

# Interventionsstudien zur Förderung der moralischen Urteilsfähigkeit; eine Meta-Analyse von Studien aus den Jahren 1985 bis 2006

WISSENSCHAFTLICHE ARBEIT  
ZUR ERLANGUNG DES AKADEMISCHEN GRADES  
'BACHELOR OF SCIENCE' (B.Sc.)  
IM FACH PSYCHOLOGIE

Fachbereich Psychologie  
der  
UNIVERSITÄT KONSTANZ

vorgelegt von

Robert Hepach  
Rheingutstraße 40 Wg 40  
78462 Konstanz

Erstgutachter: Professor Dr. Georg Lind  
Zweitgutachter: Professor Dr. Wilhelm Kempf

Konstanz, den 14.06.2007

Drei Verwandlungen nenne ich euch des Geistes: wie der Geist zum Kamele wird,  
und zum Löwen das Kamel, und zum Kinde zuletzt der Löwe.

- Friedrich Nietzsche *Also sprach Zarathustra*, Vorrede IV 29

# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>v</b>
<b>1 Einleitung</b>	<b>1</b>
<b>2 Theoretische Grundlagen</b>	<b>4</b>
2.1 Das Stufenmodell von Kohlberg und die Reifungstheorie . . . . .	4
2.2 Die Zwei-Aspekte-Theorie und die Bildungstheorie der Moral . . . . .	5
2.3 Die Messinstrumente . . . . .	7
2.3.1 Der DIT als Messinstrument des affektiven Aspekts der moralischen Urteilsfähigkeit . . . . .	7
2.3.2 Der MUT als Maß des kognitiven Aspekts der moralischen Urteilsfähigkeit . . . . .	8
2.4 Die Meta-Analyse nach Glass . . . . .	8
<b>3 Meta-Analyse</b>	<b>10</b>
3.1 Methode . . . . .	10
3.1.1 Auswahl der Studien . . . . .	10
3.1.2 Kategorisierung der Studien . . . . .	11
3.1.3 Statistische Auswertung . . . . .	12
3.2 Resultate . . . . .	17
3.2.1 Der affektive Aspekt des moralischen Urteils . . . . .	17
3.2.2 Der kognitive Aspekt des moralischen Urteils . . . . .	22
3.2.3 Gegenüberstellung der Ergebnisse zum affektiven und kognitiven Aspekt des moralischen Urteils . . . . .	25
<b>4 Diskussion</b>	<b>29</b>
4.1 Der affektive Aspekt – Prüfen der Hypothesen aus der Schlaefli et al. - Studie . . . . .	29
4.2 Der kognitive Aspekt der moralischen Urteilsfähigkeit . . . . .	32
4.3 Vergleich von DIT-Studien und MUT-Studien . . . . .	35

4.4	Schwächen dieser Arbeit und Beschränkung der Aussagekraft der Ergebnisse . . . . .	36
4.5	Kritische Auseinandersetzung mit der hier verwendeten Methode der Meta-Analyse . . . . .	38
4.6	Eine alternative Berechnung des Interventionseffekts . . . . .	39
4.7	Schlussbemerkungen . . . . .	42
<b>A</b>	<b>Tabelle</b>	<b>44</b>
<b>B</b>	<b>Verwendete Artikel</b>	<b>53</b>
<b>C</b>	<b>Listen</b>	<b>58</b>
C.1	Liste 1 Versuchspersonengruppe . . . . .	58
C.2	Liste 2 Schulklasse/Studiensemester . . . . .	58
C.3	Liste 3 Art der Intervention . . . . .	59
C.4	Liste 4 Inhalt der Intervention . . . . .	59
C.5	Liste 5 Testinstrumente . . . . .	59
C.6	Liste 7 Messskala . . . . .	60
C.7	Liste 8 Berechnung des Korrelationskoeffizienten (punkt-biserial) . .	60
C.8	Liste 10 Art der Publikation . . . . .	61
C.9	Liste 12 Designtyp . . . . .	61
	<b>Literaturverzeichnis</b>	<b>62</b>

# Zusammenfassung

In dieser Arbeit werden die Effekte von Interventionsstudien auf den affektiven und den kognitiven Aspekt der moralischen Urteilsfähigkeit untersucht. Hierfür wurden zunächst die Ergebnisse früherer Meta-Analysen zum affektiven Aspekt mit den neuen Daten verglichen. Es zeigt sich, dass der durchschnittliche Effekt von Interventionen auf den affektiven Aspekt größer geworden ist (von  $r = 0,14$  auf  $r = 0,21$ ). Im Besonderen waren die Effekte für Personen unter 16 Jahren am größten ( $r = 0,3$ ). Hinsichtlich der Intensität zeigten kurze Interventionen (bis drei Wochen) die größten Effekte ( $r = 0,25$ ).

Für den kognitiven Aspekt bestätigen sich die Ergebnisse einer früheren Meta-Analyse. Der durchschnittliche Effekt einer Intervention liegt für diesen Effekt bei  $r = 0,41$ . Besonders die Methode der Dilemmadiskussion zeigt im Vergleich mit anderen Interventionsarten den größten Effekt ( $r = 0,48$ ). Dabei haben kurze Interventionen ( $r = 0,51$ ) und Interventionen mit einer geringen wöchentlichen Intensität von weniger als einer Stunde ( $r = 0,47$ ) die deutlichsten Effekte.

Der Vergleich der Ergebnisse zu beiden Aspekten zeigt, dass die bisher durchgeführten Interventionsstudien einen unterschiedlichen Effekt auf den affektiven und den kognitiven Aspekt der moralischen Urteilsfähigkeit ausüben: Die Förderung des kognitiven Aspekts ( $r = 0,41$ ) profitiert mehr von Interventionsstudien als die des affektiven Aspekts ( $r = 0,21$ ).

# Kapitel 1

## Einleitung

Die gewaltlose und diskursive Lösung von Konflikten erfordert von Menschen hohe Fähigkeiten, die von einigen Autoren [Kohlberg, 1958, Lind, 2000] als *moralisch-demokratische* Fähigkeiten umschrieben werden. Um diese zu fördern, so legen die Befunde aus zahlreichen empirischen Interventionsstudien nahe, hat sich vor allem die Methode der Dilemmadiskussion bewährt, die von Blatt und Kohlberg [Blatt und Kohlberg, 1975] entwickelt und von Lind [Lind, 2007] zur 'Konstanzer Methode der Dilemmadiskussion' erweitert wurde.

Um die Ergebnisse der ersten Interventionsstudien aus den 70er und 80er Jahren zusammenzufassen, bedienten sich sowohl Schläefli [Schläefli et al., 1985] als auch Lind [Lind, 2000] den Methoden der Meta-Analyse nach Glass [Glass et al., 1978]. In diesen Arbeiten konnten beide Autoren verschiedene 'Moderatorvariablen' identifizieren, welche den Effekt einer Intervention auf die moralische Urteilsfähigkeit beeinflussten. Durch das Zusammentragen der bis dato bekannten Ergebnisse konnten beide Studien Hypothesen formulieren die in der vorliegenden Arbeit untersucht werden sollen.

Nach der Zwei-Aspekte Theorie [Lind, 2000] hat das moralische Urteil zwei Aspekte: den affektiven und den kognitiven Aspekt. Um die Vergleichbarkeit zwischen Studien zu erhöhen und die Datenerhebungen zu erleichtern, wurden Instrumente entwickelt, um beide Aspekte des moralischen Urteils zu untersuchen. Hierzu zählen der 'Defining Issues Test' (DIT) von James Rest und der 'Moralisches Urteil Test' (MUT) von Georg Lind. Der DIT misst den affektiven Aspekt und der MUT den kognitiven Aspekt. Beide Aspekte sind gegeneinander abgrenzbar, jedoch nicht voneinander unabhängig.

In der vorliegenden Arbeit geht es darum, unter Anwendung bewährter und neuerer meta-analytischer Techniken, die alten Ergebnisse aus den Studien von

1985 zu testen und ggf. neue Erkenntnisse zu gewinnen. Die Grundlage dafür bilden 39 Studien zum affektiven Aspekt und 12 Studien zum kognitiven Aspekt. Als relatives Effektstärkemaß dient das  $g$  nach Hedges&Oklin und der sich daraus berechnende Korrelationskoeffizient  $r$ . Beim Aufdecken von Moderatorvariablen orientiert sich diese Arbeit an den Ausführungen von Hunter&Schmidt [Hunter und Schmidt, 1990]. Der methodologische Hauptteil dieser Arbeit gliedert sich in drei Teile:

1. Der affektive Aspekt: Vergleich der neuen Ergebnisse mit der Studie von Schlaefli et al.
2. Der kognitive Aspekt: Vergleich der Ergebnisse mit der Studie von Lind
3. Vergleich der neuen Ergebnisse zu Affekt und Kognition.

Die Haupthypothese aus beiden Studien von 1985 lautet:

Interventionsstudien haben einen deutlichen Einfluss auf die moralische Urteilsfähigkeit.

Die aus dem Artikel von Schlaefli et al. zu untersuchenden Hypothesen zum affektiven Aspekt lauten wie folgt [Schlaefli et al., 1985, S. 346]:

- (1) “Moral education programs emphasizing dilemma discussion [...] produce modest but definite effects, [...].”
- (2) “Academic courses in the humanities and social studies do not seem to have an impact on moral judgement development.”
- (3) “Programs with adults (24 years and older) seem to produce larger effect sizes than programs for younger subjects [...].”
- (4) “Interventions longer than 12 weeks have no more impact than interventions of 3 to 12 weeks; however, duration less than 3 weeks tend to be ineffective [...].”

Aus der Meta-Analyse von Lind gehen folgende Hypothese zum kognitiven Aspekt hervor:

- (1) “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.”

- (2) Der deutlichste Erfolg der Blatt-Kohlberg Methode wurde in der Gruppe der 11- bis 16-Jährigen verzeichnet. Hier zeigt sich auch die größte Nachhaltigkeit.
- (3) “Bei der Dauer der Intervention scheint das Optimum bei 3 bis 10 Stunden zu liegen.“
- (4) Die beträchtliche Wirkung von Studien kürzer als 3h konnten in späteren Nachtests nicht bestätigt werden ( $r = 0,05$ ; 2 Studien).
- (5) Moralpädagogische Interventionen von längerer Dauer haben einen Effekt von  $r = 0,3$  (16 Studien). Dieser ist nicht nachhaltig ( $r = 0,1$ ; 5 Studien).
- (6) Kurze (bezüglich der Intensität) Interventionen haben bereits einen starken Einfluss.
- (7) Lange Interventionen (bezüglich der Intensität) haben einen nachhaltigeren Einfluss als kurze Interventionen.
- (8) Die Blatt-Kohlberg-Methode und andere Interventionsmethoden haben kurzfristig den gleichen Effekt, unterscheiden sich jedoch hinsichtlich der Nachhaltigkeit zugunsten der Dilemma-Diskussion.

Desweiteren wird in dieser Arbeit gefragt, wie wirksam die Dilemmadiskussion im Vergleich zu anderen Interventionsarten ist, und ob diese Methode sich auf die affektiven und kognitiven Aspekte des moralischen Urteils unterschiedlich auswirkt.



# Kapitel 2

## Theoretische Grundlagen

### 2.1 Das Stufenmodell von Kohlberg und die Reifungstheorie

Kohlberg nennt seine Theorie in Anlehnung an Jean Piaget 'kognitiv-entwicklungsorientiert'. Die Bestätigung, dass sich das moralische Urteilen sequentiell entwickelt ging aus der Längsschnittstudie von Colby&Kohlberg hervor [Colby et al., 1983]. Es zeigte sich, dass junge Menschen auf niedrigeren Stufen argumentieren und später auf höhere Stufen gelangen. Kohlberg spricht in diesem Zusammenhang von einer Reifungstheorie der Moral.

Nach Kohlberg 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 [Kohlberg, 1964, S. 425].“

Kindern fällt es zunächst schwer in Übereinstimmung mit ihren moralischen Werten zu urteilen und zu handeln. An dieser Stelle setzt der Entwicklungsprozess ein. Für jede Stufe sind unterschiedliche Affekte und Motive charakteristisch, auf Grundlage derer Menschen versuchen moralische Konflikte zu lösen.

Laut Kohlberg sind höhere Stufen gekennzeichnet durch moralphilosophisch adäquatere Aussagen. Die meisten Menschen präferieren diese Stufen unabhängig vom kulturellen Hintergrund. Dabei schlägt moralische Urteilsfähigkeit die Brücke zwischen den Werthaltungen auf der einen Seite und dem Verhalten auf der anderen. Kohlberg ergänzt seine Theorie um einen wesentlichen Punkt:

“Die Entwicklung kognitiver Strukturen ist eher das Resultat von Vorgängen der Interaktion zwischen der Struktur des Organismus und der Struktur der Umwelt als das direkte Resultat der Reifung [Kohlberg, 1984, S. 8].“

Kulturelle Faktoren können diesen Prozess beschleunigen, verlangsamen oder anhalten. Im Unterschied zur Bildungstheorie der Moral geht Kohlberg davon aus, dass sich die Abfolge der Stufen nicht ändert. Viele Forschungsergebnisse weisen jedoch daraufhin, dass Regressionen möglich sind [Lind, 2000].

## 2.2 Die Zwei-Aspekte-Theorie und die Bildungstheorie der Moral

Diese Theorie orientiert sich an den frühen Aussagen Kohlbergs zum Verhältnis des affektiven Aspekts und des kognitiven Aspekts, wonach diese zwar untrennbar sind, jedoch als parallele Aspekte des moralischen Verhaltens definiert werden [Kohlberg, 1958, S. 89].

Unter dem affektiven Aspekt verstehen wir

“[...]die im Verhalten sich zeigende, gefühlsmäßige Bindung an moralischen Maximen, Regeln oder Prinzipien [Lind, 2000, S. 52].“

Der kognitive Aspekt hingegen ist die Beziehung

“[...] zwischen Ideen und Inhalten im Denken eines Individuums [Kohlberg, 1977, S. 9].“

Jean Piaget sagt, dass:

“[...] jede Form des Verhaltens einen energiereichen oder affektiven und einen strukturellen oder kognitiven Aspekt hat [...]. Beide Seiten des menschlichen Verhaltens [sind] nicht aufeinander zurückführbar [Piaget, 1976, S. 7-8].“

Entscheidend ist, dass wir von Aspekten und nicht von Komponenten sprechen [Lind, 2000]. Blickrichtung, Ansicht oder Gesichtspunkt beschreiben den Begriff des Aspekts. Eine Komponente hingegen beschreibt den Teil eines Ganzen. Dieses kann folglich in seine Einzelteile zerlegt werden (eben eine Trennung der Komponenten). Jedoch können wir nicht den affektiven Aspekt eines Menschen getrennt von dem kognitiven Aspekt begutachten.

Im Sinne der Zwei-Aspekte-Theorie sprechen wir anstatt von Komponenten von Aspekten. Wie bereits oben erwähnt, ist der Begriff des Aspekts ein Synonym

für Blickrichtung, Ansicht oder Gesichtspunkt. Es macht also keinen Unterschied ob man seine Schwiegermutter von vorne oder von hinten betrachtet, ob man sie aus der Blickrichtung der Enkel oder einer außenstehenden Person begreift oder ob man sie unter dem Gesichtspunkt ihrer familiären Herkunft oder der ihr vermeintlich widerfahrenen Gemeinheiten in ihrem Leben versucht zu verstehen; es bleiben (höchstens) unterschiedliche Aspekte ein und derselben Person und man muss sich auch beim nächsten Kaffeetrinken im Kreis der Familie mit ihr arrangieren.

In der psychologischen Moralforschung können wir uns auf zwei Gesichtspunkte des moralischen Urteils stützen. Es lassen sich mithilfe von zwei gut erforschten Aspekten Aussagen über *ein* psychologisches Konstrukt machen. Durch die Entwicklung adäquater Messinstrumente in den letzten Jahrzehnten können sowohl Affekt als auch Kognition valide erfasst werden. Kohlberg sagt,

“[...]dass ein moralischer Akt oder eine moralische Einstellung weder anhand rein 'kognitiver' noch anhand rein 'motivationaler' Kriterien definiert werden kann [Kohlberg, 1958, S. 16].“

Georg Lind betont eine Vermeidung der

“Verwechslung der Begriffe „unterscheiden“ und „trennen“ [...] [Lind, 2000, S. 40].“

Der Begriff des kognitiven Aspekts wird verständlicher, wenn man sich von der Vorstellung löst, es handle sich dabei lediglich um Wahrnehmungen. Vielmehr versteht man in der kognitiven Entwicklungspsychologie darunter einen Prozess. Dabei geht es um die dynamische Veränderung der Strukturen, Muster und Organisationen in der mentalen Repräsentation.

Auch wenn die intrapsychischen Vorgänge bei Erwachsenen komplexer Natur sind, so bleibt folgende Kernaussage erhalten: Das kognitive System befindet sich in einem ständigen Veränderungsprozess. Der Motor ist das Individuum. Von seiner Bereitschaft zur Veränderung hängt die Dynamik im Wesentlichen ab.

“Der kognitive Aspekt ist definitionsgemäß keine Ursache für moralisches Verhalten, [...] er ist die Struktur des Urteilsprozesses, der sich im Verhalten ausdrückt [Lind, 2000, S. 53].“

Zu den Eigenschaften zählen

“[...]Universalität, innere Konsistenz und Unparteilichkeit [Kohlberg, 1977, S. 71][...].“

Um unsere moralischen Ideale mit dem Verhalten in Übereinstimmung zu bringen, benötigen wir daher bestimmte kognitive Fähigkeiten. Die unterschiedliche Ausprägung des kognitiven Aspekts führt zu unterschiedlichem Verhalten. Dies ist für den Bildungsprozess von großer Bedeutung.

