

The Relative and Absolute Effect Size
of the *Konstanz Method of Dilemma Discussion* (KMDD)¹

Georg Lind²

Abstract

In this presentation, the efficacy of the *Konstanz Method of Dilemma Discussion* KMDD is reviewed, which has been developed by Georg Lind over the past 20 years on the basis of the Blatt-Kohlberg-method of dilemma discussion. While the average effect size of successful interventions and treatments in the field of education, psychotherapy, medicine and work place improvement is about $r = 0.30$ (Lipsey & Wilson, 1993), the Blatt-Kohlberg method's mean relative effect size is $r = 0.40$ (Lind, 2002). In a series of small intervention studies, including a randomized experimental study in Thailand by Lerkiatbundit et al. (2006), the effects of the KMDD on the development of students' moral judgement competence are studied. After improvements of the method on the basis of three core didactical principles, namely a) constructivism, b) alternating phases of support and challenge, and c) free Habermas-type discourse, the relative effect sizes of the KMDD reached $r > 0.70$, and its absolute effect size grew to more than 20 C-points per semester. In contrast, the absolute effect size of regular schooling ranges from +3 points increase to 3 points decrease per year across various countries with different educational systems.

¹ This is a revised and extended version of my presentation at the annual meeting of the *American Educational Research Association*, in the symposium *The Quality of the Konstanz Method of Dilemma Discussions (KMDD) for Teaching Moral and Democratic Competencies* in Chicago, April 9 - 13, 2007.

² Contact: Prof. Dr. Georg Lind, Department of Psychology, University of Konstanz, 78457 Konstanz.
Email: georg.lind@uni-konstanz.de, <http://www.uni-konstanz.de/ag-moral/>

1. Preface

To get something done, one often needs to focus. For more than thirty years, I have been focusing on the definition and measurement of moral *competencies* (Lind, 2005) because there was no generally accepted scientific method to do so in spite of the fact that the import of such competencies has long been recognized (Darwin, 1989/1871³; Kohlberg, 1964; Lind, 2002). Moral competencies are needed to translate moral ideals into real behavior.

For more than twenty years I have been focusing on fostering moral (and democratic) competencies (Lind, 1987; 2003; 2006). I focused on the Blatt-Kohlberg method of dilemma discussion, which we tried to improve in regard to its theoretical underpinning, the method itself, and its teachability.

Yet, before improving this method, we needed to investigate what ‘improvement’ in educational science means. Traditionally, improvement meant to make an educational method better in line with a well-established, philosophically sound theory of good schooling regardless of its efficacy. Subsequently, empiricist researchers call any “significant” change on any of numerous scales an improvement if these scales were in any reasonable way associated with the terms morality and democracy or civics. To bring some order into this wide field, they suggest to distinguish between ‘behavior,’ ‘knowledge,’ ‘attitudes,’ and values’ (WWC, 2007). Yet the empiricist way of defining ‘improvement’ also has severe drawbacks. It has no linkage to any philosophical or pedagogical grounding. It lacks often internal validity because it only superficially refers to psychological research on learning and instruction, and leaves out the indispensable concept of ‘competence’ or ‘ability,’ maybe because it is more difficult to measure than,

³ “The moral faculties are generally and justly esteemed as of higher value than the intellectual powers” (P. 636, Volume 22, Part two).

e.g., attitudes and values. It operationalizes effect only in terms of the questionable concept of statistical significance, which is biased in favor of big money because any effect can be made statistically significant just by increasing the sample. Finally, it wrongly considers ‘randomized controlled trials’ as the non-plus-ultra evidence standard, as if randomized assignments of subjects to trial and control groups would solve all problems of scientific inference. Because humans are social and historical beings (their behavior is determined through social context and they learn and develop), random assignment of subjects to artificial learning groups undermines the ecological validity of the assessment (Campbell, 1976). As it turns out, the requirement of perfectly randomized trials also contributes to an bias. It can only be achieved by paying subjects and institutions huge amounts of compensations for participation.

Thus, another way of defining and assessing ‘improvement’ in the field of moral and democratic education needed to be considered and pursued, one which combines experimental thinking with theory development (Oser & Patry, 1986). Hence we focused on the *competence aspect of moral and democratic* behavior because this concept was embedded into a well-developed web of theorizing and experimental studies. Moral competencies help us to better understand the relationship between moral ideals and behavior. Moral competencies seem to play a key role in a wide arrange of behaviors – rule conformity, altruism, civic courage, decision-making, and even in academic learning (Blasi, 1980; Kohlberg & Candee, 1984; Sprinthall et al., 1994; Heidbrink, 1985; Lind, 2003). For fostering moral competencies, we focused on the improvement of an existing method, the dilemma discussion method by Blatt and Kohlberg (1975). We did this through cycles of theoretical analysis *and* experimental assessment. We wanted both to improve the efficacy of that method and to provide the ground for a better theoretical understanding. Only through an experimentally validated theory we can identify the essential principles of

effective moral and democratic teaching, and can thus adapt the method to a greater variety of contexts (e.g., students and teaching objectives). When we realized that we profoundly changed the original method, we decided to call it the ‘Konstanz Method of Dilemma Discussion.’ Eventually we learned also that the principles of effective dilemma discussions can be applied to teaching in general as well.

Finally, we trusted the discounting principle and ‘peace-meal’ science, which cannot prove absolute truth (no method can do this) but which leads, as Popper (1968) argued, to ever greater trust in theories which could be confirmed by highly falsifiable experiments by many independent researchers, while at the same time rivaling theories could be discounted by these experiments. Finally, we decided not to settle for less than perfection. We wanted to call an educational method ‘effective’ only if it could move all students’ abilities upward (this achievement is well reflected in measures of ‘relative effect size’) and not only the ability of some students, and if the absolute effect size of the method is clearly greater than that of regular school teaching.

In this paper, I will give a current account of what has been achieved so far by our efforts to create an effective method of fostering moral and democratic competencies in terms of relative and absolute effect sizes. The resulting educational method has been described at length elsewhere (Lind, 2003; 2006) and so have been the method of measurement. It should, therefore, suffice to describe both briefly.

2. The *Konstanz Method of Dilemma Discussion* (KMDD)

In the Konstanz Method of Dilemma Discussion (KMDD) has been designed to stimulate the development of moral judgment competence as defined by Kohlberg (1964) and elaborated in Lind's Dual-Aspect Theory of moral behavior (Lind, 1985, 2002, 2005 a; see also Piaget, 1976). To create an optimal environment for moral and democratic learning, the KMDD is based on three psychological-didactic principles of learning (see Lind, 2003):

First, it is based on the *constructivist* principle of self-learning through confrontation with educative tasks or problems. The KMDD employs semi-real, 'educative' moral dilemmas as tasks. Usually, semi-real dilemmas are used, which are dilemmas that do not affect anyone participating in a dilemma discussion but are likely to cause emotional real conflicts between the moral ideals of the participants, and thus trigger controversial discussions among the participants. We call a dilemma 'educative' if it triggers moral emotions enough to stimulate learning, but not too strongly to prevent learning. The role of those dilemmas then is similar to a vaccination, in which weakened viruses are injected into the body to stimulate its ability to resist the real virus. Educative moral dilemmas are not so real that they kill of moral-democratic learning, but real enough that they cause the individual to develop his or her cognitive-affective capacities (Lind, 2003; <http://www.uni-konstanz.de/ag-moral/>). In honor of the founder of cognitive-developmental moral education, we call this also the *Kohlberg-principle*. Kohlberg discovered the competence aspect of moral behavior (Kohlberg, 1963) and also insisted that morality cannot be indoctrinated but must be *stimulated* by moral tasks, i.e., moral dilemmas (Kohlberg, 1984).

Second, the KMDD is based on the principle of optimal arousal level (Yerkes & Dodson, 1908) in order to maximize the learning process and the sustainability of the learned. The optimal

excitement level, which prevents boredom and sleep during the lesson on the one hand, and over-excitement and anxiety on the other, is achievement through alternating cycles of challenge and support (see Sprinthall et al., 1994). These cycles change about every ten minutes with some range of variation in order to allow the teacher to adjust the level of excitement whenever it gets out of the optimal range. *Supportive* tasks are designed to make the student feel comfortable and reassure him or her, for example through good explanations, exchange with same-side participants, time for deliberation etc. Challenging tasks are designed to arouse the student's curiosity and attention, for example through vote-taking, decision-making, confrontation with counter arguments etc.

Third, the KMDD is based on the idea that class-room management must put into practice the very moral principles for which it is made, like mutual respect, responsibility, free speech, and equality before the law. Thus the KMDD has been designed to facilitate free moral deliberation and discourse through several features: a) sufficient time is allowed for the clarification of the dilemma; b) the authority (teacher) withholds his/her opinion on the dilemma and even leaves the moderation of the discussion to the participants; and c) participants are empowered to build self-confidence through meetings of same-side groups. In honor of one of the most prominent proponents of *discourse ethics*, Jürgen Habermas (1995), we also call this principle of moral learning the *Habermas-principle*.

