The Optimal Age of Moral Education.

A Review of Intervention Studies and an Experimental Test

of the Dual-Aspect Theory of Moral Development and Education[1]

Georg Lind

University of Constance, Germany[2]

Abstract

With regard to the optimal age of moral education, textbooks give different, and sometimes contradictory information. While cognitive-developmental theory implies that moral dilemma discussions and similar methods for fostering moral judgment competence are best used in adolescents, some authors regard such methods to be effective only in adulthood. This paper argues that Lind's Dual-Aspect-Theory lets us explain this contradiction as an methodological artifact because the two findings relate to two different aspects of morality. The former findings are based on Kohlberg's Moral Judgment Interview, which seems to tap the cognitive aspect of moral behavior. The latter findings are based on Rest's Defining Issues Test, which seems to tap the affective aspect, i.e., subjects' attitudes toward principled reasoning, and are prone to the Hawthorne-effect. A meta-analysis shows that adolescence is indeed the optimal age for fostering the cognitive aspect, moral judgment competence. This is supported by an intervention experiment that uses Lind's Moral Judgment Test, a direct cognitive measure of moral judgment competence.

---

[1] This article is based on a paper presented at the SIG Moral Development and Education meeting, convention of the American Educational Research Association, April 8 - 12, 1996.

[2] Comments and requests for revised versions of this paper should be sent to: Dr. Georg Lind, Universität Konstanz, Sozialwiss. Fakultät, D-78434 Konstanz, Germany. E-mail: georg.lind@uni-konstanz.de. For further information see: http://www.uni-konstanz.de/ag-moral/

Introduction

Should we help school-aged children to cope with the moral dilemmas of life or shouldn't we? Obviously, this is a very important question, not only for academic reasons but also for practical ones. Yet there is no clear answer to this questions. Some say, morality is already fully developed by the time children enter school either because it would be mostly determined by their genes or because morality would be formed in the very young age. Others maintain, that only adults, maybe not even all adults, are ready to learn to handle conflicts between diverging demands of their moral conscience. This latter opinion, often espoused by the character education movement, implies that during childhood and adolescence people acquire, and acquiesce to, the moral rules and values required for being a good member of society.

Depending on what we believe to be the right answer, we will conclude either a) that schools will have no impact on the moral development of children and, therefore, should skip this subject, or b) that schools should only transmit moral values and rules but should not confront students with conflicts between those values and rules, or c) that teachers (and parents) should engage in the education of children's ability to resolve moral problems themselves.

Half a century ago, the question of the optimal age of moral education was purely a matter of believe. Over the past decades so much research has been accumulated that it should be possible to turn this controversy between opposing opinions into a scientific question, and to tell educators how and when it can be nurtured. Moral-psychological research has already helped us to clarify some issues related to moral education. For example, it has clearly shown that society mostly succeeds with transmitting its moral values and rules (like "Be just," "Don't steal," "Help others") to its offsprings. Almost all children know these values and rules, distinguish them clearly from mere social conventions and personal values, and develop positive attitudes toward them.[3] But this research also shows that the transmission of moral values by no means guarantees that children will behave accordingly. It seems that moral attitudes or ideals do not automatically translate into moral behavior[4] but that, as Lind (1978; 1985; 1993) puts forth in his dual-aspect-theory, moral competencies are required for doing this. In fact, measures of moral judgment competence correlate considerably higher with moral behavior than moral attitudes do (Blasi, 1980).

---

[3] See Hartshorne & May (1928), Nucci (1995), Nunner-Winkler & Sodian (1988), Turiel (1983).

[4] Hartshorne & May (1928), Kohlberg (1984).

With regard to the optimal age of moral education, the findings are much less clear. They are even contradictory. While early intervention experiments support the cognitive-developmentalists hypothesis that adolescence is the optimal age for fostering moral competencies (Blatt & Kohlberg, 1975; Enright et al., 1981; Higgins, 1980; Leming, 1981; 1992; Lockwood, 1978), more recent studies seem to refute this hypothesis (Schläfli et al., 1985).

Which findings are valid? Is, as Piaget and Kohlberg hypothesized, late childhood and adolescence, the optimal period for moral education, or should one expect no effects until adulthood?

Some textbook authors lean toward the newer findings (Edelstein, 1987; Oser & Althof, 1992; Rest & Narvaez, 1994; Sprinthall & Collins, 1995). They seem to imply that it goes without saying that when two lines of research differ greatly in their findings, the more recent research is more trustworthy. This decision rule would be valid if we can assume a) that morality is a unitary construct and that both lines of research deal with the same construct and b) that the differences are merely of methodological nature, that is, that the more recent research is technically superior to the older one.

I will argue, however, that these two assumptions do not apply in this case and that the older findings are more valid in regard to the question of the optimal age for moral education. As Piaget and Kohlberg maintained, morality must be described in terms of not just one but of two fundamentally different aspects, namely of an affective aspect (i.e., the moral attitudes, values, ideals etc. a person holds) and a cognitive aspects (i.e., the person's moral competencies). As Lind (1978; 1995) shows, these two aspects can not only be theoretically distinguished but they can also be independently operationalized and measured. On the basis of this dual-aspect-theory, Lind (1985; 1993) hypothesizes that findings will differ depending on the research focus. Research focusing on moral competencies should show that adolescence is the optimal age for fostering the ability to apply moral rules and values to concrete decision-making processes. Research focusing on moral attitudes should reveal that adults sense better the moral expectations of their social environment than youngster do, and adapt their test scores accordingly (Emler et al., 1983; Hogan & Emler, 1995).

The Dual-Aspect-Theory of Moral Behavior

The dual-aspect-theory of moral behavior and thinking comprises four postulates, which suggest some important distinctions for discussing the optimal age of moral education. The first two postulates are based on moral philosophy, the two others on empirical research:[5]

First, morality must be described in terms of internal states rather than external states, that is, we call a person's behavior moral only if a person acts upon his or her own moral principles rather than merely complies with external rules or norms (Kohlberg, 1984; Pittel & Mendelsohn, 1966). The latter is called rule-conforming behavior. Furthermore, the description of a person's morality must involve both always both affective and cognitive aspects. In other words, the moral goodness of our behavior depends both on the moral ideals or principles to which we aspire and on our ability to judge and act accordingly.

The affective aspect of moral judgment behavior is defined as the moral ideals, values and attitudes a person holds. More specifically, it is defined as the moral perspective that marks the horizon of his or her subjective moral universe. Kohlberg's (1984) six stages of moral perspective-taking provide an appropriate scheme for typifying the affective aspect of subjects' moral thought and behavior.[6]

Moral judgment competence, the cognitive aspect of moral judgment behavior, develops through four the phases that Piaget (1965) described:[7]

- Amorality or anomia: the child lacks moral sensibility and does not care about right and wrong; he or she pays no attention at all to moral perspectives;
- Imitation: the child merely mimics moral language but sees no implications for his or her behavior;
- Heteronomy: the child subordinates his or her behavior to moral perspectives but without a clear understanding of their meaning; the child usually acts in a rigid, all-or-nothing way

---

[5] The following section is based on Lind (1985; 1993).