“Moralisches Verhalten und moralische Entwicklung ist vielmehr, so legen alle Untersuchungen nahe, das Ergebnis von Bildungsprozessen (oder dem Fehlen derselben) [Lind, 2000, S. 15].“

Ein wesentlicher Unterschied zur Reifungstheorie besteht in der möglichen Regression, wenn entsprechende Bildungsangebote fehlen.

## 2.3 Die Messinstrumente

Das von Kohlberg und Colby entwickelte Moral Judgement Interview stellte das erste psychologische Verfahren dar, mit dem die Struktur des moralischen Urteils an Probanden untersucht werden konnte. Bis in die 80er Jahre hinein war es der einzige (brauchbare) methodologische Zugang. Die mehrjährige Anwendung machte jedoch auch auf Schwächen aufmerksam: Das Verfahren ist sehr zeit- und arbeitsaufwendig. Es sollte nur von geschultem Personal ausgewertet werden und kann sich auch dabei nicht von dem Vorwurf lösen, dass gewonnene Daten durch subjektive Einschätzung und Vorgaben zugunsten der Theorie verzerrt werden [Lind, 2000, S. 10]. Den beiden im folgenden Abschnitt dargestellten Tests liegt die Methodik zugrunde, dass

“[...]psychische Dispositionen wie moralische Urteilsfähigkeit und moralische Einstellung als manifeste Eigenschaften der Verhalten begriffen und gemessen werden können [Lind, 2000, S. 23].“

### 2.3.1 Der DIT als Messinstrument des affektiven Aspekts der moralischen Urteilsfähigkeit

Beim 'Defining Issues Test' werden dem Probanden Dilemmas mit zwölf Aussagen vorgelegt. Die Aufgabe besteht darin, die Aussagen entsprechend ihrer Wichtigkeit zur Lösung des Dilemmas in eine Rangfolge zu bringen. Je nachdem auf welcher Stufe sich der Proband befindet wird er die Argumente anders ordnen, da jede Aussage von ihrer Qualität her eine andere Stufe repräsentiert.

“Therefore the way a person rates and ranks the statements can be used to locate that person's point of development in the postulated developmental sequence [Schlaefli et al., 1985, S. 312].“

### 2.3.2 Der MUT als Maß des kognitiven Aspekts der moralischen Urteilsfähigkeit

Der MUT wurde von Lind entwickelt um den kognitiven Aspekt des moralischen Urteils zu messen. Dazu muss das Messinstrument

“[...]mit der Theorie kompatibel, objektiv (replizierbar) und leistungsabfordernd sein[...] [Lind, 2000, S. 85].“

Für den MUT wird der kognitive Aspekt der moralischen Urteilsfähigkeit

“[...] als das Ausmaß, mit dem [...] sich [eine Person] bei ihren Begründungen [...] konsistent an moralischen Prinzipien – anstatt zum Beispiel an Meinungskonformität – auf einer beliebigen Stufe orientiert [Lind, 2000, S. 86].“

bezeichnet. Eine niedrige Urteilsfähigkeit zeigt sich daran, dass eine Person sich

“ [...] konsistent an der Übereinstimmung dieser Argumente mit ihrer eigenen (vorgefassten) Meinung orientiert [Lind, 2000, S. 90].“

Verwendet wird das Konzept des multivariaten Experiment. Es handelt sich beim MUT um einen experimentellen Fragebogen. Der Test besteht aus zwei Dilemmas. Zu jedem Dilemma muss die Person ihre persönlich Meinung angeben. Darauf folgen jeweils sechs Pro- und Kontra-Argumente. Hier muss der Proband für jedes Item seine Zustimmung auf einer Skala von -4 bis +4 angeben. Am Ende errechnet sich ein C-Wert, wobei hohe Scores für hohe Urteilsfähigkeit stehen und umgekehrt. Die Skala verläuft von 0 bis 100.

## 2.4 Die Meta-Analyse nach Glass

Bei einer Meta-Analyse versucht der Untersucher die Ergebnisse mehrerer Studien zusammenzufassen.

“*Meta-analysis* refers to the analysis of analyses. I use it to refer to the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings [Glass, 1976, S. 3].“

Am Anfang der Analyse werden die Hypothesen formuliert. Im Anschluss daran werden die Daten mit Hilfe einer möglichst breit angelegten Literaturrecherche zusammengetragen. Im Falle der vorliegenden Arbeit sind die Fragestellungen durch die Hypothesen der Studien von Schlaefli und Lind gegeben. In die Meta-Analyse

werden all jene Studien mit einbezogen, die den unter Abschnitt 3.1.2 aufgelisteten objektiven Kriterien genügen. Der Effekt einer Intervention wird für jede Studie in das relative Effektstärkemaß  $r_{pbs}$  (punkt-biserial) umgewandelt. Die genaue Vorgehensweise ist in Abschnitt 3.1.3 beschrieben.

Relative Effektstärken berücksichtigen bei der Darstellung der Effekte die Standardabweichung innerhalb der Stichproben. Somit wird dem Umstand Rechnung getragen, dass verschiedene Gruppen unterschiedliche Effekte zeigen können. Dieses Konzept unterscheidet sich wesentlich von statistischen Signifikanztests. Viele Studien berichten in ihren Resultaten, ob sich die Messwerte nach der Intervention *signifikant* von den Messwerten vor der Intervention unterscheiden. Ob ein Unterschied signifikant ist hängt im Wesentlichen von der Stichprobengröße ab.

“The main disadvantage with this concept is that any desired level of significance can be reached by simply increasing sample size [Lind, 2007, S. 9].“

Auf eine weitere Schwäche des Verfahrens weisen Hunter&Schmidt hin:

“[...] there is a fundamental circularity to the significance test. If you do not know whether the null hypothesis is true or false, then you do not know whether the relevant error rate is the Type 1 error or the Type 2, i.e., you do not know if your error rate is 5% or some value as high as 95% [Hunter und Schmidt, 1990, S. 30].“

Der Vorteil von relativen Effektstärken ist, dass sie sich nicht auf Aussagen der Art 'Ein (Mess-)Unterschied ist signifikant oder nicht' beschränken. Neben anderen gebräuchlichen Maßen wie Cohen's  $d$  hat sich der in dieser Arbeit verwendete Korrelationskoeffizient  $r$  als brauchbar erwiesen (siehe Abschnitt 3.1.3).

Der Nachteil von relativen Effektstärken ist, dass sie die Aussagekraft in Abhängigkeit der Varianz in den Stichprobengruppen einschränken.

“Sie sind nicht unabhängig von zufälligen, für die Aussage irrelevanten Charakteristika der jeweiligen Untersuchungsgruppe [Lind, 2000, S. 35].“

Der Effekt vergrößert sich, wenn die Varianz geringer wird (homogene Gruppen). Für gut validierte und erprobte Messverfahren können die absoluten Effektstärken ein ausreichend aussagekräftiger Indikator für einen Studieneffekt sein. Aus diesem Grund werden in dieser Arbeit auch die absoluten Effektstärken mitangegeben und im Falle des kognitiven Aspekts der moralischen Urteilsfähigkeit diskutiert<sup>1</sup>.

<sup>1</sup>Der für die Messung dieses Aspekts verwendete MUT ist ein gut validiertes Messinstrument, für das bereits frühere Forschungsergebnisse in absoluten Werten angegeben wurden,

# Kapitel 3

## Meta-Analyse

### 3.1 Methode

#### 3.1.1 Auswahl der Studien

Die Suche wurde mit Hilfe der in der Bibliothek zugänglichen elektronischen Datenbanken PSYCINFO<sup>©</sup>, PSYCHARTICLES<sup>©</sup>, PSYINDEX<sup>©</sup>, ERIC<sup>©</sup> und SCOPUS<sup>©</sup> durchgeführt. Aus weiteren Quellen konnten 35 Studien bezogen werden. Für die Datenbankenrecherche wurde zunächst anhand der 'abstracts' der Studien eine Vorauswahl getroffen.

Das Kriterium hierfür war das Vorliegen einer Interventionsstudie mit erkennbarem experimentellen Aufbau (Vorher- und Nachhermessung, Design mit mindestens einer Experimentalgruppe u.a.). Berücksichtigt wurden nicht: Übersichtsstudien oder Artikel in denen (ausschließlich) theoretische Ansätze diskutiert werden. Wichtig war der Bezug zur moralischen Urteilsfähigkeit. In vielen Studien werden Untersuchungen besprochen die sich beispielsweise mit der Atmosphäre von sozialen Kontexten befassen.

In den 'abstracts' der Studien sollte erkennbar sein, dass man sich mit dem bereits erforschten Begriff der moralischen Urteilsfähigkeit<sup>1</sup> auseinandersetzt hat. Die drei Begriffe 'MORAL', 'REASONING' und 'INTERVENTION' wurden in den elektronischen Suchmaschinen kombiniert. Alle Suchmaschinen die keine neuen Treffer brachten werden hier nicht erwähnt. Die Ergebnisse der Literaturrecherche sind in Tabelle 3.1 dargestellt.

---

<sup>1</sup>Damit ist die Erforschung des Ansatzes nach Lawrence Kohlberg gemeint, wie er in der Einleitung dargestellt wurde

**Tabelle 3.1:** Überblick über die Ergebnisse der Literaturrecherche

Suchmaschine	Keywords	Ergebnisse	Verwendet	nicht verwendet <sup>2</sup>	nicht beschaffbar
PsycINFO <sup>©</sup>	'MORAL' & 'INTERVENTION' im Titel	19	8	8	3
	'MORAL' im Titel & 'INTERVENTION' im 'abstract'	26	8	10	8
	'MORAL' & 'REASONING' & 'INTERVENTION' in 'keywords'	2	0	1	1
	'MORAL' & 'REASONING' & 'INTERVENTION' im 'abstract'	13	3	6	4
PsYINDEX <sup>©</sup>	'MORAL' & 'INTERVENTION' im Titel	5	1	3	1
	'MORAL' im Titel & 'INTERVENTION' im 'abstract'	3	0	3	0
ERIC <sup>©</sup>	'MORAL' & 'INTERVENTION' im Titel	1	1	0	0
	'MORAL' im Titel & 'INTERVENTION' im 'abstract'	8	3	3	2
	'MORAL' & 'REASONING' & 'INTERVENTION' in 'keywords'	3	1	1	1
SCOPUS <sup>©</sup>	'MORAL' im Titel & 'INTERVENTION' im 'abstract'	5	4	1	0
	'MORAL' & 'REASONING' & 'INTERVENTION' in 'keywords'	1	1	0	0
	'MORAL' & 'REASONING' & 'INTERVENTION' im 'abstract'	2	1	0	1
WEITERE QUELLEN		35	20	5	10
Σ		123	51	41	31

Die Studien wurden auf folgendem Weg beschafft:

- (1) Das Manuskript war online verfügbar.
- (2) Ein Exemplar lag in der Bibliothek vor.
- (3) Eine Fernleihe bzw. Bestellung einer Kopie war möglich.

Im zweiten Abschnitt wurden die 'abstracts' der Artikel nochmals gelesen, da nun der direkte Vergleich von mehreren Studien möglich war. Dabei erwiesen sich einige Artikel als nicht brauchbar im engeren Sinne<sup>3</sup>. In einigen Fällen blieb die Bestellung über Fernleihe ohne Erfolg. Bei den kostenpflichtigen Dissertationen wurde nur eine kleine Auswahl bestellt.

### 3.1.2 Kategorisierung der Studien

Im dritten Arbeitsschritt lagen die Studien komplett vor. Dabei stellte sich heraus, dass mehrere Studien sich auf den 'dritten' Blick als nicht verwertbar erwiesen. Folgenden Kriterien mussten die Studien bei der genaueren Durchsicht genügen:

<sup>2</sup>Hierzu zählen alle Studien, die den Kriterien aus 3.1.2 nicht genügten oder nicht genügend Daten für die Berechnung eines Effekts lieferten.

<sup>3</sup>siehe Abschnitt 3.1.2, Kategorisierung der Studien



- (1) Bezug zur Kohlberg-Theorie (manche Studien bezogen sich einzig auf Piaget, was die Vergleichbarkeit hinsichtlich der Ergebnisse vor verschiedenen theoretischen Hintergründen erschweren würde)
- (2) klar erkennbares (quasi)-experimentelles Design (mindestens eine Experimentalgruppe, Vorhermessung, Intervention, Nachhermessung)
- (3) angemessene Quantifizierung der Ergebnisse mit bewährten Messinstrumenten wie dem DIT, MJT oder MJI

Bei der Kategorisierung der Studieninhalte orientiert sich diese Arbeit im Wesentlichen an den Vorgaben von Lind [Lind, 2000, S. 201-202]. Einige Kategorien wurden aus der Studie von Schlaefli et al. übernommen [Schlaefli et al., 1985, S. 325]. Die einzelnen Kategorien sind im Folgenden dargestellt<sup>4</sup>:

1. Art der Publikation (Liste 10)
2. Welche Gruppe von Personen wurde untersucht? (Liste 1)
3. durchschnittliches Alter der Probanden
4. Schulklasse/Semester (Liste 2)
5. eventuell durchgeführter späterer Nachtest
6. Designtyp (Liste 12)
7. Versuchsleiter
8. Art der Intervention (Liste 3)
9. untersuchtes Spezialgebiet (Liste 11)
10. Inhalt der Intervention (Liste 4)
11. Dauer der Intervention
12. Intensität
13. Test (Liste 5)

Die Tabelle mit der vollständigen Darstellung der Daten findet sich im Anhang A. Im weiteren Verlauf werden die Studien der Einfachheit halber nur mit ihrer entsprechenden Nummer aufgelistet.

### 3.1.3 Statistische Auswertung

In diesen Teil der Untersuchung konnten all diejenigen Studien miteinbezogen werden, die alle notwendigen Informationen für die Berechnung des Effektmaßes hatten. Die folgenden Ausführungen orientieren sich im Wesentlichen an den Ausführungen von Hunter&Schmidt [Hunter und Schmidt, 1990].

---

<sup>4</sup>die Erklärung der einzelnen Listen finden sich im Anhang

Die Wahl des Effektmaßes erfolgt analog zu der Meta-Analyse von Lind [Lind, 2000, S. 202 - 207]. Hierfür wird der (punkt-biserial) Korrelationskoeffizient  $r$  verwendet. Er stellt eine brauchbare Alternative zu Effektstärkemaßen wie das  $d$  von Cohen bzw.  $g$  nach Glass oder Hedges&Olkin [Rustenbach, 2003, S. 68 - 70]. Diese verwenden die Differenz der Mittelwerte von Experimentalgruppe und Kontrollgruppe und teilen sie durch die Standardabweichung der Kontrollgruppe nach der Intervention (Glass) oder die gepoolte Standardabweichung (Hedges&Olkin) [Rustenbach, 2003, S. 68 - 70].

Der punkt-biserial Korrelationskoeffizient hat den Vorteil, dass er nach oben hin 'nicht offen' ist wie das  $d$ -Maß (ebenso  $g$ -Maß). In der kleinsten Ausprägung hat er den Wert -1; in der größten den Wert 1. Dies erleichtert die Vergleichbarkeit. Zudem lässt er sich problemlos aus  $g$  berechnen (siehe unten). Das relative Effektstärkemaß  $r$  korreliert die unabhängige Variable Intervention/keine Intervention mit der Mittelwertdifferenz als abhängige Variable (gemessen mit dem entsprechenden Instrument).

Wenn für die Kontrollgruppe und die Experimentalgruppe jeweils die Mittelwerte (mit entsprechender SD) vor und nach der Intervention angegeben waren, dann erfolgte die Berechnung von  $r$  wie folgt [Rustenbach, 2003, S. 70 - 71]:

1.  $s_{gepooled} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$
2.  $g_{Hedges} = \frac{x_1 - x_2}{s_{gepooled}}$
3.  $r_{pbs} = \sqrt{\frac{g^2 n_1 n_2}{g^2 n_1 n_2 + (n_1 + n_2)(n_1 + n_2 - 2)}}$

In vielen Untersuchungen wurden  $t$ -Tests oder  $F$ -Tests gerechnet. Für diese und andere Fälle konnte  $r$  aus den gegebenen Maßen berechnet werden (siehe Anhang C, Liste 8).

Um den durchschnittlichen Effekt aller Interventionsstudien zu ermitteln wurde der gewichtete Korrelationskoeffizient berechnet [Hunter und Schmidt, 1990, S. 100]:

$$\bar{r} = \frac{\sum [N_i r_i]}{\sum N_i}$$

wobei  $N_i$  und  $r_i$  der Anzahl der Versuchspersonen bzw. den Korrelationen in den einzelnen Studien entsprechen. An dieser Stelle sei bereits darauf hingewiesen,

dass viele Studien von mehr als einer Experimentalgruppe berichten. So konnten in ein und derselben Studie mehrere Effektstärken berechnet werden.

Das Ziel einer Meta-Analyse ist es neben der Bestimmung eines Gesamteffekts auch die Angabe möglicher Moderatoren dieses Effekts, sog. Moderatorvariablen. Als solche kommen in dieser Arbeit die vorher erwähnten Einzelkategorien in Frage. Hunter&Schmidt schlagen vor, bei der Suche nach Moderatorvariablen wie folgt vorzugehen:

(1)

Wir nehmen an, dass jede Studienkorrelation nicht genau der Populationskorrelation entspricht. Sie weicht von dieser durch einen Stichprobenfehler  $e$  ab [Hunter und Schmidt, 1990, S. 103]:

$$e = r - \rho$$

$$r = \rho + e$$

Stichprobenfehler entstehen immer dann, wenn aus Populationen gezogen wird. Dabei wird davon ausgegangen, dass die Untersucher den 'wahren' Zusammenhang zwischen UV und AV in der Population in ihren Experimenten 'verfehlen' (z.B. durch Messungenauigkeit, Tagesform der Probanden etc.). Weiterhin wird vorausgesetzt, dass diese Abweichungen des gemessenen Effekts zufällig vom Populationseffekt abweichen.

