit überschätzt. Dieser Befund bestätigt die Auffassung, dass das Böse in dieser Welt weniger auf das Fehlen moralischer Werte zurückgeht als auf "moralische Dummheit", d.h., auf eine im Vergleich zu der Größe der Probleme zu gering entwickelte moralische Urteils- und Diskursfähigkeit. Ich nenne ausdrücklich beide Fähigkeiten, weil ich dazwischen nur graduelle Unterschiede sehe: die "innere" Fähigkeit zur Reflexion über ein Dilemma auf der Basis eigener moralischer Prinzipien und die "äußere" Fähigkeit zur Diskussion von Konflikten auf der Basis allgemein geteilter demokratischer Normen. Beide sind Bestandteil der Kompetenz zum *vernünftigen* Handeln. Zumeist sind auch beide gleichzeitig gefordert, nur dass in manchen Situationen vielleicht die eine stärker beansprucht wird, in manchen die andere. Das Gegenteil, die Unfähigkeit zum vernünftigen Handeln, zeigt sich im Extremfall in einem völlig impulsiven, nach außen hin ziellos wirkenden Handeln. Eine Zwischenform bildet die Orientierung des eigenen Handeln an – zumeist von der Bezugsgruppe oder Autoritäten übernommenen – Meinungen über das jeweils "richtige" Verhalten. Wir haben solches Verhalten mit Piaget als *heteronom* bezeichnet (Kapitel 2). Das eigene Verhalten wird zwar mit eigenen ethischen Zielen in Verbindung gebracht, aber diese Ziele haben in Wirklichkeit kaum Macht über das Verhalten; Macht haben Konventionen und Meinungen; moralische Ziele dienen allenfalls der *Rationalisierung*, d.h., der nachträglichen Begründung der eigenen Entscheidung. Erst wenn moralische Normen das eigene Handeln wirklich anleiten, sprechen wir von vernünftigem Handeln. Dafür, dass wir unser Handeln konsistent und differenziert an Normen ausrichten können, ist

aber eine entwickelte moralische Urteilsfähigkeit notwendig. Es genügt nicht, dass wir wissen, welche Entscheidung eine bestimmte moralische Norm von uns verlangt (ein Problem, mit dem sich die Moralphilosophie vor allem befaßt); wir müssen auch in der Lage sein, im Falle eines Konfliktes zwischen dieser Norm und unserer vorgefaßten Meinung letztere zu revidieren.

Neben dieser moralischen Urteilsfähigkeit im engeren Sinne benötigen wir zahlreiche weitere Kompetenzen, um verantwortlich entscheiden zu können. Wenn zwei oder mehr Normen miteinander in Konflikt geraten, wenn also ein moralisches Dilemma vorliegt, müssen wir fähig sein, über die Rangfolge von Normen und Ausnahmen zu entscheiden. Für solche Entscheidungen benötigen wir übergeordnete Kriterien wie die Goldene Regel, Kants Kategorischen Imperativ oder das Universalisierungsprinzip, den Grundsatz der Verhältnismäßigkeit, Abstimmungs- und Entscheidungsregeln (wie Mehrheitsprinzip, Minderheitenschutz), sowie Wissen über mögliche und wahrscheinliche Folgen und Nebenfolgen unseres Handelns (vgl. Jonas 1979; Apel 1990; Habermas 1991). Hieraus wird deutlich, dass wir Bildung nicht aufteilen können in beruflich-fachliche Bildung einerseits und moralische Erziehung andererseits, sondern dass es darauf ankommt, worauf Böhler (1992) sowie Mittelstraß (1992) aufmerksam gemacht haben, beide Fähigkeitsbereiche ausgewogen und integral zu fördern. Fächer wie Chemie, Physik und Mathematik sind ebenso Teil moralisch-demokratischer Bildung wie Politik, Religion und Ethik. Ohne Fachwissen ist verantwortliches Handeln letztlich ebensowenig möglich wie ohne moralische Urteilsfähigkeit. Aber dieses Fachwissen darf nicht isoliert und nur einem Teil der Bürger vermittelt werden, sondern muß möglichst allen erschließbar gemacht und in Schule und Hochschule auf allen Ebenen der Entstehung, Vermittlung und Anwendung moral-demokratisch reflektiert werden.

Es gibt an vielen Orten erste Ansätze zu einer systematischen Umsetzung solcher Einsichten. Wie unsere Untersuchung belegt, ist Kohlberg und seinen Kollegen ein wichtiger Durchbruch gelungen bei den pädagogischen Methoden zur direkten Förderung der moralischen Urteilsfähigkeit. Für die von ihm und seinem Schüler Moshe Blatt vorgeschlagene Methode der Diskussion hypothetischer, moralischer Dilemmas konnte eindeutig nachgewiesen werden, dass sie einen deutlichen und nachhaltigen Fördereffekt hat. Dieser Effekt konnte für sehr unterschiedliche Bedingungen nachgewiesen werden, so dass wir die Kohlberg-Blatt-Methode uneingeschränkt empfehlen können, zumal ein solch eindeutiger Nachweis bislang noch für keine andere Methode gelungen ist

(Kapitel 9). Es hat sich allerdings auch gezeigt, dass die Beherrschung dieser pädagogischen Methode psychologische und moralphilosophische Kenntnisse voraussetzt, die heute in der Lehreraus- und fortbildung kaum oder nur unzureichend vermittelt werden. Lehrer haben oft falsche Vorstellungen von den moralischen Motiven ihrer Schüler (Kapitel 10), was auf solche Ausbildungsdefizite hinweist. Es erscheint daher vordringlich, dass die Lehrerausbildung (vor allem im gymnasialen Bereich) auch Methoden der Förderung moralisch-demokratischer Fähigkeiten vermittelt, und zwar nicht nur akademisch, sondern auch praktisch-anschaulich.

Bei der Förderung individueller moralischer Urteilsfähigkeit können wir allerdings nicht stehen bleiben. Es wird zunehmend wichtiger, dass die Fähigkeit gefördert wird, Konflikte mit anderen Menschen demokratisch, d.h. friedlich und gewaltfrei zu lösen, und zwar so, dass die Rechte aller Beteiligten geachtet werden und ein dauerhafter Ausgleich der Interessen geschaffen wird. Für diese Förderung haben ebenfalls Kohlberg und seine Kollegen Power, Higgins und andere eine Methode vorgeschlagen, nämlich die der *Just community* oder der sog. Demokratischen Gemeinschaft. Fritz Oser und ich haben, mit Unterstützung von Kohlberg, diesen Ansatz drei Jahre lang in drei Schulen in Nordrhein-Westfalen erprobt (Kapitel 8). Dieser Ansatz hat sich, auch wenn noch kein abschließendes Urteil möglich ist, zweifellos als sehr fruchtbar und pädagogisch wertvoll erwiesen. Die Einführung demokratischer Strukturen in der Schule bietet den Heranwachsenden die Möglichkeit, theoretisches Wissen über demokratische Entscheidungsprozesse praktisch anzuwenden, auszuprobieren und gemeinschaftlich weiterzuentwickeln. Theoretisches Wissen – Institutionenkunde, politische Konzept usw. – sind eine wichtige Voraussetzung für den Erwerb demokratischer Kompetenzen, aber Demokratie muß auch anhand ernster Entscheidungen geübt werden können, die für die Beteiligten verbindlich sind. Nur so können Vertrauen in demokratischer Prozesse und Verantwortung real vermittelt werden. Niemand würde es einfallen, beim Abschlußkochen in der Hauswirtschaftslehre anzukündigen, das Essen würde danach weggeworfen. Was sollte Schüler und Schülerinnen dann noch motivieren, das Essen genießbar, schmackhaft und attraktiv zu machen. Aber in der Politikkunde beschränken wir uns noch weitgehend darauf, Demokratie über Wertbegriffe zu vermitteln, statt sie real in der Schule zu leben. Wenn Lehrer und Schulverwaltung kein Vertrauen in Demokratie haben, wie sollen die Jugendlichen es dann entwickeln?

Es bleiben noch viele Fragen bezüglich der Bedeutung der Bildung für Moral und Demokratie offen, vor allem: Waren nicht viele Erfinder der unmenschlichen Rassentheorie und geistige Eltern der Nürnberger Rassengesetze hochgebildete Professoren? Gab es unter den erbarmungslosen Vollstreckern nicht viele, wie den Nazi-Propagandaminister *Dr.* Goebbels und den KZ-Arzt *Dr.* Mengele, mit akademischen Graden? Zeigt nicht auch die DDR-Vergangenheit, dass überdurchschnittliche Bildungserfahrungen nicht vor unmoralischem, inhumanem Verhalten schützen?

Wir benötigen nicht einfach mehr Bildung im Sinne von höheren Schulabschlüssen und mehr UniversitätsstudentInnen, sondern eine Bildung, die gezielt und methodisch moralisch-demokratische Fähigkeiten (natürlich auch bei Lehrern) fördert. Auch benötigen wir nicht die einseitige Förderung von technisch-ökonomischen Fähigkeiten, sondern eine breite Grundbildung und eine hochwertige Demokratieerziehung für alle, so dass der politisch-moralische Diskurs zwischen den Mitgliedern dieser Gesellschaft möglich bleibt bzw. dort ermöglicht wird, wo er nicht vorhanden ist. Bis zu einem gewissen Grad ist eine spezialisierte Bildung für wenige Menschen für das Funktionieren einer Gesellschaft sinnvoll. Aber leicht ist der Punkt überschritten, wo Spezialisierung und Elitisierung im Bildungssystem die Demokratie unterminieren und den Grund für ein System von Führern und Verführten, Ausbeutung und Armut, Gewalt und Gegengewalt legen.

Durch eine demokratische Gestaltung unseres Bildungssystems haben wir es weitgehend in der Hand, ob in unserer Gesellschaft Konflikte auf der Basis eines gewaltfreien Diskurses lösbar bleiben oder ob sie in offene oder verdeckte Gewalt gegeneinander münden. Dabei spielt die Fähigkeit, auf die wir in unserer Untersuchung unser besonderes Augenmerk gerichtet haben, die Fähigkeit, Konflikte mit Argumenten, statt durch Gewalt und Aggression zu lösen, eine Schlüsselrolle. Heranwachsende erwerben diese Fähigkeit nur, wenn sie lernen, dass gelebte moralische Prinzipien wie gegenseitige Achtung, Gerechtigkeit, Toleranz, Universalisierbarkeit und Wohlwollen eine wirksame Grundlage für die Auflösung von Meinungsverschiedenheiten darstellen, d.h., wenn sie selbst in der Gemeinschaft Achtung und Gerechtigkeit tatsächlich erfahren und erleben, dass ihre Meinung gehört und ihre Argumente beachtet werden. Auf dieser Einsicht muß jede Erziehung aufbauen, die auf eine Minderung von Gewalt ab-

zielt. Dass sich hierdurch tatsächlich Gewaltkulturen ändern lassen, bestätigen Projekte wie die des Bremer Pädagogen F.-J. Krafeld mit jugendlichen Randalierern (*Der Spiegel* 2/1993, S. 37). Durch Zuhören und Ernstnehmen ihrer Probleme wurde langsam, "über Monate", erreicht, dass das aggressive Verhalten abebbte und abweichende Meinungen bei diesen Jugendlichen nicht mehr, wie vorher üblich, gleich zu verbalen oder physischen Gewaltreaktionen führten, sondern Neugierde weckten: "Ich weiß, dass du anderer Meinung bist, aber gerade deshalb interessiert sie mich ja". Aus der Sicht unserer Bildungstheorie der Moral haben diese Jugendlichen einen entscheidenden Schritt in der Entwicklung ihrer moralischen Urteilsfähigkeit getan. Sie haben gelernt zuzuhören, die Meinung der anderen Person zu respektieren und an einen moralischem Dialog teilzunehmen. Diese Entwicklungsschritt wurde nicht durch Appelle, Kritik oder Gegengewalt erreicht, sondern durch eine Form moralischer Bildung.

Literaturverzeichnis

Adorno, Th.W. (1973). *Gesammelte Schriften*. Frankfurt: Suhrkamp.Allport, F.H. (1924). *Social Psychology*. Boston: Houghton Mifflin.

Allport, G.W. (1935). Attitudes. In C.M. Murchinson, Hrg., *Handbook of Social Psychology. Worchester*. Worchester, MA: Clark University.

Allport, G.W. (1961). *Pattern and Growth of Personality*. New York: Holt, Rinehart & Winston.

Althof, W. (1989). *Moralische Entwicklung und Erziehung. Heft 4: Alternative Ansätze: "Wertklärung" und "Startline"/"Lifeline"*. Fribourg/CH: Pädagogisches Institut der Universität.

Althof, W. (1990). *Teacher's moral judgment, professional responsibility, and decision making in interpersonal conflicts.* Paper presented at the International Symposium "Research on Effective and Responsible Teaching," Fribourg, Sept. 3-7, 1990.

Apel, K.-O. (1976). *Transformation der Philosophie. Band 2, Das Apriori der Kommunikationsgemeinschaft.* Frankfurt: Suhrkamp.

Apel, K.-O. (1990). *Diskurs und Verantwortung. Das Problem des Übergangs zur postkonventionellen Moral.* Frankfurt: Suhrkamp (Original 1988).