Special features of the Konstanz method are as follows: a) The KMDD leaves plenty of time at the beginning of the session to bring out and clarify the different perceptions of the presented problem by the participants. b) The KMDD contains not only a phase of controversial discussion of a dilemma, during which the participants are to articulate the reasons for their decision and listen actively to counter-arguments, but also a phase of "reconciliation," in which participants

are to nominate the “best argument of the other side.” The KMDD is always concluded by a phase of evaluation by the participants. The participants are asked: “Did you like it?” and “What have you learned?” The KMDD helps the participants reduce anxiety and any feelings of animosity against the members of the opposing group. d) Finally, *educational quality* control is built into the method at the level of participant feedback, of teacher self-controlled program evaluation, and of teacher training:

- At the end of each dilemma session, five to ten minutes are reserved for evaluation of the session by the participants (Was it fun? What did you learn?).
- Continuous self-controlled evaluation of the effectiveness of the method by the teacher is made possible through ITSE, which is a versatile and inexpensive internet-based test-administration program for the *Moral Judgment Test* and other assessment instruments (see: http://www.uni-konstanz.de/itse-projekt/itse_home.htm). The ITSE-program is open source. It is available for a handling charge, no licence fees, from the author.
- Teachers are offered a 100-hour training course (“MODEL”) and a certification process for using the method (“KMDD User Certificate”) and for training others in this method (“KMDD Trainer Certificate”).

The KMDD method is described in detail in Lind (2003) and on this web-site:

<http://www.uni-konstanz.de/ag-moral>.

Historically, the KMDD grew out of the Blatt-Kohlberg method of dilemma discussion (see Blatt & Kohlberg, 1975), yet it outgrew this method in several ways:

- Only one dilemma pro session is posed before the participants, instead of several as often is done in the Blatt-Kohlberg method. This leaves more time for the participants to prepare, to reflect, and to discuss.

- The duration is 90 minutes, as compared to 45 minutes for the Blatt-Kohlberg method. This also gives much more time to the participants to engage in own reflection and deliberation. The use of the KMDD in more than 20 years has shown that it is not possible to reduce it to 45 minutes without losing out on important phases, and its effectiveness.
- The role of the teachers is much less dominant in the KMDD. The teacher's portion of speaking is much smaller, giving the participants much more space for their own learning. The teacher also withholds his or her own arguments and does not moderate the discussion. This reduces anxiety among the participants and gives them more freedom to think and speak of the top of their heads.
- Last but not least, the KMDD uses the *counter-arguments* of other participants for stimulating moral judgment competence, rather than, as the Blatt-Kohlberg-method does, the teacher's arguments above the "stage" of the participants ("plus-one-convention"). The +1-convention has been rightly criticized for not being the most effective way of stimulating moral development and for not being compatible with cognitive-developmental theories (Berkowitz, 1981; Walker, 1983).⁴ The aim is to foster the participants' ability to apply their own moral principles in everyday life, rather than to convey particular moral orientations or values. Besides, the *plus-one-convention* is so impractical that hardly any teacher uses it. "If it was true that the '+1-stimulation' would be the only decisive strategy of Kohlbergian moral education [...] then we would have to conclude that this approach is useless for the teaching in schools." (Oser & Althof, 1992, p. 115)

⁴ "The pro/con 0 condition is, however, the condition that best distinguishes the theories. Cognitive-developmental theory is unique in predicting that higher stage reasoning will emerge from conflict and inconsistencies within one's own stage." (Walker, 1983, p. 104)

2. The criterion for efficacy: *Moral Judgment Competence*

The KMDD is to foster moral judgment competence as defined by Kohlberg (1964), as elaborated by Lind (1985; 2002) in his Dual-Aspect Theory of moral behavior, and as operationalized by Lind through the *Moral Judgment Test* (MJT; Lind, 2004; 2005). Moral judgment competence is seen as the bridge between our moral ideals on the one hand, and our every-day decision-making on the other. Kohlberg defined it as “the capacity to make decisions and judgments which are moral (i.e., based on internal principles) and to act in accordance with such judgments” (1964, p. 425). This definition is remarkable. It is distinct from other psychological definitions in three ways: first, it defines morality as moral ideals or preference but also as a *competence* or *ability*; second, it refers to *internal* (or, at least, internalized) moral principles, and not to external social norms (which we should call “norm conformity”); and third, it contains a reference to behavior or action as part of the definition. In other words, we can speak of moral competence only when the individual demonstrates his or her appreciation for a particular moral principle through his actual behavior. This means also that, in order to demonstrate his or her ability, the individual must be confronted with an appropriate moral task. Thus any valid test of moral judgment competence must contain such a moral task.

Like all competence tests, the MJT contains a difficult moral *task*. The task consists in rating a series of pro and contra arguments, which challenge the participant’s opinion on a moral deci-

on.⁵ The participant must first decide whether the action choice of the main character in the moral dilemma stories of the MJT is right or wrong. This sets the stage for the actual task of the MJT, the rating pro and contra arguments. These were chosen to represent each of Kohlberg's six "stages" of moral orientations. The respondents' rating pattern reveals whether he or she is able to discern and appreciate the moral quality of the arguments, and rates them only according to moral quality but not their opinion agreement. At the highest level of moral judgment competence (and the highest C-score on the MJT) participants discern the differences in moral quality, and judge the arguments accordingly regardless of whether they support or oppose his or her opinion on the dilemma decision (see Figure 1).

Here Figure 1

The C-score (for Competence) is a quantitative index of this competence (see Lind, 2004). It is not linked to any type of moral orientation. That is, even if a participant would prefer a low "stage" of moral reasoning as most adequate and reject higher stages, he or she could still get a high C-score. This way, the MJT had allowed us to test Piaget's (1976) hypothesis of cognitive-affective parallelism without circularity. In fact, Piaget's parallelism hypothesis has been supported in very many studies in a remarkably clear way: The higher the moral judgment competence of a person (the cognitive aspect) the stronger his or her preferences for high-stage moral reasoning and tendency to reject low-stage moral reasoning (the affective aspect; see Lind, 2002;

⁵ Not only *counter-arguments* present a challenge to most people but also supporting-arguments as they urge the proponent to justify his or her decision or opinion. Therefore, it is not surprising that some participants even refuse to respond to any argument in the test.

2005). The MJT is cannot be faked upward and participants with high C-scores are more accurate than low scorers in estimating the moral judgment competence of other people. This has been demonstrated in two experiments (Lind, 2002; Wasel, 1994).

The MJT is a unique instrument. It is an ‘experimental questionnaire’ that enables us to study the impact of dilemma discussion on moral *competencies* in a non-confounded way. In most earlier studies, the only instrument for assessing the effectiveness of dilemma discussions were tests of moral reasoning (Kohlberg’s *Moral Judgment Interview*; see Colby et al., 1987) and moral preferences (Jim Rest’s *Defining Issues Test*, John Gibbs’ *Socio-Moral Reasoning Measure*, and other instrument of moral attitudes) were available for assessing the effectiveness of dilemma discussions. Of the two, Kohlberg’s *Moral Judgment Interview* did a better job of assessing the competence aspect of moral judgment behavior. In its original version, the participants were also confronted with *counter-arguments* in order to test whether their moral orientations were consistent. Later this task was de-emphasized (for a critique see Lind, 1987), which is unfortunate. No experimental faking studies exist.⁶ Yet we believe that the MJT provides a valid proxy score for moral incompetence. None of the other tests, it seems, contain a moral task: they only test the participants’ moral preferences or orientations. Studies using them may at best reflect gains in moral judgment competence indirectly because of the parallelism phenomenon. We should, however, remember that the parallelism can break down if something is at stake, e.g., if the test-taker expects reward or punishment. The most likely reward in those studies is “social desirability,” that is, the tendency to please the instructor. Competence tests are not affected by

⁶ An exception is a small, informal faking study by Dr. Helen Haste (personal communication) who found no upward faking with the Moral Judgment Interview.

this tendency.⁷ The C-score of moral judgment competence ranges from 0 to 100. Most C-scores range from 0 to 50, with only a few people getting scores higher than 50. The MJT is not made for individual testing but for basic research and for effectiveness studies involving at least 10 measurements for each data point.

Although very high test-retest reliability coefficients have been reported (Lerkiatbundit et al., 2006), the author of the MJT advises not to use it for comparing groups of less than 10 participants. (The use of the MJT for selection purposes or other high stakes purposes is considered an abuse.) Being based on Lind's (2002) Dual-Aspect Theory of moral behavior, the MJT has not been designed to give 'Stages' (see Lind, 1989).

3. Efficacy of psychological intervention methods

Before reporting on the *efficacy* or effectiveness of the KMDD, some comments need to be made on this concept and its various meanings. Although effectiveness became a topic for psychology some time ago, it only recently became a theme in its own right in educational research (Thompson, 1996). Traditionally, psychologists were only interested in *statistical significance*, although this concept is very controversial (see, e.g., Meehl, 1978; Carver, 1993; Kromrey, 1993; Campbell, 1976). The main disadvantage with this concept is that any desired level of "significance" can be reached by simply increasing sample size.

⁷ Yet other methodological problems may arise with them, too, for example the problem of moral understatement (Kohlberg & Kramer, 1969).

