

**Favorable Learning Environments for Moral Development –
A Multiple Intervention Study With 3.000 Students in a Higher Education Context.¹**

**Georg Lind²
University of Konstanz**

Introduction

"If other sources of ethical values have declined in influence, educators have a responsibility to contribute in any way they can to the moral development of their students," writes Derek Bok, the former president of Harvard (Bok, 1976, p. 26). Most educational policy-makers and the public seem to agree that schools and colleges should promote moral development in their students. Yet, there is still a debate whether it can be taught. A quarter of a century ago, the educational researcher Theodore Newcomb had been asked: "What does college do for a person?" His answer was clear and simple: "Frankly, very little that is demonstrable" (Newcomb, 1974). The sociologist Martin Trow even believes that mass (higher) education cannot achieve this aim at all. "I believe that the development of the capacities to make moral judgements is a characteristic of higher education as we would like it to be. There is no doubt that some of the characteristics of mass higher education, especially the impersonal processing of large numbers of students through institutions where they have little close or sustained relation to any teacher, do not aid the growth of their moral capacity." (1976, p. 25)

Even Kohlbergians and neo-Kohlbergians believe that moral development does not depend on education but on chronological age: "As one would expect of a developmental variable, our data show a clear relationship between age and moral judgment stage" (Colby, Kohlberg et al., 1984). In the 1980s in their longitudinal study, then extending into the college years, Kohlberg and his associates found that students' moral reasoning in an interview situation increased considerably during their study. However, the authors believed that age, not education, was the driving force behind this: "As one would expect of a developmental variable, our data show a clear relationship between

¹ Paper presented at the annual conference of AERA in San Diego, April 13-17, 2009.

² Contact: Prof. Dr. G. Lind, Dept of Psychology, University, 78457 Konstanz, Germany. E-mail: Georg.Lind@uni-konstanz.de, web: <http://www.uni-konstanz.de/ag-moral/>

age and moral judgment stage. The correlation between age and MMS was .78." (Colby & Kohlberg, 2004, p. 47): Similarly, Rest (1986) assumed: "While I regard the demonstration of age trends in a moral judgment measure as crucial, I do not regard it as sufficient." (p. 143)

However, these authors also favor special intervention programs to enhance the ability of schools and colleges to foster moral development and democratic participation (Blatt & Kohlberg, 1975; Colby et al., 2007; Lind, 2008a). Studies also clearly show that moral development correlates more strongly with the level of education than with chronological age (Rest, 1988; Rest et al., 1999; Pascarella, 1997; Lind, 2002).

It has even been shown that moral development is not invariably moving from low to high but can regress. So far, regression has been found in low track secondary school graduates, whose moral judgment competence erodes after leaving school at age 15 (Lind, 2002), and in medical students (see Helkama et al., 2003, and Lind, 2000b). "The experience of medical education appears to inhibit the increase of moral reasoning of medical students that otherwise would be expected of young adults of that age and education level." (Self et al., 1996, p. 444). Medical education does not only seem to inhibit moral development but to cause moral regression, as Schillinger (2006) found in her cross-national study of students in the fields of medicine, psychology and business administration in Brazil, Germany and Switzerland. Hence *good* education, not age is the driving force for moral development.

Good education means, as Kohlberg (1984) and Sprinthall et al. (1993) maintain, a learning environment which provides opportunities for responsibility-taking (not role-playing) and guided reflection (Lind, 2000a; Nowak & Lind, in press). In her comparative study of university students in Brazil, Germany, and Switzerland, Schillinger (2006) shows that, indeed, the regression of moral judgment competence is due to a lack of opportunities for responsibility-taking and guided reflection. Medical, psychology and business students who report to have at least some such opportunities show moral progression. In her study of Romanian university students, Iuliana Lupu (in prep.) provides also evidence that these opportunities are causally linked to students' moral development.