(2)

Der Erwartungswert des Stichprobenfehlers ist Null. An dieser Stelle entfaltet die Meta-Analyse ihre Stärke, da mehrere Studieneffekte gemittelt werden [Hunter und Schmidt, 1990, S. 105 und 111].

$$E(r) = \rho$$

$$E(e_i) = 0$$

(3)

Die Varianz der Studienkorrelationen von der durchschnittlichen Korrelation lässt sich wie folgt berechnen [Hunter und Schmidt, 1990, S. 100]:

$$s_r^2 = \frac{\sum [N_i (r_i - \bar{r})^2]}{\sum N_i}$$

(4)

Nach Hunter&Schmidt setzt sich diese Varianz zusammen aus der tatsächlichen Populationsvarianz und der Varianz des Messfehlers [Hunter und Schmidt, 1990, S. 106]:

$$\begin{aligned}\sigma_r^2 &= \sigma_\rho^2 + \sigma_e^2 \\ \sigma_\rho^2 &= \sigma_r^2 - \sigma_e^2\end{aligned}\tag{5}$$

Die Varianz des Messfehlers berechnet sich wie folgt [Hunter und Schmidt, 1990, S. 108 - 109]:

$$\sigma_e^2 = \frac{(1-\bar{r}^2)^2}{(N-1)}$$

Dabei ist  $\bar{N}$  die durchschnittliche Stichprobengröße.  $\sigma_e^2$  ist der Anteil der oben ermittelten Varianz, der sich einzig auf die Wirkung des Stichprobenfehlers zurückführen lässt.

(6)

Eine Approximation der Populationsvarianz des Korrelationskoeffizienten

$$\sigma_\rho^2 = \sigma_r^2 - \sigma_e^2$$

gibt die tatsächliche Variation zwischen den Studienergebnissen wieder. Wenn die Standardabweichung deutlich von 0 abweicht so legen Hunter&Schmidt den Schluss nahe, dass es eine oder mehrere moderierende Variablen gibt [Hunter und Schmidt, 1990, S. 113].

An dieser Stelle möchte ich einen kurzen Einschub bzw. eine Anmerkung zur Suche nach Moderatorvariablen vornehmen. Im Sinne einer empirischen Wissenschaft müssen die eventuellen Moderatoren eines Effekts *vor* der eigentlichen Meta-Analyse angegeben werden.

Die Hypothesen werden also nicht im Laufe der statistischen Auswertung formuliert. Wir würden dann Gefahr laufen, aus unseren Daten heraus a posteriori zu Vermutungen zu gelangen. Karl Popper weist uns aber darauf hin, (riskante) Hypothesen vor dem Experiment zu formulieren und an der Empirie zu testen [Popper, 1965, S. 36-37].

In dieser Untersuchung wurden die Hypothesen theoriegeleitet aus vorhergehenden Untersuchungen übernommen, um ihnen eine 'faire' Chance zu geben sich

anhand neuer Daten zu behaupten. In der Psychologie bedienen wir uns den statistischen Methoden um unsere Schlüsse transparent und nachvollziehbar zu machen.

Hunter&Schmidt schlagen weiter vor, im Zuge der Untersuchung hinsichtlich moderierender Variablen, alle Studien in Gruppen zu unterteilen. Wenn wir also den Effekt des Moderators 'Design der Intervention' mit den Subkategorien 'randomisierte Zuordnung', 'quasiexperimentelle Zuordnung' und 'nicht quasi-experimentell' untersuchen, dann ordnen wir jede Studie einer der Gruppen (entsprechend des Designs) zu. Danach werden die oben genannten Schritte in den Gruppen durchgeführt; in diesem Fall drei 'kleine' Meta-Analysen.

Das Ergebnis sind drei Effektstärken mit den jeweiligen korrigierten Varianzen. Nach Hunter&Schmidt wird sich eine Moderatorvariable in zweierlei Hinsicht zeigen [Hunter und Schmidt, 1990, S. 112 - 114]:

- (1) Die gewichteten Korrelationen der Gruppen unterscheiden sich
- (2) Die (korrigierte) Varianz innerhalb der Gruppen ist Null (oder annähernd Null)

In den Übersichten dieser Arbeit sind für alle Korrelationskoeffizienten die 95% Konfidenzintervalle angegeben. Wird Null *nicht* von dem Intervall überlagert, so kann mit entsprechender Sicherheit gesagt werden, dass ein Effekt vorliegt.

## 3.2 Resultate

Die Darstellung der Ergebnisse der statistischen Auswertung gliedert sich in drei Teile: Im ersten Abschnitt werden alle Studien zusammengefasst, die den affektiven Aspekt des moralischen Urteils messen. Der zweite Teil stellt all jene Studienergebnisse dar, die den kognitiven Aspekt messen und im letzten Abschnitt werden die Ergebnisse für beide Aspekte verglichen und zusammengefasst.

### 3.2.1 Der affektive Aspekt des moralischen Urteils

Der affektive Aspekt der moralischen Urteilsfähigkeit lässt sich, wie in der Einleitung erwähnt, mit dem Defining Issues Test von James Rest messen. Wenn in Studien andere Instrumente eingesetzt wurden, dann sind diese in der Tabelle im Anhang A zu finden. In Tabelle 3.2 sind die Ergebnisse der DIT-Studien (der Einfachheit halber wird dieser Begriff fortan für alle Studien dieses Abschnitts verwendet) zusammengefasst. Sie sind chronologisch aufgelistet. Wenn Daten nicht erhoben werden konnten, taucht eine Lücke in der Tabelle auf. Es sind all jene Studien vermerkt, für die sich die relative Effektstärke berechnen ließ.

Im Folgenden sind die gewichtete Korrelation, sowie die Varianz (unkorrigiert), Fehlervarianz und korrigierte Varianz aufgelistet:

$$\begin{aligned}\bar{r} &= 0,21 \\ 95\%KI &: 0,18 \leq \rho \leq 0,24 \\ \sigma_{\bar{r}}^2 &= 0,0192 \\ \sigma_e^2 &= 0,0123 \\ \sigma_{\rho}^2 &= 0,0069 \\ \sigma_{\bar{r}} &= 0,08\end{aligned}$$

Das Konfidenzintervall ist sehr schmal. Der Korrelationskoeffizient der Population liegt mit entsprechender Sicherheit nicht unter 0,18 und auch nicht über 0,24. Der ermittelte Schätzer ( $\bar{r} = 0,21$ ) ist hochsignifikant<sup>5</sup>.

---

<sup>5</sup>Nullhypothese:  $\rho = 0$ ,  $\alpha = 0,05$ ,  $z(r) = 0,2132$ ,  $TW = 11,56$

Tabelle 3.2: Ergebnisse der DIT-Studien

Nr.	Studie	<i>n</i>	<i>g<sub>H&amp;G</sub></i>	<i>r</i>	95 % KI	follow-up	Zuwachs <sup>6</sup>	Messinstrument
1	2	720	0,36	0,18** <sup>7</sup>	(0,11 ≤ $\rho$ ≤ 0,25)		4,53	T(4)
2	3	50		0,21	(-0,07 ≤ $\rho$ ≤ 0,46)			T(7)
3	3	47		0,24	(-0,05 ≤ $\rho$ ≤ 0,49)			T(7)
4	4	28		0,35* <sup>8</sup>	(-0,03 ≤ $\rho$ ≤ 0,64)			T(9)
5	5	12	-0,23	-0,12	(-0,65 ≤ $\rho$ ≤ 0,49)		-2,5	T(4)
6	5	34	0,23	0,12	(-0,23 ≤ $\rho$ ≤ 0,44)		2,3	T(4)
7	6	32		0	(-0,35 ≤ $\rho$ ≤ 0,35)			T(3)
8	8	38		0,23	(-0,10 ≤ $\rho$ ≤ 0,51)			T(3)
9	9	16		0,06	(-0,45 ≤ $\rho$ ≤ 0,54)			T(4)
10	10	24	0,15	0,08	(-0,33 ≤ $\rho$ ≤ 0,47)		2,08	T(3)
11		24	0,32	0,17	(-0,25 ≤ $\rho$ ≤ 0,54)		2,58	T(3)
12	12	13		0,44	(-0,15 ≤ $\rho$ ≤ 0,80)		6,9	T(3)
13	12	15		0,46**	(-0,07 ≤ $\rho$ ≤ 0,79)		9,6	T(3)
14	12	11		0,05	(-0,57 ≤ $\rho$ ≤ 0,63)		3,7	T(3)
15	12	15		0,09	(-0,44 ≤ $\rho$ ≤ 0,58)		2,9	T(3)
16	13	26		0,39**	(0 ≤ $\rho$ ≤ 0,68)		12,18	T(3)
17	14	452		0,38**	(0,30 ≤ $\rho$ ≤ 0,46)			T(9)
18	15	83	0,02	0,01	(-0,21 ≤ $\rho$ ≤ 0,23)		6,1	T(4)
19	18	40		0,33**	(0,02 ≤ $\rho$ ≤ 0,58)	0,18		T(3)
20	20	180		0,21**	(0,07 ≤ $\rho$ ≤ 0,35)		7,6	T(4)
21	22	195	0,41	0,19**	(0,05 ≤ $\rho$ ≤ 0,32)			T(4)
22	23	48	1,32	0,56**	(0,33 ≤ $\rho$ ≤ 0,73)	0,3	7,4	T(3)
23	25	157		0,26**	(0,11 ≤ $\rho$ ≤ 0,40)			T(3)
24	28	61	0,77	0,36**	(0,12 ≤ $\rho$ ≤ 0,56)		7,69	T(4)
25	30	43		-0,17	(-0,45 ≤ $\rho$ ≤ 0,14)		1,66	T(3)
26	32	64	0,44	0,2*	(-0,03 ≤ $\rho$ ≤ 0,44)		7,78	T(3)
27	33	32		0,14	(-0,22 ≤ $\rho$ ≤ 0,47)			T(3)
28	33	39		0,16	(-0,16 ≤ $\rho$ ≤ 0,45)			T(3)
29	33	47		0,12	(-0,17 ≤ $\rho$ ≤ 0,39)			T(3)
30	34	56		0,17	(-0,10 ≤ $\rho$ ≤ 0,41)			T(7)
31	35	13	1,09	0,5*	(-0,07 ≤ $\rho$ ≤ 0,82)		9,5	T(4)
32	37	9		0,59*	(-0,12 ≤ $\rho$ ≤ 0,90)		7,69	T(3)
33	37	8		0,71*	(0,01 ≤ $\rho$ ≤ 0,94)		7,69	T(3)
34	40	35	-0,42	-0,21	(-0,51 ≤ $\rho$ ≤ 0,13)		2,89	T(4)
35	40	67	0,07	0,04	(-0,20 ≤ $\rho$ ≤ 0,28)		3,02	T(4)
36	40	14	0,47	0,24	(-0,33 ≤ $\rho$ ≤ 0,68)		10,38	T(4)
37	41			0,48**	(0,14 ≤ $\rho$ ≤ 0,72)		9,6	T(4)
38	42	34		0,05	(-0,29 ≤ $\rho$ ≤ 0,38)			T(3)
39	43	144		0,06	(-0,10 ≤ $\rho$ ≤ 0,22)			T(9)
40	44	20		0,09	(-0,37 ≤ $\rho$ ≤ 0,51)		1,21	T(4)

<sup>6</sup>in absoluten Skaleneinheiten

Die Ergebnisse können folgendermaßen zusammengefasst werden:

- (1) Die einzelnen Werte variieren beträchtlich mit  $\min(r) = -0,21$  bis  $\max(r) = 0,71$ .
- (2) Insgesamt sind fünf Studieneffekte negativ bzw. nahe 0 (auch Studie 12 mit  $r = 0,01$ ).
- (3) Sieben Studieneffekte sind sehr gering ( $r < 0,1$ ).
- (4) 14 Studieneffekte sind gering bis moderat ( $0,1 < r < 0,3$ ).
- (5) Sieben Effekte sind moderat bis stark ( $0,3 < r < 0,5$ ).
- (6) Drei Effekte sind sehr groß ( $r > 0,5$ ).
- (7) Lediglich zwei Studien liefern einen 'follow-up' - Test, der eine Aussage über die Nachhaltigkeit der Intervention zulässt.
- (8) Fünf Studien lieferten nicht genügend Informationen um die relative Effektstärke zu berechnen.
- (9) Bis auf eine Studie können alle Einzelkorrelationen in die Moderatorvariablen-Analyse miteinbezogen werden.
- (10) Aus der obigen Darstellung lässt sich schlussfolgern, dass die ermittelte Effektstärke von  $0,21$  ( $\sigma = 0,08$ ) um  $0,21/0,08 = 2,625$  Standardabweichungen von Null (kein genereller Zusammenhang) abweicht.

Um die Studienergebnisse auf Homogenität zu überprüfen schlagen Hedges&Olkin die Berechnung der 'Q - Statistik' vor [Hedges und Olkin, 1985, S. 235]:

$$Q = \sum_{i=1}^k (n_i - 3)(z_i - z_{(\bar{r})})^2$$

Dabei werden die z-transformierten Werte verwendet, da diese annähernd normalverteilt sind. Der Testwert ist  $\chi^2$  - verteilt. Mit einem Wert von  $Q = 63,03$  konnte die Nullhypothese 'Alle Korrelationen sind gleich' auf einem Niveau von  $\alpha = 0,05$  abgelehnt werden<sup>9</sup>. Dieses Ergebnis gibt zusätzlich Anlass nach moderierenden Effekten zu suchen. Tabelle 3.3 stellt die Ergebnisse der einzelnen 'kleinen' Meta-Analysen für jede (potentiell) moderierende Variable dar.

---

<sup>9</sup>KW=55,8 mit df=38



**Tabelle 3.3:** Überblick über die Ergebnisse der Moderatorvariablen-Analyse für den affektiven Aspekt

Kategorie	Subkategorie	Anzahl der Einzeleffekte	$\bar{r}$	$\sigma_{\bar{r}}^2$
Gruppe	Grundschule	1	0,36	0
	Sekundarstufe 1, Hauptschüler Junior High School	6	0,31	0,001
	Sekundarstufe 2, Mittelstufe der Realschule oder des Gymnasiums, Senior High School	4	0,27	0
	Abiturienten, High School Graduates, College undergraduates	2	0,21	0
	Studenten, College Graduates	15	0,19	0,01
	Junge Erwachsene ohne Hochschulabschluss	1	0,06	0
	Ältere Erwachsene ohne Hochschulabschluss	1	0,5	0
	Delinquente	6	0,1	0
	Alter in Jahren	10 bis 15	10	0,3
16 bis 20		10	0,21	0
21 bis 25		5	0,18	0,007
>25		9	0,15	0,012
Design		Randomisierte Zuordnung	3	0,35
	Quasiexperiment mit Kontrollgruppe	23	0,22	0
	Quasiexperiment ohne Kontrollgruppe	11	0,15	0,0025
Art der Intervention	Dilemmadiskussion nach Blatt&Kohlberg	13	0,19	0
	Just-community	1	0,19	0
	Ethikunterricht	8	0,19	0,038
	Peer-Interaktion	4	0,13	0
	Andere	7	0,21	0
Inhalt der Intervention	Dilemmata	16	0,17	0
	Formale Kommunikations- und Diskussionsfähigkeit	10	0,25	0,008
	Einführung in Kohlberg's Theorie	1	0,56	0
	Diskussion von Werten	1	0,01	0
	Andere Inhalte	12	0,13	0
Dauer der Intervention	Bis 5 Wochen	4	0,25	0
	6 bis 10 Wochen	15	0,17	0,002
	11 bis 15 Wochen	5	0,03	0
	16 bis 20 Wochen	7	0,18	0,059
	>20 Wochen	4	0,26	0,007
Intensität in h pro w	<=1	17	0,22	0,009
	<=2	10	0,27	0,013
	<=5	8	0,3	0
	>5	2	0,2	0
Testinstrument	DIT lang	17	0,17	0,002

Fortsetzung auf der folgenden Seite...

---

Fortsetzung . . .

---

DIT kurz	10	0,22	0
Andere	7	0,3	0,012

---

An dieser Stelle möchte ich kurz einen Teil aus dem Diskussionsabschnitt vorwegnehmen.

Es ist wichtig zwischen zwei Aspekten dieser Analyse der Moderatorvariablen zu unterscheiden:

1. Die Wahl der Kategorien und der Subkategorien erfolgt nach der an früherer Stelle erwähnten Kategorisierung (siehe Abschnitt 1.1.2). Die Berechnung von Effektstärken für einzelne Gruppen dient dem späteren Vergleich mit den Ergebnissen der Studie von Schlaefli et al [Schlaefli et al., 1985].
2. Die Varianzen der Gruppeneffekte dienen einer späteren statistischen Interpretation.

### 3.2.2 Der kognitive Aspekt des moralischen Urteils

Die für die Messung dieses Aspekts berücksichtigten Instrumente sind der MUT<sup>10</sup> und das MJI<sup>11</sup>. Die Ergebnisse sind in Tabelle 3.4 dargestellt. Dabei orientieren sich die Darstellungen an jenen des vorhergehenden Abschnitts.