Thus studies with different sample sizes yield different “effects” of the intervention method. APA and AERA do, therefore, now require researchers to calculate and report also measure of “effect size,” either the *d*-index, which tells us how big an effect is in terms of the pooled standard deviation of the criterion measure in the sample, or the *r*-coefficient of correlation, which tells us how strongly the intervention variable (e.g., dilemma discussion versus no dilemma discussion) correlates with the changes in the criterion measure (e.g., degree of moral judgment competence). But this index of effect size has also some draw-back since it is only a measure of *relative effect size*. It depends on the standard deviation of the criterion measure in the samples, and, therefore, results from samples with different standard deviations cannot be compared. Thus, any desired level of “effect size” can be reached by reducing the standard deviation. The same changes look the bigger, the smaller we make the variance. In this study, we nevertheless report relative effect sizes, using the *r*-coefficient, which has some nice properties: we are more familiar with it than with the *d*-index, it can be easily converted into the *d*-index, and it varies within a fixed range, namely from -1 to +1.

In developed areas of science, effect sizes are determined simply by reading the scales or by averaging several such readings. If we wish to know whether the temperature has changed from yesterday to today, we read the thermometer twice and calculate the difference. If the difference reaches a practically relevant level, we may decide to leave our coat at home or wear a sweater today depending on whether the change was positive or negative. The practically relevant level is not determined by statistical significance testing nor by the correlation coefficient but by our long-time experience of what the scale values and their differences *mean*. Thus the use of *absolute effect sizes* depends on the degree to which we have become acquainted with a particular measurement tool. Measurement error is dealt with by the design of the study, e.g., by repeating

the measurement as often as it is necessary to achieve the desired degree of accuracy. While for many practical purposes, we need only to read the scale twice, in psychological science, about ten readings are considered the minimum because of the complexity of its object.

The measurement instrument for moral judgment competence, the *Moral Judgment Test* (MJT) has been around for over 30 years (Lind, 2005) and the evidence accumulated seems so well established that we can use the MJT's C-scale for "reading off" the effect sizes in absolute terms. Informed by convention and experience, we average the reading of at least ten subjects who share the same properties in question in order to get an accurate reading of the scale value.

In this paper, we will report relative and absolute effect. We will not report *statistical significance* because of the serious flaws in this concept. We believe that we can establish confidence in the research findings by showing that a) they are replicable relevant, b) the hypothesis being tested are highly improbable and falsifiable (Popper, 1968), and c) they contribute to a consistent and coherent theory of moral behavior and development (Lind, 1985; 2002).

It should be noted that the rigorous evaluation that contributed decisively to the improvement of the KMDD method of moral education was always focused on the method and never on people who used this method. We strongly believe that only through this focus on program evaluation can we improve the efficacy of education but not by "evaluating" students and teachers.

3. What to expect?

What effect size should we expect to call it an “effect?” In order to gain a frame of reference for deciding this question we need to consider various conventions and research findings regarding “effective” interventions. In the 1980'es, a very large effectiveness study involving more than 20.000 participants was done to test the ability of the drug *Aspirin* to prevent heart attacks. After the incoming data were analyzed, the advisory committee concluded that Aspirin was “effective” and decided to stop the study prematurely since it seemed unethical to withhold Aspirin from the control group. When calculating the effect size which triggered this decision, researchers found an astonishingly low value, namely $r = 0.0011$ (Rosenthal, 1994). A correlation coefficient of this size is usually regarded as zero, and an effect size of this magnitude is not considered a significant effect (though this number may have been statistically significant because of the huge sample size). However, Aspirin reduced the number of deaths from 189 (the control group) to 104 (the experimental group) out of 10.000, and these numbers do indeed look “significant.” Note that the significance of the effect size in this study did not derive from statistical considerations but from the fact that human life was at stake.

An “effect” can also mean that a regression has been stopped. In a study of juvenile delinquents receiving treatment in form of just community experience, a method dealing with real-life dilemmas (Power et al., 1989), Stefan Glasstetter (2005) found that the experimental group showed hardly show any gains from pretest to posttest measurement using the *Moral Judgment Test*. However, at the same time the control group regressed (Figure 2). Contrary to some theories, regression of moral judgment competence has been observed in several studies (Lind, 1985 b; 2000; Schillinger, 2006).

We found a similar effect in teacher students. In general, the moral judgment competence of teacher students stagnated or regressed. Yet, those students who attended a dilemma discussion session during the semester showed no regression but rather a slight increase. While they were only 1 point apart at the beginning of the semester, at the end of the semester the two groups differed as much as ten C-points (Figure 3). Obviously, one dilemma discussion during the course made the difference.

Figures 2 and 3

Another way to get an idea of how much is “much” in regard to effect sizes, we can look at the rate of development of moral judgment competence in the absence of special intervention like the KMDD. In other words, we want to know what is the effect of an educational system that does not undertake any special efforts designed specifically to foster moral judgment competence. The secondary schools in Germany represents such a system. Drawing from various data sources, Lind (2002) estimated that these schools foster moral judgment competence by about 3,5 points per year (Lind, 2002). Figure 4 shows vocational school students’ C-score, which increases 12.1 points over a period of four years, that is, about three points per year increase.⁸ Another study by Lind (2002) comes to a similar estimate of about 3.5 points increase per year on a 100-point scale.

⁸ These data are taken from a cross-sectional by Klaus Beck (1993). I wish to thank him for letting me re-analyze the raw data of his study.

Figure 4

What happens if schooling ends before students have developed a sufficient amount of moral judgment competence that lets them become independent learners, i.e., learners who do not need the aid of parents and teachers anymore for maintaining their moral learning process? As in many other areas of learning, one can observe that a critical level of moral learning is needed for self-teaching to set in. If the learner is still below that critical level of competence development when he or she has no longer the support of an educational environment, then forgetting will set in rather than self-maintained learning. In his study of adolescents who graduate from low track schools at an early age and do not continue into college and university, Lind (2002) indeed found regression of moral judgment competence in adolescents who graduated from German lower-track secondary schools (*Hauptschule* and *Realschule*) after 9th or 10th grade. As Figure 5 shows, these graduates lose their moral judgment competence. It seems that they did not reach the level of ‘critical proficiency’ needed for self-learning to set in, as college students often do.

Figure 5

Hence, an *increase of the C-score of equal to or bigger than zero points* can already be considered an *educational effect*. Moreover, an increase of the C-score of bigger than 3.5 points can be called an *intervention effect*.

Yet, how big should the (relative) effect size be in order to call it practically significant or “effective” in a qualitative sense? As we have seen above, this question cannot be answered

without regarding the kind of effect that is involved. If life and death is at stake, or when we little is known about the method under consideration, we may consider very small relative effect sizes already as practically significant and recommend its application. As in the Aspirin-study, a $r = 0.0011$ deemed already as signifying an “effect.” Lipsey and Wilson (1993) report in their very comprehensive meta-analytical study that in the area of medicine, treatments of heart diseases and cancer need only to show an effect size of about $r > 0.15$ to be called “effective,” and to get the approval of the health authorities. Similarly, the US government seems to rest content with an effect size as small as of $r = 0.12$ for acknowledging privately developed programs for enhancing reading and mathematical literacy as “effective” and recommending its use in schools.⁹

In scientific evaluation research, a general consensus seems to exist that only an effect size of $r = 0.30$ and bigger allows one to call an intervention or treatment “effective” (Lipsey & Wilson, 1993). This consensus cuts across various fields, from medicine to works science, to psychotherapy and to education. Hence, from a relative-effect-size point of view, any effect size bigger than $r = 0.30$ is practically significant, and should lead to a recommendation of the method for practical purposes, e.g., for using it in schools for fostering moral judgment competence.

If, however, we believe that a teaching method should leave no child behind, than we should expect from a teaching method a much higher, if not ‘perfect,’ effect size of $r = 1.00$. Of course, we must keep in mind the respective efforts that are necessary to reach these effect sizes. It should not be surprising that low educational efforts in terms of time, money, teacher qualification, and research will lead to smaller effect sizes than big efforts. Hence, we should weight costs

⁹ The “Success for All” program, which is used in many US grade schools for improving reading literacy and which has been rated as one of the best programs by the *American Institutes for Research* (AIR, see <http://www.air.org/news/documents/Release200511csr.htm> , access: Dec. 12, 2005) showed “significant” gains only in one of four tests used for evaluation, and in regard to this test the effect size was only $r = 0.12$ (my calculation from the figures reported in Borman et al., 2005).

and effect size against each other and lower our expectations to a near-to-perfect effect size if this can be achieved with available resources.

3.2. How much do dilemma discussion affect moral preferences?

This paper deals only with the competence aspect of morality and not with its affective or preference aspect because we believe that moral dilemma discussion are mainly suited for fostering moral competencies rather than changing moral preferences. In fact, effectiveness studies show that dilemma discussions have little impact on moral preferences as compared, for example, to methods of “direct teaching” (Schläfli et al., 1985; Penn, 1990). According to Schläfli et al.’s meta-analysis of intervention studies, the effect size of dilemma discussions on children’s preference for principled moral reasoning is only $r = 0.11$. Only with adults the dilemma method shows a stronger effect $r = 0.29$ (see Figure 6). Some educators have concluded from this that moral education should set in only in adulthood. Yet, as we will see, these findings are not supported by studies of the competence aspect of moral behavior. The reason for the small effect size in youth could be due to the fact that in these studies the authors were only interested in the preference for principled moral reasoning (the so-called “P-score,” which is insensitive to changes at lower stages of development, or because of a social desirability effect. Possibly adults are more sensitive, and more willing to comply, to the desired outcomes of the training than adolescents.