The confusion regarding the role of education in moral development stems, I believe, from three sources: First from a lack of conceptual clarity, second from a lack of studies on the competence aspect of moral judgment behavior, and third from a lack of large scale intervention studies that can show what is possible in education:

- Although eminent scholars clearly distinguish between two *aspects* of moral behavior, namely moral ideals and preferences on the one hand and moral judgment competence on the other, both aspects of moral behavior are mostly confounded in research and measurement of moral development, or they are misunderstood as separate *components* of behavior, which can (or must) be measured by different instruments. In his dialogue with Meno, Socrates already insisted that we must distinguish between the desire to be virtuous and the ability to act accordingly. Piaget (1976) speaks about “the affective unconscious and the cognitive unconscious” as the two basic aspects of peoples’ functioning. Kohlberg (1964) also speaks of moral attitudes on one side and *moral judgment competence* on the other. The latter he defined 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" (p. 425). Lind (1982) made this definition fundamental of a new measurement methodology, the Moral Judgment Test (MJT), which allows us to assess the two aspects distinctly (producing logically independent indices for each) yet not as separate components (Lind, 2008b).
- Most studies focus on moral attitudes (e.g., the preference for principled moral reasoning) rather than on moral competencies (Rest, 1988; Pascarella, 1997). Hence these studies do not tell us whether education can foster moral *competencies*. Do schools and colleges merely change students’ moral attitudes, or do they actually help the students to develop moral-cognitive structures? Studies using tests of moral judgment competence which cannot be faked upward, like the Moral Judgment Test (Lind, 2002), are still rare.
- Studies of the correlation between existing programs of education and moral development do not tell us whether education is possible or not to foster moral development. They just tell us what can be achieved by an education that pays only lip-service to moral education but pays little, if any, attention to this when it comes down to instructional design, curriculum development, and teacher training. To clarify this question, we need large-scale intervention studies.

In this paper I will deal especially with higher education’s ability to foster moral judgment competence in its students. I will try to answer some of these questions by repor-

ting about the findings of a rather large longitudinal study, which involved 43 interventions in twelve cohorts of university students, comprising more than 3,000 participants.³

My paper will have two sections. The first section concerns the *affective* aspect of moral judgment behavior: Do university students change from preference for non-principled, “low-stage” moral reasoning to preference for principled moral reasoning? Does the preference for principled moral reasoning increase with age or education? The second section concerns the *cognitive* aspect of moral judgement behavior, namely moral judgment competence (or moral competence): Does moral competence increase with age or is it rather a function of (good) education? Can courses be value-added with special methods so that they can better stimulate moral competencies? (Lind, 2008a) Can modern principles of effective education like a democratic learning community, constructivist strategies of teaching, and cycles of challenge and support (to enhance affect control) be used for redesigning traditional college teaching? Does it have any effects and, if so, how big are they?

Hypotheses

The hypotheses for this study have been deduced from Lind’s *dual-aspect theory* of moral judgment behavior and from his *education theory* of moral development (cf. Lind, 2002; 2008a; 2008b; Nowak & Lind, in press):

- Not age but education drives moral development.
- The preference for principled moral reasoning is already very strong when students enter college. Basic moral orientations seem to be acquired at an early age, or may be even inborn (Waal, 2006). Hence, they cannot be fostered in college any further and do not need to be fostered.
- What needs to be fostered, however, is moral judgment competence, the ability to apply moral principles in every-day life (Kohlberg, 1964). We hypothesize also that education has little fostering effect because it does not exhaust the possibilities of good teaching. College teaching often seems to lack three essentials of effective and responsible teaching: a moral atmosphere of free discourse and mutual respect, opportunities for responsibility-taking and guidance for reflection and affect control (Perry, 1970; Lind, 200a; 200b; Sprinthall, Reiman & Thies-Sprinthall, 1993).

³ More studies could be added since the first draft of this paper.

- Traditional ways of teaching (lectures, seminars) can be redesigned to foster effectively moral judgment competence, not by direct teaching of psychological and philosophical theories but by practicing the three essential principles of effective moral teaching (see above, and also Lind, 2008a; in prep.).
- By using a dilemma-discussion in the way suggested by the KMDD, traditional methods of college teaching can be redesigned to enhance moral-cognitive development, even if only applied once during a term. So far, dilemma discussions have shown to be effective only if they are administered several times in a row (Blatt & Kohlberg, 1975; Schläfli et al., 1985; Lind, 2002). Yet, as the training of KMDD has become more professionalized, we expect even one single dilemma discussion to be just as effective.
- KMDD method of dilemma discussion was already optimized through continuous self-evaluation with pretest-posttest assessments⁴ in the time before this multiple interventions study began. Hence we do not expect much increase of effectiveness over the years. By looking at the variation over the years, however, we can test the robustness of the effects of these new teaching methods.