**Tabelle 3.4:** Ergebnisse der Studien zum kognitiven Aspekt der moralischen Urteilsfähigkeit.

Nr.	Studie	<i>n</i>	<i>g<sub>H&amp;G</sub></i>	<i>r</i>	95 % KI	follow-up	Zuwachs	Messinstrument
1	1	34		0,62** <sup>12</sup>	(0,36 ≤ ρ ≤ 0,79)	0,73		T(1)
2	11	120	0,83	0,39**	(0,23 ≤ ρ ≤ 0,53)		8,37	T(6)
3	19	242		0,56**	(0,47 ≤ ρ ≤ 0,64)		25,46	T(6)
4	24	83	0,84	0,39**	(0,19 ≤ ρ ≤ 0,56)	0,37	14	T(6)
5	26	48		0,24* <sup>13</sup>	(-0,05 ≤ ρ ≤ 0,49)		1,6	T(6)
6	27	123		0,6**	(0,47 ≤ ρ ≤ 0,7)		13,1	T(6)
7	27	71		0,11	(-0,13 ≤ ρ ≤ 0,33)		5,09	T(6)
8	27	6		0,52	(-0,5 ≤ ρ ≤ 0,94)		21,65	T(6)
9	31	157		0,18**	(0,02 ≤ ρ ≤ 0,33)		3,37	T(6)
10	36	21		0,72**	(0,42 ≤ ρ ≤ 0,88)		0,75	T(1)
11	45	41		0,53**	(0,27 ≤ ρ ≤ 0,72)		13,3	T(1)
12	46	40		0,24	(-0,08 ≤ ρ ≤ 0,51)		10,7	T(6)

Im Folgenden sind die gewichtete Korrelation, sowie die Varianz (unkorrigiert), Fehlervarianz und korrigierte Varianz aufgelistet:

$$\begin{aligned} \bar{r} &= 0,41 \\ 95\%KI &: 0,36 \leq \rho \leq 0,46 \\ \sigma_{\bar{r}}^2 &= 0,0307 \\ \sigma_e^2 &= 0,0063 \\ \sigma_{\rho}^2 &= 0,0243 \\ \sigma_{\bar{r}} &= 0,16 \end{aligned}$$

Das Konfidenzintervall gibt an, dass der Korrelationskoeffizient in der Population mit entsprechender Sicherheit nicht kleiner als 0,36 ist und nicht größer als 0,46.

<sup>10</sup>Moralisches Urteil Test nach Lind

<sup>11</sup>Moral Judgement Interview (Form A und B) nach Kohlberg

Die Ergebnisse können, entsprechend denen der DIT-Studie, folgendermaßen zusammengefasst werden:

- (1) Die einzelnen Werte variieren mit  $\min(r) = 0,11$  bis  $\max(r) = 0,72$ .
- (2) Keiner der Studieneffekte ist negativ.
- (3) Vier Studieneffekte sind gering bis moderat ( $0,1 < r < 0,3$ ).
- (4) Zwei Studieneffekte sind moderat bis stark ( $0,3 < r < 0,5$ ).
- (5) Sechs Studieneffekte sind sehr groß ( $r > 0,5$ ).
- (6) Zwei von 12 Studien berichten follow-up Tests mit moderat bis sehr großen Effekten.
- (7) Zwei von 14 Studien lieferten nicht genügend Informationen um die Effektstärken zu berechnen.
- (8) Alle aufgelisteten Studien können in die Analyse nach Moderatorvariablen miteinbezogen werden.
- (9) Aus der obigen Darstellung lässt sich schlussfolgern, dass die ermittelte Effektstärke von 0,41 ( $\sigma = 0,16$ ) um  $0,41/0,16 = 2,6$  Standardabweichungen von Null (kein genereller Zusammenhang) abweicht.

Die  $Q$  - Statistik für die Einzelstudien ist  $Q = 44,81$  ( $p < 0,01$ )<sup>14</sup>. Dieses Ergebnis untermauert die Berechtigung nach moderierenden Effekten zu suchen.

---

<sup>14</sup>KW=24,73, df=11

Die nachstehende Tabelle 3.5 gibt einen Überblick über die Analyse nach Moderatorvariablen.

**Tabelle 3.5:** Überblick über die Moderatorvariablen-Analyse für den kognitiven Aspekt der moralischen Urteils

Kategorie	Subkategorie	Anzahl der Einzeleffekte	$\bar{r}$	$\sigma_{\bar{r}}^2$
Gruppe	Schüler bis 10 Jahre	1	0,24	0
	Studenten, College Graduates	9	0,46	0,0078
	Delinquente	2	0,57	0
Alter in Jahren	10 bis 15	2	0,26	0,0193
	16 bis 20	5	0,5	0
	21 bis 25	4	0,41	0,0254
Art der Intervention	Dilemmadiskussion nach Blatt&Kohlberg	6	0,48	0,0091
	andere	6	0,35	0,0298
Inhalt der Intervention	Dilemmas	6	0,48	0,0091
	Formale Kommunikations- und Diskussionsfähigkeit	1	0,39	0,01
	Diskussion von Werten	5	0,32	0,13
Dauer der Intervention	Bis 5 Wochen	3	0,51	0,0074
	6 bis 10 Wochen	2	0,34	0
	16 bis 20 Wochen	7	0,37	0,0311
Intensität in h pro w	<=1	6	0,47	0,0087
	<=2	4	0,41	0,0288
	<=5	2	0,24	0,0044

### 3.2.3 Gegenüberstellung der Ergebnisse zum affektiven und kognitiven Aspekt des moralischen Urteils

Insgesamt konnten 51 Studien in die Meta-Analyse miteinbezogen werden ( $n_{gesamt} = 3932$ ). In Tabelle 3.6 sind die Ergebnisse für den affektiven und den kognitiven Aspekt des moralischen Urteils dargestellt. Auffällig ist, dass der durchschnittliche Effekt für den kognitiven Aspekt doppelt so groß ist wie für den affektiven Aspekt. Interessant ist weiterhin, dass die obere Grenze des 95% Konfidenzintervalls für die DIT-Studien um 0,12 Einheiten unter der unteren Grenze des 95% Konfidenzintervalls für die MUT-Studien liegt.

**Tabelle 3.6:** Vergleich der Ergebnisse der DIT-Studien und der MUT-Studien zum affektiven und zum kognitiven Aspekt der moralischen Urteilsfähigkeit.

	Affekt	Kognition
$\bar{r}$	0,21	0,41
95 %	$0,18 \leq \rho \leq 0,24$	$0,36 \leq \rho \leq 0,46$
$\sigma_{\bar{r}}^2$	0,0192	0,0307
$\sigma_e^2$	0,0123	0,0063
$\sigma_{\bar{\rho}}^2$	0,0069	0,0243
$\sigma_{\bar{r}}$	0,08	0,16
Anzahl Einzeleffekte	39	12
$n_{gesamt}$	2946	986

In Abbildung 3.1 sind die die Effektstärken für unterschiedliche Altersgruppen jeweils für DIT-Studien und MUT-Studien dargestellt<sup>15</sup>. Für alle Altersgruppen älter als 16 Jahre liegen die Korrelationskoeffizienten für den kognitiven Aspekt deutlich über denen des affektiven Aspekts. Lediglich für die Altersgruppe 10 bis 15 Jahre ist der Effekt für DIT-Studien größer. Desweiteren ist festzuhalten, dass der durchschnittliche Korrelationskoeffizient für DIT-Studien für ältere Altersgruppen abnimmt. Der größte Effekt in MUT-Studien zeigt sich in der Altersgruppe 16 bis 20 Jahre.

<sup>15</sup>Der Einfachheit halber werden im Folgenden und in den Abbildungen Studien zum affektiven Aspekt als DIT-Studien bezeichnet und Studien zum kognitiven Aspekt als MUT-Studien.

In Abbildung 3.2 sind die Effektstärken für die unterschiedliche Dauer der Intervention jeweils für DIT-Studien und MUT-Studien dargestellt. Für beiden Studientypen zeigt sich der gleiche Verlauf: Der Effekt ist bei Interventionen bis 8 Wochen am stärksten; bei längerer Dauer nimmt der Effekt ab. Es zeigt sich fast kein Unterschied für die Dauer von 8 bis 10 Wochen und 15 bis 20 Wochen. Dabei fällt auf, dass die Effekte für den kognitiven Aspekt deutlich über denen des affektiven Aspekts liegen.

In Abbildung 3.3 sind die Effektstärken für die unterschiedliche Intensitäten der Intervention jeweils für DIT-Studien und MUT-Studien dargestellt. Auffällig ist, dass der Effekt für DIT-Studien mit zunehmender Intensität zunimmt, wohingegen der Effekt für MUT-Studien abnimmt. Während Intensitäten von weniger als einer Stunde und von einer bis zwei Stunden einen deutlicheren Effekt auf den kognitiven Aspekt des moralischen Urteils haben, ist der Effekt für DIT-Studien deren Intensität zwischen zwei und fünf Stunden liegt größer als für entsprechende MUT-Studien.

Abbildung 3.4 stellt die Effektstärken für die unterschiedlichen Interventionsarten jeweils für DIT-Studien und MUT-Studien dar. Die Methode der Dilemmadiskussion übt auf den kognitiven Aspekt des moralischen Urteils einen deutlich stärkeren Einfluss aus ( $r = 48$ ), als auf den affektiven Aspekt ( $r = 0, 19$ ). Für andere Interventionsmethoden nimmt die Effektstärke für MUT-Studien ab ( $r = 0, 35$ ), wohingegen sie für DIT-Studie steigt ( $r = 0, 22$ ).

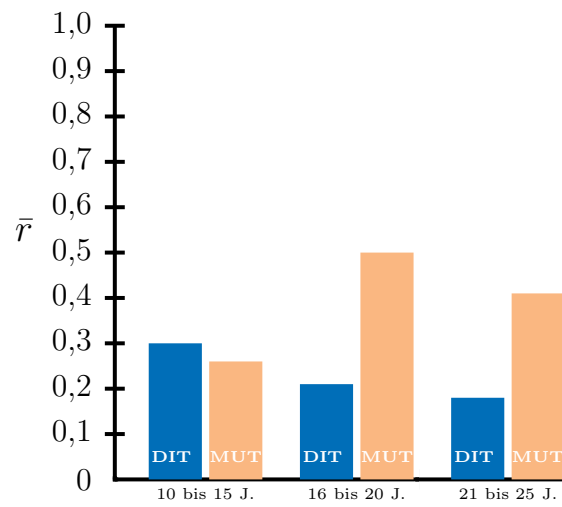


Abbildung 3.1: Interventionseffekte für unterschiedliche Altersgruppen

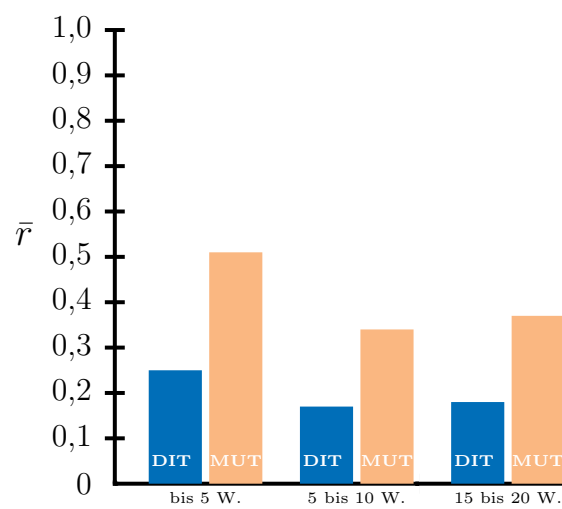
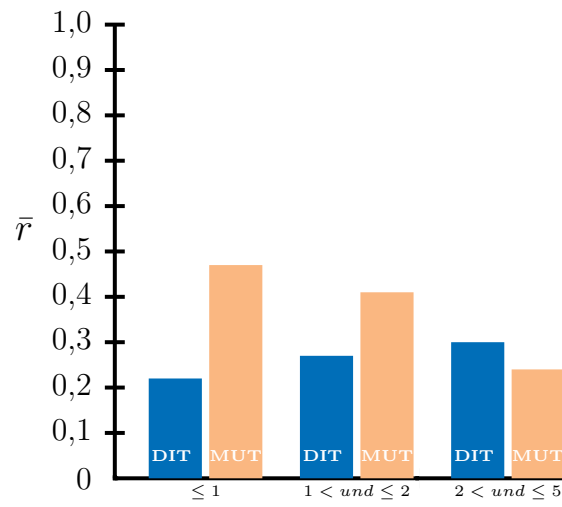
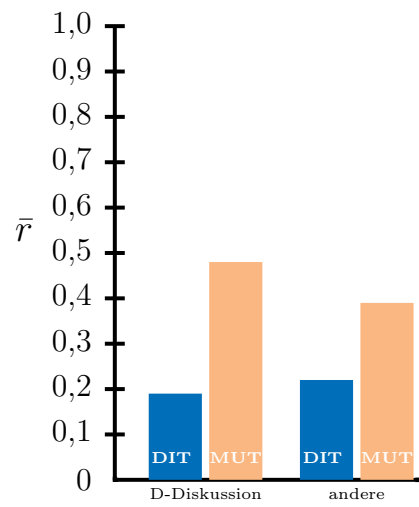


Abbildung 3.2: Effektstärken für die unterschiedliche Dauer der Intervention





**Abbildung 3.3:** Effektstärken in Abhängigkeit von der Intensität der Intervention in h pro Woche



**Abbildung 3.4:** Effektstärken für verschiedene Interventionsarten

# Kapitel 4

## Diskussion

Dieser Teil der Arbeit ist in sieben Abschnitte unterteilt: Um die Wirkung von Interventionsstudien auf die zwei unterschiedlichen Aspekte der moralischen Urteilsfähigkeit getrennt zu behandeln, werden im ersten Abschnitt dieser Diskussion die Ergebnisse zum affektiven Aspekt mit den Hypothesen aus der Studie von Schlaefli et al. [Schlaefli et al., 1985] verglichen. Im Anschluss daran werden die Hypothesen der Studie von Lind [Lind, 2000] mit den vorliegenden Daten verglichen. Daran schließt sich eine Diskussion der Ergebnisse zwischen affektivem und kognitivem Aspekt an. Der vierte und fünfte Abschnitt enthält eine kritische Auseinandersetzung mit der Aussagekraft dieser Arbeit und den hier verwendeten Methoden. Im Anschluss wird ein alternativer Berechnungsweg der relativen Effektstärke kurz vorgestellt. Das Ende bilden einige abschließende Bemerkungen.

### 4.1 Der affektive Aspekt – Prüfen der Hypothesen aus der Schlaefli et al. - Studie

Die durchschnittliche Effektstärke der DIT-Studien ist mit  $r = 0,21$  in einem moderaten Bereich. Der Signifikanztest aus Abschnitt 3.2.1 bringt uns zu dem Schluss, dass dieser Effekt auf einem Niveau von  $\alpha = 0,05$  überzufällig von Null abweicht. Die Daten zeichnen ein heterogenes Bild: Das Minimum reicht von  $\min(r)=-0,21$  bis  $\max(r)=0,71$ . Dabei bleibt zu notieren, dass fünf Studien negative bzw. keine Effekte von Interventionen berichten. Der Großteil der Effekte bewegt sich in einem Bereich zwischen  $r = 0,1$  und  $r = 0,5$ . Drei Studien berichten von Effektstärken von  $r > 0,5$ .

Sehr bedauerlich ist der Umstand, dass nur zwei Artikel einen 'follow-up' - Test berichten, denn gerade diese Messungen können angemessene Aussagen über den nachhaltigen Effekt der jeweiligen Intervention machen. Die reinen 'Posttest'

**Tabelle 4.1:** Vergleich der Ergebnisse aus der Schlaefli et al. - Studie mit den neuen Resultaten

	Schlaefli et al. 1985	Meta-Analyse 2007
Genereller Effekt von Interventionstudien	0,14	0,21
Effekt von Dilemmadiskussionen	0,2	0,19
Effekt von Unterricht und Wertediskussionen	0,09	0,22
Interventionseffekt für Vpn <16 Jahre	0,11	0,3
Interventionseffekt für Vpn bis 18 Jahre	0,11	0,21
Interventionseffekt bei Studenten <25 Jahre	0,13	0,18
Interventionseffekt bei Erwachsenen >25	0,29	0,15
Kurze Interventionen (0 bis 3w)	0,04	0,25
Interventionen von 4 bis 12 Wochen	0,16	0,1 bis 0,17
Interventionen von 13 bis 28w	0,15	0,03 bis 0,26

- Werte können einen Scheineffekt widerspiegeln: Die Versuchspersonen verhalten sich während des Programms und der abschließenden Messung den Erwartungen ihrer Versuchsleiter entsprechend.

Es kann nicht ausgeschlossen werden, dass die Versuchspersonen den Leitern einen Gefallen tun möchten und die Effekte künstlich nach oben 'verfälschen'. Erst ein späterer Nachtest (nachdem die Intervention für die Betroffenen ein Stück weit entfernt liegt) kann derartige Spekulationen gehaltlos machen. In Tabelle 4.1 sind die zentralen Effektstärken der Schlaefli - Studie den eigenen Ergebnissen gegenübergestellt. Da Schlaefli und Kollgegen  $d$ -Maße als Effektstärken angegeben haben, wurden diese zur besseren Vergleichbarkeit in  $r$ -Maße umgewandelt (siehe Anhang C, Liste 8).

Der deutliche Unterschied zwischen den Gesamtkorrelationen ( $r = 0,14$  und  $r = 0,21$ ) ist nicht von der Hand zu weisen. Er lässt den Schluss zu, dass die Interventionsmethodik effizienter geworden ist (da die Messinstrumente die gleichen sind). Das könnte u.a. darin zu sehen sein, dass die Effekte für Dilemmadiskussionen (Methodik wenig veränderbar) gleich geblieben sind und die der Unterrichtsinterventionen (relativ viel Raum zu Veränderung) zugenommen haben.

Vermutlich führt die intensivere Auseinandersetzung mit den Theorien Kohlbergs zu einem besseren Verständniss der Kernaussagen. Desweiteren werden Interventionsstudien in verschiedenen Bereichen durchgeführt (siehe Tabelle im Anhang A) was von zunehmender Popularität der Psychologie der Moralentwicklung

zeugt. Interessant sind die Effekte für die unterschiedlichen Altersgruppen: Während in der Studie von 1985 die Effekte mit dem Alter zunehmen ist dies in der vorliegenden Meta-Analyse genau umgekehrt. Dabei sind die Effekte für die Altersgruppen von 16 bis 18 Jahre und 19 bis 25 Jahre, ähnlich den Ergebnissen der Schläfli et al. - Studie, ungefähr gleich ( $r = 0,21$  und  $r = 0,18$ ). Aus den neuen Daten wird ersichtlich, dass auch kurze Interventionen bedeutsame Effekte haben können ( $r = 0,25$ ).