[6] Thus, as Lind (1985) argues, Kohlberg's six (or five) stages of moral judgment do not extend and supersede Piaget's (1965) cognitive phases of moral development, but supplement them to form a two-dimensional model of affective-cognitive moral development. Lind's dual-aspect-model avoids the conceptual and methodological difficulties of Kohlberg's attempt to combine both aspects into one single dimension (see also Lind, 1993).

[7] For more details on this adaption of Piaget's (1965) phases of moral-cognitive development, see Lind (1993). The last two phases, heteronomy and autonomy, correspond also to Kohlberg's substage A and substage B, a distinction he makes for all major stages of moral orientation (Kohlberg, 1984).

because he or she makes decisions under only one moral perspective but is not able to integrate different moral perspectives and make judgments in a differentiated way; he or she relies mostly on external authorities;

- Autonomy: the child is able to integrate many moral perspectives and make an integrated and differentiated moral judgment that does justice both to the abstract principles and the concrete situation involved. Moral autonomy is not absolute but always in regard to a particular moral perspective (e.g., in regard to Kohlberg's Stage 3, "good boy/nice girl" perspective). That is, a child is call morally autonomous in regard to a particular level of moral perspective-taking; he or she considers all moral perspectives available to him or her on that stage for making a decision, even though they may conflict with his or her most salient one.

Second, while we can clearly distinguish both aspects of morality and, therefore, can describe and measure both the cognitive and the affective aspect in a logically independent way (the degree of moral judgment competence needs not be linked with particular moral principles), both aspects are not separable, that is, they cannot be conceived of as being two different substances, components or sets of behavior. Hence, we must measure a person's moral attitudes and competencies with one and the same measuring instrument, but must not confuse both aspects.

Third, the dual-aspect-theory implies that development of both aspects is time-lagged parallel. "Time-lagged" means that is the acquisition of certain moral perspective or orientations will precede the development of the cognitive competencies needed for autonomous moral action. In other words, moral perspectives serve as goals or ideals for the development of the moral competencies needed to approximate them in everyday-action. "Parallel" means that, while both aspects can be logically distinguished and independently measured, they will correlate empirically. That is, the greater a child's moral judgment competence the more he or she will prefer higher over lower moral perspectives, whereas "higher" and "lower" refers here to Kohlberg's Stage model. Empirical studies using Lind's Moral Judgment Test, which lets us measure both aspects independently, unanimously support this hypothesis.[8]

---

[8] This includes studies of various age-education groups and of various cultural backgrounds. The typical correlations between the C index of moral judgment competence on the one side and the indices for subjects' attitudes toward each of the six Kohlbergian stages are: Stage 1 -.55, Stage 2 -.55, Stage 3 -.40, Stage 4 ±0.00, Stage 5 +.40, Stage 6 +.45. For summaries of research on the parallelism hypothesis, see Lind (1985, pp. 45-46; 1993).

Fourth, the dual-aspect-theory suggests that both aspects of moral behavior and development require distinct methods for their change or enhancement. While methods of direct teaching or transmittance of moral values seems to be well suited for changing children's moral attitudes, they do not seem appropriate for enhancing the development of moral competencies. Teaching methods, which require children to solve concrete behavioral problems, seem more appropriate. Such teaching methods have been suggested by Kohlberg and his colleagues, e.g., the Moral Dilemma Discussion (MDD) method (Blatt & Kohlberg, 1975).[9] The MDD method works similar to medical immunization. Whereas medical treatment stimulates the body's immune system by introducing a mildly toxic dose of a virus, the MDD method confronts the child's cognitive system with moral dilemmas that are semi-real, that is, dilemmas which are real enough to arouse a child's moral emotions but are not so real as to have real consequences for the child.

This dual-aspect-theory builds largely upon, but also differs significantly from, the theories of Piaget (1965), Kohlberg (1984) and Rest (1986).[10] Like Piaget and Kohlberg, this theory postulates that the cognitive and the affective aspect of moral behavior are parallel. In contrast to both these theorists, we believe that these two aspects cannot only be theoretically distinguished but can also be measured independently from each other. Like Piaget we also believe that the affective and the cognitive side of human behavior cannot, and should not, be viewed as separate components but as distinguishable aspects or properties of behavior. Lind's (1978; 1995) Moral Judgment Test (MJT) is proof that it is possible, to measure the two aspects of moral judgment behavior independently, without separating them into different components or test items. Through this achievement, now we can evaluate the effects of moral education programs on the competence aspect of moral judgment without confusing them with the effects on moral affects or attitudes.

Given these clarifications, we assume that, if taught and measured appropriately, moral education should show its greatest effects during adolescence when children develop their capacity of abstract reasoning (Kuhn et al., 1977). Blatt and Kohlberg (1975) suggested that the optimal age for applying the MDD method is the age between 10 and 16 years of age. Drawing upon Piaget's theory of cognitive development (Piaget & Inhelder, 1969), they argue that "pre-

---

[9] For an excellent description such teaching methods see DeVries and Zan (1994). A summary of the basic principles of the Moral Dilemma Discussion method is given by Lind (1996b), assessable on the Internet. An outgrowth of the MDD method is the Just Community (JC) method described by Power et al. (1989).

[10] The similarities and dissimilarities of the dual-aspect-theory on the one side and its predecessors on the other side are explained in Lind (1985a; 1985b; 1993; 1995).

adolescence is a favored or 'critical' period for intervention into moral development" (p. 132). Therefore, if adolescents do not receive educational aid, some "might be expected to stabilize at the preconventional level, as do most adolescent delinquents" (p. 132). "With the adolescent only beginning to acquire formal operational capacities," Enright et al. (1983) noted, "during this period, a program designed to induce new cognitive acquisitions could be expected maximally effective" (p. 135).

Thus, there should be no, or only spurious, effect of moral teaching if teaching is aimed at changing moral values or if the measurement taps largely the affective rather than the cognitive aspect of moral judgment behavior.

Recently, these two conditions of effective moral education have been questioned by Rest and his associates. They conclude a) that non-developmental methods of direct teaching have greater effects than cognitive-developmental methods (Penn, 1990; Rest & Narvaez, 1994; Schläfli et al., 1985) and b) that adulthood, not adolescence, is the optimal age of education (Schläfli et al., 1985). A growing number of textbook authors accept this conclusion as state of the art. So, for example, Sprinthall and Collins (1995) argue that "programs with adults (24 years of age and older), . . . are much more effective than those with adolescents, probably because young adults have a larger capacity to examine moral dilemmas from a personal standpoint" (p. 231; see also , Edelstein 1987, p. 188; Oser & Althof, 1992, p. 155).

Astonishingly, this assertion is accepted without discussing the fact that it fundamentally contradicts cognitive-developmental theory and most earlier research findings (Enright et al., 1983; Higgins, 1980; Leming, 1981; Power et al., 1989). In this paper I shall show that this contradiction can be explained by the dual-aspect-theory and that the resolution of this contradiction on these grounds leads to a affirmation of the views of Piaget and Kohlberg and a rejection of Rest's assertion. Our argument goes like this: The dual-aspect-theory implies that moral competencies and moral attitudes cannot only be distinguished (though not separated) from one another, they also must be measured independently when questions of educational effects are at stake.

