

## **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 verändert würde, wie auch gegen Theorien, die beobachtete Veränderungen allein auf das Wirken genetisch gesteuerter Reifungsprozesse zurückführen. Zwar bestimmen beide Einflußfaktoren, in welchem Rahmen Moralentwicklung stattfinden kann: Soziale Normen und Institutionen sowie genetische Faktoren setzen Grenzen. Aber 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 zu sagen, welche Bedingungen deren Effektivität begünstigt oder hemmt. 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 Experimente 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 aber auch 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 Interventions-

ansatzes 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.<sup>1</sup> 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 soll. Das eigene Urteil aufgrund persönlicher Kenntnis kann durch Experimente oder Statistiken nur schwer ersetzt werden. Für Eltern, Lehrer, Schulleiter, 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 zu umfas-

1 So konnte für den moralpädagogischen Wertklärungsansatz 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).

send zu evaluieren, daß 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 da versagt; er kann aber auch unberechtigterweise akzeptiert werden, nur weil die Umstände ein günstiges Ergebnis zustande gebracht haben, das 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, das heißt, der Ansatz muß mehrmals unter verschiedenen Bedingungen wiederholt werden, und zwar so, daß vermutete andere Ursachen für die festgestellte 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- und 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 macht. 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.<sup>2</sup>

Nehmen wir beispielsweise an, daß der Erfolg einer moralpädagogischen Intervention durch die Zusammensetzung der Schülerschaft in einer Klasse bedingt ist. Gemäß dem Maßnahmenkatalog zur Kontrolle von "Gefahren" für die Validität 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 Alternativklärung 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 eine 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 gedacht war, Verzerrung im Hinblick auf den Untersuchungszweck auszuschalten. Da die Förderung des moralischen Urteils in intakten Klassengruppen stattfinden soll, würde die zufällige Aufteilung der Schüler auf die Klassen die Validität der Inter-

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

ventionsstudie vermindern. Dieses Beispiel zeigt, daß eine rein "methodologische" Experimentalistik ihren eigenen Ansprüchen nicht genügen kann.<sup>3</sup>

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 solcher 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

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

tungen lassen sich durchaus wissenschaftlich streng überprüfen. Wahrheiten gelten so lange, wie sie nicht als trivial überführt 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 ihre Teile 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".<sup>4</sup> Die Dilemma-Diskussion besteht im Kern in der Diskussion von kurzen (fiktiven) Geschichten, in denen zwei etwa gleich wichtige Normen und Werte aufeinandertreffen, so daß der Handelnde, 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.<sup>5</sup>

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 betreffen.

4 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.

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

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-demographischem und ethnischem 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?<sup>6</sup>

#### 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  $\chi^2$  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.<sup>7</sup> 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 sagen daher - für sich genommen - nichts über die Wirksamkeit von pädagogischen Interventionen aus.

6 Die Quellen für die Untersuchungen, auf die sich die nachfolgenden Ergebnisse der Meta-Analyse beziehen, sind am ende dieses Kapitels aufgelistet. Eine tabellarische Übersicht über die Ergebnisse im einzelnen kann vom Autor angefordert werden.

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

- *Differenzwerte zwischen Vor- und Nachtest:* Der Vergleich von Nach- und Vortests hebt die Nachteile der Nachtestwerte auf. Die Differenzwerte 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 ( $\chi^2$ ) 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  $\chi^2$ -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 dem berichteten Signifikanz-Kennwerten ( $\chi^2$ ,  $t$ ,  $F$ ) oder, wo diese fehlen, aus Mittelwertdifferenzen und Standardabweichungen berechnet werden mußten.<sup>8</sup> Die  $r$ -Werte werden direkt, ohne Umrechnung in Fishers  $z$  verglichen und zusammengefaßt, da beide Verfahren praktisch zu den gleichen Ergebnissen

8 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<br>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<br>$QS_w =$ Quadratsumme <i>within groups</i> und<br>$QS_b =$ Quadratsumme <i>between groups</i> |
| (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<br>Irrtumswahrscheinlichkeit $p$  |

**Abb. 9-1** Formeln für die Umrechnung von Statistiken in Effektstärke  $r^2$

kommen (Glass et al. 1978, S. 147). Als Zwischenstufe werden hier die quadrierten Korrelationskoeffizienten, die sogenannten Determinationskoeffizienten  $r^2$  berechnet.<sup>9</sup> 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

<sup>9</sup> 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.).