Im Folgenden werden die Hypothesen besprochen [Schlaefli et al., 1985, S. 346 - 347]:

(1)

“Moral education programs emphasizing dilemma discussion [...] produce modest but definite effects, [...]“

Diese Hypothese lässt sich mit den Daten der vorliegenden Arbeit bestätigen. Die Effektstärken sind moderat von  $r = 0,18$  (ausschließlich Dilemmadiskussionen in der Intervention) und  $r = 0,21$  (Dilemmadiskussionen gepaart mit Unterricht).

(2)

“Academic courses in the humanities and social studies do not seem to have an impact on moral judgement development.“

Insofern 'academic courses' mit der hier verwendeten Kodierung 'Ethikunterricht' verglichen werden kann, ist diese Hypothese auf Grundlage der derzeitigen Daten widerlegt. Der Effekt für Unterricht liegt bei  $r = 0,19$  und für andere Methoden sogar bei  $r = 0,22$ .

(3)

“Programs with adults (24 years and older) seem to produce larger effect sizes than programs for younger subjects [...]“

Wie in Tabelle 4.1 dargestellt, widerlegen die aktuellen Daten diese Hypothese. Erwachsene haben mit  $r = 0,15$  kleiner Effekte als alle anderen Gruppen  $0,18 < r < 0,3$ .

(4)

“Interventions longer than 12 weeks have no more impact than interventions of 3 to 12 weeks; however, duration less than 3 weeks tend to be ineffective [...]“

Diese Hypothese lässt sich teilweise bestätigen. Studien von kurzer Dauer (bis 3 Wochen) haben mit  $r=0,25$  einen deutlichen Effekt. Die Effekte für Studien von 4 bis 12 Wochen sowie 12 Wochen und länger unterscheiden sich kaum ( $r = 0,1$  bis  $r = 0,17$  bzw.  $r = 0,03$  bis  $r = 0,26$ <sup>1</sup>).

## 4.2 Der kognitive Aspekt der moralischen Urteilsfähigkeit

Die durchschnittliche Effektstärke von  $r = 0,41$  entspricht dem Ergebnis der Meta-Analyse von 1985 mit  $r = 0,42$ . In Tabelle 4.2 sind die Befunde der Studie von Lind [Lind, 2000] und der Meta-Analyse von 2007 einander gegenübergestellt. Es zeigt sich, dass kurze und mittlere Interventionen die gleichen Effektstärken haben wie in der Studie von 1985. Längere Interventionen haben einen größeren Effekt auf den kognitiven Aspekt der moralischen Urteilsfähigkeit als 1985 vermutet. Dieser ist mit  $r = 0,47$  fast genauso groß wie jener für kurze Interventionen ( $r = 0,51$ ). Bei der Intensität von Interventionen kommt die vorliegende Arbeit zu fast gleichen Ergebnissen wie die Meta-Analyse von 1985.

Der Effekt der Dilemmadiskussionsmethode steigert sich fast auf  $r = 0,5$ . Der Effekt für Personen im Alter von 11 bis 16 Jahren ist mit  $r = 0,26$  geringer als 1985 ( $r = 0,52$ ). Dabei ist zu beachten, dass in der vorliegenden Arbeit lediglich zwei Studien für diese Altersgruppe berücksichtigt werden konnten. Für die anderen Altersgruppen zeigen sich ähnliche Effekte wie 1985.

---

<sup>1</sup>Meines Erachtens nach lässt die relativ hohe Restvarianz den Schluss, dass auch diese Werte sich deutlich unterscheiden, nicht zu.

**Tabelle 4.2:** Vergleich der Ergebnisse der Studie von Lind und der Meta-Analyse von 2007

	Lind 1985	Meta-Analyse 2007
Genereller Effekt	0,42	0,41
Kurze Interventionen	0,52	0,51
Mittlere Interventionen	0,37	0,34
Längere Interventionen	0,3	0,47
Intensität der Intervention $\leq 1h$	0,48	0,47
Intensität der Intervention $\geq 1h$	0,32	0,24 bis 0,41
Dilemmadiskussion	0,44	0,48
Keine Dilemmadiskussion	0,43	0,39
Alter der Vpn 11 bis 16	0,52	0,26
Alter der Vpn 17 bis 20	0,48	0,5
Alter $\geq 21$	0,47	0,41

Die geringe Anzahl an 'follow-up' - Tests lässt keinen Vergleich mit der Lind-Studie bezüglich der Nachhaltigkeit der Interventionen zu. Im Folgenden wird auf die Hypothesen eingegangen:

(1)

“Die auf Kohlberg gründende Dilemma-Diskussions-Methode [...] 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“

Diese Hypothese hat sich bestätigt. Die durchschnittliche Effektstärke von  $\bar{r} = 0,41$  zeigt einen deutlichen Effekt auf. Besonders für Studenten ist der Effekt mit  $r = 0,46$  groß. Die Dilemmadiskussionsmethode zeigt einen überdurchschnittlichen Effekt auf den kognitiven Aspekt der moralischen Urteilsfähigkeit mit  $r = 0,48$ .

(2)

Der deutlichste Erfolg der Blatt-Kohlberg Methode wurde in der Gruppe der 11 bis 16 Jährigen verzeichnet. Hier zeigt sich auch die größte Nachhaltigkeit.

Diese Hypothese hat sich nicht bestätigt. In Die Gruppe der 10 bis 15 Jährigen hatte mit  $r = 0,26$  den geringsten Effekt. Wie bereits erwähnt lässt die geringe Anzahl an miteinbezogenen Studien keine zuverlässigen Schlüsse zu. Der größte Effekt zeigte sich für die Gruppe der 16 bis 20 Jährigen ( $r = 0,5$ ).

(3)

„Bei der Dauer der Intervention scheint das Optimum bei 3 bis 10 Stunden zu liegen.“

Diese Hypothese hat sich bestätigt. Sowohl in der Dauer als auch der Intensität zeigten sich die größten Effekte für kurze Intervention ( $r = 0,51$  bzw.  $r = 0,47$ ).

(4)

Die beträchtliche Wirkung von Studien kürzer als 3h konnten in späteren Nachtests nicht bestätigt werden ( $r = 0,05$ ; 2 Studien).

Über diese Hypothese kann anhand der vorliegenden Daten keine Aussage gemacht werden.

(5)

Moralpädagogische Interventionen von längerer Dauer haben einen Effekt von  $r = 0,3$  (16 Studien). Dieser ist nicht nachhaltig ( $r = 0,1$  (5 Studien)).

Diese Hypothese hat sich teilweise bestätigt. Die Effektstärke für Interventionen mit einer Dauer von 16 bis 20 Wochen liegt sogar bei  $r = 0,37$ .

(6)

Kurze (bezüglich der Intensität) Interventionen haben bereits einen starken Einfluss.

Diese Hypothese hat sich bestätigt (siehe (3)).

(7)

Lange Interventionen (bez. der Intensität) haben einen nachhaltigeren Einfluss als kurze.

Anhand der vorliegenden Daten kann über diese Hypothese keine Aussage gemacht werden.

(8)

Die Blatt-Kohlberg-Methode und andere Interventionsmethoden haben kurzfristig den gleichen Effekt, unterscheiden sich jedoch hinsichtlich der Nachhaltigkeit zugunsten der Dilemma-Diskussion.

Zur Überprüfung dieser Hypothese liefert die vorliegende Studie keine Informationen.

In der Forschung haben sich die Zuwächse für Interventionen gemessen in Absolutwerten der C-Skala des MUT als guter Indikator für die Stärke des Effekts erwiesen; unabhängig von der Varianz der Stichproben. Der in der vorliegenden Arbeit ermittelte durchschnittliche Zuwachs ist 12,88 Punkte.

### 4.3 Vergleich von DIT-Studien und MUT-Studien

Interventionsstudien zur Förderung der moralischen Urteilsfähigkeit haben einen deutlichen jedoch klar unterscheidbaren Effekt auf den affektiven und den kognitiven Aspekt des moralischen Urteilens. Die Übereinstimmung mit den Ergebnissen von Lind aus dem Jahr 1985 scheint eine langfristige Bestätigung der Zwei-Aspekte Theorie der Moralentwicklung zu sein. Aufgrund der breiten Datenlage ließe sich die Vermutung aufstellen, dass ein und dieselbe Intervention unterschiedliche Effekte auf die zwei Aspekte des moralischen Urteils hat.

Machen wir uns bewusst, dass die moralische Einstellung einer Person zunächst von seinem Wissen und seiner Erfahrung abhängt. Welche Einstellung bzw Motivation moralisch „gut“ oder „schlecht“ ist, wissen wir meist nur aus eigener Erfahrung oder jemand bringt es uns bei. Im Laufe seiner Entwicklung sieht sich der Mensch vielen unterschiedlichen Situationen ausgesetzt, in denen seine moralische Urteilsfähigkeit gefragt ist. Während Kleinkinder über den Interaktionsrahmen des engen Familienkreises und der Spielgefährten nicht hinausreicht, so muss der heranwachsende Mensch zunehmend mehr Situationen bewältigen, in denen er seine Einstellungen überdenkt.

Viele Forscher haben darauf aufmerksam gemacht, dass es Phasen in der Entwicklung des Menschen gibt, die durch die unterschiedliche Bereitschaft zur Einstellungsänderung gekennzeichnet sind. Interventionsstudien können helfen, Menschen auch ohne direkte Erfahrung in ihrer moralischen Orientierung zu verändern. Über das Vermitteln von Wissen und die Erprobung in Peer-Interaktionen (beides Interventionsarten) kann die moralische Einstellung gefördert werden. Im Sinne Kohlbergs würden dies bedeuten, die Person wird den Ideen höherer moralischer Stufen ausgesetzt, erprobt sie in Diskussionen, gelangt zu Einsicht und passt ihre Einstellung an.

Es ist daher nicht verwunderlich, dass beide Meta-Analysen (1985 und 2007) trotz unterschiedlicher Teilergebnisse (Altersabhängigkeit, Dauer der Intervention) zu moderaten (0,14 und 0,21) Ergebnissen kommen. Die moralische Einstellung eines Menschen ist etwas sehr Persönliches. Die Chancen, dass kurze Interventionen die richtige 'Passung' zwischen Interventionsinhalt und Werterhaltung der einzelnen Personen finden, scheinen zunächst gering. Dies spiegelt sich in der Bandbreite der Effekte wieder (von negativ bis positiv). Die höheren Effekte der neuen Studie zeigen zwar, dass die Methoden sich weiterentwickelt haben, jedoch ist der Effekt nicht mehr als morderat.



Dies lässt die Schlussfolgerung zu, dass Interventionsstudien wie wir sie bisher kennen, keinen großen Einfluss auf den affektiven Aspekt haben. Es ist der andere Aspekt des moralischen Urteils, bei dem die Wirkung deutlicher zu sehen ist. Lind schreibt in seinem Buch 'Ist Moral lehrbar?' [Lind, 2000], dass der kognitive Aspekt des moralischen Urteils in erster Linie eine Fähigkeit ist. Dabei schließt er sich den Überlegungen Kohlbergs an, wonach das moralische Urteil die Brücke schlägt zwischen den inneren Werten einer Person einerseits und ihrem (moralischen) Verhalten andererseits.

Eine Person kann hohe moralische Prinzipien verfolgen; wenn ihr jedoch die kognitiven Möglichkeit fehlen, kann sie nicht entsprechend dieser Einstellungen handeln. Eine Fertigkeit muss geübt werden. Dieses Einüben kann im Rahmen von Interventionen erfolgen. Der generelle Effekt von  $r = 0,41$  zeigt deutlich, dass diese Formen des Eingreifens im Sinne Linds (MUT) aber auch Kohlbergs (MJI) einen erheblichen Einfluss haben. Die Schlussfolgerung ist, dass eine Person nach einer Intervention fähiger ist moralisch zu urteilen. Es scheint daher nicht verwunderlich, dass Interventionen in denen Fertigkeiten vermittelt werden einen erheblichen Einfluss auf den kognitiven Aspekt, nicht aber auf den affektiven Aspekt haben.

Die Forschung der letzten Jahre scheint dazu geführt zu haben, dass die Methoden weiterentwickelt wurden, um durchweg positive Ergebnisse zu erzielen. Dabei muss darauf hingewiesen werden, dass diese Förderung scheinbar unabhängig von den eigenen moralischen Erfahrungen und Kenntnissen der Probanden verläuft: Junge Versuchspersonen, die offensichtlich weniger Erfahrungen gesammelt haben als Erwachsene profitieren nicht weniger von Interventionen. Im Gegenteil: Es zeigen sich hier deutlichere Effekte. Findet sich hierin eine Bestätigung der Annahme, dass kognitive Strukturen in jungen Jahren leichter 'umzubauen' sind?

#### **4.4 Schwächen dieser Arbeit und Beschränkung der Aussagekraft der Ergebnisse**

Eine Meta-Analyse ist zunächst nur so gut, wie die in ihr analysierten Studien. Wenn diese bereits enorme methodologische Schwächen aufweisen, dann ist auch die Analyse der Daten im Voraus verzehrt. Ähnlich der digitalen Bearbeitung von Bildern kann aus einer Aufnahme mit geringer Auflösung (größerer Unschärfe) im Nachhinein kein schärferes Bild gezeichnet werden.

Die Stärke der Meta-Analyse liegt u.a. darin, durch die relativ große Anzahl der Einzeleffekte, Gebrauch von statistischen Verfahren zu machen. Die vorliegende Arbeit bezieht sich (noch) auf eine geringe Anzahl von Studien. Dies ist zunächst durch den knappen Bearbeitungszeitraum für die Erstellung bedingt. Es ist nicht auszuschließen, dass unter der Berücksichtigung aller theoretisch verfügbaren Studien wesentlich andere Ergebnisse zu erwarten wären; vielleicht sogar gegenteilige. Das Problem der selektiven Auswahl ist in der vorliegenden Arbeit offensichtlich. Es wurden zur Untersuchung des affektiven Aspekts ausschließlich in Journals veröffentlichte Studien berücksichtigt. Die miteinbezogenen Dissertationen sind ebenfalls veröffentlicht sowie einige publizierte Dissertationen. Da sich in der Regel nur Studien veröffentlichen lassen, in denen Effekte berichtet werden, fehlen in dieser Arbeit all jene die vermutlich zu keinen positiven Ergebnissen gekommen waren.

Es ist also wieder nicht auszuschließen, dass zu jedem Satz an Interventionen (z.B. jene die Dilemmadiskussionen verwendeten) eine gleiche Anzahl von Studien durchgeführt wurden, in denen keine Effekte berichtet wurden. Trotz dieser Limitierungen lieferten die meisten Studien genug Informationen, um alternativ zu den angegebenen statistischen Effektmaßen das  $g$  nach Hedges&Olkin aus den Standardabweichungen und Mittelwerten zu berechnen. Desweiteren ist zu beachten, dass bereits während einer Intervention nicht alle Einsteiger auch an den Nacherhebungen teilnehmen. Dies ist wohlmöglich bereits ein besonderer Anteil an Teilnehmern, die sich vermutlich schon im 'commitment' von den anderen unterscheiden.

Auf der Ebene der Meta-Analyse können oftmals nicht alle Studien miteinbezogen werden. So lieferten z.B. zwei Studien lediglich Signifikanzniveaus als Ergebnis. Wieder andere stellten ihre Ergebnisse in einer Art und Weise dar, die eine Umrechnung in einen repräsentativen Effekt im Sinne des Korrelationseffekts nicht möglich machten.

Es ist auffällig, dass lediglich eine Studie ein Effektmaß in den Resultaten angab. Die anderen verwiesen stets auf Signifikanztests. Im Laufe der Analyse konnten nicht immer alle Studien den Untergruppen der Moderatorsuche zugeordnet werden, da sie für die jeweilige Kategorie keine entsprechenden Informationen lieferten. Diese wurden automatisch nicht in die Analyse miteinbezogen. Eine Übersicht aller Studien findet sich im Anhang A. Dort ist auch nachzulesen, welche Kategorien kodiert werden konnten. Desweiteren sind die Studiennummerierungen aus dem Resultateteil eindeutig denen im Anhang zuzuordnen.

Prinzipiell wurden alle Effekte in die Analyse miteinbezogen, unabhängig von den jeweiligen methodologischen Schwächen der Studien. Somit ist der 'Pool' so heterogen wie möglich: gute sowie schlechte Studien, kleine und große Effekte wurden berücksichtigt. Dies soll der besseren Generalisierung der Ergebnisse dienen. Viele Studien lieferten nicht die Informationen, um allen Hypothesen aus den früheren Artikeln nachzugehen. Die bei Lind gefundenen Ergebnisse, dass geschultes Personal bei der Durchführung der Intervention nachhaltigere Effekte bewirkt als nicht-geschultes, konnten in dieser Arbeit nicht untersucht werden. Meist wurden Angaben in den Artikeln nur dann gemacht, wenn geschultes Personal eingesetzt wurde. Es wäre meiner Ansicht nach wenig brauchbar gewesen, davon auszugehen dass, alle Studien, die diese Angaben nicht machten, kein geschultes Personal verwendeten.