The high correlation so often been found between both aspects vanishes in educational settings because the distinctive features of competencies and attitudes become highly salient. Subjects attitude scores can change in any direction within a very short time, either because their attitudes really change or because the subjects have reasons to simulate such change. In contrast, subjects can acquires competencies, skills, abilities and alike only slowly (as it takes much exercise to develop them) and they will lose them also slowly rather than abruptly. Experimen-

tally, competencies and attitudes can be clearly distinguished when we instruct people to simulate them "upward". While people should be able to exhibit any attitude that we instruct them to simulate, they should not be able to simulate only a lower, not a higher, level of moral competency than they actually possess.

Thus, while intervention aiming at a change of students' moral values might show impressive changes due to students desire to please their teacher, such interventions should show no increase in students' moral competencies. Therefore, from the dual-aspect point of view, we hypothesize that the vast differences in findings concerning the optimal age (and best method) of moral education can be mostly accounted for by the confounding of moral attitudes and moral competencies in the measurement of moral development. If the aim of moral education is the enhancement of moral competencies and if these are taught with appropriate methods like the MDD method, then the optimal age for it should be adolescence. In contrast, if the teacher wants, for some reason, for students to get high scores on moral attitudes scales and he or she makes sure that the students know what they are expected to say, then higher-age-groups might produce better results. Adults might be more "able" to acquiesce than adolescents, who might have difficulties to discern what attitudes their teachers expect from them.

To substantiate the dual-aspect-theory regarding the optimal age for enhancing students' moral competencies, I will juxtapose the major empirical source for the adulthood-hypothesis, namely the meta-analysis of many intervention studies using the DIT by Schläfli et al. (1985) with Lind's (1993) meta-analysis of 141 intervention studies using Kohlberg's Moral Judgment Interview (MJI).

Intervention studies based on Kohlberg's MJI appear to give the most accurate and defendable picture of educational effects as the MJI is widely regarded as the best "cognitive" measure of moral development. Yet MJT-based studies leave also some questions open because the MJI is not a pure competence measure but confounds to some extend moral attitudes and competencies (a subject must simultaneously have a high judgment competence and also prefer Stage-6-Morality to get a high score on the MJI) and it has not yet been thoroughly tested for fakability (Lind, 1995). These open questions can be answered by MJT-based studies because the MJT has been designed to measure moral attitudes and moral competence simultaneously without confouding them, and which has withstood rigorous faking experiments (Lind, 1978; 1993; Wasel, 1993).

A Note on Reporting Findings and Statistical Significance.

In agreement with expert's judgment, we will report here figures of practical significance or effect size but not of statistical significance. For many decades statisticians are castigating the use of significance tests for being mostly out of place and as impeding rather than aiding the progress of educational science.[11]

I will report here two measures of effect size, a) the correlational coefficient r, and, where possible, b) the raw-score gains as a measure of absolute effect sizes. The r-coefficient tells us how much gain (or loss, if negative) the intervention has produced relative to the variation of scores in the sample. This index does not depend on the sample size as significance tests do. The r-index also varies within known boundaries, from -1.00 to +1.00, in contrast to the d-index, which has no upper boundary and is less known outside meta-analytical studies. Both indices can be easily converted into each other.

Absolute raw-score gains give an even better idea of the effect size index because they depend neither on sample size nor on (accidental) variation in the samples studied. Absolute raw score gains can be made comparable for studies using different types of tests by considering the maximum length of the scales.[12]

MJI-based findings: Pre-adolescence is the optimal age

Blatt and Kohlberg (1975), who used the Moral Judgment Interview for evaluating effects, give us the first experimental data on the validity of the cognitive-developmental hypothesis regarding the optimal age for using the Moral Dilemma Discussion method. In their first study of 11 to 12-year-old students, they found that the MDD method was highly effective. These pre-adolescents gained on average 66 MMS points, that is, more than a half stage, on a scale from 100 (equivalent to Kohlberg's Stage 1) to 600 (Stage 6), whereas the control group showed no gains.

---

[11] Fortunately, the most important professional organizations for educational research, APA and AERA, do not request significance testing any longer (Thompson, 1996). Here is a sample of just a few expert quotes. Cohen (1994) argues "that [statistical significance testing] has not only failed to support the advance of psychology as a science but also has seriously impeded it" (p. 997). Shaver (1993) says: "The dominance of statistical significance testing is dysfunctional, because such tests do not provide the information that many researchers assume they do. Statistical significance testing also diverts attention and energy from more appropriate strategies, such as replication and consideration of the practical or theoretical significance of results" (p. 294).

[12] For example, a gain of 50 MMS in the Moral Judgment Interview is equivalent to 10 percent of the total scale range, from 100 to 600. Such gain is comparable to 10 points on the P-scale of the Defining Issues Test, which ranges from zero to 100, and to 10 points of the C-scale of the Moral Judgment Test, which also ranges from zero to 100.

They also found that this effect was still discernible after one year when they followed up their subjects. After one year, the children's MMS scores dropped a little, but only 11 points. This contrasts with the changes in a control group, which showed hardly any gains (only four points), but a drop of scores in the follow-up assessment (26 points).

In their second study, Blatt and Kohlberg (1975) found that older children, ages 15 to 16, also gained markedly (36 MMS points).[13] In their follow-up assessment one year after the intervention experiment, they found a <u>further</u> increase of approximately 30 points.[14] So the gross increase for this older group was also more than 60 points.[15]

Since then, the effect of the MDD method on moral development has been tested in several hundreds intervention studies, reviewed by Enright et al. (1983), Higgins (1980), Leming (1981), Lind (1993), Lockwood (1978), Schläfli et al. (1985) and others. Many of these studies involved adolescents from 10 to 16 years of age, and of these most found positive effects for the MDD method. So, in contrast to recent statements by Emler (1996) and Damon and Gregory (1997), we do have a wide and well-researched basis for evaluating cognitive-developmental programs of moral education.

Many of these reviews and meta-analyses, however, do not systematically differentiate between the subjects' level of moral development or at least their age (which is only a very rough indicator of the developmental level). Moreover, they rely exclusively on significance testing, so we can hardly say how big the effects of MJI-based intervention studies have been, how the effects in the various developmental levels compare to each other, and how these studies compare as a whole to DIT-based studies.

---

[13] This gain has been calculated by the author on the basis of Blatt and Kohlberg's (1975, p. 148) data, which show these scores only separated by socio-economic status.

[14] This number is only an estimate as it has been taken from Blatt and Kohlberg's (1975) Table 6, on p. 146, in which they report the findings from the follow-up study only for the 11 and 15-year-old children combined.

[15] The differences between the first and the second study call for some speculations about the reasons for this differential process. The sample of the first study was drawn from an affluent, upper-middle-class population, while the second sample represents lower class and lower middle-class children. Possibly, the latter showed some "sleeper-effect," that is, the cognitive-developmental method appears to have stimulated their development even <u>after</u> the intervention was over. This interpretation is supported by the fact that, within the intervention, the so-called "disadvantaged" gained somewhat less (29 MMS points) than the "common man" (38.5 MMS points). So, if tested immediately after the intervention, the gains were higher the higher the social status of the children was. But when followed-up later, the effect-sizes were about equal.