Arbuthnot, J. (1979). Error in self-assessment of moral judgment stages. *The Journal of Social Psychology 107*, 289-290.

Arbuthnot, J., Gordon, D.A., & Jurkovic, G.J. (1987). Personality. In H.C. Quay, Hrg., *Handbook of juvenile delinquency*. New York: Wiley, S. 139-183.

Arbuthnot, J. & Gordon. D.A. (1988). Crime and cognition. Community applications of sociomoral reasoning development. *Criminal Justice and Behavior 15*, 379-393.

Aufenanger, S., Garz, D. & Zutavern, M. (1981). *Erziehung zur Gerechtigkeit.* München: Kösel.

Bargel, T. & Kuthe, M. (1992). *Band 1: Schullandschaft in Unordnung; Band 2: Wege aus der Schulkrise.* Tübingen: Thalheimer Verlag.

Bargel, T., Markiewicz, W. & Peisert, H. (1982). University Graduates: Study experience and social role - Empirical findings of a comparative study in five European countries (FORM-Project) In: M. Niessen & J. Peschar, Hrg., *Comparative research on education.* Oxford: Pergamon, S. 55-78.

Beck, K. (1995). Aspeckt der moralischen Urteilsbildung bei kaufmännischen Lehrlingen – Methodologische Probleme und empirische Befunde. In: E.H. Hoff & L. Lappe, Hrg., Verantwortung im Arbeitsleben. S. 114-128. Heidelberg: Roland Asanger Verlag.

Belenky, M. & Gilligan, C. (1983). Der Einfluß einer Abtreibungskrise auf die Moralentwicklung. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moralisches Urteilen und soziale Umwelt.* Weinheim: Beltz, S. 211-222.

Berliner, D.C. (1987). Ways of thinking about students and classroom by more and less experienced teachers. In J. Calderhead, Hrg., *Exploring teachers' thinking*. London: Cassell Educational Ltd, S. 60-83.

Bierhoff, H.W. (1980). *Hilfreiches Verhalten*. Darmstadt: Steinkopff.

Biggs, D., Colesante, R., Agafonov, A., Felming, J., & Campbell, K. (1999). Education about controversial issues: the role of tolerance, moral judgment, and course involvement. Symposium at the AME conference in Minneapolis, Nov. 1999.

Binford, M. (1983). The democratic political personality: Functions of attitudes and styles of reasoning. *Political Psychology 4,* 663-684.

Blasi, A. (1980). Bridging moral cognition and moral action. A critical review of literature. *Psychological Bulletin 88,* 1-45.

Bock, R.D. (1975). *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill.

Böhler, D. (1992). Für die Zukunft studieren. *Mittelungen des Hochschulverbandes 6,* 347-349.

Boesch, E.E. (1984). The development of affective schemata. *Human Development 27,* 173-183.

Boom, J. & Molenaar, P.C.M. (1989). A developmental model of hierarchical stage structure of objective moral judgments. *Developmental Review 9,* 133-145.

Bozarth, J.D. & Roberts, R.R. (1972). Signifying significant significance. *American Psychologist 27,* 774-775.

Braunmühl, E.v. (1975). *Anti-pädagogik. Studien zur Abschaffung der Erziehung*. Weinheim: Beltz.

Bredenkamp, J. (1970). Über Maße der praktischen Signifikanz. *Zeitschrift für Psychologie 177,* 310-318.

Breed, W. & Ktsanes, T. (1961). Pluralistic ignorance in the process of opinion formation. *Public Opinion Quarterly 25,* 382-392.

Brezinka, W. (1988). Competence as an aim of education. In B. Spieker & R. Straughton, Hrg., *Philosophical issues in moral education and development.* Philadelphia, S. 75-98.

Bromme, R. (1987). Die Untersuchung des professionellen Wissens von Lehrern: Zum Inhalt und zur psychologischen Struktur von Wissen. In G.-A. Eckerle & J.-L. Patry, Hrg., *Theorie und Praxis des Theorie- Praxis-Bezugs in der empirischen Pädagogik.* Baden-Baden: Nomos.

Broughton, J. (1970). *'College regression' in moral development.* Unveröffentlichtes Manuskript, Columbia University, Teachers College, New York.

Broughton, J. (1978). The cognitive-developmental approach to morality: A reply to Kurtines and Greif. *Journal of Moral Education 1*, 81-96.

Brugman, D. & Dohle, M.J.M. (1990). *Elementary school teachers' cognitions about their pupils's moral reasoning development.* Paper presented at the International Symposium “Research on Effective and Responsible Teaching,” Fribourg, September 3-7, 1990.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review 62,* 193-217.

Budner, S. (1969). Intolerance of ambiguity as a personality variable. *Journal of Personality 30,* 29-50.

Bull, N.J. (1969). *Moral judgment from childhood to adolescence.* London: Routledge & Kegan.

Burton, R.U. (1963). Generality of honesty reconsidered. *Psychological Review 70,* 481-499.

Campbell, D.T. & Stanley, J (1966). *Experimental and Quasi-experimental Designs for Research.* Chicago: Rand McNally.

Candee, D., Graham, R. & Kohlberg, L. (1978). *Moral development and life outcomes.* Harvard University: National Institute of Education (manuscript).

Carver, R.P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48,* 378-399.

Carver, R.P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education, 61*(4), 287-292

Child, I. (1954). Socialization. In G. Lindzey, Hrg., *Handbook of Social Psychology. Vol. II.* Cambridge MA: Addison-Wesley, S. 655-692.

Chomsky, N. (1973). *Aspekte der Syntax-Theorie.* Frankfurt: Suhrkamp.

Cohen, J. (1969). A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin 71,* 281-284.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press, Second Edition.

Cohen, R. (1969). *Systematische Tendenzen bei Persönlichkeitsbeurteilungen. Eine empirische Untersuchung.* Stuttgart: Huber.

Colby, A., Kohlberg, L., Gibbs, J. & Liebermann, M. (1983). Moral development: A longitudinal study of moral development. *Child Development Monographs 48,* 1-96.

Colby, A. & Kohlberg, L. (1987). *The Measurement of Moral Judgment. Volume I and II.* New York: Columbia University Press.

Comunian, A.L. & Gielen, U. (1999). Social role-taking and moral judgment competence improve after group discussion: An Italian intervention study. Unpublished paper, Padua, New York.

Converse, P.E. (1970). Attitudes and Non-Attitudes: Continuation of a dialoque. In E.R. Tufte, *The Quantitative Analysis of Social Problems.* Reading MA: Addison-Wesley, S. 168-189.

Cramér, H. (1966). *Mathematical Methods of Statistics.* New York: Macmillan.

Damon, W. (1984). *Die soziale Welt des Kindes.* Frankfurt: Suhrkamp.

Dann, H.-D. (1989a). Subjektive Theorien als Basis erfolgreichen Handelns von Lehrkräften. *Beiträge zur Lehrerbildung 7,* 247-254.

Dann, H.-D. (1989b). Was geht im Kopf der Lehrer vor? Lehrerkognitionen und erfolgreiches pädagogisches Handeln. *Psychologie in Erziehung und Unterricht 36,* 81-90.

Dann, H.-D. (1990). Subjective theories: A new approach to psychological research and educational practice. In G.R. Semin & K.J. Gergen, Hrg., *Everyday understanding. Social and scientific implications.* London: Sage Publications, S. 227-243.

Dann, H.-D., Cloetta, B., Müller-Fohrbrodt, G. & Helmreich, R. (1978). *Umweltbedingungen innovativer Kompetenz. Eine Längsschnittuntersuchung zur Sozialisation von Lehrern in Ausbildung und Beruf.* Stuttgart: Klett.

Dann, H.-D., Tennstädt, K.-Ch., Humpert, W. & Krause, F. (1987). Subjektive Theorien und erfolgreiches Handeln von Lehrern/-innen bei Unterrichtskonflikten. *Unterrichtswissenschaft 15,* 306-320.

Dewey, J. (1964). *Demokratie und Erziehung. Eine Einleitung in die philosophische Pädagogik.* Braunschweig: Georg Westermann Verlag (Original 1915).

Dewey, J. (1976). *Lectures on psychological and political ethics.* New York: Hafner Press (Original 1898).

Dippelhofer-Stiem, B. (1983). *Hochschule als Umwelt. Probleme der Konzeptualisierung, Komponenten des methodischen Zugangs und ausgewählte empirische Befunde.* Weinheim: Beltz.

Dippelhofer-Stiem, B. & Lind, G., Hrg. (1987). *Studentisches Lernen im Kulturvergleich. Ergebnisse einer international vergleichenden Längsschnittstudie zur Hochschulsozialisation.* Weinheim: Deutscher Studienverlag.

Döbert, R. (1987). Horizonte der an Kohlberg orientierten Moralforschung. *Zeitschrift für Pädagogik 33,* 491-511.

Döbert, R. & Nunner-Winkler, G. (1975). *Adoleszenzkrise und Identitätsbildung.* Frankfurt: Suhrkamp.

Döbert, R. & Nunner-Winkler, G. (1978). Performanzbestimmende Aspekte des moralischen Bewußtseins. In G. Portele, Hrg., *Sozialisation und Moral. Neuere Ansätze zur moralischen Entwicklung und Erziehung.* Weinheim: Beltz, S. 101-121.

Dobbelstein-Osthoff, P. (1990). *Just community – basisdemokratische Spielwiese oder Chance zur Demokratisierung der Schule?* Vortrag für das International Colloquium "Responsible and effective teaching," Fribourg/ Schweiz, September 1990. (Landesinstitut für Schule und Weiterbildung, Soest.)

DuBois, J. (1997). *The moral judgment of medical personnel. A cross-cultural study about brain death and organ explantation*. Unpublished dissertation, Universität Wien. Österreich.

Durkheim, E. (1976). *Soziologie und Philosophie* Frankfurt: Suhrkamp (Original 1924).

Durkheim, E. (1984). *Erziehung, Moral und Gesellschaft.* Frankfurt: Suhrkamp (Original 1902).

Eckensberger, L.H. (1983). Research on moral development in Germany. *The German Journal of Psychology 7,* 195-244.

Eckensberger, L.H. (1984). *On structure and content in moral development.* Saarbrücken: Universität des Saarlandes (Arbeiten der Fachrichtung Psychologie).

Eckensberger, L.H. & Reinshagen, H. (1980). Kohlbergs Stufentheorie der Entwicklung des Moralischen Urteils: Ein Versuch der Reinterpretation im Bezugsrahmen handlungstheoretischer Konzepte. In L.H. Eckensberger &

R. Silbereisen, Hrg., *Entwicklung sozialer Kognitionen.* Stuttgart: Klett, S. 65-131.

Edelstein, W. (1986). Moralische Intervention der Schule. Skeptische Überlegungen. In F. Oser, R. Fatke & O. Höffe, Hrg., *Transformation und Entwicklung.* Frankfurt: Suhrkamp, S. 327-349.

Edelstein, W. (1987). Förderung der moralischen Entwicklung in der Schule. Möglichkeiten und Grenzen. *Zeitschrift für Pädagogik 33,* 185-205.

Edelstein, W., Keller, M., Essen, C.v. & Mönnig, M. (1986). Moralische Sensibilität, Handlungsentscheidung und moralische Konsistenz. In F. Oser, W. Althof & D. Garz, Hrg., *Moralische Zugänge zum Menschen – Zugänge zum moralischen Menschen.* München: Peter Kindt Verlag, S. 44-66.

Edwards, C.P. (1975). Societal complexity and moral development. A Kenyan study. *Ethos 3,* 505-527.

Edwards, C.P. (1981). The comparative study of the development of moral judgment and reasoning. In R.H. Munroe, R.L. Munroe & B.B. Whiting, Hrg., *Handbook of Cross-Cultural Human Development.* New York: Garland Publishing, S. 501-527.

Edwards, C.P. (1986). Cross-cultural research on Kohlberg's stages. A basis for consensus. In S. Modgil & C. Modgil, Hrg., *Lawrence Kohlberg: Consensus and Controversy.* Brighton GB: Palmer Press.

Emler, N. (1983). Morality and politics: The ideological dimension in the theory of moral development. In H. Weinreich-Haste & D. Locke, Hrg., *Morality in the Making.* Chichester, GB: Wiley, S. 47-71.

Emler, N., Renwick, S. & Malone, B. (1983). The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology 45,* 1073-80.

Enright, R.D., Lapsley, D.K., Harris, D.J. & Schaver, D.J. (1983). Moral development interventions in early adolescence. *Theory into Practice 22,* 134-144.

Erikson, E. (1966). *Identität und Lebenszyklus.* Frankfurt: Suhrkamp.

Eyferth, K. (1959). Integration und Starrheit. *Psychologische Rundschau,* 159-169.

Fend, H. (1969). *Sozialisierung und Erziehung.* Weinheim: Beltz.

Fend, H. (1977). *Schulklima: Soziale Einflußprozesse in der Schule.* Weinheim: Beltz

Fend, H. (1990). *Vom Kind zum Jugendlichen.* Bern: Huber.