Abschließend noch eine Bemerkung: Die Suche nach Schwächen, Tücken und Limitierungen von Meta-analysen kann beliebig bis ins Detail ausgedehnt werden. Dabei läuft man Gefahr von den eigentlichen Fragestellungen abzukommen; also den Blick für das Wesentliche zu verlieren. Halten wir uns vor Augen: Wir suchen nicht die Nadel im Heuhaufen oder das Zünglein an der Waage. Wir suchen nach Trends oder Tendenzen. Aussagen der Art: 'Wenn ein Intervention von 45 Minuten auf 1h verlängert wird und nur Versuchspersonen jünger als 16 Jahre mit zwei hypothetischen Dilemmas und drei realen konfrontiert werden, dann hat man den maximalen Effekt.' interessieren uns nicht. Vielmehr beschäftigt uns die Frage ob Interventionsstudien einen Effekt haben und welche Faktoren diesen beeinflussen könnten?

## 4.5 Kritische Auseinandersetzung mit der hier verwendeten Methode der Meta-Analyse

Bei der Auswertung der Daten orientiert sich diese Arbeit an den Ausführungen von Hunter&Schmidt. Sie gehen generell von 12 Studienartefakten aus, von denen 11 auf der Stufe der Meta-Analyse korrigierbar sind. Zu diesen 'study artifacts' gehören der Stichprobenfehler, Messgenauigkeit der Instrumente, deren Konstruktvalidität u.a. Hunter&Schmidt nennen eine Meta-Analyse, die nur den Stichprobenfehler korrigiert eine „BARE BONES Meta-Analysis“ [Hunter und Schmidt, 1990, S. 100]. Angesichts der vielen Korrekturmöglichkeiten des Korrelationskoeffizienten, die dem Untersucher zur Verfügung stehen, verhielte sich dieser wie ein Strauß der den Sand in den Kopf steckt, würde er nicht von den anderen Artefaktkorrekturen Gebrauch machen [Hunter und Schmidt, 1990, S. 100]. Die Korrektur eines Artefakts führt zu einer Verbesserung der Korrelation 'nach oben'. Damit

soll dem Umstand Rechnung getragen werden, dass die Effekte in Einzelstudien systematisch unterschätzt werden. Ich habe in dieser Meta-Analyse lediglich von der Stichprobenfehlerkorrektur Gebrauch gemacht, aus den folgenden Gründen:

1. Diese Artefaktkorrektur verändert nicht die durchschnittliche Effektstärke.
2. Die korrigierte Varianz des gemittelten Korrelationskoeffizienten gibt einen statistisch nachvollziehbaren Anlass für die Suche nach moderierenden Variablen.
3. Eine Variable kann dann als Moderator des Effekts gewertet werden, wenn die Analyse der einzelnen Gruppen der Kategorien unterschiedliche Korrelationen ergibt und wenn die Varianz innerhalb der Stichproben (wieder korrigiert nach Stichprobenfehler) gleich und nahezu Null ist.

Von den anderen Korrekturen wurde nicht Gebrauch gemacht, da sie für die eigentliche Fragestellung, ob es einen Effekt gibt, nicht entscheidend sind. Ein Effektmaß von 0,21 oder 0,41 führt zu der gleichen Schlussfolgerung wie eventuell nach oben korrigierbare Pendant.

Es ist kritisch anzumerken, dass Hunter&Schmidt in ihrer Methode die Korrelationskoeffizienten nicht in  $z$ -Werte (nach Fisher) transformieren. Dieser Schritt ist eigentlich unumgänglich, da erst die  $z$ -Werte<sup>2</sup> annähernd normalverteilt sind. Es wäre zu argumentieren, dass sich  $z$ -Werte und die entsprechenden Korrelationskoeffizienten bis zu einem Wert von 0,5 kaum unterscheiden. Jedoch zeigen die Daten, dass es durchaus Korrelationen gibt, die jenseits von  $r = 0,5$  liegen. Wenn z.B. alle Studienkorrelationen in  $z$ -Werte umgerechnet werden und daraufhin von diesen die durchschnittliche Korrelation berechnet wird, dann ergeben sich für den affektiven Aspekt  $r = 0,22$  und für den kognitiven Aspekt  $r = 0,43$ . Diese unterscheiden sich zwar nicht wesentlich von den bisher ermittelten, jedoch sind für beide Aspekte Studien mit Korrelationen über  $r = 0,5$  berücksichtigt wurden. Es ist nicht auszuschließen, dass andere Ergebnisse bei der Moderatorvariablenanalyse zustande kommen, wenn nicht mit  $z$ -transformierten Werten gerechnet wird.

## 4.6 Eine alternative Berechnung des Interventionseffekts

Betrachten wir folgendes Beispiel: Eine Studie berichtet in ihrer Untersuchung folgende Werte:

---

<sup>2</sup>Hierfür verwendete Formeln:  $z(r) = 0,5 \ln \frac{1+r}{1-r}$  und  $r(z) = \frac{e^{2z} - 1}{e^{2z} + 1}$

**Tabelle 4.3:** Beispiel 1

$n_{EG}$	$n_{KG}$	$EG_{pr}(\sigma)$	$EG_{post}(\sigma)$	$KG_{pr}(\sigma)$	$KG_{post}(\sigma)$	$\sigma_{gepoolt}$
55	28	39,8(14,1)	45,97(15,5)	45,68(14,2)	45,68(15,5)	15,5

Der Effekt der vorliegenden Intervention scheint recht groß zu sein. Die EG hat einen absoluten Punktzuwachs von über sechs Punkten, die Kontrollgruppe hingegen stagniert. Die Standardabweichungen weichen in allen vier Fällen kaum voneinander ab;  $\sigma_{gepoolt}$  scheint diese in den Gruppen gut anzunähern.

Bei der Berechnung des Effektmaßes erhalten wir  $g = 0,02$  und demnach ein  $r = 0,01$ . Folglich wäre diese Studie nahezu wirkungslos. Das Problem ist, dass die KG am Anfang der Intervention viel höher abschneidet als die EG. Das Effektmaß (und auch  $d$ ) berechnen jedoch den Effekt aus der Differenz der Gruppenmittelwerte nach der Intervention. Folglich sind sie nicht sensitiv für die unterschiedlichen Ausgangslagen. Da sich EG nach und KG nach der Intervention kaum unterscheiden, scheint der Effekt gleich Null zu sein.

Während der Untersuchungen hat sich ein alternativer Berechnungsweg als brauchbar(er) erwiesen: Anstatt nur die Mittelwerte am Ende der Intervention in den Effekt einfließen zu lassen, werden auch die Mittelwerte vor dem Beginn der Intervention berücksichtigt. Die Formel lautet dann wie folgt:

$$h = \frac{(\bar{X}_2 - \bar{X}_1) - (\bar{Y}_2 - \bar{Y}_1)}{s_{gepoolt}}$$

Dabei sind  $\bar{X}$  und  $\bar{Y}$  die Mittelwerte in der EG bzw. der KG. Für das obere Beispiel erhalten wir dann ein  $g$  von 0,4 und  $r = 0,19$ . Dieser Effekt scheint die Wirkung der Intervention besser wiederzugeben als der obere.

Dass es sich bei dieser Form der Berechnung nicht nur um eine Korrektur nach oben handelt, soll das folgende Beispiel zeigen:

**Tabelle 4.4:** Beispiel 2

$n_{EG}$	$n_{KG}$	$EG_{pr}(\sigma)$	$EG_{post}(\sigma)$	$KG_{pr}(\sigma)$	$KG_{post}(\sigma)$	$\sigma_{gepoolt}$
12	12	24,33(18,49)	26,91(19,95)	19,25(12,41)	21,5(13,13)	16,89

Hier scheint die Intervention offensichtlich keinen Effekt gehabt zu haben: Die Experimentalgruppe legt ca. 2,5 Punkte. Die KG erreicht einen ähnlichen Zuwachs

auch ohne Intervention. Die Schlussfolgerung wäre, dass die Intervention praktisch gar nicht für den Punktezuwachs verantwortlich gemacht werden kann. Wenn wir das Effektmaß wie oben zunächst über die Differenz der Mittelwerte nach der Intervention berechnen erhalten wir für  $g = 0,32$  und  $r = 0,17$ .

Dieses Ergebnis müsste den Leser stutzig machen. Beim näheren Betrachten fällt auf, dass die EG vor und nach der Intervention größere Werte hat als die KG. Die Differenz von  $26,91 - 21,5 = 5,41$  täuscht einen Interventionseffekt vor. Die alternative Berechnung ergibt ein  $g$  von  $0,02$  und  $r = 0,01$ . Dies scheint das Ergebnis der Intervention besser wieder zugeben.

Diese Art der Berechnung hat jedoch nicht nur den Vorteil, dass es die Unterschiede vor einem 'treatment' berücksichtigt. Betrachten wir folgendes Beispiel:

**Tabelle 4.5:** Beispiel 3

$n_{EG}$	$n_{KG}$	$EG_{pr}(\sigma)$	$EG_{post}(\sigma)$	$KG_{pr}(\sigma)$	$KG_{post}(\sigma)$	$\sigma_{gepoolt}$
32	32	23,97(12,8)	31,75(15,45)	27,25(13,36)	25,38(13,51)	14,51

In diesem Fall gibt das Effektmaß  $g = 0,44$  (bzw.  $r = 0,22$ ) den deutlichen Effekt der Intervention wieder. Dennoch zeigt dieses Beispiel, dass das 'treatment' in doppelter Hinsicht effektiv war: Bei seiner Anwendung führt es zu einem Punktezuwachs, beim Nicht-Anwenden zu einer Regression! Die alternative Berechnung ergibt:  $g = 0,67$  und  $r = 0,32$ .

Die drei Beispiele sind nicht fiktiv; es handelt sich um Studien aus dieser Arbeit. Allgemein lässt sich die alternative Berechnung wie folgt darstellen:

$$h = \frac{(\mu_{x2} - \mu_{x1}) - (\mu_{y2} - \mu_{y1})}{\sigma}$$

Es sei angemerkt, dass  $h$  und  $g$  nicht identische Dinge messen. Bei  $h$  betrachten wir die Differenz von Differenzen während für  $g$  die Differenz von Mittelwerten in das Effektmaß einfließen. Aus statistischer Sicht werden dabei zwei verschiedene Dinge miteinander verglichen. Dabei ist es legitim (wie beim Effektmaß  $d$  ebenfalls) von einem  $\sigma$  auszugehen, welches für beide Gruppen vorher und nachher repräsentativ (gleich) ist. Von den vorliegenden Studien berichteten nur wenige die Werte für die alternative Berechnung. Demnach ist es mir in dieser Arbeit nicht möglich gewesen, die generellen Effekte und Moderatorvariablenenergebnisse auf dem alternativen Weg zu berechnen.

## 4.7 Schlussbemerkungen

Interventionsstudien haben unabhängig von der verwendeten Methodik und den Studienteilnehmern einen moderaten Einfluss auf den affektiven Aspekt und einen großen Einfluss auf den kognitiven Aspekt der moralischen Urteilsfähigkeit. Auf der Grundlage der vorliegenden Daten können zusätzlich folgende Hypothesen aufgestellt werden:

- (1) Der Effekt einer Intervention auf den affektiven Aspekt ist für Personen über 16 Jahre geringer als für jüngere Personen.
- (2) Der Effekt auf den kognitiven Aspekt ist für kurze Interventionen (bis 5 Wochen) am größten.
- (3) Der Effekt einer Intervention auf den kognitiven Aspekt ist für eine Intensität von weniger als einer Stunde am größten und nimmt bei einer Zunahme der Intensität (eine bis fünf Stunden) ab.
- (4) Der Interventionseffekt auf den affektiven Aspekt nimmt mit zunehmender Intensität zu.
- (5) Die Dilemmadiskussionsmethode hat einen größeren Effekt auf den kognitiven Aspekt als andere Interventionsarten. Der Effekt auf den affektiven Aspekt ist für beide Bedingungen gleich, wobei der Effekt der Dilemmadiskussion etwas geringer ist.

Die Ergebnisse der Meta-Analyse für den kognitiven Aspekt bestätigen die bisherige Forschung [Lind, 2007]. Insbesondere die Steigerung der Absolutwerte weist mit durchschnittlich 12,88 Punkten auf einen, unabhängig von der Stichprobenvarianz, sehr starken Effekt hin.

Wie lassen sich die Ergebnisse erklären? Die Bildungstheorie der Moral sagt voraus, dass gezielte Interventionen die moralische Urteilsfähigkeit von Menschen erhöhen. Die unterschiedlichen Effekte für Affekt und Kognition lassen den Schluss zu, dass Interventionen wie sie bislang durchgeführt wurden, vor allem den kognitiven Aspekt beeinflussen. Für eine Verbesserung des Effekts für den affektiven Aspekt scheinen andere Interventionen nötig zu sein. Halten wir uns vor Augen, dass die moralische Motivation und Einstellung sehr viel stärker von Reifungsprozessen abhängig ist als die kognitiven Fähigkeiten. Während letztere durch Übung, wie beispielsweise bei der Konstanzer Methode der Dilemmadiskussion gefördert werden, scheint der moralische Affekt einer Person anders beeinflusst zu werden. Es wäre zu vermuten, dass erst langjährige Programme z.B. an Universitäten oder

Schulen die moralische Einstellung verändern, da sich dieser Aspekt der moralischen Urteilsfähigkeit *langsamer* entwickelt.

Für die Förderung des kognitiven Aspekts der moralischen Urteilsfähigkeit hat sich die Dilemmadiskussion als die effektivste Interventionsart erwiesen. Besonders der große Effekt von kurzen und wenig (zeitlich) intensiven Interventionen macht die Umsetzung z.B. im schulischen oder universitären Betrieb praktikabel. In seinen Untersuchungen konnte Lind die deutlichen Effekte von einmalig durchgeführten Diskussionen nach dem Vorbild der Konstanzer Methode der Dilemmadiskussion belegen [Lind, 2000].

Die weitere Forschung sollte sich darauf konzentrieren, in Interventionsstudien gezielt den kognitiven Aspekt zu untersuchen. Dabei wäre es wichtig weitere Erfahrungen bei der Verwendung der Dilemmadiskussionsmethode unter verschiedenen Bedingungen und in unterschiedlichen Kulturkreisen zu sammeln. Hinsichtlich des affektiven Aspekts sollten andere Interventionsarten entwickelt und die Beeinflussungsfaktoren genauer untersucht werden.

Für die weitere Forschung ist eine Verbesserung der Forschungsmethoden von großer Bedeutung. Insbesondere in der psychologischen Moralforschung haben die sozialwissenschaftlichen Methoden in den vergangenen Jahrzehnten, unabhängig von den Entwicklungen in den Neurowissenschaften, zu einem besseren Verständnis und zahlreichen neuen Erkenntnissen geführt.



# Anhang A

## Tabelle

Die folgende Tabelle listet alle in dieser Arbeit verwendeten Studien auf. Diese sind alphabetisch geordnet. In den Spaltenköpfen wird ggf. auf die Listen im Anhang C verwiesen, wo sich weitere Informationen befinden. Beispielsweise verweist 'L2' auf Liste 2. Längere Passagen wurden als Fußnote unterhalb der Tabelle eingefügt. Die Abkürzung VL steht für Versuchsleiter.

**Tabelle A.1:** Tabelle mit den Einzelergebnissen; alphabetisch geordnet

Quelle		Versuchspersonen					Expe. Design		Intervention					Messinstr.		Ergebnisse					
Nr.	Au- tor/Jahr	Pub- Art (L8)	Grup- pe (L1)	Alter	Schul- klas- se/Sem- ester (L2)	bes. Merk- male	spät. Nach- test	Design (L12)	VL	Art der In- ter- ven- tion (L3)	Inhalt der In- ter- ven- tion (L4)	Dauer (Wo- chen)	Inten- sität (h pro Woche)	Test (L5)	Mess- ska- la (L7)	Zu- wachs (ab- so- lut)	$n$	$F, t$ etc.	$r$ (L8)	$p$	
1	Arbuthnot (1986)	J(1)	G(9)	14,54	k.A.	Personen aus länd- lichen Gebie- ten	Ja (ein Jahr)	Q(2)	Ge- schul- tes Per- so- nal	A(1)	I(1) I(2) <sup>1</sup>	16w - 20w	45 min	T(1)	MMS	k.A.	$n_{EG} = 18$ $n_{KG} = 16$	$F(1, 32)$ $20, 36$	=	$r = 0, 62$ $r_{fu} = 0, 73$	$p <$ $0, 0001$
2	Bebeau (1994)	J(1)	G(5)	23	1. bis 4. Jahr	Keine	Nein	Q(3)	College Leh- rer	A(1)	I(1) I(3)	4 Jahre mit je 39h	<1h	T(3)	P - Score	4,53	$n_{EG} = 720$	k.A.	$r = 0, 18$	k.A.	
3	Binfet (2004)	J(1)	G(2)	11,6	6. bis 7. Klasse	Kanada	Nein	Q(1)	gesch. Leh- rer	A(1) A(6) <sup>2</sup>	I(1)	8w	40 min	T(7)	eigene	k.a.	$n = 97$	$t1=1,83$ und $t1=2$	$r = 0, 21$ und $r = 0, 24$	beide $p < 0, 05$	
4	Brugman (2003)	J(1)	G(4)	14,23	8. bis 10. Klasse	Russische Schüler aus Moskau und Um- gebung	Ja (3 Mo- nate)	Q(2)	Junior- psychologen	A(3)	I(3) I(7) <sup>3</sup>	4w	2h	MJST	k.A.	k.A.	$n_{EG} = 28$	$F(1, 26)$ $3, 74$	=	$r = 0, 35$	$p <$ $0, 05$

Fortsetzung...