3.3 What is the effect size of the Blatt-Kohlberg method on moral reasoning?

For a long time, the Blatt-Kohlberg method of dilemma discussion was one of the few methods of moral education which was thoroughly studied. Hundreds of interventions studies were conducted in the 70ies and 80ies, most of them reported statistically “significant” increases of the MMS score, the index of the *Moral Judgment Interview* (Lockwood 1978; Leming 1981; Higgins, 1980). Yet, besides statistical significance, none of these studies reported effect sizes and many were not documented sufficiently in order to calculate effect sizes *post hoc*. Lind (2002) retrieved 48 publications (mostly gray literature like dissertations and research reports) between 1970 and 1984 which did report all the numbers needed. These publications reported about 141 single intervention studies. For all these interventions, Lind calculated r as an effect size measure using all information (pretest scores, repeated measurement, control group comparison) available for each study. The formula for transforming statistical significance coefficients like t , F , or *proportion of variance* into the effect size measure r were taken from Glass et al. (1978) and Cohen (1988).

Because the intervention studies different widely in many respects, the result of this meta-analysis gives a very robust estimate of the average effect size of the Blatt-Kohlberg method on moral reasoning measures. This turned out to be $r = 0.40$ (see Figures 6). Thus the effect size of the Blatt-Kohlberg method is clearly above zero (none of the interventions produced negative effects!) and also above the usual criterion of $r = 0.30$ for calling an intervention “effective.” This meta-analysis showed that, in contrast to interventions aiming at changing moral orientations, the Blatt-Kohlberg method was most effective in the group of the 10 to 16 year old participants. It also showed that interventions of more than 45 minutes had more lasting effects on

moral reasoning (Lind, 2002). An adaption of the method of dilemma discussion and *just community* in Germany produced similar results (Lind & Althof, 1992; Oser & Althof, 1992).

3.4 What is the efficacy of the KMDD on moral judgment competence?

If it had been only for the numerical success of this method, there would have been hardly an good reason to revise and improve it – except that a r of 0.40 still leaves room for larger effects.

For three reasons, we had felt that the method of dilemma discussion still could be improved:

- First, because it was not clear what aspects of the method made it really work and would thus be indispensable if the method needed to be adapted to various curricular settings. Was it the “plus-one-convention”, i.e., the modeling of better moral reasoning by the teacher, or was it the stimulation of cognitive-affective processes through the confrontation with (counter-)arguments? We felt that the latter explanation was more in line with cognitive-developmental theory and also more practical and could thus help to make the method better teachable. A clarification of this question, we felt, was also necessary if the method was to be included in teacher education programs.
- Second, the instrument used for assessing the outcomes, Kohlberg’s *Moral Judgment Interview* confounded affective and cognitive aspects. Since education, in contrast to indoctrination or brainwashing, has little impact on individuals’ moral orientations and more impact on the competence aspect of moral behavior, we believed that intervention studies with more proper instrument, like the MJT, could reveal that the effects were even bigger yet.

- Third, we felt that one could really increase the efficacy of the dilemma method by giving the participants more space for moral deliberation and discourse by reducing the number of dilemmas in one session to one and by increasing the overall time from 45 minutes to 90 minutes.

When we started our intervention studies using a revised Blatt-Kohlberg method and the MJT, we had zero effects. We hypothesized that the MJT, which has shown to be non-fakeable, was perhaps a too difficult task for the respondents and, therefore, not sensitive enough for education-induced changes. Then we discovered that the C-scores of about two thirds of the sample were nicely increasing, but the C-scores of those who initially had the highest C-scores were sharply decreasing, thus averaging out as zero change. Accidentally, we learned that these participants were irritated because they suspected bad intentions on the side of the researcher (“Do they want to check on my trustworthiness?”). After we included an explanation in our instructions (“These questions are exactly the same as the ones you had answered some time ago because we are interested in their changes...”), virtually all negative trends disappeared.

We also learned over the years which psychological mechanisms really caused the changes, and how we can increase the efficacy of the dilemma discussion method. At some point, we felt that the changes became so substantial that we decided to give it a distinct name, the *Konstanz Method of Dilemma Discussion* (KMDD), describe in more detail above. Besides theoretical considerations, it was also the continuous use of self-evaluation, which Kohlberg has once called “bootstrapping,” that drove the development of this new method. Over the years, this process drove us away from Kohlberg’s method, but also brought us back closer to his basic pedagogy, as I had the privilege to observe in action: rarely intervening, modeling, or stage-typing.

This process of refining and also of changing the original method of dilemma paid off well. With the Konstanz method of dilemma discussion it is possible now to get a relative effect size of $r = 0.70$ and more, and an absolute effect size (gains) of 15 C-points and more. In two seminars (2004 and 2005-06), in which the psychological-didactic principles of the KMDD have been applied, this author got an relative effect size of $r > 0.70$ and more (see Figures 6 and 9, "KMDD 1" and "KMDD 2"). It is noteworthy that in the second seminar, we even did not use the format of the KMDD but applied its principles as faithfully as possible. In absolute terms, the C-score gains per *semester* were 15 and 20 points, respectively. These gains due to the KMDD compare very well with the gains of 3 - 4 points per *year* typically achieved by schools and universities with good teaching practices, and with the gains of about 6 points achieved by the Blatt-Kohlberg method of dilemma discussion.

These strong effects of the KMDD have been fully corroborated by a randomized experimental intervention study of colleagues from the *Prince Songkla University* in Thailand (Lerkiatbundit et al., 2006). They ran dilemma discussions using the KMDD for six weeks, one discussion every week. The participants were assigned at random to the experimental and the control group. Moreover, the authors also carried out a follow-up study (second posttest) six months later. In this Thai study, also very strong effects of the KMDD were found. The relative effect size was $r = 0.70$, and the absolute effect size (gain) was 14 C-points in only six weeks (Figure 8).

Figures 6 to 9

3.6 Can teachers be trained to use the KMDD effectively?

Yes, they can. However, their training needs to be of high quality and comprise about 100 hours of learning and tutored practicing (Koszinoffski, 2006). This study comprised initially 29 teachers, of whom five went through the full training program, which stretched over almost one school year. Though these five teachers did not get any relief from their work during this training, two succeeded to utilize most of the training program, three could not for various reasons. The main problem those three teachers was the heavy workload they had at school which prevented them from developing proper dilemmas for their classes. The effects of this training were assessed on the level of the students of the participating teachers, who were asked to nominate one of their classes for testing with the MJT and other instruments at the beginning and the end of the training program.

The findings were very clear-cut. The students of the two teachers who could make use of the full training were able to increase the moral judgment competence of their students by 8 to 14 C-points and more. The students of the three teachers who could only make use of parts of the program showed no or only minimal C-score gains (Koszinoffski, 2006, p. 30). Note that the *a priori* probability of this findings was only $p = 1/2^5 = 0,031$. That is, even though the sample size was merely $n = 5$, this finding is not only practically significant but also statistically significant.

The C-score gains are masked by the initial high-scorers (as they were in the early times of KMDD research). In the low-scoring group the students showed the most marked differences. The students of the two successful teachers increased their C-score by 15 to 18 points (p. 32), whereas the C-scores of the students of the unsuccessful teachers increased only between zero and eight points. In the initially high-scoring group most students' C-scores regressed. Because

of our experience with this phenomenon some years ago, we assume that this is a methodological artefact. The teachers may have been not able to convince those students of the need to be re-tested. Future studies will have to attend to this problem.

In the study by Zierer (2005), the KMDD and the MJT were adapted by the teacher to be used in grade school with 3rd graders. Zierer convened five dilemma discussions over a period of six weeks in an experimental class, and tested also a control group of 3rd graders, who did not get the treatment. His findings clearly show that he was able to use the KMDD effectively. The C-scores of the students in the experimental group increased by 5.5 points, whereas the C-scores in the control group remained unchanged (Zierer, 2006, p. 27; quotes taken from the draft 2005-report).

Over the past decades, we have trained about one thousand teachers in Germany, Colombia, Mexico, Italy and other countries to use the KMDD. Anecdotal reports show that in all instances, their students have welcomed this method and shown positive effects beyond increase of moral judgment competence. Many reports agree that the students who had an opportunity to participate in a KMDD-type discussion increased their desire for learning not only about moral matters but also about academic matters. As a biology-teacher wrote: “One such session is worth three weeks of biology-classes.” She and others report that the whole climate of learning in their classes changed for the better after they have started to use the KMDD method. In current studies we are more systematically investigating this surprising impact of dilemma discussions on the students.