Fend, H., Knörzer, W., Nagl, W., Specht, W. & Väth-Szusdziara, R. (1976). *Sozialisationseffekte der Schule*. Weinheim: Beltz.

Fetz, R.L. (1982). Pour une ontologie génétique. Jean Piaget et la philosophie moderne. *Revue Internationale de Philosophie 36,* 409-434.

Fetz, R.L. & Oser, F. (1986). Weltbildentwicklung, moralisches und religiöses Urteil. In W. Edelstein & G. Nunner-Winkler, Hrg., *Zur Bestimmung der Moral. Philosophische und sozialwissenschaftliche Beiträge zur Moralforschung.* Frankfurt: Suhrkamp, S. 443-469.

Fischer, G. & Roppert, J. (1964). Bemerkungen zu einem Verfahren der Transformationsanalyse. *Archiv der Gesamten Psychologie 116,* 98-100.

Fishbein, M. & Ajzen, I. (1975). *Attitudes, Beliefs, Intentions and Behavior.* New York: Addison-Wesley.

Fishkin, J. Keniston, K. & McKinnon, C. (1973). Moral reasoning and political ideology. *Journal of Personality and Social Psychology 27,* 109-119.

Foerster, F.W. (1918). *Politische Ethik und politische Pädagogik.* München: Ernst Reinhardt.

Framhein, G. & Langer, J., Hrg. (1984). *Student und Studium im internationalen Vergleich*. Klagenfurt: Kärntner Druck- u. Verlagsgesellschaft.

Fraser M.J., Walberg, H.J., Welch, W.W., & Hattie, J.A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11,* 145-252.

Gadenne, V. (1976). *Die Gültigkeit psychologischer Untersuchungen.* Stuttgart: Kohlhammer.

Gadenne, V. (1990). Methoden als Hilfsmittel für heuristische Forschungsentscheidungen - Bemerkungen zu Wottawas Abhandlung der psychologischen Methodenlehre. *Psychologische Rundschau, 41,* 84-97.

Garz, D. (1989). *Theorie der Moral und gerechte Praxis.* Wiesbaden: Deutscher Universitätsverlag.

Garz, D., Oser, F., & Althof, W. Hg. (1999). *Moralisches Urteil und Handeln*. Frankfurt: Suhrkamp.

Gebhardt, F. (1968). Über die Ähnlichkeit von Faktormatrizen. *Psychologische Beiträge 10,* 591-599.

Gielen, U. (1984). Cross-cultural research on moral reasoning and Stage 5: Comments on Ian Vine's paper. In G. Lind, *Beiträge zum Zweiten Konstanzer Werkstattgespräch über Moral und Umwelt.* Universität Konstanz: Bericht, S. 77-84.

Gielen, U. (1986). Moral reasoning in radical and non-radical German students. *Behavior Science Research 20,* 71-109.

Gibbs, J.C. (1977). Kohlberg's stages of moral judgment. A constructive critique. *Harvard Educational Review 47,* 43-61.

Gibbs, J.C., Widaman, K.F., & Colby, A. (1982). Construction and validation of a simplified, group-administrable equivalent to the Moral Judgment Interview. *Child Development 53,* 895-910.

Gilgenmann, K. (1984). Zur Diskussion über Werte, Wertwandel, Moralerziehung und Stufen der Moralentwicklung. In A. Regenbogen, *Moral und Politik - soziales Bewußtsein als Lernprozeß.* Köln: Pahl-Rugenstein, S. 10-26.

Glass, G.V., McGaw, B. & Smith, M.L. (1978). *Meta-Analysis in Social Research.* London: Sage Publications.

Grabow, M. (1990). *Die Entwicklung moralischer Urteilsfähigkeit in der Erzieherausbildung.* Frankfurt: Lang.

Graumann, C.-F. (1988). Der Kognitivismus in der Sozialpsychologie – Die Kehrseite der 'Wende'. *Psychologische Rundschau 39,* 83-90.

Gross, M.L. (1997). Ethics and activism: The theory and practice of political morality. Cambridge, MA: Cambridge University Press.

Haan, N. (1977). *Coping and Defending. Process of Self-environment Organization.* New York: Academic Press.

Haan, N. (1978). The two moralities in action context: Relationship to thought, ego regulation and development. *Journal of Personality and Social Psychology 36,* 286-305.

Habermas, J. (1983). *Moralbewußtsein und kommunikatives Handeln.* Frankfurt: Suhrkamp.

Habermas, J. (1985). Philosophical notes on moral judgment theory. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral Development and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education.* Chicago: Precedent Publishing Inc., S. 3-20.

Habermas, J. (1991). *Erläuterungen zur Diskursethik.* Frankfurt: Suhrkamp.

Hagemann, W., Heidbrink, H. & Schneider, M.M., Hrg. (1982). *Kognition und Moralität in politischen Lernprozessen.* Opladen: Leske Verlag.

Harnatt, J. (1979). Anmerkungen zur Kritik psychologischer Forschungsmethoden. *Psychologische Beiträge 21,* 203-211.

Hartmann, H.A. (1985). What is social about morality? Morals, morality, and ethics in social science perspectives. In G. Lind, H.A. Hartmann & R. Wa-

kenhut, Hg., *Moral Development and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education.* Weinheim: Beltz, S. 275-295.

Hartmann, Heinz (1972). *Ich-Psychologie. Studien zur psychoanalytischen Theorie.* Stuttgart: Klett.

Hartshorne, H. & May, M.A. (1928-30). *Studies in the Nature of Character. Vol.I: Studies in deceit. Vol. II: Studies in service and self-control. Vol. III: Studies in organisation of character.* New York: Macmillan.

Hays, W. (1963). *Statistics.* New York: Holt.

Heidbrink, H. (1985). Moral judgment competence and political learning. In G. Lind, H.A. Hartmann & R. Wakenhut, *Moral Development and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education.* Chicago: Precedent Publishing Inc, S. 259-271.

Heidbrink, H. (1989). Moralpsychologie: Die Reanalyse einer Untersuchung von 1913. *Geschichte der Psychologie 17,* 4-9 (FernUniversität Hagen).

Heidbrink, H. (1991). *Stufen der Moral. Zur Gültigkeit der kognitiven Entwicklungstheorie Lawrence Kohlbergs.* München: Quintessenz.

Heidbrink, H. (1995). Zur Existenz von Moralstufen. Eine strukturelle Validitätsanalyse. In: E.H. Witte, Hrg., Sozial Kognition und empirische Ethikforschung, S. 77-96. Lengerich: Pabst.

Heidbrink, H. (1996). *Einführung in die Moralpsychologie.* München: Beltz/ Psychologie Verlags Union.

Heider, F. (1958). *The Psychology of Interpersonal Behavior.* New York: Wiley.

Helkama, K. & Ikonen, M. (1986). Some correlates of maturity of moral reasoning in Finland. *Behavior Science Research 20,* 110-131.

Higgins, A. (1980). Research and measurement issues in moral education interventions. In R. Mosher, Hrg., *Moral education. A first generation of research and development.* New York: Praeger, S. 92-107.

Higgins, A. (1987). Moralische Erziehung in der Gerechte Gemeinschaft-Schule. Über praktische Erfahrungen in den USA. In G. Lind & J. Raschert, Hrg., *Moralische Urteilsfähigkeit. Eine Auseinandersetzung mit Lawrence Kohlberg über Moral, Erziehung und Demokratie.* Weinheim: Beltz, S. 54-72.

Higgins, A. (1989). Das Erziehungsprogramm der Gerechten Gemeinschaft: Die Entwicklung moralischer Sensibilität als Ausdruck von Gerechtigkeit und

Fürsorge. In G. Lind & G. Pollitt-Gerlach, Hrg., *Moral in 'unmoralischer' Zeit. Zu einer partnerschaftlichen Ethik in Erziehung und Gesellschaft.* Heidelberg: Asanger, S. 101-128.

Higgins, A., Power, C., & Kohlberg, L. (1984). The relationship of moral atmosphere to judgments of responsibility. In W.H. Kurtines & J.L. Gewirtz, eds., *Morality, moral behavior and moral development.* New York: Wiley.

Hofer, M., Pekrun, R. & Zielinski, W. (1986). Die Psychologie des Lerners. B. Weidenmann & A. Krapp, Hrg., *Pädagogische Psychologie.* München: Psychologie Verlags Union, S. 219-275.

Höffe, O. (1986). Autonomie und Verallgemeinerung als Moralprinzipien. In F. Oser, R. Fatke & O. Höffe, Hrg., *Transformation und Entwicklung*. Frankfurt: Suhrkamp, S. 56-86.

Hofstätter, P. (1973). *Sozialpsychologie.* Berlin: Walter de Gruyter.

Hogan, R. (1970). A dimension of moral judgment. *Journal of Consulting and Clinical Psychology 33,* 307-316.

Holstein, C. (1976). Development of moral judgment: A longitudinal study of males and females. *Child Development 47,* 51-61.

Holt, R. (1986). Bridging the rift in Political Psychology: An open letter to Stanley Hoffman. *Political Psychology 7,* 234-244.

Jonas, H. (1979). *Das Prinzip Verantwortung.* Frankfurt: Suhrkamp.

Johnson, J.A. & Hogan, R. (1981). Moral judgments and self-presentation. *Journal of Research in Personality 15,* 57-63.

Jöreskog, K.G. (1970). Estimation and testing of simplex models. *The British Journal of Mathematical and Statistical Psychology 23,* 131-145.

Keasey, C.B. (1971). Social participation as a factor in the moral development of preadolescents. *Developmental Psychology 52,* 216-220.

Keasey, C.B. (1974). The influence of opinion-agreement and quality of supportive reasoning in the evaluation of moral judgments. *Journal of Personality and Social Psychology 30,* 477-482.

Keasey, C.B. (1975). Implicators of cognitive development of moral reasoning. In D. Palma & J. Foley, Hrg., *Moral development.* New York: Wiley, S. 39-56.

Keller, M. (1990). Zur Entwicklung moralischer Reflexion: Eine Kritik und Rekonzeptualisierung der Stufen des präkonventionellen moralischen Urteils in der Theorie von L. Kohlberg. In M. Knopf & W. Schneider, Hrg., *Entwick-*

lung. Allgemeine Verläufe - Individuelle Unterschiede. Pädagogische Konsequenzen. Göttingen: Hogrefe, S. 19-44.

Keller, M. & Edelstein, W. (1986). Beziehungsverständnis und moralische Reflexion. In W. Edelstein & G. Nunner-Winkler, Hrg., *Zur Bestimmung der Moral. Philosophische und sozialwissenschaftliche Beiträge zur Moralforschung.* Frankfurt: Suhrkamp, S. 321-346.

Klockhaus, R. & Habermann-Morbey, B. (1986). *Psychologie des Schulvandalismus.* Göttingen: Hogrefe.

Koch, S. (1963). Epilogue In S. Koch, Hrg., *Psychology: A study of a science. Volume 3: Formulations of the Person and the Social Context.* New York: McGraw-Hill, S. 729-788.

Koch-Priewe, B. (1986). *Subjektive didaktische Theorien von Lehrern*. Frankfurt: Haag und Herchen.

Kohlberg, L. (1958). *The development of modes of moral thinking and choice in the years 10 to 16.* University of Chicago: Unveröffentlichte Dissertation.

Kohlberg, L. (1963). The development of children's orientations toward a moral order. *Vita Humana 6*, 11-33.

Kohlberg, L. (1964). Development of moral character and moral ideology. In M.L. Hoffman & L.W. Hoffman, Hrg., *Review of child development research, Vol. I.* New York: Russel Sage Foundation, S. 381-431.

Kohlberg, L. (1969). Stage and sequence. The cognitive-developmental approach to socialization. In D. Goslin, Hrg., *Handbook of socialization theory and research*. Chicago: Rand McNally, S. 347-480.

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel, Hrg., *Cognitive Development and epistemology.* New York: Academic Press, S. 151-235.

Kohlberg, L. (1973). Continuities in childhood and adult moral development revisited. In P.B. Baltes & W. Schaie, Hrg., *Life-span developmental psychology, research and theory.* New York: Academic Press, S. 179-204.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In T. Lickona, *Moral Development and Behavior: Theory, Research and Social Issues*. New York: Holt, Rinehart & Winston, S. 31-53.

Kohlberg, L. (1977). *Assessing moral stages: A manual, Part I, Introduction.* Cambridge MA: Center for Moral Education, Harvard University.

Kohlberg, L. (1981). *The meaning and measurement of moral development*. Worcester, MA: Clark University Press.

Kohlberg, L. (1984). *The Psychology of Moral Development. Volume II, Essays on Moral Development.* San Francisco: Harper & Row.

Kohlberg, L. (1984/1969). Stage and sequence: The cognitive-developmental approach to socialization. In L. Kohlberg, *The Psychology of Moral Development. Volume II, Essays on moral development.* San Francisco: Harper & Row. (Original 1969), S. 1-169.

Kohlberg, L. (1985). Foreword. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral Development and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education.* Chicago: Precedent Publishing Inc, S. xv-xvii.

Kohlberg, L. (1987). Moralische Entwicklung und demokratische Erziehung. In G. Lind & J. Raschert, Hrg., *Moralische Urteilsfähigkeit. Eine Auseinandersetzung mit Lawrence Kohlberg über Moral, Erziehung und Demokratie*. Weinheim: Beltz, S. 25-43.