"grauen" Literatur konnten in unserem Fall nur bei 74 von 141 Studien alle Informationen ermittelt werden, die für die Berechnung eines Effektstärkemaßes notwendig sind. Bei den anderen Studien fehlten einige oder alle wichtigen Informationen ( $n$ ,  $\chi^2$ ,  $t$  oder  $F$ ).<sup>10</sup> 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. Da das Effektstärkemaß  $r_{xy}$  auf die Variation der Meßwerte in einer Untersuchungsgruppe standardisiert wird, führt diese Nachlässigkeit meistens zu einer *Unterschätzung* der Effektstärke. Dort, 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 Abb. 9-1 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 vorgeschlagen) keine Bewertungen der Güte der Untersuchungen vor. Es kommt uns 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

10 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.

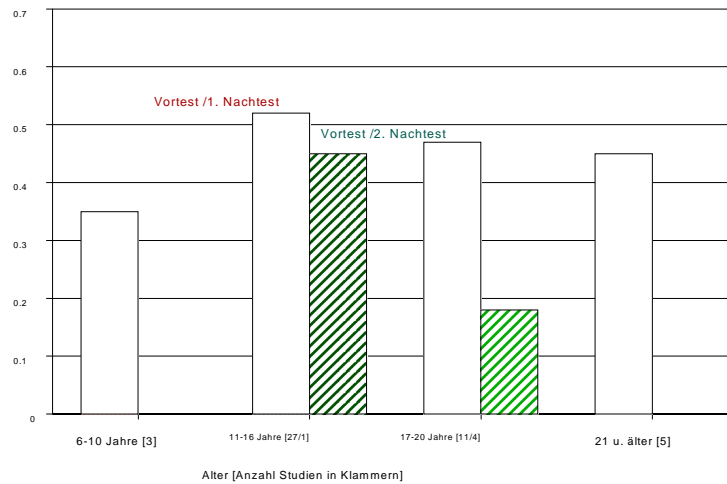
berücksichtigen, stellt für die Prüfung der Kohlberg-Methode eine zusätzliche Bürde dar. Geringe oder negative Effekte können auf methodische Mängel der Untersuchungsdesigns statt auf die Kohlberg-Theorie zurückgehen. Das nehmen wir in Kauf, da wir auch bei der praktischen Anwendung des Kohlberg-Ansatzes von sehr unterschiedlichen Bedingungen ausgehen müssen.

## 9.5 Ergebnisse

Das Hauptergebnis unserer Meta-Analyse ist: *Die auf Kohlberg gründenden 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.

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:

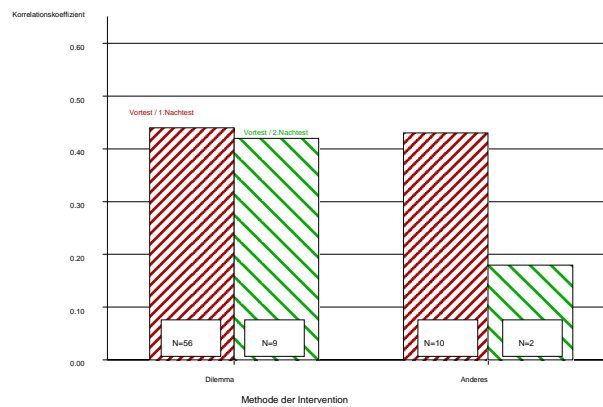
Das *Alter der Probanden* spielte nur insofern eine Rolle, als die deutlichsten Erfolge der Kohlberg-Methode in der Gruppe der 11-16jährigen Schüler gefunden wurden. Eine Studie erbrachte, daß in dieser Gruppe auch mit der größten Dauerhaftigkeit des Interventionseffekts zu rechnen ist (Abb. 9-2). Dieser Befund steht in Kontrast zu der weit verbreiteten Meinung, daß der



**Abb. 9-2** Effekt der Intervention in verschiedenen Altersgruppen

Effekt in diesem Alter geringer sei als im frühen Erwachsenenalter.