<sup>1</sup>Leben, Eigentum, Wahrheit, Autorität, Recht

<sup>2</sup>individuelles, schriftliches Reflektieren

<sup>3</sup>Internalisierung von Werten zur generellen Verbesserung der moralischen Atmosphäre

Tabelle A.1 – Fortsetzung

5	Buttel (2003)	J(1)	G(9)	31	k.A.	Unter- teilung in Afro- amerika- ner und weiße Ameri- kaner	nein	Q(3)	k.A.	A(5) A(6) <sup>4</sup>	I(7) <sup>5</sup>	10w	2h	T(3)	P- Score	k.A.	$n_{EG1} = 12$ $n_{EG2} = 24$	$T(46)$ - ,46	=	$r_1 = 0,12$ $r_2 = 0,12$	$p <$ 0,65
6	Buttel (2001)	J(1)	G(9)	31	k.A.	Afro- amerika- ner und kau- kasische Ameri- kanische	nein	Q(3)	Straf- vollzugsbeamte	A(6) <sup>6</sup>	I(7) <sup>7</sup>	12w	2h	T(3)	P- Score	k.A.	$n_{EG} = 32$	$t(30)=0,98$	$r = 0$	$p <$ 0,26	
7	Canary (2007)	J(1)	G(5)	22,38	3. und 4. Jahr	Süd- Westen der USA	nein	Q(2)	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.
8	Cecil (1997)	J(2)	G(9)	13	k.A.	k.A.	nein	Q(2)	geschult	A(4)	I(3)	10w	1h	T(3)	P- Score	k.A.	$n = 16$	k.A.	$r = 0,06$	$p >$ 0,05	
9	Chase (1998)	J(2)	G(5)	35	k.A.	Virginia	nein	Q(2)	geschult	A(3)	I(7) 8	15w	2,5h	T(3)	P- Score	k.A.	$n = 38$	$F(1,34)=1,92$	$r = 0,23$	$p >$ 0,05	
10	Claypoole (2000)	J(1)	G(9)	14,5	k.A.	<sup>9</sup>	nein	Q(2)	k.A.	A(1)	I(1) I(2)	5w	2hj	T(4)	P- Score	2,08 und 2,58	$n_{EG1} = 12$ $n_{KG1} = 12$ $n : KG2 =$ 12	$F(3,47)$ 2,8	=	$r_1 = 0,08$ $r_2 = 0,17$	$p >$ 0,05
11	Comunian (2006)	J(1)	G(5)	24,17	Post- graduate stu- dent	keine	nein	Q(2)	Ge- schultes Per- so- nal	A(3) A(4) A(5)	I(3) I(5)	16w	1h	MUT	C- Score	8,37	$n_{EG} = 61$ $n_{KG} = 59$	$F(1;58)$ 1,88) $F(1;60)$ 10,4	=	$r = 0,39$	$p_1 <$ 0,17 $p_2 <$ 0,002

Fortsetzung...

<sup>4</sup>kognitiv-behaviorale Intervention

<sup>5</sup>psychoedukativer Unterricht

<sup>6</sup>kognitiv-behavioral; Gruppenübung, Wutmanagement, Fertigkeitentraining, 20 Personen mit 2 Gruppenführer

<sup>7</sup>Wutmanagement, Fertigkeitentraining

<sup>8</sup>Rollenspiele, Gruppenarbeit u.a.

<sup>9</sup>'schwererziehbare' Jugendliche in einer Besserungsanstalt

Tabelle A.1 – Fortsetzung

12	Dehaan et al. (1997)	J(1)	G(3)	16,3	7. bis 9. Klasse	„upper/middle class“	nein	Q(2)	Ge-schulte Lehrer	A(1) A(3)	I(1) I(2) I(3)	16w	5h	T(4)	P-Score	6,9; 9,6;3,7; 2,9	$n_{EG1} = 13$ $n_{EG2} = 15$ $n_{EG3} = 11$ $n_{KG} = 15$	$t_1 = 2,03, t_2 = 2,23, t_3 = -0,68, t_4 = 1,06$	=	$r_1 = 0,44,$ $r_2 = 0,46,$ $r_3 = 0,05,$ $r_4 = 0,09$	$p_1 < 0,05,$ $p < 0,05$ sonst n.s.
13	Dellaportas (2006)	J(1)	G(5)	22	4. Jahr	Keine	nein	Q(3)	Ge-schultes Personal	A(1) A(3)	I(3) I(6) I(7)	12w	1h	T(4)	P-Score	12,18	$n_{EG} = 26$	$T(26) = 2,38$	=	$r = 0,39$	$p < 0,02$
14	Gibbons (1995)	J(1)	G(2)	k.A.	4. bis 6. Klasse	Kanada „middle class“	nein	Q(1)	(Sport)-A(5) Lehrer	A(6) <sup>10</sup>	I(3)	7m	1,5h	T(7) <sup>11</sup>	eigene	0,32	$n_{KG} = 140$ $n_{EG1} = 146$ $n_{EG2} = 166$	<sup>12</sup>	$r = 0,38$	$p_1 < 0,0001;$ $p_2 < 0,0001;$ $p_3 < 0,0001;$ $p_4 < 0,0001$	
15	Glazer-Waldmann (1990)	J(1)	k.A.	34,6 & 30,3(KG)	k.A.	k.A.	nein	Q(2)	k.A.	A(4)	I(3) I(5) I(6) <sup>13</sup>	8w	2h	T(3)	P-Score	6,1 & 4,69	$F(1; 81) = 55$ $n_{EG} = 55$ $n_{KD} = 28$	=	$r = 0,01$	$P < 0,033$	
16	Grier et al. (1998)	J(1)	G(2)	11,26	k.A.	öffentliche Schule in Detroit	nein	Q(2)	k.A.	A(1)	I(1) I(2)	5w	0,5h	T(1)	k.A.	k.A.	$n_{EG} = 19$ $n_{KG} = 11$	$z(EG) = 0,52$ $z(KG) = 0,15$	=	$k, A,$	beide n.s.

Fortsetzung...

<sup>10</sup>'fair-play' im Sportunterricht oder in anderen Fächern implementiert.

<sup>11</sup>Horrock's Prosocial Play Behavior Inventory HPPBI entspricht Test mit 'judgement', 'intention' und 'reason'

<sup>12</sup> $F(\text{judgement}(2;449))=76,4$   $F(\text{reason}(2;449))=220,3$   $F(\text{intention}(2;449))=58,4$ ,  $F(\text{behavior}(2;449))=96,7$

<sup>13</sup>Verschwiegenheit, Wahrheit

Tabelle A.1 – Fortsetzung

17	Hößle (2001)	J(2)	G(4)	17- 19	11.&12. Klasse	k.A.	nein	Q(2)	Lehrer un- ter Su- per- vi- sion	A(3) <sup>14</sup>	I(1) I(2) I(5) I(6)	12w	3h	T(7)	Eigener	k.A.	$n_{EG1} = 15$ $n_{EG2} = 19$ $n_{EG3} = 9$	k.A.	k.A.	
18	Hoffman (1994)	J(2)	G(5)	18,5	first year	Maryland	Ja (8w)	Q(2)	geschult	A(4)	I(7) 15 I(6) <sup>16</sup>	3 Tage	6h	T(3)	P- Score	k.A.	$n = 40$	$T=2,43$	0,33	$p < 0,05$
19	Kim (2006)	J(2)	G(5)	18 bis 22	sophomore und junior	College sou- theast US	nein	Q(2)	nicht ge- schult	A(1) A(6) <sup>17</sup>	I(1) I(7)	2 bzw. 3 Wochen	1h bzw. 2h	T(6)	C- Score	k.A.	$n = 242$	k.A.	0,56	$p < 0,05$
20	Krawczyk (1997)	J(1)	G(5)	k.A.	k.A.	US Ameri- kaner, New England	nein	Q(3)	<sup>18</sup>	A(1) A(3)	I(1) I(3)	(a) 14w	1h	T(3)	P- Score	7,6	$n = 180$	$F$ (Postwerte) = 7,95	$r = 0,21$	$p <$ 0,001
21	Kruger (1992)	J(2)	G(1)	8,6	k.A.	„girl scouts in urban Atlanta“	nein	k.A.	k.A.	A(1) A(4)	k.A.	1 Tag	k.A.	k.A.	k.A.	k.A.	$n_{EG1} = 24$ $n_{EG2} = 24$	k.A.	k,A,	k.A.
22	Kuther et al. (1997)	J(1)	G(4)	k.A.	10. bis 12. Klasse	k.A.	nein	Q(2)	Lehrer	A(2)	k.A.	k.A.	k.A.	T(3)	P- Score	k.A.	$n_{EG} = 68$ $n_{KG} = 127$	k.A.	$r = 0,19$	
23	LaGrone (1996)	J(1)	G(5)	EG:31 KG:27,4	k.A.	„master degree pro- gramm“	Ja (6 Mo- nate)	Q(2)	k.A.	A(3)	I(4)	16w	1h	<sup>19</sup>	eigener	k.A.	$n_{EG} = 23$ $n_{KG} = 25$	k.A.	$r = 0,56$	$p <$ 0,022

Fortsetzung...

<sup>14</sup>Thema: Gentechnik

<sup>15</sup>gemeinsames Beten

<sup>16</sup>Wer bin ich? Religion

<sup>17</sup>semiotische Mediation am Computer

<sup>18</sup>(a) Doktorrand der Theologie (b) PhD-Gruppe (c) verschieden

<sup>19</sup>Ähnlich DIT nach Welton et al. (1994)

Tabelle A.1 – Fortsetzung

24	Lerkiat-bundit et al. (2006)	J(1)	G(5)	k.A.	k.A.	Medizinstudentin und Schwester auszubildende	ja	Q(1)	k.A.	A(1) <sup>20</sup>	I(1) I(2)	6w	1,5h	MUT	C-Score	14	$n_{EG} = 42$ $n_{KG} = 41$	k.A.	$r = 0,39$ $r_{post} = 0,37$	k.A.
25	Lies (2005)	J(2)	G(5)	8-22	undergraduate	Latina	nein	Q(2)	k.A.	A(6) <sup>21</sup>	I(2) I(7) <sup>22</sup>	8w	k.A.	T(3)	P-Score	k.A.	$n = 157$	k.A.	$r = 0,26$	k.A.
26	Lind (2003)	J(3)	G(5)	k.A.	k.A.	Lehramtstudentin	nein	Q(2)	geschult	A(1)	I(1) I(2)	1 Tag	1,5 h	T(6)	C-Score	1,6	$n = 48$	$F(1,41)=1,47$	$r = 0,24$	$P < 0,2316$
27	Lind (2007)	J(3)	G(5)	22	k.A.	Lehramtstudentin	nein	Q(2)	geschult	A(3)	I(7) Verantwortungsübernahme	12w	1,5 h	T(6)	C-Score	13,9 bzw. 5,9 bzw. 21,65	$n = 123$ bzw. $n = 71$ bzw. $n = 6$	k.A.	$r = 0,6$ bzw. $r = 0,11$ bzw. $r = 0,52$	k.A.
28	Lopez & Garcia (1998)	J(1)	G(1)	13	k.A.	„middle, lower-middle socio-economic background“	nein	Q(2)	Lehrer	A(6) <sup>23</sup>	k.A.	1 Jahr a 20 sessions	1h	T(3)	P-Score	7,69	$n_{EG} = 30$ $n_{KG} = 31$	$t(51,76) = 2,86$	$r = 0,36$	$p < 0,01$
29	Medrano (1994)	J(1)	G(1)	9 bis 14	5. bis 8. Klasse	Staatl. Schule in Vizcaya	nein	Q(3)	Geschulte Lehrer	A(1)	I(2) I(6) 24	6w	1h	T(1)	MMS	k.A.	$n_{EG} = 94$	k.A.	k.A.	k.A.
30	Mills (2005)	J(2)	G(5)	21,11	undergraduate	Latines Frauen	nein	Q(2)	geschult	A(3) A(4)	I(1) I(7) <sup>25</sup>	15w	k.A.	T(3)	P-Score	k.A.	$n = 43$	k.A.	-0,17	k.A.

Fortsetzung...

<sup>20</sup>KMDD: Arzt und Organspende; Student und drogenabhängiger Freund; u.a.

<sup>21</sup>Verantwortungsübernahme im sozialen Dienst

<sup>22</sup>Reflektieren der Erfahrungen in Kleingruppen

<sup>23</sup>selbstständiges, reflektierendes Lösen von Aufgaben

<sup>24</sup>Schulkonflikte, Diebstahl, Vandalismus

<sup>25</sup>Rollenspiele, guided reflection

Tabelle A.1 – Fortsetzung

31	Mouratidou (2007)	J(1)	G(3)	12,3	7. und 8. Klasse	Griechenland	nein	Q(1)	Sportlehrer	A(4)	I(7) <sup>26</sup>	16w	3h	T(6)	C-Score	3,37	$n = 157$	$F(1,152) = 5,31$	0,18	$p < 0,05$
32	Morgan (2000)	J(1)	G(5) G(6) G(8)	PA:30 KS:29	k.A.	Polizisten- anwärter und Krimi- nologie- studien- Virginia	nein	Q(2)	Lehr- stuhl- angestell- [Rollen- über- nahme]	A(1) A(6)	I(1) I(2)	10w	1h	T(4)	P-Score	7,78	$n_{EG} = 32$ $n_{KG} = 32$	$F = 8,88$	$r = 0,22$	$P < 0,05$
33	Munson (1988)	J(2)	G(2)	k.A.	k.A.	USA	nein	Q(2)	nein	A(1) A(6)	I(1)	8w	0,5h	T(3)	P-Score	k.A.	$n = 32$ bzw. $n = 39$ bzw. $n = 47$	k.A.	$r = 0,14$ bzw. $r = 0,16$ bzw. $r = 0,12$	$p < 0,05$
34	Nas (2005)	J(1)	G(9)	16	k.A.	Niederlande	nein	Q(2)	geschult	A(4)	I(7) <sup>27</sup>	8w	3h	T(7)	eigener	k.A.	$n = 56$	$F(1,45) = 1,3$	$r = 0,14$ bzw. $r = 0,16$ bzw. $r = 0,17$	n.s.
35	Reiman et al. (2002)	J(1)	G(7)	44	k.A.	Lehrerinnen aus amerika- nischen Vorstäd- ten mit 5 bis 30 Jahren Erfah- rung	nein	Q(2)	k.A.	A(3) A(5)	I(3)	7 Mo- nate mit 10 „ses- si- ons“	4h	T(3)	P-Score	9,5	$n_{EG} = 8$ $n_{KG} = 5$	$t(11) = 3,66$	$r = 0,5$	$P < 0,01$
36	Ries (1992)	J(1)	G(5)	18	k.A.	Illinois; Psycho- logiestu- denten	nein	Q(2)	k.A.	A(3)	I(5) I(6)	18w	3h	T(1)	Stufenscore	0,75	$n_{EG} = 12$ $n_{KG} = 9$	$t() = 4,89$	$r = 0,72$	$p < 0,001$

Fortsetzung...

<sup>26</sup>reziprokes Feedback-Lernen im Sportunterricht mit Gleichaltrigen

<sup>27</sup>Verhaltenstraining, guided reflection

Tabelle A.1 – Fortsetzung

37	Santilli et al. (1992)	J(1)	G(5)	22	„under-graduate“	k.A.	nein	Q(3)	Nicht geschult	A(3) A(4) <sup>28</sup>	I(3)	8w	3h 20min	T(4)	P-Score	10,56 und 8,38	$n_{EG1} = 9$ $n_{EG2} = 8$	$T_1(7) = 2,41, t(8) = 3,02, t(15) = 3,45$	$r_1 = 0,59$ $r_2 = 0,71$	$p < 0,025,$ $p < 0,01$ $p < 0,005$
38	Self et al (1993)	J(1)	G(5)	k.A.	Freshman	Texas A&M Uni Medicine Students	nein	Q(2)	k.A.	A(4) A(6) [Film- dis- kus- sion]	k.A.	8w	1h	T(3)	P-Score	6,23 und 6,18	$n_{EG1} = 48$ $n_{EG2} = 37$ $n_{KG} = 29$	k.A.	k.A.	$p < 0,002$ $p < 0,007$
39	Self et al. (1998)	J(1)	G(5)	k.A.	Medizin- studenten der Jahre 1991 bis 1998	Texas A&M Uni Medicine Students	nein	Q(3)	Nicht geschult	A(1) A(4)	k.A.	18w	1h	T(3)	P-Score	$\delta_1 = 0,65$ $\delta_2 = 0,14$ $\delta_3 = 4,3$ $\delta_4 = 6,25$ $\delta_5 = 5,34$	$n_{EG1} = 480$ $n_{EG2} = 101$ $n_{EG3} = 76$ $n_{EG4} = 47$ $n_{EG5} = 25$	k.A.	k.A.	$p_1 < 0,17$ $p_2 < 0,8548$ $p_3 < 0,01$ $p_4 < 0,01$ $p_5 < 0,04$
40	Senne et al. (2003)	J(1)	G(5)	k.a.	k.a.	Lehramt- studenten	nein	Q(3)	k.A.	A(5)	I(7) <sup>29</sup>	16w, EG3: 48w	1h	T(3)	P-Score	EG1: 2,89 KG1: 4,12 EG2: 3,02 KG2: 2,81; EG3: 10,38; KG3: -3	$n_{EG1} = 16$ $n_{KG1} = 19$ $n_{EG2} = 34$ $n_{KG2} = 33$ $n_{EG3} = 9$ $n_{KG3} = 5$	$t(14) = 2,415$	$r_1 = -0,21$ $r_2 = 0,04$ $r_3 = 0,24$	$P < 0,025$
41	Stoll et al. (1993)	J(1)	G(5)	k.A.	k.A.	Sport- studenten	nein	Q(2)	Lehrer	A(1) A(3) A(6) <sup>30</sup>	I(1) I(2)	18w	1h	T(3)	P-Score	9,6	$k, A,$	$d_{Glass}(EG) = 1,1$	$r = 0,48$	k.A.

Fortsetzung...