Methods

This study involved 3,102 student participants from one German university, who were enrolled in psychology or teacher education programs. About one third was male, two thirds female. All participants were interviewed online before and after each course. The courses were mostly two-hours courses over one semester; some were compact courses of one week during vacation. All courses were taught by the same teacher. So there is no variation due to teacher characteristics. The participation in these interviews was mandatory as they were part of the teacher's quality management of his courses. Return rates were mostly above 90 percent. Participants who dropped out before the end of a course, or whose data from pretest and posttest could not be matched, were dropped from longitudinal analysis. This accounts for varying degrees of freedom in analysis of variance.

⁴ For this we have developed the internet-based program ITSE (Improvement of Teaching through Self-Evaluation). See: http://www.uni-konstanz.de/itse-projekt/itse_home.htm .

The experimental design of this study allows for testing the above hypotheses in a good, if not in the best possible, way:

- The very large experimental sample (N = 500) and control sample (N = 2,602) allows us to generalize the findings and to break down the sample into subsamples, for more complex analyses.
- Teacher quality was held high and constant as all interventions were done by the author, who has more than 20 years of training and experience in the KMDD.
- In pretest and posttest an identical test was used for measurement (see below), so that there is no measurement error and validity problem due to parallel test forms. The repeated use of the same test does not seem to inflate scores but to deflate them because of test-taking weariness. That problem could be solved though an improved test instruction.
- Though participants were not randomly assigned by the experimenter, self-assignment and administrative assignments did not lead to any distortions of the input. The degree of moral judgment competence was nearly the same in the experimental and the control groups (see Figures 5 and 8). In this way we could avoid the many unsolved problems of randomly assigned participants. Random assignment severely impedes the ecological validity of intervention studies because assignment by the researcher instead of a natural agent undermines study motivation and makes students feel like human guinea pigs. Random assignment does not guarantee unambiguous causal attribution; the study results may be explained by side-conditions which cannot be assigned at random; cross-over effects can counter-act random assignment: a teacher can hardly refrain from applying the effective methods that he or she has learned through the experiment to the control group. Last but not least, in the social sciences random experiments have not shown any superiority over non-random experiments. In their review of hundreds of meta-analyses in the social and health sciences, Lipsey and Wilson (1993) conclude that non-random study designs affect study outcomes only marginally. If an effect is substantial, it shows in random and in non-random experiments.
- The intervention studies were completely anonymous. Hence the participants had no incentive to fake their scores, as far as this is possible with an ability test like the MJT (Lind, 2002). In contrast to tests of moral preferences, the MJT does not lend itself to social desirability effects.

Measurement

All participants completed an online version of the Moral Judgment Test (MJT) by Lind (2008b) before and after each intervention. The standard MJT has been translated and validated in more than 30 languages, has been used in numerous research and educational evaluation projects, and has shown to be highly valid (Lind, 2008b). A brain study by Prehn et al. (2008) shows that moral judgment competence as measured with the MJT is strongly correlated with activities in the right dorsolateral prefrontal cortex (DLPFC).

The MJT is a theory-based instrument of experimental measurement of moral-cognitive functions. It has been constructed on the basis of Lind's (2008b) *dual-aspect-theory* of moral behavior, which is grounded on Piaget's (1976) aspect-theory, and Kohlberg's (1964) definition of moral judgment competence. The items of the MJT form a multivariate N=1 experiment (Lind, 1982). The standard MJT consists of two dilemma stories and arguments pro and contra the decision made by the protagonist in the story. Each argument is to represent one of the six basic moral orientations described by Kohlberg (1984). Thus the items form a 2 x 2 x 6 experimental design (2 dilemmas x 2 opinion-agreement x 6 moral orientations). Its items were constructed to represent the six types of moral reasoning described by Kohlberg (1984). The items' validity was rated by experts and was improved to increase validity only. There was no item "analysis" or item selection in order to maximize correlation with age, and to maximize internal consistency. Neither are there safeguards in the tests to detect and eliminate participants with low response consistency. In contrast to psychometric tests, functional measurement depends on individual response consistency as an important source of measurement.⁵ Response consistency information is used for indexing moral-cognitive functioning.