Therefore, some time ago, Leonore Link and I collected all information that we could get about cognitive-developmental moral intervention studies conducted until the mid Eighties (see Lind, 1993). We analyzed 48 publications and research reports that documented interventions in 141 experimental samples. For this presentation, I re-analyzed those interventions that used the MJI for measuring the effects.

The MJI-based studies clearly support the assumption that the optimal age for the MDD method is adolescence and not adulthood. Although, the MDD method has also a developmental effect in younger and in older age-groups, the adolescents aging from 11 to 16 gain most. **Figure 1** gives a summary of this analysis. It depicts graphically the effects of MDD in various age groups, expressed as the correlation coefficient r.[16]

One should also note that, on average, these studies found big effect sizes. The average effect was greater than r = .50, which means that it was almost twice as much as the pooled standard deviation of the MMS scores in these samples (pre- and posttest combined). In the control groups much smaller or no effects were found.

DIT-based Findings: Adulthood is the Optimal Age

In contrast to these findings, intervention studies evaluated with Rest's (1979) Defining Issues Test come to a quite different conclusion. To make the findings in these studies better comparable with the Kohlbergian studies, the d-indices, reported by Schläfli, Rest and Thomas (1985) are converted here into correlation coefficients (see **Figure 2**).

When comparing these findings with MJI-based studies (**Figure 1**), two findings are most noteworthy. First, in DIT-based studies, adults gain considerably (effect size r = .28) from moral cognitive interventions. Second, in DIT-based studies, the MDD method seems to produce hardly any effects in adolescents and young adults (.11 < r < .14). The effect sizes in these age groups are greater than the pre-posttest changes in control-groups that did not receive such a treatment ® = .05). However, note that all effects are generally much smaller than in MJI-based studies ® = .40).

---

[16] Note that a r = 0.00 means that there is no effect, a high r means that there is an extremely large effect. A r = .45 is approximately equivalent to d = 1.0, meaning that the size of the effect was as great as one standard deviation of the pooled pre- and posttest scores. Conventionally, a r > .10 is considered as a notable effect size, a r > .30 as a moderate effect size, and a r > .50 as a strong effect (Cohen, 1977). The correlation coefficients have not been converted into Fischer's z-scores prior to aggregation because, as Glass et al. (1978, p. 147) have shown, the aggregation of unconverted rs leads to almost identical results.

How can this difference be explained? Are these real differences, as some authors seem to belief, or are they methodological artifacts?

Rest and his colleagues suggest three possible explanations for the age differences in effect sizes. First, they hypothesize that because adult subjects usually volunteered rather than were assigned to the moral education program, they may show a greater willingness to participate in it. However, this explanation does not account for the fact that this phenomenon appeared only in DIT-based intervention studies but not in MJI-based studies.

Second, they argued that interventions with adults might be more powerful because adults can draw on more experiences. Again, this argument cannot account for the fact that in MJI-based studies adults gained less than adolescents. Moreover, this argument works both ways: we may also expect that adults gain less from MDD method because, as Kohlberg would argue, this method foremost stimulates competencies that the highly developed subjects already have. I will return to this point later.

Third, they suggest that the DIT might be relatively insensitive effects of the MDD and other methods of moral education in adolescents (Schläfli et al., 1985, p. 344). Indeed, the P index used with the DIT for measuring moral development, is designed to measure only an increase of preferences for stage five and six reasoning, but not shifts within lower stage reasoning. “The P score indicates,” as Rest and Narvaez (1994) point out, “the extent to which the subject considers the 'Principled’ items (items from stages five and six) as most important [. . .] One way to interpret the DIT's P score is the extent to which the subject makes moral judgements like a moral philosopher” (p. 234).

Besides the explanations which the authors considered, there is still another methodological explanation for the overall low effect-sizes. The DIT was developed on the basis of classical test theory, which suggests item selection procedures to maximize both a) inter-individual variation and b) individual stability of scores, to get a high “reliability” coefficient for the test. This strategy of test construction, as Wohlwill (1977) pointed out, partly undermines the purpose of developmental and educational research, namely to measure change as an consequence of educational efforts. Making a test highly stable or “reliable” is good for the purpose of sorting and selecting people but not for the aims usually pursued in education.

Thus, the lack of effects shown by adolescent samples may be indeed a methodological artifact. The DIT shows the MDD to be effective in adults only a) because it may reflect mostly, if not exclusively changes in moral attitudes and, therefore, be susceptible to the so-called “Hawthorne-effect” (Emler et al., 1983), and b) because it is insensitive to changes in the range

of moral-cognitive development for which the MDD method is designed. This could mean, by the intervention program, the adult subjects might not become more competent to make moral judgments but might try to please the moral expectations of their teachers who want them to make moral judgments like a moral philosopher. With adolescents, either the DIT is not sensitive enough to show such effects, or adolescents are less inclined to acquiesce.

Several lines of evidence support this hypothesis. Emler et al. (1983) showed in an experimental study that subjects can simulate or fake DIT-scores upward.[17] Schläfli et al. (1985) report that in DIT-based studies, direct teaching methods, like exposing the subjects to Kohlberg's stage theory, produce much higher effect size (d = .56; r= .27) than the MDD method (d = .25, r= .12; p. 344, Table VI). They also report, that in adults this effect of direct moral value teaching is even more pronounced: "The effect size of the non-exposed adult sample is [d= .23; r = .11], but the effect size of the exposed adult sample is .71 [r = .33]-- about three times greater" (p. 344). Finally, Penn (1990) shows that the "direct teaching of fundamentals," that is, the indoctrination of moral values, has a stronger effect on P-score gains (d = .65, r = .31) than the moral dilemma discussion method.

All these studies have no bearing on Kohlberg's claim that the MDD method can effectively simulate moral judgment competence, because these effects have not been measured with tests of moral competence. In studies using the Kohlberg interview method, direct teaching methods hardly showed any effects and they were less persistent than the effects of the MDD method (Lind, 1993). So the empirical ground is still scarce for making a judgment on Schläfli et al.'s (1985) hypothesis that, to profit from the MDD method, a higher level of moral judgment competence is necessary than adolescents usually possess.

---

[17] Their finding has been replicated in an independent experiment by Markoulis (1989). An even the faking experiment by Barnett et al. (1995), which was designed to show the opposite, clearly supports the hypothesis that the DIT is fakable. The P-score in the experimental group increases markedly as a function of simple instruction, and only because of an unusual scoring this fact is disguised in the tables (for reanalysis, see Lind, 1996b). Though other studies showed that some kind of instructions do not show this effect (for summary, see Rest, 1986), the conclusion remains valid that the DIT can be simulated upward and, therefore, should be considered a measure of moral attitudes.

Testing the Limits of the Dilemma-Discussion Method

To get a more valid idea of the normal range and the limits of moral education programs using the dilemma discussion method, we conducted a series of studies using Lind's (1985; 1995) test of moral judgment competence, the Moral Judgment Test (MJT). In contrast to the DIT, the MJT's C index allows us to measure subjects' moral judgment competence independently from their moral attitudes. Faking experiments showed that the C index of the MJT cannot be simulated upward. Neither the instructions that proved to work with the DIT, nor other instructions enabled the subjects to fake the C score upward (Bühn, 1995; Lind, 1993; Wasel, 1993).