Kohlberg, L., Boyd, D.R. & Levine, C. (1986). Die Wiederkehr der sechsten Stufe: Gerechtigkeit, Wohlwollen und der Standpunkt der Moral. In W. Edelstein & G. Nunner-Winkler, Hrg., *Zur Bestimmung der Moral. Philosophische und sozialwissenschaftliche Beiträge zur Moralforschung*. Frankfurt: Suhrkamp, S. 205-240.

Kohlberg, L. & Candee, D. (1984). The relationship of moral judgment to moral action. In L. Kohlberg, *Essays on Moral Development. Vol II: The psychology of moral development.* San Francisco: Harper & Row, S. 498-582.

Kohlberg, L., Colby, A., Gibbs, J., Speicher-Dubin, B., & Power, C. (1977). *Assessing Moral Stages: A Manual. Part I, Introduction.* Harvard University (Manuskript).

Kohlberg, L. & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development 12,* 93-120.

Kramis, J. (1990). *Quellen des Wissens über guten Unterricht und deren Stärken und Schwächen*. Berichte zur Erziehungswissenschaft, Universität Fribourg/Schweiz.

Krathwohl, D., Bloom, B. & Masia, B. (1964). *Taxonomy of Educational Objectives. Handbook II: Affective Domain.* New York: McKay.

Krämer-Badoni, T. & Wakenhut, R. (1985). Morality and the military lifeworld. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral Development*

and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education. Chicago: Precedent Publishing Inc, S. 205-220.

Krebs, D., Denton, K.L., Vermeulen, S.C., Carpendale, J.I. & Bush, A. (1990). The structural flexiblity of moral judgment. *Journal of Personality and Social Psychology* (im Druck).

Krech, D. & Crutchfield, R.S. (1948). *Theory and problems of social psychology.* New York: McGraw-Hill.

Kruskal, W. (1958). Ordinal measures of association. *Journal of the American Statistical Association, 53,* 814-861.

Kuhn, D., Langer, J., Kohlberg, L., & Haan, N. (1977). The development of formal operations in logical and moral judgments. *Genetic Psychological Monographs, 95,* 97-188.

Kumpf, M. (1990). Geschichte der Sozialpsychologie. In E.G. Wehner, Hrg., *Geschichte der Psychologie. Eine Einführung.* Darmstadt: Wissenschaftliche Verlagsgesellschaft, S. 103-128.

Kurtines, W.M. & Greif, E.B. (1974). The development of moral thought: Review and evaluation of Kohlberg's approach. *Psychological Bulletin, 81,* 453-470.

Latané & Darley (1970). *The unresponsive bystander: why doesn't he help?* New York: Appleton.

Lawrence, J. (1980). Moral judgment intervention studies using the Defining Issues Test. *Journal of Moral Education, 9,* 178-191.

Lazarus, R.S. (1966). *Psychological stress and the coping process.* New York: McGraw-Hill.

Leming, J. (1981). Curricular effectiveness in moral/values education: A review of research. *Journal of Moral Education, 10,* 147-164.

Lempert, W. (1988). Soziobiographische Bedingungen der Entwicklung moralischer Urteilsfähigkeit. *Kölner Zeitschrift für Soziologie und Sozialpsychologie 40,* 40-61.

Levy-Suhl, M. (1912). Die Prüfung der sittlichen Reife jugendlicher Angeklagter und die Reformvorschläge zum § 56 des deutschen Strafgesetzbuches. *Zeitschrift für Psychotherapie,* 232-254.

Lewin, K., Lippitt, R. & White, R.K. (1939). Patterns of aggressive behavior in experimentally created 'social climates'. *The Journal of Social Psychology 10,* 271-299.

Lienert, G. (1967). *Testtaufbau und Testanalyse.* Weinheim: Beltz.

Lind, G. (1978). Wie mißt man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts. In Portele, G., Hrg. *Sozialisation und Moral.* Weinheim: Beltz, S. 171-201.

Lind, G. (1982). Experimental Questionnaires: A new approach to personality research. In A. Kossakowski & K. Obuchowski, Hrg., *Progress in the Psychology of Personality*. Amsterdam: North-Holland, S. 132-144.

Lind, G. (1984). Theorie und Validität des Moralisches-Urteil-Tests. Zur Erfassung kognitiv-struktureller Effekte der Sozialisation. In G. Framhein & J. Langer, Hrg., *Student und Studium im interkulturellen Vergleich.* Kärntner Druck- und Verlagsanstalt, S. 166-187.

Lind, G. (1985a). *Inhalt und Struktur des moralischen Urteilens. Theoretische, methodologische und empirische Untersuchungen zur Urteils- und Demokratiekompetenz bei Studierenden.* Dissertation, Sozialwiss. Fakultät der Universität Konstanz. http://www.uni-konstanz.de/ag-moral/

Lind, G. (1985b). The theory of moral-cognitive judgment: A socio-psychological assessment. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral Development and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education.* Chicago: Precedents Publishing Inc., S. 21-53. http://www.uni-konstanz.de/ag-moral/

Lind, G. (1985c). Attitude change or cognitive-moral development? How to conceive of socialization at the university. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral development and the social environment. Studies in the philosophy and psychology of moral judgment and education.* Chicago: Precedent Publishing Inc., S. 173-192.

Lind, G. (1985d). Moral judgment competency and education in democratic society. In P. Weingartner & G. Zecha, Hrg., *Conscience: An interdisciplinary approach*. Dordrecht: Reidel, S. 91-121.

Lind, G. (1986a). Parallelität von Affekt und Kognition in der moralischen Entwicklung. In F. Oser, W. Althof & D. Garz, Hrg., *Moralische Zugänge zum Menschen. Zugänge zum moralischen Menschen.* München: Peter Kindt Verlag, S. 158-179.

Lind, G. (1986b). Cultural differences in moral judgment? A study of West and East European University Students. *Behavioral Science Research, 20*, 208-225.

Lind, G. (1987). Ansätze und Ergebnisse der “Just Community”-Schule. *Die Deutsche Schule 1/87*, 4-12.

Lind, G. (1989a). Die Entwicklung moralischer Gefühle durch Vernunft und Dialog. In G. Lind & G. Pollitt-Gerlach, Hrg., *Moral in 'unmoralischer' Zeit. Zu einer partnerschaftlichen Ethik in Erziehung und Gesellschaft.* Heidelberg: Asanger, S. 7-32. (http://www.uni-konstanz.de/ag-moral/).

Lind, G. (1989b). Moralische Entwicklung in betrieblichen Organisationen. In H. Steinmann & A. Löhr, Hrg., *Unternehmensethik*. Stuttgart: Poeschel, S. 299-314.

Lind, G. (1990). Demokratie und Erziehung in der Schule: Ausgangslage und Ziele des DES-Projekts in Nordrhein-Westfalen. In U. Steffens & T. Bargel, Hrg., *Arbeitskreis “Qualität von Schule”. Heft 7: Erziehung und Demokratie in der Schule - Konzepte und Erfahrungen zum Ansatz von Kohlberg und Oser*. Wiesbaden: Hessisches Institut für Bildungsplanung und Schulentwicklung.

Lind, G. (1993). Moralerziehung als demokratische Bildung. *Politisches Lernen 2/1993*, 20-26.

Lind, G. (1994). *Why do juvenile delinquents gain little from moral discussion programs?* Paper presented at the Annual Convention of the American Educational Research Association in New Orleans, 1994.

Lind, G. (1995). What do moral educators know about their students? A study of pluralistic ignorance. In: R. Olechowski & G. Svik, Hrg., *Experimental Research on Teaching and Learning*, S. 221-243. Frankfurt/New York: Peter Lang.

Lind, G. (1996a). *Educational environments which promote self-sustaining moral-development.* Paper presented at the meeting of the American Educational Research Association, New York, 1996 (http://www.uni-konstanz.de/ag-moral/)

Lind, G. (1996b). *The optimal age of moral education: A review of intervention studies and an experimental test of the dual-aspect-theory of moral development and education*. Paper presented at the meeting of the American Educational Research Association, New York, 1996 (http://www.uni-konstanz.de/ag-moral/).

Lind, G. (1996c).Testing moral judgment competence. A response to Rest, Thoma and Edwards (http://www.uni-konstanz.de/ag-moral/).

Lind, G. (1999c). Moralisches Dilemma-Diskussion. Ziele, Aktivitäten und Zeitplan für eine Schulstunde (http://www.uni-konstanz.de/ag-moral/).

Lind, G. (2000a). Are helpers always moral? Empirical findings from a longitudinal study of medical students in Germany. In: A.L. Comunian & U. Gielen, eds., *Human development in international perspective, S.* 463-477. Lengerich: Papst Science Publishers.

Lind, G. (2000b). Moral Regression in Medical Students and Their Learning Environment. *Revista Brasileira de Educacao Médica, 24*(3), 24-33.

Lind, G. (2000c). The importance of role-taking opportunities for self-sustaining moral development. *Journal of Research in Education, 10,* 9-15.

Lind, G. (2002, im Druck). The meaning and measurement of moral judgment competence – A dual-aspect theory. In D. Fasko & W. Willis, eds., *Philosophical and psychological perspectives of moral development and education*. Cresskill, N.J.: Hampton Press.

Lind, G. & Althof, W. (1992). Does the Just Community program make a difference? Measuring and evaluating the effect of the DES project. *Moral Education Forum, 17*, 19-28.

Lind, G., Grocholewska, K. & Langer, J. (1987). Haben Frauen eine andere Moral? Eine empirische Untersuchung von Studentinnen und Studenten in Österreich, Deutschland und Polen." In L. Unterkirchner & I. Wagner, Hrg., *Die andere Hälfte der Gesellschaft. Soziologische Befunde zu geschlechtsspezifischen Formen der Lebensbewältigung*. Wien: Verlag des Österr. Gewerkschaftsbundes, S. 394-406.

Lind, G. & Link, L. (1986). *Moralische Atmosphäre Fragebogen (MAF). Theoretische Grundlagen und Instrument.* Universität Konstanz.

Lind, G. & Raschert, J., Hrg. (1987). *Moralische Urteilsfähigkeit. Eine Auseinandersetzung mit Lawrence Kohlberg über Moral, Erziehung und Demokratie*. Weinheim: Beltz.

Lind, G., Sandberger, J.-U. & Bargel, T. (1985). Moral competence and democratic personality. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral development and the social environment. Studies in the philosophy and psychology of moral judgment and education.* Chicago: Precedent Publ. Inc, S. 55-78.

Lind, G. & Wakenhut, R. (1985). Testing for moral judgment competence. G.Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral development and the*

social environment. Studies in the philosophy and psychology of moral judgment and education. Chicago: Precedent Publishing Inc, S. 79-105.

Link, L. & Lind, G. (1988). *Entwicklung der sozialen Verantwortung in der Schule.* Studien zur moralisch-demokratischen Urteilsfähigkeit, Nr. 1. Manuskript-Druck, Konstanz.

Lippitt, R. & White, R.K. (1973). Eine experimentelle Untersuchung über Führungsstil und Gruppenverhalten. In C.-F. Graumann & H. Heckhausen, Hrg., *Pädagogische Psychologie 1, Entwicklung und Sozialisation.* Frankfurt: Fischer Taschenbuch Verlag (Original 1947), S. 324-347.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181-209.

Locke, D. (1983). Moral reasons and moral action. In H. Weinreich-Haste & D. Locke, Hrg., *Morality in the making.* Chichester, GB: Wiley, S. 111-124.

Lockwood, A.L. (1978). The effects of values clarification and moral development curricula on school-age subjects: a critical review of recent research. *Review of Educational Research, 48,* 325-364.

Loevinger, J. (1976). *Ego development.* San Francisco CA: Jossey-Bass.

Lumsden, J. (1976). Test theory. *Annual Review of Psychology, 27,* 251-280.

Markoulis, D. (1989). Political involvement and socio-moral reasoning: Testing Emler's interpretation. *British Journal of Social Psychology, 28*, 203-212.

Mauermann, L. (1988), Ethische Grundlagen aktueller angloamerikanischer Erziehungskonzepte. In C. Günter, L. Kerstiens, L. Mauermann, F. Pöggeler & H.-J. Werner, Hrg., *Ethik und Erziehung*. Stuttgart: Kohlhammer, S. 141-170.

May, M.A. & Hartshorne, H. (1926). First steps toward a scale for measuring attitudes. *The Journal of Educational Psychology, 17,* 145-162.

McGeorge, C. (1975). The susceptibility to faking of the Defining Issues Test of moral development. *Developmental Psychology, 11,* 108.

McGuire, W.J. (1966). The current status of cognitive consistency theories. In S. Feldman, Hrg., *Cognitive consistency.* New York: Academic Press.

Meehan, K. A., Woll, S. B., & Abbott, R. D. (1979). The role of dissimulation and social desirability in the measurement of moral reasoning. *Journal of Research and Personality, 13*, 25-38.

Meeus, W. & Raaijmakers, Q. (1989). Autoritätsgehorsam in Experimenten des Milgram-Typs: Eine Forschungsübersicht. *Zeitschrift für Sozialpsychologie, 20,* 70-85.