4. How effective is the *Konstanz Method of Dilemma Discussion*?

The findings reported above let us clearly answer this questions with Yes, the KMDD is *highly effective*. Participants seem already to benefit considerably from this method when only one discussion session is held. The benefit of one session is about as big as the attendance of a full year of good schooling. If several KMDD-sessions have been held, the gains in moral judgment competence are about five times as high. This also holds true if the three basic psychological principles underlying the KMDD method are applied to the class in general. This means that the KMDD can be integrated well into any academic curriculum.

Since many years, the KMDD is now used in schools (grade 3 through 12), universities and many other institutions of education with good success. Most, if not all, students and teachers using this method enjoy it and profit from it. Teacher training programs in some German states teach it. The KMDD has been recommended by the Minister of Education of Colombia to all teachers in her country. The *Medical School of Monterrey, Mexico*, the *Tecnológico de Monterrey* (Hernandez & Medina, 2005), and the German Armed Forces are about to implement this method for teaching ethical and civic competencies throughout their system. Moreover, after twenty years in use, new benefits from using the KMDD and its underlying psychological teaching principles are still being discovered.

This success of the KMDD comes with a price. In particular, teachers must be thoroughly trained. A short introduction or the reading of a manual usually does not suffice. After many years of experience, we estimate that a 100-hour training program with a certification is required to prepare teachers for using this method effectively (see MODEL program, Notes, below). Yet,

this prize seems to be quite worth its benefits. Even with this caveat in mind, the *Konstanz Method of Dilemma Discussion* has clearly shown its strength as a method of moral and democratic education. As we have seen, only a few sessions per year already have an effect on moral learning which is sometimes bigger than the effect of a year of regular teaching. The KMDD is especially suited to be integrated into academic subjects teaching. Teachers who used it recommend to start a new topic with a dilemma discussion because of its positive impact on learning motivation and learning ability (Heidbrink, 1985). We have good reasons to believe that students' higher moral judgment competence will affect also their rule-conformity, altruism, and decision-making capacity (Kohlberg & Candee, 1984; Sprinthall et al., 1994; Mansbart, 2001; Lind, 2003).

In general, the intervention research on the KMDD has taught us some important lessons. First, adolescents, and even grade-school and kindergarten kids, in fact do benefit much from moral education if it is engaging. We need not, and should not, wait until adulthood. Second, evaluation can be a powerful tool for developing effective teaching methods, but only if it is used for improving methods and *not* for selecting and sanctioning people, and if the measurement instrument is fully valid. For this reason, we started with developing an adequate measure of moral judgment competence before we entered the field of moral education. Finally, we learned that education can be made much more effective if we utilize the findings of psychological research and bring back moral emotions into moral education, though in a carefully designed 'educative' dose.

Notes

The web-sites referred to in this paper are:

- KMDD – The Konstanz Method of Dilemma Discussion:
<http://www.uni-konstanz.de/ag-moral/moral/dildisk.htm> (in German, English, and Spanish – only partially synchronous; with local Google search tool).
- MJT – *Moral Judgment Test*: <http://www.uni-konstanz.de/ag-moral/mut/mjt-engl.htm> .
- ITSE – *Improvement of Teaching Through Self-Controlled Evaluation*:
http://www.uni-konstanz.de/itse-projekt/itse_home.htm .
- *Moral and Democracy Psychology & Education*: <http://www.uni-konstanz.de/ag-moral/> (in German, English, Spanish, and french – only partially synchronous; with local Google search tool). Also accessible through: <http://democracy-education.net> .
- MODEL – *Moral and Democratic Learning* courses and certification:
http://www.uni-konstanz.de/ag-moral/moral/workshop/kurse/_projekt_home.htm (in German, to be offered also in English).

References

- Beck, K. (1993). *Dimensionen der ökonomischen Bildung. Messinstrumente und Befunde*. Nürnberg: Universität Erlangen-Nürnberg, Lehrstuhl für Wirtschaftspädagogik. Unpublished report.

- Berkowitz, M.W. (1981). A critical appraisal of the 'plus-one' convention in moral education. *Phi Delta Kappan*, March 1981, 488-489.
- Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, 88, 1-45.
- 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.
- Borman, G.D., Slavin, R.E., Cheung, A., Chamberlain, A.M., Madden, N.A., & Chambers, B. (2005). Success for All: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis*, 27, 1, 1-20.
- Campbell, D.T. (1976). *Assessing the impact of planned social change*. Occasional papers # 8. Social research and public policies: The Dartmouth/OECD Conference. Hanover, NH: Dartmouth College, The Public Affairs Center. (Source: <http://www.wmich.edu/evalctr/pubs/ops/ops08.html>)
- Carver, R.P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education*, 61(4), 287-292.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press, Second Edition.
- Colby, A. & Kohlberg, L. (1987). *The measurement of moral judgment. Volume I, Theoretical foundations and research validation*. Volume I: Theoretical Foundations and Research Vali-

dition;¹⁰ Volume II Standard Issues Scoring Manual.¹¹ New York: Columbia University Press.

Darwin, Ch. (1989). The descent of man, and selection in relation to sex. *The works of Charles*

Darwin, edited by Paul H. Barrett & R. B. Freeman. London: Pickering (Original 1871).

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

Habermas, J. (1995). *Moral consciousness and communicative action*. Cambridge, MA: The MIT Press. (Originally published 1983)

Heidbrink, H. (1985). Moral judgment competence and political learning. 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*, pp. 259-271. Chicago: Precedent Publishing Inc.

Hernandez, C. & Medina, G., 2005. *Ethics and professionalism in medicine: Cross-curricular integration of ethical basis for medicine students*. Presentation at the AME meeting in Cambridge, MA, Nov. 2005.

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.

¹⁰ Anat Abrahami, John Gibbs, Ann Higgins, Kelsey Kauffman, Marcus Lieberman, Mordechai Nisan, Joseph Reimer, Dawn Schrader, John Snarney, and Mark Tappan.

¹¹ In collaboration with Betsy Speicher, Alexandra Hower, Daniel Candee, John Gibbs, and Clark Power.

- Kohlberg, L. (1964). Development of moral character and moral ideology. In: M.L. Hoffman & L.W. Hoffman, eds., *Review of Child Development Research*, Vol. I, pp. 381-431. New York: Russel Sage Foundation.
- Kohlberg, L. (1984). *Essays on moral development, Vol. II, The psychology of moral development*. San Francisco, CA: Harper & Row.
- Kohlberg, L. & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development*, 12, 93-120.
- Kohlberg, L. & Candee, D. (1984). The relationship of moral judgment to moral action. In: L. Kohlberg, *Essays on moral development, Vol. II, The psychology of moral development*, pp. 498-582. San Francisco: Harper & Row.
- Koszinoffski, R. (2006). *Überprüfung der pädagogisch-didaktischen Lehrkompetenz von Lehrpersonen bezüglich der Konstanzer Methode der Dilemma-Diskussion*. [Assessment of the pedagogical-didactic teaching competence of teachers in regard to the Konstanz Method of Dilemma Discussion.] Diploma (master) thesis, Department of Psychology, University of Konstanz: <http://www.ub.uni-konstanz.de/kops/volltexte/2007/2270/>
- Kromrey, J.D. (1993). Ethics and data analysis. *Educational Researcher*, 22, 24-27.
- Leming, J.S. (1981). Curricular effectiveness in moral/values education: A review of research. *Journal of Moral Education*, 10(3), 147-164.
- Lerkiatbundit, S., Utaipan, P., Laohawiriyanon, C., & Teo, A., 2006. Randomized controlled study of the impact of the Konstanz method of dilemma discussion on moral judgement. *Journal of Allied Health*, 35(2), 101-108.
- Lind, G. (1985 a). The theory of moral-cognitive judgment: A socio-psychological assessment. In G. Lind, H.A. Hartmann & R. Wakenhut, eds., *Moral development and the social environ-*

- ment. Studies in the philosophy and psychology of moral judgment and education*, pp. 21-53.
Chicago: Precedents Publishing Inc.
- Lind, G. (1985 b). Growth and regression in cognitive-moral development. In C. Harding, Ed.,
Moral Dilemmas. Chicago: Precedent Publishing Inc., pp. 99-114.
- Lind, G. (1987). Moral competence and education in democratic society. In G. Zecha & P. Weingartner, Eds., *Conscience: An interdisciplinary approach*. Dordrecht: Reidel, pp. 91-122.
- Lind, G. (1989). Measuring moral judgment: A review of 'The Measurement of Moral Judgment' by Anne Colby et al. *Human Development*, 32, 388-397.
- Lind, G. (2000). Moral regression in medical students and their learning environment. *Revista Brasileira de Educacao Médica*, 24(3), 24-33.
- Lind, G. (2002). *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung* [Can Morality be Taught?...]. Berlin: Logos.
- Lind, G. (2003). *Moral ist lehrbar. Handbuch zur Theorie und Praxis der moralischen und demokratischen Bildung* [Morality can be Taught. Handbook of theory and practice of moral and democratic education]. München: Oldenbourg.
- Lind, G. (2004). The meaning and measurement of moral judgment competence – A dual aspect theory. To be published in D. Fasko & Willis, eds., *Contemporary Philosophical and Psychological Perspectives on Moral Development and Education*. Creskill, NJ: Hampton Press, in press.
- Lind, G. (2005). *30 years of MJT-research: Strong support for the cognitive-developmental theory of moral development and education*. Presentation at the Conference of the Association for Moral Education (AME), Nov. 3.-5. 11. 2005, Cambridge, MA.