<sup>28</sup> Rollenspielübung

<sup>29</sup> praktisches und theoretisches Fähigkeitentraining für den Beruf

<sup>30</sup> moralisches Wissen, Wertschätzen und Handeln



Tabelle A.1 – Fortsetzung

42	Thompson-Grim (1999)	J(2)	G(3) G(4)	16	9.- 12.Klasse	Gefährdete Jugendliche	nein	Q(2)	geschult	A(6) <sup>31</sup>	I(7) <sup>32</sup>	9w	7,5h	T(3)	P-Score	k.A.	$n = 34$	k.A.	$r = 0,05$	n.s.
43	Turner et al. (1998)	J(1)	G(6)	26	k.A.	Schwestern- auszubildende	nein	Q(3)	Aus- bilder	A(3)	I(7) <sup>33</sup>	k.A.	1h	JAND	eigene	k.A.	$n_{EG} = 144$	$F = 0,57$	$r = 0,06$	$p < 0,569$
44	Wandzilak (1988)	J(1)	G(5)	k.A.	10. bis 12. Klasse	k.A.	nein	Q(2)	Sport- trainer	A(1)	I(1) I(2)	9w	15 min	T(3)	P-Score	1,21	$n_{EG} = 10$ $n_{KG} = 10$	$F(1, 18) = 0,15$	$r = 0,09$	$p > 0,05$
45	Walker & Taylor (1997)	J(1)	G(9)	16,5	k.A.	Männlich, alle Teil- nehmer hatte eine mind. eine Monats- Strafe	nein	Q(1)	Ge- schultes Per- so- nal	A(1) A(4)	I(1) I(2)	3 Tage	20 bis 40 Minuten	T(1)	WAS	13,3	$n = 41$	$F(1, 41) = 15,20$	$r = 0,53$	$p < 0,001$
46	Zierer (2006)	J(1)	G(1)	9	3.- 4.Klasse	k.A.	nein	Q(2)	nicht geschult	A(1)	I(1)	6w	1h	T(6)	C-Score	10,7	$n = 40$	$F(1, 38) = 4,42$	$r = 0,24$	$p < 0,0422$

<sup>31</sup>körperliche Gruppenbetätigungen

<sup>32</sup>physische Aktivitäten zur Stärkung der Gruppendynamik

<sup>33</sup>Unterricht in Schwesternpraxis

# Anhang B

## Verwendete Artikel

1. Arbuthnot, J. Gordon, D. A. (1986). Behavioral and Cognitive Effects of a Moral Reasoning Development Intervention for High-Risk Behavior-Disordered Adolescents. *Journal of Consulting and Clinical Psychology*, 54(2), 208
2. Bebeau, M.J., Thoma, S.J. (1994). The impact of a dental ethics curriculum on moral reasoning. *Journal of dental education*, 58(9), 684-692
3. Binfet, Tyler (2004). It's All in Their Heads: Reflective Abstraction as an Alternative to the Moral Discussion Group. *Merill-Palmer Quarterly*, 50(2), 181-200
4. Brugman, D., Podolskij, A.I., Heyman, P.G., Boom, J., Karabanova, O., Idobaeva, O. (2003). Perception of moral atmosphere in school and norm transgressive behaviour in adolescents: An intervention study. *International Journal of Behavioral Development*, 27(4), 289-300
5. Buttell, F.P. (2001). Moral Development among Court-Ordered Batterers: Evaluating the Impact of Treatment. *Research on Social Work Practice*, 11(1), 93-107
6. Buttell, F.P. (2003). Exploring the relevance of moral development as a treatment issue in batterer intervention. *Social Work Research*, 27(4), 232-241
7. Canary, H. E. (2007). Teaching ethics in communication courses: an investigation of instructional methods, course foci, and student outcomes. *Communication Education*, 56(2), 193-208
8. Cecil, D. (1997). A study of psychoeducational programs for troubled youth. Dissertation

9. Chase, N. (1998) A cognitive development approach to professional ethics training for counselor education students. Dissertation
10. Claypoole, S., Moody, E., Peace, S. (2000). Moral dilemma discussions: An effective group intervention for juvenile offenders. *Journal for Specialists in Group Work*, 25(4), 394-411
11. Comunian, A.L., Gielen, U.P., (2006). Promotion of moral judgement maturity through stimulation of social role-taking and social reflection: an Italian intervention study. *Journal of Moral Education*, 35(1), 51-69
12. Dehaan, R., Hanford, R., Kinlaw, K., Philler, D., Snarey, J. (1997). Promoting ethical reasoning, affect and behaviour among high school students: an evaluation of three teaching strategies. *Journal of Moral Education*
13. Dellaportas, S. (2006). Making a difference with a discrete course on accounting ethics. *Journal of Business Ethics*, 65(4), 391-404
14. Gibbons, S.L., Ebbeck, V., Weiss, M.R. (1995). Fair Play for Kids: effects on the moral development of children in physical education. *Research Quarterly for Exercise and Sport*, 66(3), 247-255
15. Grier, L.K., Firestone, I.J., (1998). The effects of an intervention to advance moral reasoning and efficacy. *Child Study Journal*, 28(4), 267-290
16. Hedl, J., Glazer, H., Chan, F. (2005). Improving the Moral Reasoning of Allied Health Students. *Journal of Allied Health*, 34(2), 121-122
17. Hößle, Corinna (1999) Untersuchung zur moralischen Urteilsfähigkeit von Schülern zum Thema Gentechnik: Eine Interventionsstudie
18. Hoffman, D. (1994) Nurturing spirituality: An assessment of the cognitive, moral judgment, and faith development of first-year college students participating in a developmental retreat program. Dissertation
19. Kim, Soyoung (2006) Juxtaposition of semiotic mediation with social mediation: the effect of text types and social interaction on moral judgement. Dissertation
20. Kuther, T.L., Higgins-D'Alessandro, A. (1997). The Role of Moral Reasoning and Domain Judgement in Adolescent Risk Engagement.
21. Krawczyk, R. (1997). Teaching Ethics: Effect on Moral Development. *Nursing Ethics*, 4(1), 57-65

22. Lerkiatbundit, S., Utaipan, P., Laohawiriyanon, C., Teo, A. (2006). Impact of the Konstanz method of dilemma discussion on moral judgment in allied health students: A randomized controlled study. *Journal of Allied Health*, 35(2), 101-108.
23. LaGrone, R.M., Welton, R.E., Davis, J.R. (1996) Are the effects of accounting ethics interventions transitory or persistent? *Journal of Accounting Education*, 14(3), 259-276
24. Lies, J. (2006). Effects of service-learning on college student moral reasoning. Dissertation
25. Lind (2007). The Relative and Absolute Effect Size of the Konstanz Method of Dilemma Discussion (KMDD). unveröffentlichtes Manuskript
26. Lopez, J.G., Garcia, R. (1998). The improvement of moral development through an increase in reflection. A training programme. *Journal of moral education*
27. Medrano, C. eta Caba, M.A. De la (1994). A model of Intervention for improving Moral reasoning: An experiment in the Basque Country. *Journal of Moral Education*, 23(4), 427-437.
28. Mills, J. (2005). The effect of a deliberate psychological education model on the ego development, moral development, and sexual assertiveness of college women. Dissertation
29. Morgan, B., Morgan, F., Foster, V., Kolbert, J. (2000). Promoting the moral and conceptual development of law enforcement trainees: a deliberate psychological educational approach. *Journal of Moral Education*
30. Mouratitdou, K. (2007). Physical education and moral development: An intervention programme to promote moral reasoning through physical education in high school students. *European Physical Education Review*, 13(1). 41-56
31. Munson, B. (1989). The relationship among religious orientation, level of moral reasoning, and selective educational interventions of students attending a Christian high school. Dissertation
32. Nas, C. (2005) Effects of the EQUIP programme on the moral judgement, cognitive distortion, and social skills of juvenile delinquents. *Psychology, Crime & Law*, 11(4), 421-434

33. Reiman, A.J., DeAngelis Peace, S. (2002). Promoting teachers' moral reasoning and collaborative inquiry performance: a developmental role-taking and guided inquiry study, *Journal of Moral Education*, 31(1), 51-66
34. Ries, S.I. An intervention curriculum for moral development. *Journal of Moral Education*
35. Santilli, N.R., Hudson, L.M. (1992). Enhancing moral growth: is communication the key? *Adolescence*, 27(105), 145-160
36. Self, D.J., Baldwin, D.C., Olivarez, M. (1993). Teaching medical ethics to first year students by using film discussion to develop their moral reasoning. *Academic medicine, Journal of the Association of American Medical Colleges*, 68, 383-385
37. Self, D.J., Baldwin, D.C., Olivarez, M. (1998). The amount of small-group case study discussion needed to improve moral reasoning skills of medical students. *Academic medicine, Journal of the Association of American Medical Colleges*, 73(5), 521-523
38. Senne, T.A., Rikard, G.L. (2003). The teaching portfolio as a developmental intervention: promoting developmental stage growth in physical education teacher candidates.
39. Stoll, S.K., Beller, J.K. (1993). The Effect of a longitudinal teaching methodology and classroom environment on both cognitive and behavioral moral development.
40. Thompson-Grim, L. (1999). Ropes programming as a means toward improving moral reasoning in at-risk youth. (alternative high schools, outdoor education). Dissertation
41. Turner, S.L., Bechtel, G.A. (1998) The effectiveness of guided design on ethical decision making and moral reasoning among community nursing students. *Nursing Connections*, 11(1), 69-74
42. Walker, L.J., Taylor, J.H. (1997) Moral climate and the development of moral reasoning: the effects of dyadic discussions between young offenders. *Journal of Moral Education*
43. Wandzilak, Thomas (1988). Values Development through Physical Activity: Promoting Sportsmanlike Behaviors, Perceptions, and Moral Reasoning. *Journal of Teaching in Physical Education*, 8(1), 13-22

44. Zierer, K. (2006) Eine Erprobung der Konstanzer Methode der Dilemmadiskussion in einer 3. Grundschulklasse. Regensburg. Unveröffentlichte Dissertation.

# Anhang C

## Listen

### C.1 Liste 1 Versuchspersonengruppe

- G(1) Kinder bis 10 Jahre, Grundschule bzw. amerik. System: Elementary School
- G(2) Kinder zwischen 10 und 14 Jahren, Sekundarstufe 1, Hauptschüler, bzw. amerik. System: Junior High School
- G(3) Jugendliche zwischen 14 und 17 Jahren, Sekundarstufe 2, Mittelstufe der Realschule und des Gymnasiums, Fachschule bzw. amerik. System: Senior High School
- G(4) Jugendliche zwischen 17 und 19 Jahren, Abiturienten, Meisterschüler bzw. amerika. System: High School Graduates, College Undergraduates
- G(5) Studenten an Universitäten und Fachhochschulen bzw. amerik. System: College Graduates
- G(6) Junge Erwachsene zwischen 19 und 25 Jahren ohne Hochschulausbildung
- G(7) Ältere Erwachsene *mit* Hochschulausbildung (auch College)
- G(8) Ältere Erwachsene *ohne* Hochschulausbildung
- G(9) Delinquente

### C.2 Liste 2 Schulklasse/Studiensemester

im amerikanischen System: High School (Gesamtschule):

- 7. bis 9. Klasse Junior High School
- 10. bis 12. Klasse Senior High School
- 10. Klasse Sophomores
- 11. Klasse Juniors
- 12. Klasse Seniors

College:

1. Jahr Freshman
2. Jahr Sophomore
3. Jahr Junior
4. Jahr Senior

### **C.3 Liste 3 Art der Intervention**

- A(1) Standard-Intervention; Diskussion von moralischen Werten, hypothetischen moralischen Dilemmas (Intervention wie in der Studie von Blatt); Stimulation durch Versuchsleiter/Lehrer
- A(2) 'Just Community Approach' (nach Power, Higgins, Scharf, Hickey); Mitverwaltung, Demokratisierung der institutionellen Entscheidung; Diskussion 'realer' Dilemmas der Beteiligten
- A(3) Ethik-Unterricht (im Rahmen bestehender Lehrpläne/Curricula)
- A(4) Peer-Interaktion, kleine Gruppen diskutieren moralische Probleme
- A(5) Indirekte Intervention über das Training von Ausbildern, Lehrern und Eltern
- A(6) andere Art der Intervention (jeweils in den Fußnoten vermerkt)

### **C.4 Liste 4 Inhalt der Intervention**

- I(1) 'Hypothetische', vom Versuchsleiter eingebrachte moralische Dilemmas (ggf. in der Fußnote notiert)
- I(2) 'Reale', von den beteiligten erlebte moralische Dilemmas
- I(3) 'Formale' Kommunikations- und Diskussionsfähigkeiten und -regeln
- I(4) Einführung in Kohlbergs Theorie (Meta-Kogniton)
- I(5) Diskussion abstrakter, moralischer Werte wie z.B. 'Gleichheit', 'Freiheit' usw. (ggf. in der Fußnote notiert)
- I(6) Diskussion anderer Werte (ggf. in der Fußnote notiert)
- I(7) andere Inhalte (ggf. in der Fußnote notiert)

### **C.5 Liste 5 Testinstrumente**

- T(1) Moral Judgement Interview (MJI) von Laurence Kohlberg
- T(2) MJI (alt) wie T(1)
- T(3) Defining Issues Test (DIT) von James Rest; Standard Version mit sechs Dilemmas



- T(4) DIT kurze Fassung mit drei Dilemmas  
 T(5) Moralisches Urteil Test (MUT) Standardversion mit beiden Dilemmas 'Arbeiter/Diebstahl' und 'Arzt/Euthanasie'  
 T(6) MUT andere Dilemmas (ggf. in der Fußnote notiert welche)  
 T(7) anderer Test (ggf. in der Fußnote notiert)

## C.6 Liste 7 Messkala

Moral Judgement Interview:

- MMS Moral Judgement Score (100 bis 600)
- STS Stage Score (1 bis 6)

Defining Issues Test:

- P-Score Anteil postkonventioneller Präferenzen (0 bis 100)
- D-Score Neue Version des P-Score

Moral Judgement Interview:

- C-Score moralische Kompetenz (0 bis 100)

## C.7 Liste 8 Berechnung des Korrelationskoeffizienten (punkt-biserial)

$$r_{pbs} = \sqrt{\frac{chi^2}{n}}$$

$$r_{pbs} = \sqrt{\frac{t^2-1}{t^2-1+n}}$$

$$r_{pbs} = \sqrt{\frac{F}{F+n-2}}$$

$$r_{pbs} = \sqrt{\frac{d^2}{(d^2+a)}}$$

$$s_{gepoolt} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

$$g_{Hedges} = \frac{\bar{X}_1 - \bar{X}_2}{s_{gepoolt}}$$

$$r = \sqrt{\frac{g^2 n_1 n_2}{g^2 n_1 n_2 + (n_1 + n_2)(n_1 + n_2 - 2)}}$$

## **C.8 Liste 10 Art der Publikation**

- J(1) Journal
- J(2) Dissertation
- J(3) nicht-publiziertes Manuskript

## **C.9 Liste 12 Designtyp**

- Q(1) klassisches experimentelles Design (rand. Zuordnung)
- Q(2) quasi-experimentelles Design mit Kontrollgruppe
- Q(3) nur 'Posttest' oder keine Kontrollgruppe

# Literaturverzeichnis

- [Blatt und Kohlberg, 1975] Blatt, M. M. und Kohlberg, L. (1975). The effects of classroom moral discussion upon children's level of moral judgment. *Journal of Moral Education*, 4 (2):129–161.
- [Colby et al., 1983] Colby, A., Kohlberg, L., Gibbs, J., und Lieberman, M. (1983). A longitudinal study of moral judgment. *Monographs of the Society for Research in Child Development*, 48.
- [Glass, 1976] Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5:3–8.
- [Glass et al., 1978] Glass, G. V., McGaw, B., und Smith, M. L. . (1978). *Meta-Analysis in Social Research*. London: Sage Publications.
- [Hedges und Olkin, 1985] Hedges, L. V. und Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, Orlando, Florida.
- [Hunter und Schmidt, 1990] Hunter, John, E. und Schmidt, Frank, L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Verlag Hans Huber, Länggasse-Strasse 76, CH - 300 Bern 9.
- [Kohlberg, 1958] Kohlberg, L. (1958). *The development of modes of moral thinking and choice in the years 10 to 16*. PhD thesis, University of Chicago.
- [Kohlberg, 1964] Kohlberg, L. (1964). Development of moral character and moral ideology. In Hoffman, M. und Hoffman, L., editors, *Review of child development research Volume 1*, pages 381–431. Russel Sage Foundation, New York.
- [Kohlberg, 1977] Kohlberg, L. (1977). *Assessing moral stages: A manual Part 1, Introduction*. Cambridge MA, Center for Moral Education, Harvard University.
- [Kohlberg, 1984] Kohlberg, L. (1984). *The psychology of moral development. Volume 2, Essays on Moral Development*. Harper & Row, San Francisco.

- [Lind, 2000] Lind, G. (2000). *Ist Moral lehrbar?: Ergebnisse der modernen moralpsychologischen Forschung*. Logos Verlag Berlin, Michaelkirchstr. 13, 10179 Berlin.
- [Lind, 2007] Lind, G. (2007). The relative and absolute effect size of the konstanz method of dilemma discussion (kmdd). *unveröffentlichtes Manuskript*.
- [Piaget, 1976] Piaget, J. (1976). *Psychologie der Intelligenz*. Kindler, München.
- [Popper, 1965] Popper, Karl, R. (1965). *Conjectures and refutations*. Routledge & Kegan Paul Limited, Broadway House, 68-74 Carter Lane.
- [Rustenbach, 2003] Rustenbach, Stephan, R. (2003). *Metaanalyse: Eine anwendungsorientierte Einführung*. SAGE Publications, 2455 Teller Road Newbury Park, California 91320.
- [Schlaefli et al., 1985] Schlaefli, A., Rest, J. R., und Thoma, S. J. (1985). Does moral education improve moral judgement? a meta-analysis of intervention studies using the defining issues test. *Review of Educational Research*, 55.