With the MJT, the affective aspect is indexed by summated ratings of the arguments for each of the six stages of moral orientations. The cognitive aspect, moral judgment competence, is indexed by the so-called C-score (C for competence), which ranges from 0 to 100 points. This index reflects the degree to which a participant rates given moral

⁵ Classical test theory and item response theory are not suited for constructing measures for testing psychological functioning because of their (often hidden) assumptions. They postulate that response consistency is only a property of the test ("reliability," "measurement error") not of the respondent and his or her psychological functioning. Moreover, the empirical item selection suggested by both test theories prevents the falsification of these theories. If data contradict the theory the data are eliminated, not the theory.

arguments (pro and contra a certain dilemma solution) in regard to their moral quality instead of other properties like opinion-agreement. It is calculated using intra-individual analysis of variance. The MJT is administered twice, as pretest and posttest. There is no sign of learning effects of the repeated administration of the test if the test is administered a couple of weeks apart. Because there is nothing at stake, the participants have no incentive to rehearse their answers. Score-lowering effects due to test weariness are eliminated through proper instruction.

Independent variables

The variable ‘term’ designates the term or semester between the years 2003 and 2009 when the interventions took place. In Germany, there is a winter term from mid October to mid February, and a summer term from mid April to mid July. The variable ‘course’ designated a regular teaching module like a lecture, seminar, or workshop of about 35 hours or a 2-hour course over one semester. It also includes some compact courses of 5 days or, in one instance, of 3 days. All courses have been taught by the author. So teachers’ characteristics were eliminated as source of variance. Most courses were part of the psychology and the teacher education program of one German university; this university is highly competitive and has been recognized as one of five ‘excellent’ universities in Germany. Two courses were taught to education officers of the German Armed Forces and one course at a Mexican university. To some of the teacher education courses, external students with a professional background were admitted as guest students.

“KMMD” is an abbreviation for the *Konstanz Method of Dilemma Discussion* by Lind (2008a; in prep.). The KMDD is a two-hour intervention with a moral dilemma discussion. In this study, the KMDD method of dilemma discussion has been used in some of the courses (seminars and lectures) one or two times. The method originates from the Blatt-Kohlberg-method (Blatt & Kohlberg, 1975) but has been considerably modified over the past two decades to enhance its efficacy. It turned out that three principles of teaching were essential for this: 1. The principle of democratic learning community of teachers and learners based on mutual respect and justice; 2. the constructivist principle of learning through stimulating cognitive growth by confronting the learner with adequate tasks (rather than imposing knowledge on him or her); and 3. the principle of affect control through cycles of challenge and support.

In many of the courses, one KMDD-dilemma session was run, usually at the beginning of the course in order to demonstrate the method and in some cases just to improve the learning climate of the course.

The three KMDD-teaching principles were also used to re-design the courses and make them a more favorable learning environment. As it turned out, seminars lend themselves more easily to such redesign than lectures (at least when one deals with the German higher education context). This variable is called “course-type.” By comparing moral development scores of students who attended such KMDD-base seminars with students who attended traditional courses (lectures) we can estimate the efficacy of this didactic innovation.

To keep teacher influence constant, all courses were taught by the author of the KMDD, who has more than 20 years of experience in this method. He is also “KMDD-trainer” who trains students to become Senior KMDD-teachers (see <http://www.uni-konstanz.de/ag-moral/>). To be effective, a teacher needs to have about 120 hours of intensive training in the method (Lind, 2008a).