- Lind, G., 2006. Effective moral education: The Konstanz Method of Dilemma Discussion. *Hellenic Journal of Psychology*, 3, 189-196.
- Lind, G. & Althof, W. (1992). Does the Just Community experience make a difference? Evaluation and evidence of the effect size of the German program. *Moral Education Forum*, pp. 19-28.
- Lipsey, M.W. & Wilson, D. B. (1993). The efficacy of psychological, educational and behavioral treatment. Confirmation from meta-analysis. *American Psychologist*, 48, 1181-09.
- Lockwood, A.L. (1978). The effects of value clarification and moral development curricula on school-age subjects: A critical review of recent research. *Review of Educational Research*, 48, 325-364.
- Mansbart, F.-J. (2001). *Motivationaler Einfluss der moralischen Urteilsfähigkeit auf die Bildung von Vorsätzen*. (Motivational influence of moral judgment competence on the formation of intentions.) Unpublished master thesis. University of Konstanz, Germany, Dept of Psychology.
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft Psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Oser, F. & Althof, W. (1992). *Moralische Selbstbestimmung. Modelle der Entwicklung und Erziehung im Wertebereich*. Stuttgart: Klett-Cotta.
- Oser, F. & Patry, J.-L. (1986). Interventionsstudien für sozial-kognitive Kompetenz: Beispiele und theoretische Überlegungen zur Transformations-Validität. [Intervention studies for socio-cognitive competencies: Examples and theoretical considerations on 'transformation validity.'] *Unterrichtswissenschaft*, 14, 254-268.

- Penn, W. (1990). Teaching ethics – A direct approach. *Journal of Moral Education*, 19(2), 124-138.
- Piaget, J. (1976). The affective unconscious and the cognitive unconscious. In: B. Inhelder & H.H. Chipman, eds., *Piaget and his school*, pp. 63-71. New York: Springer.
- Popper, K.R. (1968). *Conjectures and Refutations*. New York: Harper & Row.
- Power, C., Higgins, A., Kohlberg, L. (1989) *Lawrence Kohlberg's Approach to Moral Education*. New York: Columbia University Press.
- Rosenthal, R. (1994). Parametric measures of effect size. F. Cooper & L.V. Hedges, eds., *The handbook of research synthesis*, pp. 232-240. New York: Russell Sage Foundation.
- Schillinger, M. (2006). *Learning environment and moral development: How university education fosters moral judgment competence in brazil and two German-speaking countries*. Aachen: Shaker Verlag.
- 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.
- Sprinthall, N.A., Sprinthall, R.C. & Oja, S.N. (1994). *Educational Psychology. A Developmental Approach. Sixth Edition*. New York: McGraw-Hill.
- Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: three suggested reforms. *Educational Researcher*, 25(2), 26-30.
- Walker, L.J. (1983). Sources of cognitive conflict for stage transition in moral development. *Developmental Psychology*, 19, 103-110.
- Wasel, W. (1994). *Simulation moralischer Urteilsfähigkeit. Moralentwicklung: eine kognitiv-strukturelle Veränderung oder ein affektives Phänomen?* [Simulation of moral judgment

competence. Moral development: a cognitive-structural or an affective phenomenon?]

Unpublished Master thesis, Department of Psychology, University of Konstanz.

WWW – What Works Clearinghouse (2007). Web-site of the *What Works Clearinghouse* of the

U.S. Department of Education. [Http://www.whatworks.ed.gov/](http://www.whatworks.ed.gov/) (Access: April 15, 2007).

Yerkes, R.M. & Dodson, J.D. (1908) The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.

Zierer, K. (2006). *Können Kinder Moral lernen? Eine Evaluationsstudie zur Moralerziehung in der Grundschule*. Baltmannsweiler Schneider Verlag Hohengehren (Can children learn morality? Results from an evaluation study in grade school.) Quotes in the text are taken from Zierer's (2005) original research report.

Figures

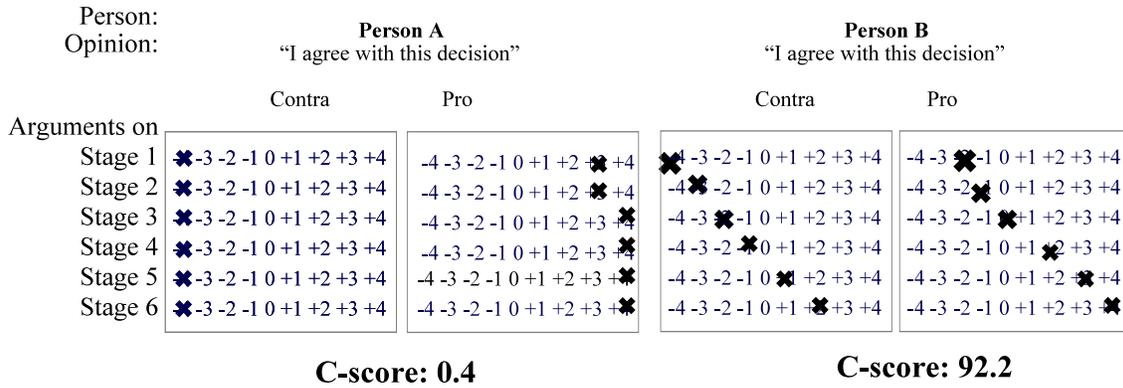


Figure 1 Two fictitious response pattern to one of the dilemma of the MJT illustrating low and high moral judgment competence (C-scores).

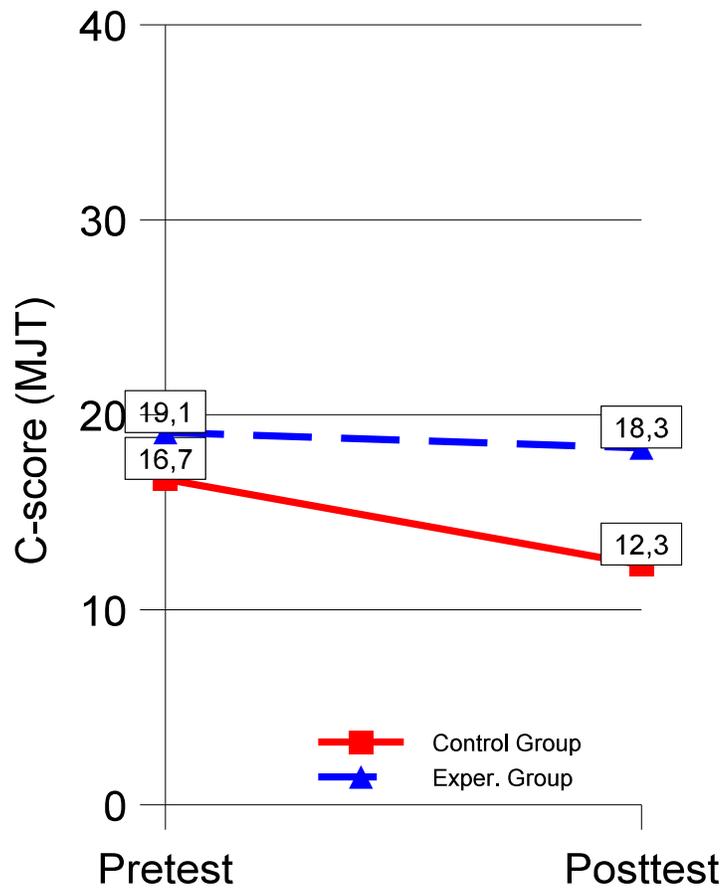


Figure 2 Impact of a Just Community program on juvenile delinquents' moral judgment competence: comparison of an experimental (n = 26) and a control group (n = 13). Source: Glasstetter, 2005, p. 194 ff.