Milgram, S. (1974). *Obedience to authority*. New York.

Mills, C.W. (1956). *White Collar*. New York: Oxford University Press.

Mittelstraß, J. (1992). *Leonardo-Welt. Über Wissenschaft, Forschung und Verantwortung*. Frankfurt: Suhrkamp.

Moers, M. (1930). Zur Prüfung des sittlichen Verständnisses Jugendlicher. *Zeitschrift für angewandte Psychologie 34,* 431-511.

Moon, Y.-L. (1986). A review of cross-cultural studies on moral judgment development using the Defining Issues Test. *Behavior Science Research 22,* 147-177.

Mowrer, O.H. (1972). Conscience and the unconscious. In R.C. Johnson, P.R. Dodecki & O.H. Mowrer, Hrg., *Conscience, Contract and Social Reality.* New York: Holt, S. 349-371.

Murphy, J.M. & Gilligan, C. (1980). Moral development in the late adolescence and adulthood. A critique and reconstruction of Kohlberg's theory. *Human Development 23,* 77-104.

Nagl, W., Walter, H.-H. & Staud, J. (1986). *Statistische Verfahren der empirischen Sozialforschung in einem Programmpaket. Das Konstanzer Statistische Analysesystem KOSTAS.* Zentrum I Bildungsforschung. Universität Konstanz.

Neill, A.S. (1969). *Theorie und Praxis der antiautoritären Erziehung. Das Beispiel Summerhill.* Reinbek b. Hamburg: Rowohlt (Original 1960).

Neisser, U. (1966). *Cognitive Psychology.* New York: Appleton-Century-Crofts.

Neisser, U. (1980). On 'social knowing'. *Personality and Social Psychology Bulletin 6,* 601-604.

Newcomb, Th.M. (1974). What does college do for a person? Frankly very little. *Psychology Today, 1974.*

Niemczyński, A., Czyżowska, D., Pourkos, M. & Mirski, A. (1988). The Cracow study with Kohlberg's moral judgment interview: Data pertaining to the assumption of cross-cultural validity. *Polish Psychological Bulletin 19,* 43-53.

Norman, D.A. (1980). Twelve issues for cognitive science. *Cognitive Science, 4,* 1-32.

Nunner-Winkler, G. (1978). Probleme bei der Messung des moralischen Urteils mit standardisierten Verfahren. In L.H. Eckensberger, Hrg., *Entwicklung des moralischen Urteilens.* Saarbrücken: Universitätsdruck, S. 337-358.

Nunner-Winkler, G. (1980). *Zugang zum moralischen Urteilen: Qualitative Forschung oder standardisierte Verfahren?* Max-Planck-Institut, Starnberg: Manuskript.

Nunner-Winkler, G. (1987). Was bedeutet Kohlbergs Theorieansatz für die moderne bildungspolitische Situation in der Bundesrepublik? In G. Lind & J. Raschert, Hrg., *Moralische Urteilsfähigkeit. Eine Auseinandersetzung mit Lawrence Kohlberg über Moral, Erziehung und Demokratie*. Weinheim: Beltz, S. 16-24.

Nunner-Winkler, G. (1989a). Wissen und Wollen. Ein Beitrag zur frühkindlichen Moralentwicklung. In A. Honneth, Th. McCarthy, C. Offe & A. Wellmer, Hrg., *Zwischenbetrachtungen. Im Prozeß der Aufklärung*. Frankfurt: Suhrkamp, S. 574-600.

Nunner-Winkler, G. (1989b). Verantwortung für andere: Das Problem der 'positiven' Pflichten. In G. Lind & G. Pollitt-Gerlach, Hrg., *Moral in 'unmoralischer' Zeit. Zu einer partnerschaftlichen Ethik in Gesellschaft und Erziehung.* Heidelberg: Asanger, S. 33-54.

Nunner-Winkler, G. (1990a). *The moral learning process: cognitive and motivational learning.* Paper presented at the Ringberg meeting.

Nunner-Winkler, G. (1990b). *Altruism as a basis of morality?* Paper prepared for the International Symposion: Moral Education for the New Millennium, Notre Dame University, Nov. 7-10, 1990.

Nunner-Winkler, G. & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development, 59*, 123-133.

Oesterreich, G. (1989). Handlungsfähigkeit: Über die Wurzeln von Angst, Aggression und Kriminalität. In S. Bäuerle, Hrg., *Kriminalität bei Schülern*. Stuttgart: Verlag für angewandte Psychologie, S. 36-63.

Oser, F. (1981a). *Moralisches Urteil in Gruppen, Soziales Handeln, Verteilungsgerechtigkeit – Stufen der interaktiven Entwicklung und ihre erzieherische Stimulation.* Frankfurt: Suhrkamp.

Oser, F. (1981b). Die Theorie von Lawrence Kohlberg im Kreuzfeuer der Kritik – eine Verteidigung. *Bildungsforschung und Bildungspraxis, 3,* 51-64.

Oser, F. (1986). Moral education and values education: the discourse perspective. In M.C. Wittrock, Hrg., *Handbook of Research on Teaching. Third Edition.* New York: Macmillan.

Oser, F. (1987a). Können Lehrer durch ihr Studium Experten werden? Ein Reformkonzept der Lehrerbildung. *Zeitschrift für Pädagogik ,33,* 805-822.

Oser, F. (1987b). Möglichkeiten und Grenzen der Anwendung des Kohlbergschen Konzept der Erziehung in unseren Schulen. In G. Lind & J. Raschert, Hrg., *Moralische Urteilsfähigkeit. Eine Auseinandersetzung mit Lawrence Kohlberg über Moral, Erziehung und Demokratie*. Weinheim: Beltz, S. 44-53.

Oser, F. (1988a). Verlust im Gewinn: Biographie und Determination des Entwicklungsmodells von Lawrence Kohlberg. *Schweizerische Zeitschrift für Psychologie, 47*, 109-119.

Oser, F. (1988b). Die gerechte Gemeinschaft und die Demokratisierung der Schulwelt: Der Kohlbergansatz, eine Herausforderung für die Erziehung. *Vierteljahresschrift für wissenschaftliche Pädagogik, 1*, 59-79.

Oser, F. (1990). *Teaching and professional morality: A discourse approach.* Paper presented at the International Symposium "Research on Effective and Responsible Teaching," Fribourg, September 3-7, 1990.

Oser, F. & Althof, W. (1989). *Moralische Entwicklung und Erziehung.* Studienbrief der FernUniversität Hagen (Pädagogisches Institut Freiburg, Schweiz).

Oser, F. & Althof, W. (1992). *Moralische Selbstbestimmung. Modelle der Entwicklung und Erziehung im Wertebereich.* Stuttgart: Klett-Cotta.

Oser, F. & Schläfli, A. (1986). Und sie bewegt sich doch. Zur Schwierigkeit der stufenmäßigen Veränderung des moralischen Urteils am Beispiel von Schweizer Banklehrlingen. In F. Oser, R. Fatke & Höffe, O., Hrg., *Transformation und Entwicklung*. Suhrkamp, S. 217-251.

Peck, R.F. & Havighurst, R.J. (1962). *The psychology of character development.* New York: Wiley.

Perret-Clermont, A.-N. (1990). *Social psychology of transmission: Negotiations of knowledge in relationships between teacher and students.* Paper presented at the International Symposium "Research on Effective and Responsible Teaching," Fribourg, September 3-7, 1990.

Perry, W.G. (1970). *Forms of Intellectual and Ethical Development in the College Years.* New York: Holt.

Piaget, J. (1951). *Play, Dreams and Imitation in Childhood.* New York: Norton (Original 1928).

Piaget, J. (1973a). *Das moralische Urteil beim Kinde*. Frankfurt: Suhrkamp (Original 1932).

Piaget, J. (1973b). *Erkenntnistheorie der Wissenschaften vom Menschen.* Frankfurt: Ullstein (Original 1970).

Piaget, J. (1976). *Psychologie der Intelligenz.* München: Kindler.

Pittel, S.M. & Mendelsohn, G.A. (1966). Measurement of moral values: a review and critique. *Psychological Bulletin 66,* 22-35.

Plack, A. (1969). *Die Gesellschaft und das Böse. Eine Kritik der herrschenden Moral (Vierte Auflage).* München: List.

Popper, K.R. (1968). *The Logic of Scientific Discovery.* London: Hutchinson (Original 1934).

Portele, G. (1978). Du sollst das wollen! Zum Paradox der Sozialisation. In G. Portele, *Sozialisation und Moral.* Weinheim: Beltz, S. 147-168.

Portele, G. (1985). How scientists think about science and morality. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral development and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education*. Chicago: Precedent Publishing Inc., S. 193-204.

Posner, M. (1976). *Kognitive Psychologie*. München: Juventa.

Power, C., Higgins, A. & Kohlberg, L. (1989). *Lawrence Kohlberg's approach to moral education.* New York: Columbia University Press.

Presley, S.L. (1985). Moral judgment and attitude toward authority of political resisters. *Journal of Research in Personality 19,* 135-151.

Raths, L.E., Harmin, M. & Simon, S. (1978). *Values and Teaching.* (2nd edition) Columbus: Charles E. Merrill.

Raven, J. (1973). The attainment of non-academic educational objectives. *International Review of Education 19,* 305-344.

Reichenbach, H. (1969). Die Suche nach ethischen Leitsätzen und der kognitivethische Parallelismus. In H. Albert & E. Topitsch, Hrg., *Werturteilsstreit.* Darmstadt: Wissenschaftliche Buchgesellschaft, S. 455-471.

Reinhardt, S. (1980). Moralisches Urteil im politischen Unterricht. *Gegenwartskunde 29,* 449-460.

Reinhardt, S. (1990). *Notes on the morality of teaching: How to handle conflicts.* Poster presented at the International Symposium "Research on Effective and Responsible Teaching," Fribourg, September 3-7, 1990.

Rest, J.R. (1973). The hierarchical pattern of moral judgment: A study of patterns of comprehension and preference of moral stages. *Journal of Personality, 41,* 86-109.

Rest, J.R. (1974). Developmental psychology as a guide to value education: A review of "Kohlbergian" programs. *Review of Educational Research 44,* 241-259.

Rest, J.R. (1979). *Development in judging moral issues.* Minneapolis, MI: University of Minnesota Press.

Rest, J.R. (1984). The major components of morality. In W.M. Kurtines & J.L. Gewirtz, Hrg., *Morality, Moral Behavior, and Moral Development.* New York: Wiley, S. 24-38.

Rest, J.R. (1986a). Ein interdisziplinärer Ansatz zur Moralerziehung und eine Vierkomponenten-Modell der Entstehung moralischer Handlungen. In F. Oser, W. Althof & D. Garz, Hrg., *Moralische Zugänge zum Menschen. Zugänge zum moralischen Menschen.* München: Peter Kindt Verlag, S. 20-41.

Rest, J.R. (1986b). *Moral Development. Advances in Research and Theory.* New York: Praeger.

Rest, J.R. (1988). Why does college promote development in moral judgment? *Journal of Moral Education, 17,* 183-194.

Rest, J.R., Thoma, S.J., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology, 89*(1), 5-28.

Rest, J.R., Narvaez, D., Bebeau, M.J. & Thoma, S.J. (1999). Postconventional moral thinking. A Neo-Kohlbergian approach. Mahwah, NJ: Erlbaum.

Ringer, F. (1969). *The Decline of German Mandarins: The German Academic Community 1890-1933.* Cambridge MA: Harvard University Press.

Rosenberg, S. (1988). *Reasons, Ideology and Politics.* Princeton, NJ: Princeton University Press.

Rousseau, J.-J. (1971). *Emil oder Über die Erziehung.* Paderborn: Schöningh (Original 1762).

Rulon, B. (1990). *The Just Community as a model for staff development: The socio-moral complexity of the workplace.* Paper presented at the International Symposium "Research on Effective and Responsible Teaching," Fribourg, September 3-7, 1990.

Russell, B. (1972). *A History of Western Philosophy.* New York: Simon and Schuster (Original 1945).

Sandberger, J.-U., Jetten, E., Južnič, Langer, J., Bromberek, B., Framhein, G., Percy, K.A. & Walter, H.G. (1982). Between uncertainty and commitment - A comparison of first semester university students' occupational perspectives and values in five European countries (FORM-Project). In M. Nießen & J. Peschar, Hrg., *Comparative research on education.* Oxford: Pergamon, S. 79-122.

Schachter, S. & Singer, J.E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69,* 379-399.

Schirp, H. (1990). Förderung der Urteilsfähigkeit durch Öffnung von Schule – Gestaltung des Schullebens durch Förderung der Urteilsfähigkeit. In U. Steffens & T. Bargel, Hrg., Arbeitskreis "Qualität von Schule". Heft 7: *Erziehung und Demokratie in der Schule – Konzepte und Erfahrungen zum Ansatz von Kohlberg und Oser.* Wiesbaden: HIBS.

Schläfli, A., Rest, J.R. & Thoma, S.J. (1985). Does moral education improve moral judgment? A meta-analysis of intervention studies using the Defining Issues Test. *Review of Education Research 55,* no. 3, 319-352.