Analysis

If our research shall have an impact on educational practice and policy-making, and if we need to quantify our findings, we should choose quantitative indicators which are unbiased, comparable to other studies, and are also comprehensible. The best way to do this is, I believe, the calculation of absolute effect sizes as it is done in the natural sciences. In the applied natural sciences we accomplish measurement by reading a scale like meter or centigrade (in some locations yards and degree Fahrenheit). If you ask somebody how much warmer it is today then yesterday you will hardly get answers like “it is significantly warmer today on a $p = .05$ -level,” or “the effect size of the climate change is $d = 0.12$ or $r = 0.08$.”⁶

Statistical significance tests tell us how small a difference is given a certain sample size; it does *not* tell us how effective an intervention is. If (ab)used as an index of effect size, it tells us only how large the samples were. With statistical significance tests, any

⁶ In his article “The Earth Is Round ($p < .05$)” the eminent statistician Jacob Cohen criticizes: “After 4 decades of severe criticism, the ritual of null hypothesis significance testing – mechanical and dichotomous decisions around $\alpha = .05$ criterion – still persists ” (1994, p. 997). The former president of APA, Paul Meehl holds statistical significance testing responsible for the “slow progress of soft psychology.” (Meehl, 1978).

intervention can be made looking effective by enlarging the sample. Meanwhile, AERA and APA did not abandon significance testing, but they require authors to report relative effect size indices like r (correlation coefficient) or d (standardized mean score difference). Researchers now chose either of the two, unfortunately, often without telling us which of the two they use and why they use it. While r is the more familiar statistics for educational researchers, the d index is more often used for reporting relative effect sizes. Both give different results. While r is limited to the range between -1 and +1, d varies from 0 to ∞ (unlimited). By a simple formula they can be equated, however. To give some examples, when looking at the gains of “moral maturity” (as measured with Kohlberg’s *Moral Judgment Interview*), the dilemma discussion method by Blatt and Kohlberg (1975) has, calculated from about 141 interventions, a mean relative effect size of $r = 0.40$ in secondary schools children (Lind, 2002, p. 204). When looking at students’ preferences for principled moral reasoning (as measured with Rest’s *Defining Issues Test*), this intervention has a relative effect size of only $r = 0.11$ (Schläfli et al., 1985). When looking at students’ moral judgment *competence* (as measured with the MJT), dilemma discussions have shown much higher effect sizes, of $r = 0.50$ and higher. In their randomized intervention study using six dilemma discussion sessions, Lerkiatbundit et al., (2006) have found an effect of $r = 0.70$ (my calculation). The effect of the KMDD could be very stable. In a post-posttest after six month they found no decrease of the C-score in the experimental group.

Essentially, both statistics (r and d) are standardized means scores over the pooled standard deviation of the samples. Because they do not depend on samples sizes, they are less biased and better comparable to other studies. However, because they depend on empirical standard deviations, they are still biased and not fully comparable and are also hard to understand by practitioners. Standard deviations vary from study to study to a large degree, though this variation is hardly ever reported. This creates a substantial bias, which is mostly unintended but can also be used for “beautifying” the effects of an educational intervention. By keeping standard deviation small in selecting the experimental samples, one can make the effects look big. Moreover, most educational practitioners find it difficult to interpret such effect sizes – and keep looking out helplessly for the asterisks which allegedly signal “significance.”

Absolute effect size statistics (aES) simply compare mean scores. This makes them unbiased against sample size and standard deviations. However, in order to be comparable and meaningful, absolute effect sizes should be calculated for scales based on

objective tests that have been rigorously validated and whose properties are well known. Absolute effect sizes can be related to the amount of time and effort invested to create changes.

The *Moral Judgment Test*, MJT, provides a suitable basis for estimating absolute effect sizes (Lind, 2008b). It is objective and unbiased (no purposeful item selection was used for its construction; see above). Identical test forms are used for pretest and posttest, letting us compare scores without auxiliary assumptions. After more than 30 years of use in many studies comprising more than 300,000 participants, the MJT's properties are well known. Especially we know much about the typical effect sizes of various educational means and interventions upon students' moral judgment competence. Lind (2002) estimated that the average secondary school in Germany has an impact of about 3.5 C-points per year or about 2 C-points per semester (one teaching term at college level). In other countries similar gains per year have been found, yet in many countries and also in some university programs also losses (or regressions) of two and more points have been found (Lind, 2008b).