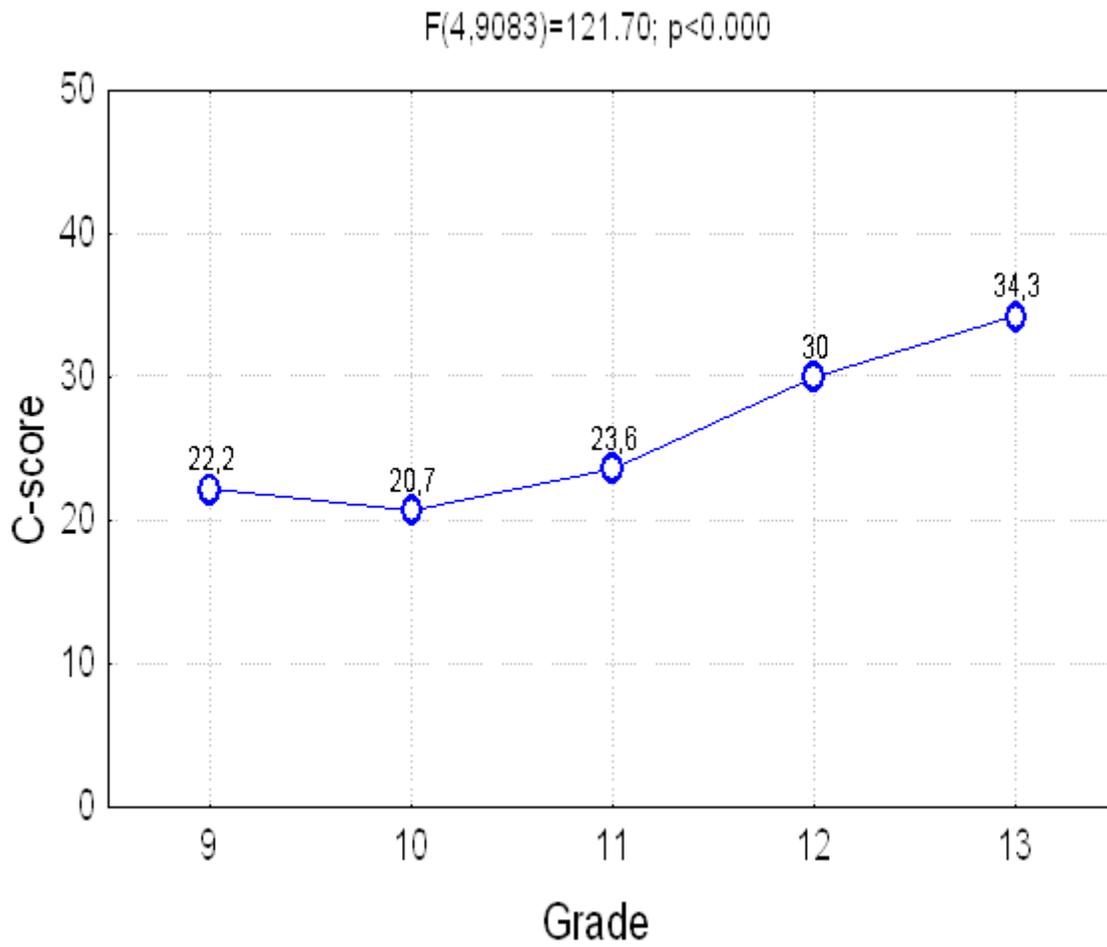


Figure 4 Moral Judgment Competence (C-score) of vocational school students grade 9 to 13. Data source: Klaus Beck (1993), vocational school study of 1993, n = 9084. Scoring and data analysis by GL.

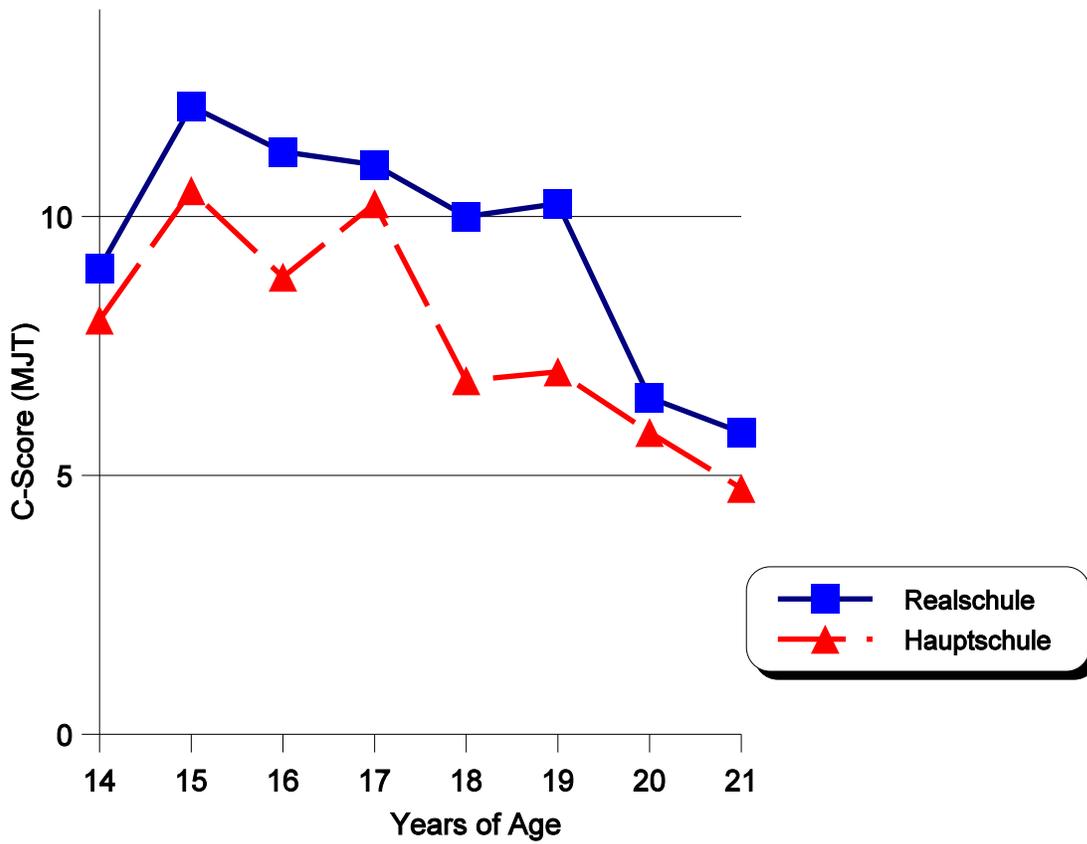


Figure 5 Moral judgment competence (MJT’s C-Score) of graduates from German “Hauptschule” and “Realschule,” two lower track secondary school-systems. While in the last year of schooling there is still some increase, the C-scores drop after leaving school. Data source: EMNID study of 1991 analyzed by Lind, 2002.

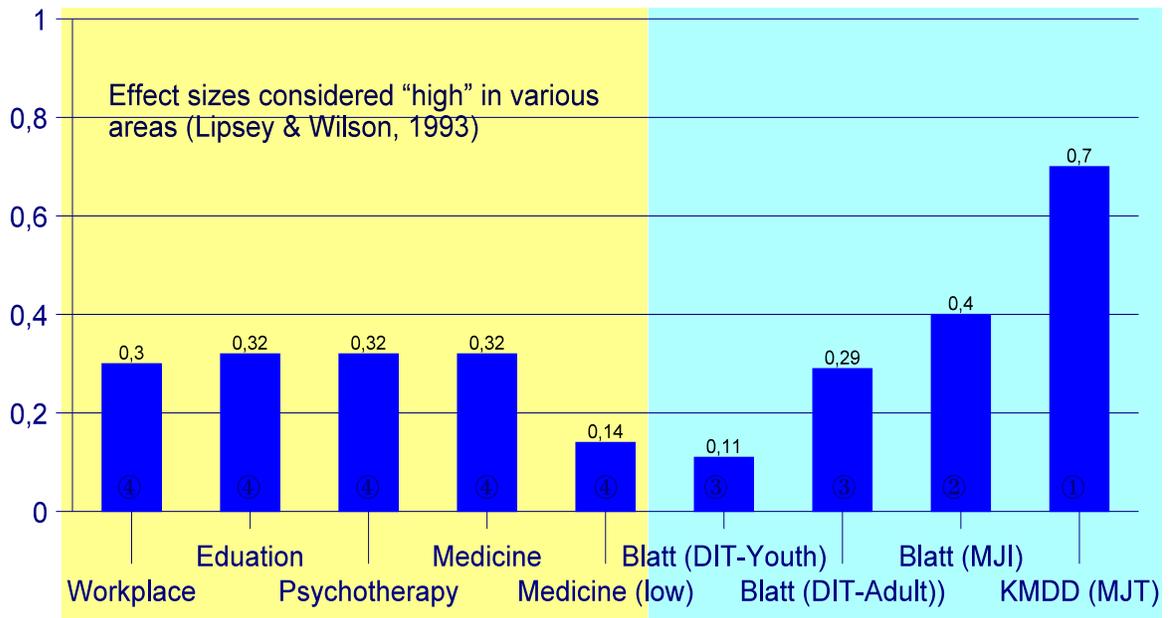


Figure 6 Relative effect sizes in various fields of intervention.

Notes

- ① The effects of the Konstanz-method on moral judgment competence (assessed with *the Moral Judgment Test*, MJT) is based on unpublished analyses of pretest-posttest intervention study by Lind; N = 42, and the study by Lerkiatbundit et al. (2006).
- ② The effects of the Blatt-Kohlberg dilemma discussion method on stage of moral reasoning (Moral Judgment Interview, MJI): Meta-analysis of 141 intervention studies by Lind (2002).
- ③ The effects of the dilemma discussion method (Blatt-Kohlberg) on moral preferences (Defining-Issues- Test, DIT) estimated by Schläfli et al. (1985). “DIT-Youth” depicts the effect sizes in groups of adolescents below the age of 20, “DIT-Adult” for older subjects.
- ④ The effects of medical, psychological, educational & vocational interventions (Lipsey, M.W. & Wilson, D. B., 1993).