Schmied, D. (1981). Standardisierte Fragebogen zur Erfassung des Entwicklungsstandes der moralischen Urteilskompetenz. *Diagnostica, 27,* 51-65.

Schmitt, M. (1982). *Zur Erfassung des moralischen Urteils: Zwei standardisierte objektive Verfahren im Vergleich.* Universität Trier: Bericht.

Schmitt, M. (1988). *Über die Konsistenzannahme des Eigenschaftsmodells und über Konsistenz als Eigenschaft: Theoretische, methodische und empirische Untersuchungen.* Universität Trier: Fachbereich I – Psychologie.

Schmitt, M. (1989). *Ipsative Konsistenz (Kohärenz) als Profilähnlichkeit.* Universität Trier: Trierer Psychologische Berichte, Band 16, Heft 2.

Schreiner, G. (1981). Politische Bildung und moralische Erziehung. *Deutsche Schule 7/8,* 395-414.

Schreiner, G. (1989). Demokratische Erziehung muß über den Horizont der Kleingruppe hinausgehen! Zu einem Vergleich des Team-Kleingruppen-Modells mit der Gerechten Schulgemeinschaft. In G. Lind & G. Pollitt-Gerlach, Hrg., *Moral in 'unmoralischer' Zeit. Zu einer partnerschaftlichen Ethik in Erziehung und Gesellschaft.* Heidelberg: Asanger, S. 129-152.

Schreiner, G. (1992). Was kann die Kohlberg-Theorie für eine pädgogische Handlungstheorie leisten? *Die Deutsche Schule 4/92,* 423-435.

Scott, W.A. (1968). Attitude Measurement. In G. Lindzey & E. Aronson, *Handbook of Social Psychology.* Reading MA: Addison-Wesley, S. 205-270.

Seitz, C. (1991). *Demokratiefähigkeit und moralisches Urteilen bei Schülern - Ein Vergleich von einheimischen und Aussiedlerkindern.* Konstanz: Diplomarbeit im Fach Psychologie.

Senger, R. (1985). Segmentation of soldiers' life world. In G. Lind, H.A. Hartmann & R. Wakenhut, Hrg., *Moral development and the Social Environment.*

Studies in the Philosophy and Psychology of Moral Judgment and Education. Chicago: Precedent Publishing Inc., S. 221-242.

Shaw, M. & Wright, J. (1967). *Scales for the Measurement of Attitudes.* New York: McGraw-Hill.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the New Reform. *Harvard Educational Review 57,* 1-22.

Shweder, R. (1982). Review of Lawrence Kohlberg's "Essay on moral development, Volume I.: The philosophy of moral development." Liberalism as destiny. *Contemporary Psychology, June 1982,* 421-424.

Sidanius, J. (1985). Cognitive functioning and sociopolitical ideology revisited. *Political Psychology 6,* 637-661.

Simon, H. (1982). Comments. In M. Clark & S. Fiske, Hrg., *Affect and cognition.* S. 333-342. Hillsdale, NJ: Erlbaum.

Simpson, E.L. (1974). Moral development research: A case study of scientific cultural bias. *Human Development, 17,* 81-106.

Skinner, B.F. (1938). *The Behavior of Organism.* New York: Appleton-Century-Crofts.

Snarey, J. (1985). The cross-cultural universality of socio-moral development: A critical review of Kohlbergian Research. *Psychological Bulletin, 97,* 202-232.

Stolz, G.E. (1987). *Das Zusammenwirken von Elternhaus, Schule und Freundschaftsgruppen in der Erziehung.* Weinheim: Deutscher Studienverlag.

Stuhr, D.E. (1978). Moral education with children: An examination of related studies. *Character Potential: A Record of Research, 8,* 182-187.

Sullivan, E.V. (1977). A study of Kohlberg's structural theory of moral development: A critique of liberal social science ideology. *Human Development 20,* 352-376.

Tesser, A. & Shaffer, D.R. (1990). Attitude and attidue change. *Annual Review of Psychology, 41,* 479-523.

Tocqueville, A. de (1976). *Über die Demokratie in Amerika.* München: Deutscher Taschenbuchverlag (Original 1835).

Tolman, E. (1932). *Purposive Behavior in Animals and Men.* New York: Century.

Tomlinson-Keasey, C. & Eisert, D.C. (1981). From a "Structure d'Ensemble" to separate organizations for cognitive and affective development. *Human Development, 5,* 1-19.

Torgerson, W.S. (1967). *Theory and Methods of Scaling.* New York: Wiley (Original 1958).

Turiel, E. (1983). *The development of social knowledge. Morality and convention.* Cambridge: University Press.

Vasudev, J. (1986). Kohlbergs Universalitätspostulat aus indischer Sicht. In W. Edelstein & G. Nunner-Winkler, Hrg., *Zur Bestimmung der Moral. Philosophische und sozialwissenschaftliche Beiträge zur Moralforschung.* Frankfurt: Suhrkamp, S. 145-177.

Villenave-Cremer, S. & Eckensberger, L. (1986). Zur Rolle affektiver Prozesse im moralischen Urteil. In F. Oser, W. Althof & D. Garz, *Moralische Zugänge zum Menschen -Zugänge zum moralischen Menschen.* München: Peter Kindt, S. 180-204.

Vilmar, F. (1973). *Strategien der Demokratisierung. Band I: Theorie der Praxis*. Neuwied: Luchterhand.

Vine, I. (1984). Moral maturity in socio-cultural perspective: Are Kohlberg's stages universal? In S. Modgil & C. Modgil, Hrg., *Lawrence Kohlberg: Consensus and Controversy*. Brighton GB: Falmer Press.

Wakenhut, R. (1978). *Über die Einbeziehung von Situationen in psychologische Messungen. Ein Beitrag zur interaktionistischen Persönlichkeitsforschung*. Frankfurt: Peter Lang.

Walker, L.J. (1986). Cognitive processes in moral development. In G.L. Sapp, Hrg., *Handbook of Moral Development.* Birmingham, AL: Religious Education Press, S. 109-145.

Walker, L.J., DeVries, R. & Bichard, S.L. (1984). The hierarchical nature of stages of moral development. *Developmental Psychology 20,* 960-966.

Wasel, W. (1994). *Simulation moralischer Urteilsfähigkeit. Moralentwicklung: eine kognitiv-strukturelle Veränderung oder ein affektives Phänomen*? Konstanz: Unveröffentlichte Diplomarbeit, Fachgruppe Psychologie, Universität Konstanz.

Watson, J.B. (1970). *Behaviorism.* New York: Norton (Orginal publiziert 1924).

Wechsler, D. (1961). *Die Messung der Intelligenz Erwachsener*. Deutsche Bearbeitung von A. Hardesty und H. Lauber. Bern: Huber.

Weiner, B. (1976). *Theorien der Motivation*. Stuttgart: Klett.

Weinreich-Haste, H. (1986). Kohlberg's contribution to political psychology: A positive view. In S. Modgil & C. Modgil, Hrg., *Kohlberg: Consensus and Controversy*. Brighton: Falmer Press.

Weinreich-Haste, H., Adams, C. & Clay, A. (1986). *Trying to be morally right or morally left: Testing the relationship between moral reasoning and political orientation.* Paper presented to the Annual Scientific Meeting of ISPP, Washington, Juni 1986.

Wellmer, A. (1986). *Ethik und Dialog.* Frankfurt: Suhrkamp.

Westmeyer, H., Hannemann, J., Nell, V., Völkel, U. & Winkelmann, K. (1987). Eine Monotheory-Multimethod Analyse: Plädoyer für einen deduktivistischen Multiplismus. *Diagnostica, 33,* 227-242.

Wischka, B. (1982). *Moralisches Bewußtsein und Empathie bei Strafgefangenen.* Universität Marburg: Unveröff. Diplomarbeit.

Wohlwill, J.F. (1977). *Strategien entwicklungspsychologischer Forschung.* Stuttgart: Klett.

Wolfle, (1963). Preface. In S. Koch, Hrg., *Psychology: A Study of a Science*. Formulations of the person and the social context, Volume III. McGraw-Hill: New York, S. v-vii.

Wottowa, H. (1990). Einige Überlegungen zu (Fehl-) Entwicklungen der psychologischen Methodenlehre. *Psychologische Rundschau, 41,* 84-107.

Wren, T. (1990). *Caring About Morality. Philosophical Perspectives in Moral Psychology.* Cambridge, MA: MIT-Press.

Yussen, S. (1977). Characteristics of moral dilemmas written by adolescents. *Developmental Psychology, 12,* 162-163.

Zajonc, R. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151-175.

Quellen für die analysierten Untersuchungen

Arbuthnot, J. (1975). Modification on moral judgment through role playing. *Developmental Psychology, 11,* 319-324.

Arbuthnot, J. (1984). Moral reasoning development programs in prison: Cognitive-developmental and critical reasoning approaches. *Journal of Moral Education, 13,* 113-123.

Beck, C.M., Sullivan, E. & Taylor, N. (1972). Stimulating transition to post-conventional morality: The Pickering High School today. *Interchange 3,* 28-37.

Berkowitz, M.W., Gibbs, J. & Broughton, J. (1980). The relation of moral judgment stage disparity to developmental effects of peer dialogues. *Merrill-Palmer Quarterly, 26 (4),* 341-357.

Bernier, J.E. & Rustad, K. (1977). Psychology of counseling curriculum: a follow-up study. *The Counseling Psychologist, 6,* 18-21.

Biskin, D. & Hoskisson, K. (1977). An experimental test of the effects of structural discussions of moral dilemmas found in children's literature on moral reasoning. *Elementary School Journal, 77,* 404-416.

Boyd, D. (1980). The condition of sophomoritis and its educational cure. *Journal of Moral Education, 10,* 24-39.

Cognetta, P. (1977). Deliberate psychological education: a high school cross-age teaching model. *The Counseling Psychologist, 6,* 22-24.

Copeland, T.F. (1979). *An attempt to enhance the moral development of offenders.* Kansas State University: Dissertation.

Copeland, T.F. & Parish, T.S. (1979). An attempt to enhance moral judgment of offenders. *Psychological Reports, 45,* 831-834.

Crockenberg, S.B. & Nicolayev, J. (1979). Stage transition in moral reasoning as related to conflict experience in naturalistic settings. *Merrill-Palmer Quarterly, 25,* 185-192.

Doornink, J.D. & Hamm, N.H. (1978). Generalized and non-generalized changes in moral judgment. *Journal of Genetic Psychology, 132,* 277-282.

Dowell, R.C. (1971). *Adolescents as peer counselors*. Cambridge, MA. Unpublished doctoral dissertation. Harvard University.

Enright, R.D. (1981). A classroom discipline model for promoting social cognitive development in early childhood. *Journal of Moral Education ,11,* 47-60.

Enright, R.D. & Hendel, D.D. (1979). An evaluation of growth in a university program. *Journal of Moral Education, 9,* 50-52.

Erickson, V.L. (1977). Deliberate psychological education for women: a curriculum follow-up study. *The Counseling Psychologist, 6,* 25-29.

Faust, D. & Arbuthnot, J. (1978). Relationship between moral and Piagetian reasoning and the effectiveness of moral education. *Developmental Psychology, 14,* 435-436.

Fleetwood, R. & Parish, T. (1976). Relationship between moral development test scores of juvenile delinquents and their inclusion in an moral dilemma discussion group. *Psychological Reports, 39,* 1075-1080.

Galbraith, R.E. & Jones, T.A. (1977). Teaching for moral reasoning in the Social Studies: a research report. *The Counseling Psychologist, 6,* 60-63.

Gibbs, J.C., Arnold, K.D., Ahlborn, H. & Cheesman, F.L. (1984). Facilitation of sociomoral reasoning in delinquents. *Journal of Consulting and Clinical Psychology, 52,* 37-45.

Harris, D.E. (1977). A curriculum sequence for moral development. *Theory and Research in Social Education, 5,* 1-21.

Hayden, B. & Pickar, D. (1981). The impact of moral discussion on children's level of moral reasoning. *Journal of Moral Education, 10,* 131-134.

Jennings, W.S. (1979). *The juvenile delinquent as a moral philosopher: the effect of rehabilitation programs on the moral reasoning and behavior of male youthful offenders.* Harvard University: Dissertation.

Jennings, W.S. & Kohlberg, L. (1983). Effects of a just community program on the moral development of youthful offenders. *Journal of Moral Education, 12,* 33-50.

Jensen, L.C. & Larm, C. (1970). Effects of two training procedures on intentionality in moral judgments among children. *Developmental Psychology, 2,* 310.

Keasey, C.B. (1973). Experimentally induced changes in moral opinions and reasoning. *Journal of Personality and Social Psychology, 26,* 30-38.

Lorimer, R. (1971). Change in the development of moral judgment in adolescence: the effect of a structured exposition vs a film and discussion. *Canadian Journal of Behavioral Science, 3,* 1-10.

Maitland, K.A. & Goldman, J.R. (1974). Moral judgment as a function of peer group interaction. *Journal of Personality and Social Psychology, 30,* 699-704.

Maqsud, M. (1982). Effects of Nigerian children's discussion on their moral progression. *Journal of Moral Education, 11 (3),* 181-187.