We calculate the "absolute effect size" (aES) of the KMDD-method by assessing the posttest-pretest gains of the C-score (MJT) during one teaching term (or an equivalent compact course) and compare it with the gains found in traditional non-KMDD teaching using this formula:

$$\text{aES} = (C_2 - C_1)_{\text{KMDD}} - (C_2 - C_1)_{\text{trad}}$$

In plain language this formula states that the "pure" effect size (or added value) of the treatment is calculated by subtracting the increase of moral judgment competence (C-score) in the traditional teaching groups $[(C_2 - C_1)_{\text{trad}}]$ from the increase of moral judgment competence (C-score) in the intervention groups $[(C_2 - C_1)_{\text{KMDD}}]$.

Because our multiple intervention studies involve more than 3.000 participants, even very small effects would become "significant," and because these statistics are biased and not comparable with other studies, we omit them from our report. It would be inappropriate to calculate these insignificant and biased statistics for conventional reasons only. The aES shows the real, unbiased effect sizes of our educational intervention and can be easily compared to any other study using the MJT as a measurement instrument. With some caution, this statistic can even be used to compare our

results with studies using other measurement instruments, if their scales are projected on a common scale from 0 to 100.

Findings

The affective aspect: Moral orientations

Moral orientations are extremely universal and stable (Rest, 1969; Lind, 2002). Most, if not all, students prefer moral reasoning of type V and VI according to Kohlberg's (1984) categories of moral reasoning.⁷ Principled moral orientations are accepted by far the most. Moral orientations are preferred the less the lower they are on Kohlberg's scale.

Because this preference hierarchy has already been found in very young children, we did not expect any correlation with age. Indeed, there is no correlation with age (see Figure 1).

There is also no evidence of a change of moral orientations due to education. When moral orientations are compared before and after a course, the hierarchy (or ranking) of types of moral orientations remain stable, although there are small shifts in the judgment behavior of students who took part in the seminars (Figure 2).

As the figure shows, students' judgment behavior is mainly impacted by the type of moral orientation that they represent. The effect size of the type of moral orientation is very high, namely $r = 0.62$. Seminars and workshops using the KMDD-didactic have a small impact ($r = 0.19$) on moral orientations. However, there is no inversion of the rank order or preference hierarchy of the six types but a stronger accentuation (Figure 3). This accentuation hints to gains in moral judgment competence as both aspects do correlate highly with one another when nothing is at stake (Lind, 2002).

⁷ Even though Kohlberg's "Stage-Theory" has not been supported by empirical research (Rest, 1986; Lind, 2002) his Stages should not be used to categorize people, but they are very helpful to categorize types of reasoning (Lind, 2008b).

The cognitive aspect: Moral judgment competence

Biological age is not correlated with moral judgment competence (Figure 4). The correlation is slightly negative. This confirms findings from other studies that show that at least in adulthood, age can be negatively associated with moral development (Schillinger, 2006). The high positive correlation of age with moral development in adolescence (Colby et al., 1984, p. 47) may be actually due to education, whose impact on moral development is confounded with age (Lind, 2002). So we can no longer assert that moral competencies develop in an invariant sequence.

Although in the past several studies of adolescent students report moderate to strong correlations of level of education with moral development (Kohlberg, 1986; Rest, 1986; Rest et al., 1999; Lind, 2002), we found that higher education nowadays seems to have a smaller impact on moral judgment competence. In traditional lectures moral judgment competence increases only 1 C-point per semester (lower, solid line in Figure 5). While a few lectures produced small increases of moral judgment competence, others showed no effect or a negative effect (Figure 6). The average increase is lower than the approximately 3.5 points found in secondary school students in Germany (Lind, 2002). The 2008 lecture sticks out with its absolute effect size of $aES = 6.0$ ($N = 54$). It was a “compact lecture,” that is it was a five-day course from 8.30 a.m. to 3 p.m., which consisted of three parts: a demonstration, a traditional lecture and a small-group activity each day. In spite of its high attendance rate of about 100 students (of which 54 completed both surveys), it was the first lecture which embraced some elements of KMDD-didactics like the seminars. This finding shows that even lectures may be successfully revised in order to improve of the learning environment. In order to present the lecture in a compact form, it had been re-designed fundamentally: Each day, about one third of the time was reserved for work in small groups of four students; each day these groups had to apply what they have learned in the lecture for two hours. In exchange, the amount of subject matter had to be reduced by one third. The participants were very enthusiastic about this set-up and showed, besides substantial increase in moral judgment competence, greater learning motivation and attendance rates than they did ever before.