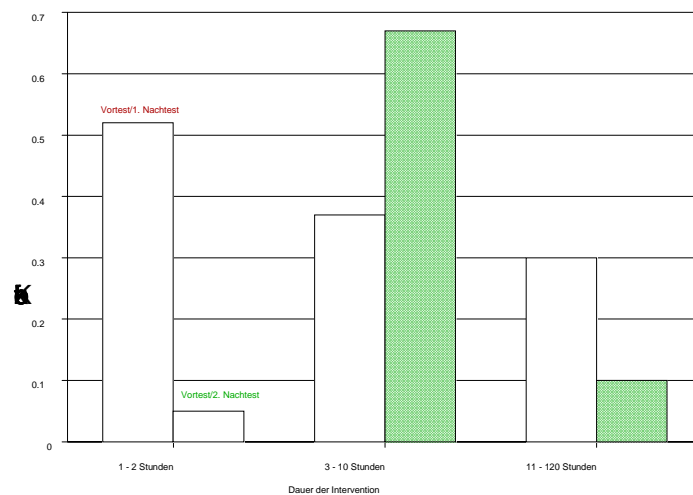
Bei der *Dauer der Intervention* scheint das Optimum bei 3 bis 10 Stunden zu liegen. Kürzere Interventionen zeigen zwar auch eine beträchtliche Wirkung ( $r = .52$ ; Basis: 17 Studien), der Effekt scheint jedoch von geringerer



**Abb. 9-3** Interventionsmethode: Dilemma versus andere

Dauer zu sein. Studien, die einen zweiten Nachtest durchführten, zeigen, daß die Effektstärke bald rapide abfällt ( $r = .05$ ; 2 Studien). Moralpädagogische Interventionen von längerer Dauer haben etwa 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; s. Abb. 9-4).<sup>11</sup>

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 (Abb. 9-5). Auf den ersten Blick erstaunlich ist die Tatsache, daß Ethik-Unterricht und Interventionen, die mit allgemeinen, moralbezogenen Geschichten oder mit Rollenspiel-Aufgaben die Diskussion



**Abb. 9-4** Einfluß der Interventionsdauer auf die Effektstärke

11 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.

in den Schülergruppen in Gang brachten, fast ebenso effektiv sind wie klassische Dilemma-Diskussionen. Langfristig erwies sich die Dilemma-Methode aber als die wirkungsvollere Methode (Abb. 9-3).

Schließlich analysierten wir, 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, daß die Lehrer in den meisten Fällen auch noch nach dem Interventionsexperiment mit den Schülern in Kontakt blieben und dieser Kontakt für ein weiteres Ansteigen der Urteilsfähigkeit verantwortlich gemacht werden kann. 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.<sup>12</sup> Unsere detaillierten Analysen der Studien ergaben aber weiterführende Befunde, die zum Teil gängigen Annahmen widersprechen: a) Die vielzitierte These, wonach Erwachsene stärker von moral-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 zu den 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 einen viel größeren 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 Interventions-Studien, 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 *Bildungs-*

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

*theorie* der Moralentwicklung, wenn man davon ausgeht, daß die - mit Abstand am häufigsten evaluierte - Dilemma-Diskussions-Methode für diese Altersgruppe am angemessensten ist.

Wie kommt diese Diskrepanz zustande und welche Aussage ist 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 Moralischen 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 *Sozialisierungstheorie* 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  $r = 0,26$ .<sup>13</sup> 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

13 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 (schriftlich dargebotenen) Kohlberg-Interviews *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.

Überzeugung kommen, daß die reifungstheoretisch inspirierte These vom spezifischen Nachholbedarf an moralischer Entwicklung im Erwachsenenalter nicht zutrifft, und daß die 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 handelt, 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 wie handlungsorientierte Dilemma-Diskussionen oder Just community-Programme. Anderslautende Befunde zu a) bis c) sind auf Studien begrenzt, die mit dem *DIT* von Rest (1986) evaluiert wurden und müssen den methodologischen Unzulänglichkeiten dieses Meßverfahrens zugerechnet werden. Der *DIT* mißt stärker als der *MJI* und insbesondere der *MUT* die *affektive* Seite des moralischen Urteils, also moralische Einstellungen (Kapitel 6). *DIT*-Effekte spiegeln daher eher moralische Einstellungsänderungen wieder, wie sie beispielsweise bei Erwachsenen durch theoretische Schulung erreicht werden können, als eine Entwicklung der moralischen *Urteilsfähigkeit*. *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.

## 9.7 Quellen für die analysierten Untersuchungen

(Die Nachweise für die im Text zitierte Literatur befinden sich im Anhang)

- 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 sophomoris 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 inmate'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.