To test the lower limits, Rudi Hubert and I have conducted a series of dilemma discussions in a prison for juvenile delinquents, sentenced for serious misdemeanor like robbery and murder (Lind, 1994). We confronted these (voluntary) participants in this intervention with dilemmas believed to be very realistic for them. The first dilemma dealt with a delinquent who was promised a shortening of the sentence if he would reveal the name of a second person involved in the case. The participants were very reluctant to discuss this 'realistic' dilemma because they believed that their answers would be taken as real and would probably cause them great troubles in the prison. Though the other dilemmas were less realistic, we never succeeded in getting them involved in a discussion of these dilemmas the way we did with the same-age groups whom we pretested in schools. We found that the initially very low competence scores of these convicts did not increase through this intervention. The effect size was close to zero.

To see whether such intervention programs can foster the development of moral judgment competence in normal youth, we evaluated a one-week program conducted by Fritz Oser and his colleagues with Swiss bank apprentices (Lind, 1993; for a description of this intervention project, see Oser & Schläfli, 1985). Using the MJT as a measure of the competence aspect of the apprentices´ moral judgment behavior, we found an effect size of $r = .26$ ($d = .53$). Interestingly, according to Oser and Althof (1992, p. 106) the evaluation with Kohlberg Moral Judgment Interview showed no effect size.

Finally, to test the upper limits of the effectiveness of moral dilemma discussion programs Manfred Scholz and I conducted an intervention study with high school students who took an extra course in Psychology as their special field. We supposed that these students had a high moral judgment competence. If adulthood is the optimal age for the MDD method, we hypothesized a) that persons with initially high C scores should show substantial gains, and b) that the gains should be the higher the higher the initial C score were. If, however, average moral judgement competence si the best conditions for effective programs, the findings should be reverse:

then we should expect only little, if any, gain in this sample. Only the initially low scoring subjects, who resemble more the typical adolescent, should gain substantially from the MDD method (for more details, see Scholz, 1996). Our measurements showed that these subjects indeed had a very high moral judgment competence. Their initial average C score was 39.0, which is well above average adolescents' C score; it resembled more average C-score of German university students (Lind, 1993, pp. 251ff.).

We ran three sessions of dilemma discussion, each lasting 90 minutes because MJI-based intervention studies showed that as few as three to four dilemma discussion sessions can produce substantial effect sizes (Lind, 1993, p. 200). Experimenter effects and sequencing effects were controlled for through assigning at random the three trained discussion leaders and the dilemmas used for discussion to the participating classes. The subjects were pre- and posttest ed (but not followed-up).

The findings of this intervention show clearly that students with such high competence scores profit less from moral dilemma discussions than younger and less developed subjects. Our highly developed subjects showed no gains at all; the C scores of both experimental groups were lower at the end, though not very markedly (36.7 and 38.6) than at the beginning of the intervention (39.0 and 41.6). However, those subjects who initially scored lowest gained substantially (6.4 C points).[18]

Discussion

Kohlberg's theory implies that adolescence is the optimal age for using the Moral Dilemma Discussion method in moral education programs. This conclusion is well supported by studies that were evaluated using tests of moral judgment competence like Kohlberg's Moral Judgment Interview and Lind's Moral Judgment Test.

The contention that the MDD method has substantial effects only in adulthood seems to be based on a methodological artifact. It is supported only by intervention studies using Rest's Defining Issues Test. The DIT is hardly sensitive to changes on lower level of moral reasoning as are typically found in adolescents, the main addressees of the MDD method. The positive effects of the MDD method in adult samples can be explained through the fact that the DIT is prone to

[18] Gains of more than 5 C points are considered as "substantial." They roughly correspond to ¼ Stage in Kohlberg's 5-Stage-model. Five C points represent 5 percent of the entire scale of possible scores. Figures of statistical significance are inappropriate here.

the Hawthorne-effect because, like most attitude measures, it can be faked upward.[19] The adults subjects may readily guess the intentions of the experimenter and respond to the DIT in a way that please him or her. This explanation is reinforced by the fact that, in adult samples, direct methods of value teaching have a greater effect size than cognitive methods like the MDD method.

The differential findings of studies using measures of moral competence versus measures moral attitudes, and using cognitive-developmental versus direct moral teaching methods, reinforce the distinctions made in Lind's dual-aspect-theory (Lind, 1985; 1993; 1995). The effects of moral education differ greatly depending on which aspect we address in a moral education program and which aspects we actually measure. If these distinctions are not made, the findings may look confusing, if not contradictory.

In respect to moral education, this finding adds to a bulk of literature that shows that the Moral Dilemma Discussion method is a powerful method for fostering moral development in adolescents. In contrast, methods like value indoctrination and moral behavior drills have shown to be little effective in fostering moral development. If new improved variants of these methods are suggested, they should be submitted to the same kind of rigorous empirical testing as the MDD method has been.

One caveat is in turn. This paper dealt only with one method of moral education, the Moral Dilemma Discussion method, and also only with one type of outcome, namely increase in moral judgment competence. We can speculate that our conclusion apply also to other types of moral-cognitive methods like the Just Community method (Power et al., 1989). However, there is much less research available on other methods so they cannot be tested as rigorously as the MDD method.

We can also speculate that the MDD method affects not only the level of moral judgment competence but other competencies and motivations, and, therefore, MDD methods are also beneficial for adults and for children prior to their adolescence. Research on those speculations are still scarce. But studies reviewed here and my extensive experiences with moral dilemma

---

[19] This does not mean that DIT-based findings are irrelevant for developmental theories. Lind (1993) demonstrates that, under most conditions, moral attitudes correlate very highly with moral competencies and thus support Piaget's Parallelism theorem, and the claim that the DIT validly indicates moral development. In the Social Sciences, we often use such indirect indicators with great success, if a direct measure of the construct in question is not feasible. Many studies show, that, usually, the DIT is well suited for studying moral development and the research and has paved the way for a great wealth of findings that are fully in line with other cognitive-developmental research (Rest, 1986).

discussion in kindergarten, college, university and adult groups, suggests that they may be correct. Through the MDD method, developmentally advance subjects seem to gain new insights into the "actual" meaning of moral values and explore a wide range of action-consequences of otherwise abstract moral principles. And, they also learn to appreciate the MDD method as an effective method for moral learning and dilemma solution.

In the long run, this insight may foster the ability of self-sustaining moral development (Lind, 1996a). People may look up or construct MDD situations for themselves without the help of teachers and educators. In a discussion about abortion, one student remarked: "I like to watch pro-con discussions on TV, but this here is much better, it lets us much deeper think through the arguments for and against an issue." When I asked the participants to evaluate the discussion class, most say they would sign up for it again."

Such self-sustaining moral development perhaps takes place in teachers who become involved in this method, as our experience with a school-democracy-project shows that I conducted together with Fritz Oser, Ann Higgins, Larry Kohlberg, Heinz Schirp, Sibylle Reinhardt and others, in schools in Germany. In this project, the participating teachers told us afterwards that they changed considerably as a consequence of conducting MDD classes and JC meetings. They said their thinking about the students changed much as they got to know them better than during regular classes. One teacher, who had applied for early retirement, gained a new insight into his role as an educator and withdrew his application (for additional information on this project see Lind & Althof, 1992; Oser & Althof, 1992).