Matefy, R. & Acksen, B. (1976). The effect of role-playing discrepant positions on change in moral judgments and attitudes. *Journal of Genetic Psychology 128,* 189-200.

Maul, J.P. (1979). *A study of the moral atmosphere in and the development of moral reasoning in a school with intensive education.* Rutgers University, New Jersey: Dissertation.

Maul, J.P. (1980). A high school with intensive education: Moral atmosphere and reasoning. *Journal of Moral Education, 10,* 9-17.

Mosher, R. & Sullivan, P. (1976). A curriculum in moral education for adolescents. *Journal of Moral Education, 5,* 159-172.

Oja, S.N. & Sprinthall, N.A. (1978). Psychological and moral development for teachers? Can you teach old dogs? *Character Potential: A Record of Research, 8,* 218-225.

Piburn, M. (1977). Teaching about science and society: Moral judgment and the Prisoner's Dilemma. *Theory and Research in Social Education, 5,* 20-30.

Rholes, W., Bailey, S. & McMillan, L. (1982). Experiences that motivate moral development: The Role of cognitive dissonance. *Journal of Experimental Social Psychology, 18,* 524-536.

Rosenkoetter, L.J., Alderman, M., Nelson, R. & Ottaviano, M. (1982). An evaluation of the effect of moral dilemma discussions for sixth graders. *Journal of Early Adolescence, 2,* 75-82.

Rosenkoetter, L.J., Landman, S. & Mazak, S. (1980). Use of moral discussion as an intervention with delinquents. *Psychological Reports, 46,* 91-94.

Rustad, K. & Rogers, C. (1975). Promoting psychological growth in a high school class. *Counselor Education and Supervision, 14,* 227-285

Scharf, P. & Hickey, J. (1976). The prison and the innate's conception of legal justice experiment in democratic education. *Criminal Justice and Behavior, 3,* 107-122.

Selman, R. & Lieberman, M. (1975). An evaluation of a cognitive developmental curriculum on primary grade children. *Journal of Educational Psychology, 67,* 712-716.

Sequin-Tremblay, G. & Kiely, M. (1979). Development du jugement moral chez l'adolescente en reeducation. *Canadian Journal of Behavioral Science, 11,* 32-44.

Stanley, S.F. (1976). *A curriculum to affect the moral atmosphere of the family and the moral development of adolescents.* Boston University School of Education: Dissertation.

Stanley, S.F. (1978). Family education to enhance the moral atmosphere of the family and the moral development of adolescents. *Journal of Counseling Psychology, 25,* 110-118.

Sullivan, P.J. & Sullivan, M. (1980). Establishing moral education programs: A priority for guidance. *The Personnel and Guidance Journal,* 63-73.

Tracy, J. & Cross, H. (1973). Antecedents of shift in moral judgment. *Journal of Personality and Social Psychology, 26,* 238-244.

Walker, L.J. (1982). The sequentiality of Kohlberg's stages of moral development. *Child Development, 53,* 1330-1336.

Walker, L.J. (1983). Sources of cognitive conflict for stage transition in moral development. *Developmental Psychology, 19,* 103-110.

Walker, L.J. & Richards, B. (1979). Stimulating transitions in moral reasoning as a function of stage of cognitive development. *Developmental Psychology, 15,* 95-103.

Zahner, C. & McDavis, R. (1980). Moral development of professional and paraprofessional counselors and trainees. *Counselor Education and Supervision, 19,* 243-251.

Anhang

Moralische Atmosphäre Fragebogen (MAF)[162]

In diesem Fragebogen stellen wir Dir wieder eine ganze Reihe Fragen. Diesmal geht es allerdings nicht mehr um Deine Meinung zu irgendwelchen Geschichten. Wir wollen gerne mehr darüber erfahren, wie Du Deine Schule und Deine Lehrer oder Lehrerinnen siehst. Dann interessiert uns auch noch Deine Meinung über Mogeln oder über gegenseitiges Helfen.
Wie das hier abläuft, weißt Du ja nun schon.
Deshalb nur noch mal GANZ KURZ

Beantworte bitte ALLE Fragen.
Verstehst Du eine Frage nicht, dann schreib das einfach daneben! Willst du sonst noch etwas dazu sagen, kannst Du das auch daneben schreiben! Auch hier wieder nur EINE Antwort ankreuzen. Findest du eine andere Antwort genauso richtig, dann schreib das bitte auch einfach neben die Frage.

Wir werden Dich auch mal nach der Meinung Deiner Mitschüler oder deiner Lehrer fragen. Es macht dann nichts, wenn Du das nicht so genau weißt. Wir interessieren uns nämlich dafür, was die anderen DEINER MEINUNG nach sagen würden. Wenn Du mal etwas nicht so genau weißt, dann kreuze das an, was Deiner Meinung am nächsten kommt.

Nun brauchen wir wieder Deine Geheimnummer!
Schreib also Deinen Namen nicht auf den Fragebogen, sondern mach' bitte wieder die folgenden Angaben:

In welchem Monat bist Du geboren? .....
Erster Buchstabe des Vornamens Deiner Mutter: .....
Erster Buchstabe des Vornamens Deines Vaters: .....

o Mädchen o Junge

162 MAF Copyright (1987) G. Lind & L. Link

```
                                  Stimmt           Stimmt
                                  gar nicht  voll und ganz
Die Schüler und Schülerinnen
an unserer Schule gehen freund-   (1) (2) (3) (4) (5)
lich miteinander um.
```

A. *In diesem Teil des Fragebogens findest Du viele Aussagen über das Leben in der Schule. Da steht zum Beispiel:*
Wenn Du glaubst, das "stimmt voll und ganz", dann kreuze bitte die "5" an. Denkst Du das "stimmt gar nicht", dann mußt Du die "1" ankreuzen. Liegt deine Meinung irgendwo dazwischen, mußt Du dich für "2", "3" oder "4" entscheiden.

Hier geht es um die Schüler und Schülerinnen

A1 Die Schüler und Schülerinnen an unserer Schule gehen freundlich miteinander um.
"Stimmt gar nicht / Stimmt nur wenig / Stimmt teils-teils / Stimmt weitgehend / Stimmt voll und ganz" (1-5)

A2 Sie helfen sich auch dann gegenseitig, wenn sie nicht befreundet sind.

A3 Sie kommen auch dann miteinander aus, wenn sie verschiedenen Gruppen angehören.

A4 Es gibt Gruppen, die gegenüber anderen zusammenhalten. Da muß man zur Gruppe gehören, um Hilfe zu bekommen.

Hier geht es um die Lehrer und Lehrerinnen

A5 Die Lehrerinnen und Lehrer an unserer Schule gehen mit uns meistens gerecht um und achten uns.
"Stimmt gar nicht / Stimmt nur wenig / Stimmt teils-teils / Stimmt weitgehend / Stimmt voll und ganz" (1-5)

A6 Sie sprechen untereinander und mit uns offen über Probleme. Sie suchen mit uns nach Lösungen.

A7 Sie wollen die Schulordnung immer durchsetzen.

A8 Die Schulordnung ist im großen und ganzen gut und gerecht.

Hier geht es um die Schule

A9 Wir Schülerinnen und Schüler können Verhaltensregeln und Entscheidungen in der Schule mitbestimmen.
"Stimmt gar nicht / Stimmt nur wenig / Stimmt teils-teils / Stimmt weitgehend / Stimmt voll und ganz" (1-5)

A10 Manche Entscheidungen werden abgestimmt. Dabei dürfen alle mitmachen. Jeder hat eine Stimme.

A11 Wir fühlen uns als Teil der Schule und sind stolz auf unsere Schule.

A12 Unsere Schule ist eine richtige Gemeinschaft. Alle kümmern sich um einander.

A13 Man kann hier viel lernen und bekommt eine gute Ausbildung.

A14 In unserer Schule wollen einige immer besser sein als andere.

A15 Es wird wenig abgeschrieben.

A16 Es gibt selten Schlägereien.

A17 Es wird nur selten geschwänzt.

A18 Es wird wenig gestohlen.

A19 In unserer Schule wird kaum etwas zerstört.

B. *In diesem Abschnitt findest Du nochmals die gleichen Aussagen. Diesmal interessiert uns allerdings etwas anderes. Wie WICHTIG sind Dir diese Dinge für Deine Schule?*

B1 Die Schülerinnen und Schüler sollten freundlich miteinander umgehen.
Für mich ist das:
"Unwichtig / Wenig wichtig / Teils-teils / Wichtig / Sehr wichtig" (1-5)

B2 Sie sollten sich auch dann gegenseitig helfen, wenn sie nicht miteinander befreundet sind.

B3 Sie sollten auch miteinander auskommen, wenn sie zu verschiedenen Gruppen gehören.

B4 Es sollte für mich eine Gruppe geben, in der man gegenüber anderen zusammenhält.

B5 Die Lehrerinnen und Lehrer sollten mit uns gerecht und respektvoll umgehen.

B6 Alle sollten miteinander offen über Probleme sprechen.

B7 Die Vorschriften der Schule sollten gerecht und gut sein.

B8 Die Lehrerinnen und Lehrer sollten die Schulordnung immer durchsetzen.
B9 Wir Schülerinnen und Schüler sollten Verhaltensregeln und Entscheidungen in der Schule mitbestimmen können.
B10 Wichtige Entscheidungen sollten abgestimmt werden und alle Lehrer und Schüler sollten dabei eine Stimme haben.
B11 Wir sollten uns der Schule zugehörig fühlen können und stolz auf unsere Schule sein können.
B12 Unsere Schule sollte eine richtige Gemeinschaft sein, in der sich alle um einander kümmern.
B13 Man sollte hier viel lernen können und eine gute Ausbildung bekommen.
B14 Es sollte bei uns wenig Konkurrenz untereinander geben.
B15 Es sollte wenig abgeschrieben und gemogelt werden.
B16 Es sollte selten Schlägereien geben.
B17 Es sollte nur selten geschwänzt werden.
B18 Es sollte nur selten gestohlen werden.
B19 In unserer Schule sollte nichts zerstört werden.

C. *Lies Dir bitte die Aussagen noch einmal kurz durch. Vielleicht sind da Themen bei, die Du gerne einmal in der Klasse oder in einer Schulversammlung besprechen willst. Schreibe sie hier bitte auf. Du kannst natürlich auch andere Themen nennen, die Du für interessant und wichtig hältst.* .....(Forts. Rückseite)

D. *Mitbestimmung in der Schule – Was würdest du davon halten?*
D1 Solltet Ihr darüber mitbestimmen, was in der Schule erlaubt und verboten wird?
Ich finde das: “Ganz falsch / Eher falsch / Teil-teils / Eher richtig / Ganz richtig”
D2 Solltet Ihr auch mitbestimmen können, was bei Verstößen gegen die Schulordnung erfolgen soll?
D3 Sollten Schülerinnen und Schüler mitbestimmen können, wie jemand bestraft wird, wenn er gegen die Schulordnung verstoßen hat?

E. *An jeder Schule gibt es viele feste Gruppen. Die Schülerinnen oder Schüler, die zu so einer Gruppe gehören, sind oft zusammen und machen viel gemeinsam. Das hat verschiedene Gründe. Vielleicht haben sie gemeinsame Interessen, machen Sport im selben Verein oder vielleicht kommen sie aus dem gleichen Viertel oder der gleichen Gegend.*

E1 Gibt es in Deiner Klasse solche Gruppen?
"Ja / Nein" (1-2)

E2 Gehörst Du zu so einer Gruppe?
"Ja, zu einer / Ja, zu mehreren / Nein" (1-3)

Wie häufig sprichst Du über Deine Hausaufgaben oder was Du in der Schule lernst

E31 ... mit Schülern oder Schülerinnen aus Deiner Klasse oder Deinem Jahrgang?
"Nie / Selten / Manchmal / Häufig / Sehr häufig" (1-5)

E32 ... mit anderen Schülern oder Schülerinnen?

E33 ... mit Lehrerinnen oder Lehrern?

E34 ... mit Deinen Eltern?

Wie häufig sprichst du über persönliche Dinge

E41 ... mit Schülern oder Schülerinnen aus Deiner Klasse oder aus Deinem Jahrgang? "Nie / Selten / Manchmal / Häufig / Sehr häufig" (1-5)

E42 ... mit anderen Schülern oder Schülerinnen?

E43 ... mit Lehrerinnen oder Lehrern?

E44 ... mit Deinen Eltern?

F. *Es gibt Regeln, was alles bei Klassenarbeiten verboten und erlaubt ist. Ihr alle wißt zum Beispiel:*

VERBOTEN ist das ABSCHREIBEN und MOGELN

Aber, Ihr wißt, manche mogeln natürlich trotzdem und erhalten dadurch eine bessere Note. Wir wollen jetzt Deine Meinung zum Thema "MOGELN" hören. Antworte auch bei diesem Thema bitte ehrlich!
Wie stehst Du selbst zum Mogeln und Abschreiben und wie sehen das wohl deine Mitschüler und Mitschülerinnen?

F1 Ich halte das für:
"Ganz falsch / Eher falsch / Teil-teils / Eher richtig / Ganz richtig" (1-5)

Wie sehen das wohl Deine Mitschülerinnen und Mitschüler?