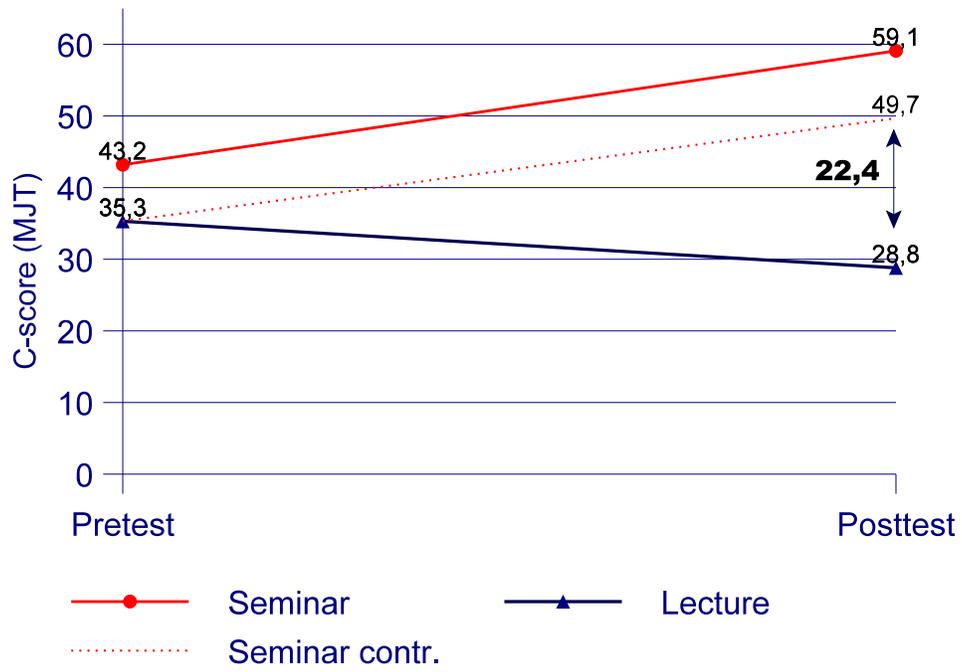


Figure 7 Absolute effect size = 22,4 C-points of the KMDD (as part of a seminar, n= 19 psychology students) in comparison to regular lecture without KMDD (teacher students, n = 48).

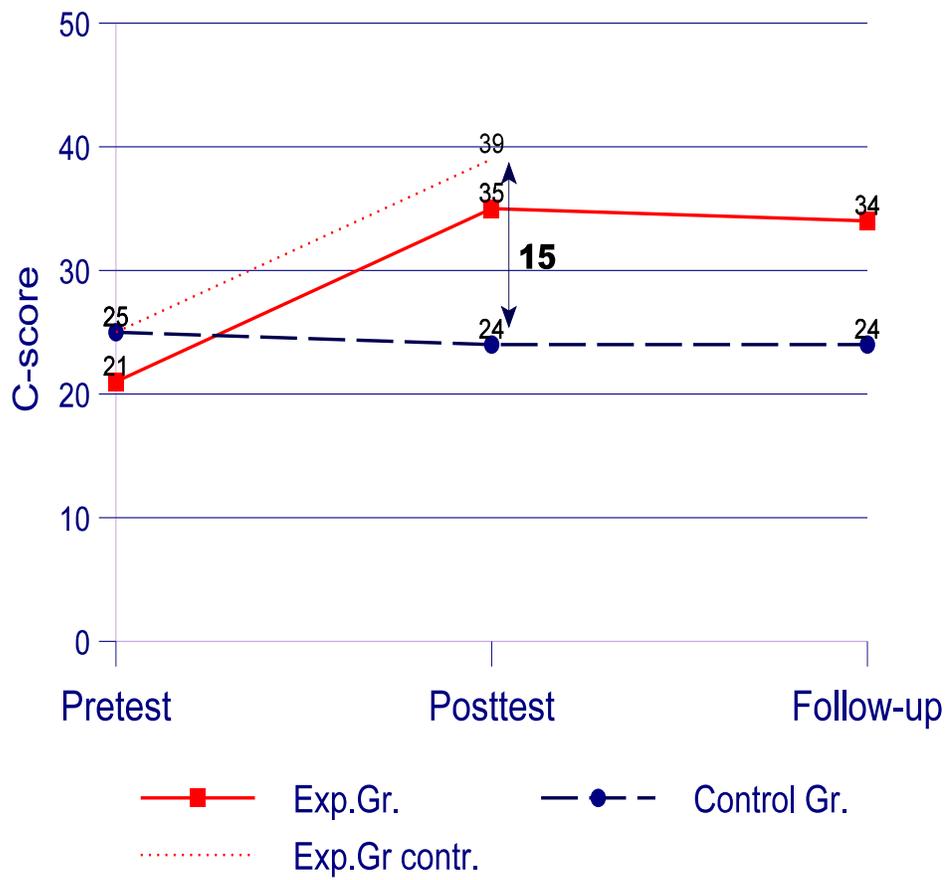


Figure 8 Randomized experimental interventions study by Lerkiatbundit et al. (2006). The intervention consisted of six dilemma discussions, one every week, using the KMDD. The participants were 38 pharmacy and 45 dental nurse students. Relative effect size $r = 0.70$, absolute effect size = 15 C-points.

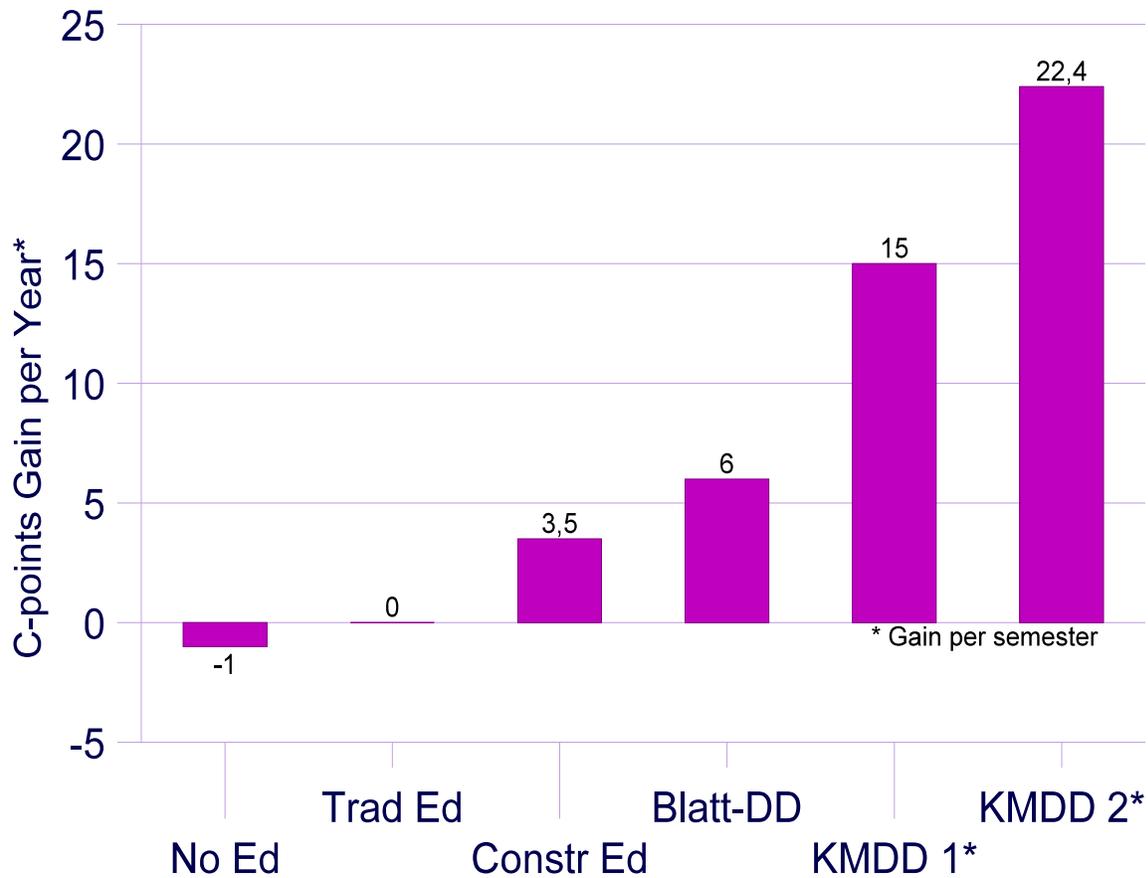


Figure 9 Absolute effect size of the KMDD in comparison to the Blatt-Kohlberg method and regular schooling.

Notes

“No Ed” = No education in late adolescence because of early graduation from lower track schools (after 9th grade).

“Trad ed” = Traditional lecture-style teaching, favoring rote learning.

“Constr ed” = Constructivist type of “teaching,” stimulation of learning and understandig.

“Blatt-DD” = Blatt-Kohlberg method of dilemma discussion (cf. Lind, 2002).

“KMDD-1” = Seminar in 2004, constructivist teaching and one dilemma discussion (KMDD).

“KMDD-2” = Seminar in 2006-05, constructivist teaching based on the KMDD principles.