Other authors have argued that also kindergarten and elementary school children profit from cognitive-developmental methods of moral teaching at school (Shaheen & Kuhmerker, 1991), and at home (Speicher, 1994). However, for children of this age, we still lack the kind of rigorous measures of moral judgment competence that we have for studying adolescents and adults. So we need more intervention studies in this age groups before we can settle this dispute on empirical grounds.

References

Blatt, M. & Kohlberg, L. (1975). The effect of classroom moral discussion upon children's level of moral judgment. Journal of Moral Education, 4, 129-161.

Bühn, A. (1995). Die Rolle von moralischer Urteils- und Wahrnehmungsfähigkeit im Prozeß der Fach- und Berufswahl von Studierenden. [The role of moral judgment and perception competence in choosing the field of study and of career.] University of Konstanz, unpublished diploma research.

Cohen, J. (1977). Statistical power analysis. (2nd edition). New York: Academic Press.

Cohen, J. (1994). The Earth Is Round ($p < .05$). American Psychologist, 49(12), 997-1003.

Colby, A. & Kohlberg, L. (1987). The Measurement of moral judgment. Volume I and II. Cambridge, MA: Cambridge University Press.

Damon, W. & Gregory, A. (1997). The youth charter: towards the formation of adolescent moral identity. Journal of Moral Education, 26, 117-130.

Davison, M. (1979). The internal structure and the psychometric properties of the Defining Issues Test. In J.R. Rest (1979), Development in Judging Moral Issues, pp. 223-245. Minneapolis, MN: University of Minnesota Press.

DeVries, R., & Zan, B. (1994). Moral classrooms, moral children. Creating a constructivist atmosphere in early education. New York: Teacher College Press.

Edelstein, W. (1987). Förderung der moralischen Entwicklung in der Schule. Möglichkeiten und Grenzen. Zeitschrift für Pädagogik, 33, 185-205.

Emler, N. (1996). How can we decide whether moral education works? Journal of Moral Education, 25, 117-126.

Enright, R.D., Lapsley, D.K., Harris, D.J. & Schawver, D.J. (1983). Moral development interventions in early adolescence. Theory into Practice, 22, 134-144.

Glass, G.V., McGaw, B. & Smith, M.L. (1978). Meta-Analysis in social research. London: Publications.

Hartshorne, N. & May, M.A. (1928). Studies in the nature of character. New York: Macmillan.

Higgins, A. (1980). Research and measurement issues in moral education interventions. In R. Mosher, Ed., Moral education. A first generation of research and development, pp. 92-107. New York: Praeger.

Kohlberg, L. & Higgins, A. (1984). Continuities and discontinuities in childhood and adult development revisited -- again. In L. Kohlberg, The Psychology of Moral Development, pp. 426-497. San Francisco: Harper & Row Publisher.

Kuhn, D., Langer, J., Kohlberg, L. & Haan, N. (1977). The development of formal operations in logical and moral judgments. Genetic Psychological Monographs, 95, 97-188.

Leming, J.S. (1981). Curricular effectiveness in moral/values education: A review of research. Journal of Moral Education, 10, 147-164.

Lind, G. (1978). Wie mißt man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts. In G. Portele, Sozialisation und Moral, pp. 171-201. Weinheim: Beltz

Lind, G. (1985a). Growth and regression in moral-cognitive development. In C. Harding, Ed., Moral dilemmas. Philosophical and psychological issues in the development of moral reasoning, pp. 99-114. Chicago: Precedent Publishing Inc.

Lind, G. (1985b). The theory of moral-cognitive judgment: A socio-psychological assessment, pp. 21-53. In G. Lind, H.A. Hartmann & R. Wakenhut, Eds., Moral development and the social environment. Studies in the philosophy and psychology of moral judgment and education. Chicago: Precedents Publishing Inc.

Lind, G. (1993). Moral und Bildung. Zur Kritik von Kohlbergs Theorie der moralisch-kognitiven Entwicklung. Heidelberg: Asanger. New edition: Lind, G. (2002). Ist Moral lehrbar? Berlin: Logos.

Lind, G. (1994). Why do juvenile delinquents gain little from moral discussion programs. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 1994.

Lind, G. (1995). The meaning and measurement of moral competence revisited--A dual-aspect model. Paper presented to the conference of the American Educational Research Association, in San Francisco, April 1995. Replacement: Lind, G. (2019). How to teach morality. Berlin: Logos.

~~Lind, G. (in press). The Psychology of Moral Competence. Hillsdale, NJ: L. Erlbaum.~~

Lind, G. (1996a). Educational environments which promote self-sustained moral development. Paper presented at the meeting of Division E, American Educational Research Association, New York, April 1996.

Lind, G. (1996b). A re-analysis of the Barnett, Evens and Rest (1995) DIT-faking-study. Unpublished manuscript.

Lind, G. (1996c). Leading moral dilemma discussions. In L. Nucci's World-wide-web page "http://www.uic.edu:80/~lnucci/MoralEd/practices.html"

Lind, G. & Althof, W. (1992). Does the Just Community program make a difference? Measuring and evaluating the effect of the DES project. Moral Education Forum, 17, 19-28.

Niemczynski, A., Czyzowska, D., Pourkos, M. & Mirski, A. (1988). The Cracow study with Kohlberg's moral judgment interview: Data pertaining to the assumption of cross-cultural validity. Polish Psychological Bulletin, 19, 43-53.

Nucci, L. & Weber, E. (1991). The domain approach to values education: From theory to practice. In W. Kurtines & J. Gewirtz, eds., Handbook of moral behavior and development: Theory, research, and application, Vol. 3, pp. 251-266. Hillsdale, NJ: L. Erlbaum.

Nunner-Winkler, G. & Sodian, B. (1988). Children's understanding of moral emotions. Child Development, 59, 1323-33.

Oser, F. & Althof, W. (1992). Moralische Selbstbestimmung. Modelle der Entwicklung und Erziehung im Wertebereich. Stuttgart: Klett-Cotta.

Oser, F. & Schläfli, A. (1985). The thin line phenomenon. In G. Lind, H.A. Hartmann & R. Wakenhut, Eds., Moral development and the social environment. Studies in the philosophy and psychology of moral judgment and education. Chicago: Precedents Publishing Inc., pp. Xxx-xxx.

Penn, W. (1990). Teaching ethics--A direct approach. Journal of Moral Education, 19(2), 124-138.

Piaget, J. & Inhelder, B. (1969). The psychology of the child. New York: Basic Books.

Power, C., Higgins, C. & Kohlberg, L. (1989). Lawrence Kohlberg's Approach to Moral Education. New York: Columbia University Press.

Rest, J.R. (1979). Development in judging moral issues. Minneapolis, MI: University of Minnesota.

Rest, J.R. & Narvaez, D. (1994). Summary: What's possible? In J. R. Rest, & D. Narvárez, Eds., Moral Development in the Professions: Psychology and Applied Ethics, pp. 213-224. Hillsdale, NJ: Erlbaum.

Schläfli, A., Rest, J.R. & Thoma, S.J. (1985). Does moral education improve moral judgment? A meta-analysis of intervention studies using the Defining Issues Test. Review of Educational Research, 55, 319-352.