F2 Ich glaube, die meisten aus meiner Klasse halten das für:
"Ganz falsch / Eher falsch / Teil-teils / Eher richtig / Ganz richtig" (1-5)

Sollte man bei einer Klassenarbeit jemandem helfen und abschreiben lassen?

F3 Ich halte das für:
"Ganz falsch / Eher falsch / Teil-teils / Eher richtig / Ganz richtig" (1-5)

F4 Ich glaube, die meisten aus meiner Klasse halten das für:
"Ganz falsch / Eher falsch / Teil-teils / Eher richtig / Ganz richtig" (1-5)

F5 Würdest Du selbst bei einer Klassenarbeit mogeln?
"Klar! / Nein! / Weiß nicht" (1-3)

F6 Würden das die meisten anderen wohl tun?
"Klar! / Nein! / Weiß nicht" (1-3)

F7 Würdest Du anderen beim Mogeln helfen?
"Klar! / Nein! / Weiß nicht" (1-3)

F8 Würden wohl die meisten anderen beim Mogeln helfen?
"Klar! / Nein! / Weiß nicht" (1-3)

G. *Ati* und *Bugi*

Oft kommt es auf die Umstände an, ob man jemanden abschreiben läßt. Stelle Dir bitte einmal die folgende Situation vor:
Du hast zwei Freunde oder Freundinnen in Deiner Klasse. Wir wollen sie mit ihrem Spitznamen Ati und Bugi nennen. Beide magst Du eigentlich gleich gern. Beide stehen in einem Fach zwischen 4 und 5. Du hast in diesem Fach allerdings keine Probleme und stehst recht gut. Da beide alleine nicht zurechtkommen, bietest Du ihnen Deine Hilfe an. Vor der nächsten Klassenarbeit willst Du mit ihnen zusammen lernen.

Ati lehnt Dein Angebot ab und hat keine Lust vorher so viel zu lernen.
Bugi geht auf Dein Angebot ein und paukt zusammen mit Dir für die Arbeit. Bei der Klassenarbeit setzen sich nun beide in Deine Nähe. Während der Arbeit haben beide mehr und mehr Schwierigkeiten und können die Aufgaben nicht lösen. Sie geben Dir zu verstehen, daß sie von Dir abschreiben wollen. Die Lehrerin beobachtet Euch in diesem Moment gerade nicht.
Zunächst zu Ati:

G1 Würdest Du *Ati* abschreiben lassen, obwohl Ati doch keine Lust hatte, mit dir zu lernen?
"Ja / Nein"

G2 Hättest Du ein schlechtes Gefühl, es so zu machen?
"Ja / Etwas / Nein"
Und nun zu *Bugi*:

G3 Würdest Du Bugi abschreiben lassen? Bugi kommt ja trotz "Paukerei" in der Arbeit nicht klar. "Ja / Nein"

G4 Hättest Du ein schlechtes Gefühl dabei, so zu handeln? "Ja / Etwas / Nein"

H. *Freunde und Freundinnen helfen sich meistens gegenseitig. Manchmal gibt es aber einzelne Schülerinnen und Schüler, die zu keinem Freundeskreis gehören und die in der ganzen Klasse unbeliebt sind. Sollte man denen genauso helfen?*

Stelle Dir dazu folgende Situation vor: Chris ist ein ziemlich unbeliebter Mitschüler und hat in seiner Klasse keinen Freund. In einem Fach, in dem Du besonders gut bist, steht Chris "mangelhaft". Er muß die Klasse wiederholen, wenn das so weitergeht. Alle anderen kommen in diesem Fach gut mit.

Die Lehrerin bittet Euch nun, dem Chris beim Lernen zu helfen. Er soll doch auch, wie die anderen, das Klassenziel erreichen.

H1 Sollte man Chris in dieser Situation helfen? "Ja / Nein / Weiß nicht"

H2 Würdest Du ihm in diesem Fall auch wirklich helfen? "Ja / Nein / Weiß nicht"

H3 Was glaubst Du, tun Deine Mitschülerinnen und Mitschüler? Wie viele würden dem Chris wohl in dieser Situation helfen? "Keiner / Wenige / Einige / Viele"

H4 Jetzt denken wir den Fall einmal ein wenig weiter. Nimm an, Du hast Chris nicht geholfen. Der ist nun sitzen geblieben und muß die Klasse wiederholen. Wie stark würdest Du Dich dafür verantwortlich fühlen? "Gar nicht / Etwas / Teilweise / Stark / Sehr stark"

I. *Wir haben schon einmal bei anderen Schülern und Schülerinnen herumgefragt:*

"Sollte man einem unbeliebten Mitschüler helfen?"

Einige Ansichten dazu wollen wir Dir hier auf der nächsten Seite zeigen.

I1 Das ist pure Gefühlsduselei, wenn man einem hilft, mit dem man nicht befreundet ist!

"Das sehe ich auch so / Das sehe ich nicht so" (1-2)

I2 Es ist mir egal, ob unbeliebten Schülern geholfen wird oder nicht!

I3 Ich finde es *für mich* besser, auch unbeliebten Schülern zu helfen. Was andere aber in dieser Situation tun, das geht mich nichts an!

I4 Ich ärger mich und bin enttäuscht, wenn einem unbeliebten Mitschüler nicht geholfen wird. Ich weiß allerdings nicht, wie man das ändern kann!

I5 Ich versuche die anderen zu überzeugen. Sie sollen auch unbeliebten Schülern helfen!

I6 Ich wünsche mir, daß sich die Lehrer darum kümmern. Sie sollten dafür sorgen, daß solchen Schülern geholfen wird!

I7 Ich wünsche mir, daß Lehrer und Schüler dieses Thema miteinander besprechen. Vielleicht finden wir dann zusammen eine Lösung!

J. *Wie stehen die anderen dazu?*
Was glaubst Du, denken Deine Mitschülerinnen und Mitschüler?

J1 Sagen viele, daß man einem Mitschüler nicht helfen soll, wenn der in der Klasse unbeliebt ist?

J2 Ist vielen egal, ob unbeliebten Mitschülern geholfen wird?

J3 Wieviele Deiner Mitschülerinnen und Mitschüler setzen sich dafür ein, daß auch unbeliebten Schülern geholfen wird?

J4 Haben wohl viele Interesse daran, das Thema mit Euren Lehrerinnen und Lehrern zu besprechen?

Was glaubst Du, wie das Deine Lehrerinnen und Lehrer sehen?

J5 Wievielen Deiner Lehrerinnen und Lehrern ist es egal, ob ihr unbeliebten Mitschülern in einer Notlage helft oder nicht?

J6 Wieviele Lehrerinnen und Lehrer verlangen von Euch, solchen Mitschülern zu helfen?

J7 Was glaubst Du? Wieviele Deiner Lehrerinnen und Lehrer möchten das Thema wohl gerne gemeinsam mit Euch besprechen?

K. *Wir haben Schülerinnen und Schüler gefragt: "Welche Gründe sprechen dafür, dem Chris zu helfen. Einige Gründe, die uns genannt wurden, haben wir hier aufgeschrieben. Uns interessiert nun, was Du von diesen Gründen hältst.*
Kreuze bitte wieder an, wie weit die Aussage für dich stimmt!

K1 Es ist eine Sache des Anstandes und der guten Manieren, ob man auch unbeliebten Mitschülern in einer Notlage hilft.
"Stimmt gar nicht / Stimmt nur wenig / Stimmt teil-teils / Stimmt weitgehend / Stimmt voll und ganz" (1-5)

K2 Die Klasse ist eine Gemeinschaft. Dazu gehören alle, auch die, die man nicht so gut leiden kann. Deshalb muß man auch allen helfen.

K3 Vielleicht kommt man selbst einmal in eine Notlage. Es ist daher im eigenen Interesse, wenn sich Schülerinnen und Schüler gegenseitig helfen.

K4 Wenn ich einem unbeliebten Schüler in einer Notlage helfe, dann habe ich ein besseres Gewissen.

K5 Damit sich alle wohlfühlen können ist es richtig, in einer Notlage zu helfen.

K6 Wenn der Lehrer oder die Lehrerin das will, sollte man einem unbeliebten Mitschüler helfen.

L. Nun einige Gründe die *dagegen* sprechen, dem Chris zu helfen.

L1 Wenn ich einem unbeliebten Schüler helfe, werde ich vielleicht deshalb von den anderen abgelehnt. [Antwortmöglichkeiten wie oben bei Frage K]

L2 Der Chris hat sich bei den anderen unbeliebt gemacht. Er muß die Konsequenzen nun auch selber tragen.

L3 Man muß einem unbeliebten Mitschüler nur helfen, wenn es der Lehrer oder die Lehrerin verlangt.

L4 Wenn man anderen hilft, dann macht man sie vielleicht nur abhängig. Darauf sollte man achten.

L5 Schüler, die sich unbeliebt machen, gehören nicht zur Klassengemeinschaft. Sie können daher auch keine Hilfe erwarten.

L6 Es ist unmöglich, zu allen eine gute Beziehung zu haben. Man kann deshalb nur seinen Freunden helfen.

M. *Einige sagen: "Es ist ungerecht, einem Mitschüler nicht zu helfen, bloß weil man ihn nicht leiden kann!" Wie findest Du ihre Gründe?*

M1 Es ist ungerecht, weil man mit jedem in Not Mitleid haben sollte.
"Stimmt gar nicht / Stimmt nur wenig / Stimmt teil-teils / Stimmt weitgehend / Stimmt voll und ganz" (1-5)

M2 Es ist ungerecht, weil jeder Mensch mit gleicher Achtung behandelt werden sollte. Man muß alle gleich respektieren.

M3 Es ist ungerecht, denn Du willst in so einer Situation selber Hilfe bekommen.

N. *Einige sagen: "Man sollte jedem helfen. Sonst geht doch der Zusammenhalt in der Klasse verloren." Entscheide bitte wieder, inwieweit die Gründe für dich stimmen!*

N1 Man sollte helfen, weil die Gemeinschaft kaputt geht, wenn alle nur auf den eigenen Vorteil achten.
"Stimmt gar nicht / Stimmt nur wenig / Stimmt teil-teils / Stimmt weitgehend / Stimmt voll und ganz" (1-5)

N2 Man sollte helfen, weil man nur miteinander auskommt, wenn man sich gegenseitig auch mal einen Gefallen tut.

N3 Man sollte helfen, weil jeder für jeden anderen auch verantwortlich ist. Sonst gibt es keine Gemeinschaft.

O. *Nun geht es darum, wie Du Deine jetzigen Lehrerinnen und Lehrer in Deiner Klasse und Deiner Schule erlebst. Uns interessiert, wie häufig Du Deine Lehrerinnen und Lehrer so erlebst, wie es in den folgenden Aussagen beschrieben wird. Es geht also bei allen Fragen darum: Wie häufig erlebst Du Deine Lehrerinnen und Lehrer so? Auch hier solltest Du wieder alles ehrlich und offen beantworten. Keiner erfährt, wer was geschrieben hat!*

O1 Sie versuchen, meine Gefühle zu verstehen.
"Nie / Selten / Manchmal / Häufig / Sehr häufig" (1-5)

O2 Sie blamieren Schüler und Schülerinnen vor der ganzen Klasse.

O3 Sie hören sich meine Meinung immer an - auch wenn meine Meinung nicht mit ihrer übereinstimmt.

O4 Sie sind für mich da. Ich kann mich immer an sie wenden, wenn ich das brauche.

O5 Sie gehen auf die Eigenarten und Bedürfnisse ihrer Schüler/innen ein.

O6 Sie machen mir Mut beim Lernen.

O7 Sie interessieren sich für meine Probleme, auch wenn sie nichts mit der Schule zu tun haben.

O8 Sie sind zu den guten Schülern und Schülerinnen freundlicher als zu den schlechten.

O9 Sie können Fehler nicht zugeben.
O10 Meine Lehrerinnen und Lehrer versuchen, meine Stärken zu betonen - und nicht meine Schwächen.
O11 Meine Lehrerinnen und Lehrer lassen mich spüren, daß ich ihnen eher gleichgültig bin.
O12 Sie nehmen meine Ideen und Meinungen ernst.
O13 Sie zeigen, daß sie von jedem/r Schüler/in eine feste Meinung haben.
O14 Sie kümmern sich darum, wie ich mich fühle.
O15 Sie vertrauen mir.
O16 Sie versuchen, meine Ideen und meine Meinung zu verstehen.
O17 Sie machen sich über uns lustig.
O18 Es ist ihnen egal, ob sie mich beleidigen oder nicht.
O19 Sie nehmen mich als Person ernst.

P. *Nun noch ein paar Fragen, wie Du den Unterricht erlebst.*
P1 Meine Lehrerinnen und Lehrer unterrichten den Stoff Schritt für Schritt. "Nie / Selten / Manchmal / Häufig / Sehr häufig" (1-5)
P2 Sie machen einen interessanten Unterricht.
P3 Sie nutzen die Unterrichtszeit gut aus, um uns eine Menge beizubringen.
P4 Ihre Erklärungen sind meistens verständlich und klar.
P5 Sie unterrichten so, daß mich das zum Nachdenken herausfordert.