Scholz, M. (1996). Experimentelle Überprüfung der Auswirkungen von Dilemma-Diskussionen auf die moralische Urteilsfähigkeit von Gymnasiasten [Experimental test of the effect of dilemma-discussion on the moral judgment competence of Gymnasium-students.] Universität Konstanz, Department of Psychology, unpublished diploma research.

Shaheen, J. & Kuhmerker, L. (1991). Free to learn, free to teach. Manhatten, KS: Master teacher (P.O. Box 1207, Manhatten, KS 55502).

Shaver, J.P. (1993). What statistical significance testing is, and what it is not. Journal of Experimental Education, 61, 293-316.

Speicher, B. (1994). Family patterns of moral judgment during adolescence and early adulthood. Developmental Psychology, 30, 624-632.

Sprinthall, N.A. (1994). Social role taking and counseling. In J. Rest & D.F. Narváez, Eds., Moral Development in the professions: Psychology and applied ethics. Hillsdale, NJ: Erlbaum.

Sprinthall, N.A. & Collins, W. A. (1995). Adolescent Psychology. A Developmental View. Third Edition. New York: McGraw-Hill.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: three suggested reforms. Educational Researcher, 25(2), 26-30.

Turiel, E. (1983). The development of social knowledge. Morality and convention. Cambridge: University Press.

Wasel, W. (1993). Simulation moralischer Urteilsfähigkeit. Moralentwicklung: eine kognitiv-sturkturelle Veränderung oder ein affektives Phänomen? [Simulation of moral judgment competence. Is moral development a cognitive-structural change or an affective phenomenon?]University Konstanz, unpublished diploma research.

Wohlwill, J.F. (1977). Strategien entwicklungspsychologischer Forschung. Stuttgart: Klett.

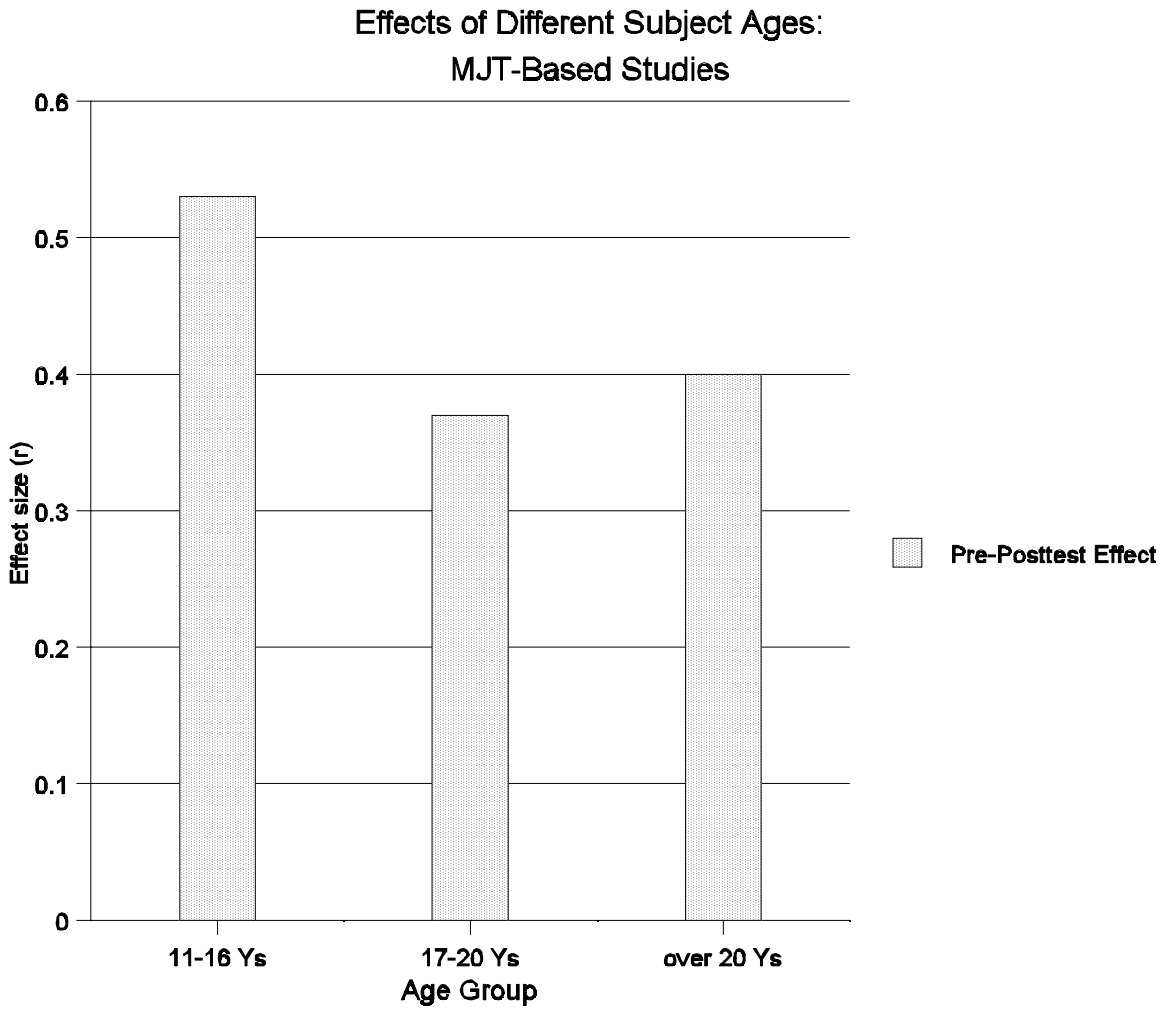


Figure 1 The basis for this meta-analysis are 48 intervention samples, in which the Moral Dilemma Discussion Method was used for treatment, and in which the effect was measured through the Moral Judgment Interview (MJI). Effect size is indexed by the correlation coefficients, r. For more information on this meta-analysis and the sources used, see by Lind (1993, pp. 193 ff.).

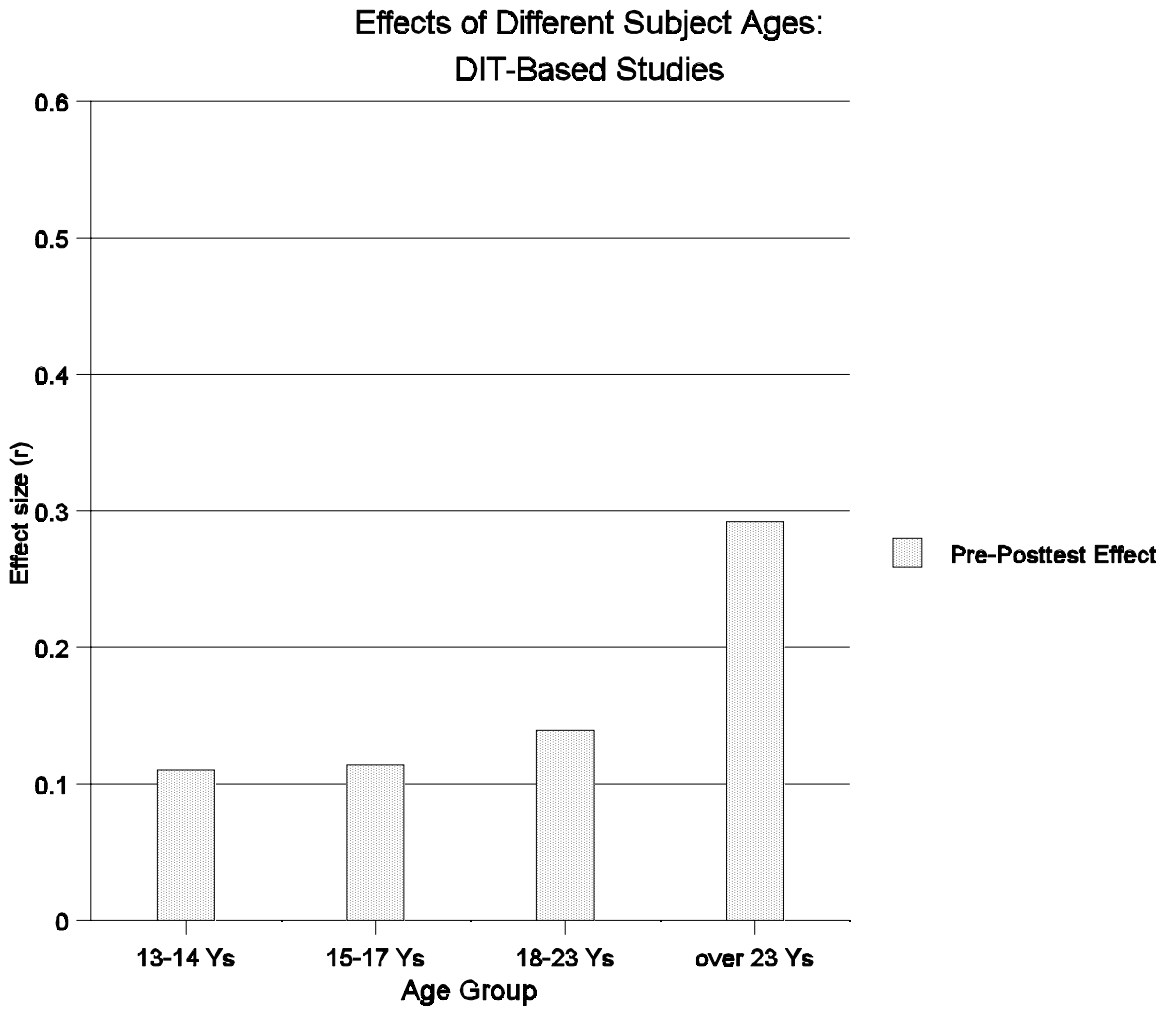
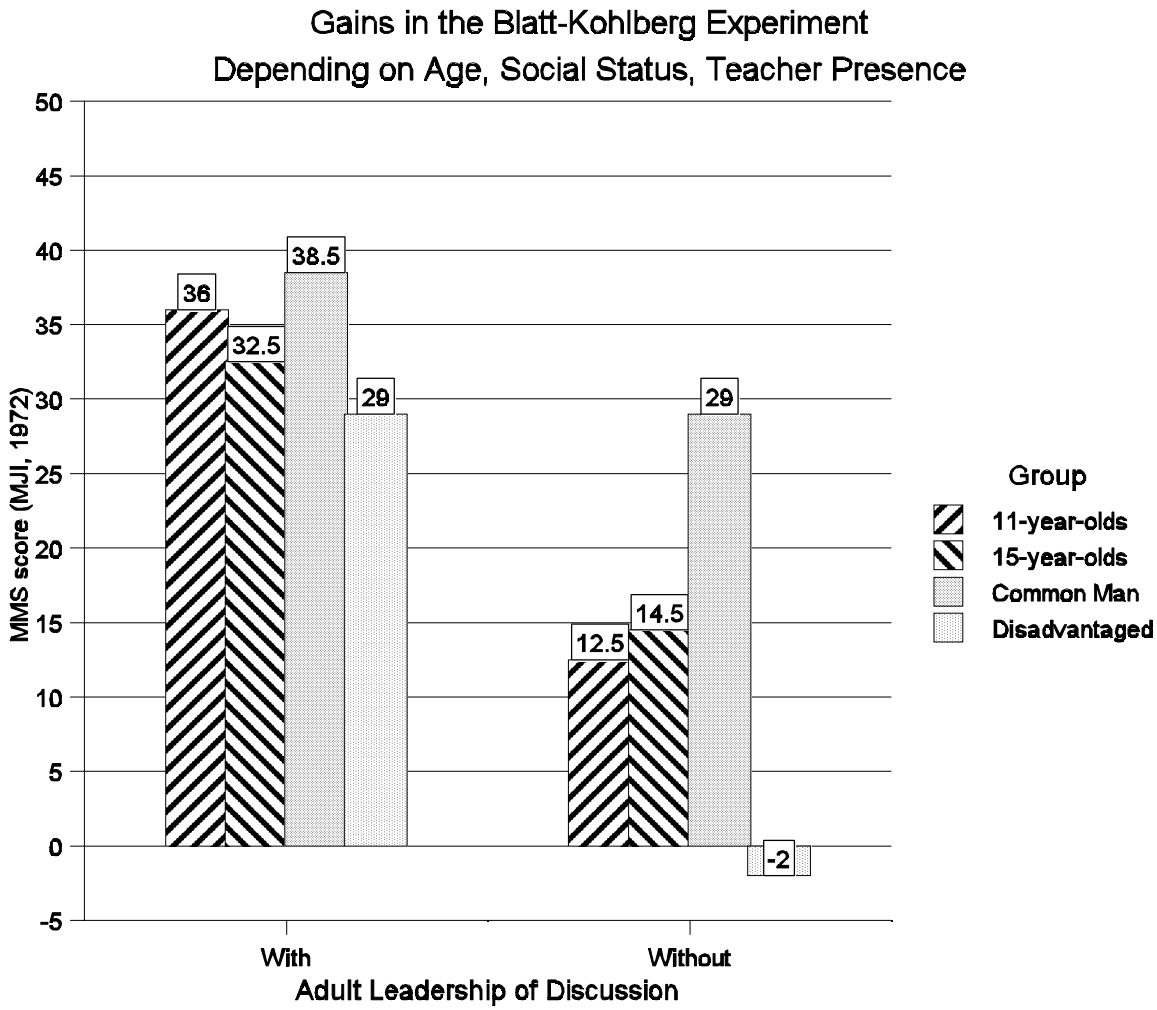


Figure 2 This meta-analysis by Schläfli, Rest and Thoma (1985) is based on 68 intervention samples, which were evaluated with Rest's (1979) Defining Issues Test (DIT). For better comparability, the d-values reported by Schläfli et al. (1985, p. 345, Table VII) are converted here into correlation coefficients by the following formula: $r = \sqrt{d^2 / (d^2 + 4)}$ (see Glass et al., 1978, pp. 148 ff.).


Gains in the Blatt-Kohlberg Experiment
Depending on Age, Social Status, Teacher Presence
MMS score (MJI, 1972)
50
45
40
35
30
25
20
15
10
5
0
-5
36
32.5
38.5
29
12.5
14.5
29
-2
With
Without
Adult Leadership of Discussion
Group
11-year-olds
15-year-olds
Common Man
Disadvantaged