In contrast to most traditional lectures, seminars and workshops using the KMDD or its teaching principles do have a high and consistent impact on students' moral

judgment competence (Figure 5). In each term these seminars were highly effective in stimulating moral growth. The gains were between 7.9 and 15.7 C-points (Figure 7).

In the previous analyses, the effects of dilemma discussion were confounded with the overall effects of seminar (which was also based on similar teaching principles). To be able to discern the “pure” impact of a single 2-hour dilemma discussion, we can contrast courses with and without a dilemma discussion. As Figure 8 shows, a single dilemma discussion adds 2.8 C-points increase to traditional lectures, and 3.2 C-points increase to the KMDD-based courses. Hence, both methods together (KMDD-didactic plus KMDD dilemma discussion) increases the absolute effect size of traditional college teaching of about 0.5 to 1.0 point to 13.6 C-points per term, which is about 50% of the moral growth that had taken place during about 20 years before these students entered university. Of course, such numerical calculations need to be taken with a grain of salt, showing how much we can achieve in the domain of moral education if we use effective methods.

In Germany, psychology programs are much more competitive (to get accepted almost a grand point average of A is needed) than teacher education programs. It seems that grand point average is positively correlated with moral judgment competence. As a consequence, first year psychology students have a higher moral judgment competence than teacher students. Interestingly, psychology students also gain more from their study than do teacher students (Figure 9). Because the German psychology students in Schillinger’s (2006, p. 83) study, who attended a different university, do not show any increase of moral judgment competence, one can assume that the gain found in the students of our study is due to a more favorable learning environment rather than due their greater eagerness to learn.

These differences between the two programs account for the large initial differences between student participants in regard to their moral judgment competence depicted in Figure 7. The seminars included in this intervention study attracted mostly fourth year psychology students, while most lecture participants were first year teacher students.

Summary

The findings from this probably largest intervention study in the moral domain, helps to clarify some of the questions raised in the beginning: Quality of learning environment seems to be the most important factor in moral development. In detail, the findings show:

- Higher education does not change moral orientations. Even the interventions do not cause substantial changes but accentuate a tendency which exists in the students already when they begin their higher education (or probably a long time before that). Actually students' moral orientations do not need to be changed because, as our data indicate, most if not all students bring along a strong desire to be moral. When being asked to discuss a moral dilemma, they clearly prefer principled moral reasoning over other "stages" of reasoning as the most adequate level of moral discourse.
- Age is not the driving force behind the development of moral judgment competence. We even found a small negative correlation with age, meaning that moral judgment competence can decrease with age. This competence does not decrease from cohort to cohort over the years (Figure 6), thus cohort effects cannot account for these negative age trends. Rather they could be due to the fact that the older students in this study were "late-comers," as most lectures were intended for first year students.⁸
- At average, higher education seems to have a modest impact on moral development. If the data from two programs, psychology and teacher education, can be taken as representative, higher education produces a gain of about 2.5 C-points per semester. This could mean that ten semesters of study can increase moral judgment competence by 25 points, if such an extrapolation of data would be permissible. In reality, the typical increase from first year to fourth year of study is about 4 points (Lind, 2000a; Schillinger, 2006). This is not really much when we think of the high expectations connected with higher education. Even worse, some programs like medical education produce a regression of moral competencies rather than a progression (Lind, 2000b; Helkama et al., 2003; Schillinger, 2006).

⁸ The German universities used to have only "Diplom"-programs which do not distinguish between undergraduates and graduate students. Only recently, this distinction was introduced together with the creation of BA and MA study programs.

- Even in modern mass education, moral competencies can be effectively fostered. With as little effort as one 2-hour session during a semester, using the *Konstanz Method of Dilemma Discussion* (KMDD), we can achieve a 3 to 4 points gain of moral judgment competence. When this session is embedded in a *seminar* which is taught on the basis of the same teaching principles as the KMDD, we can achieve an increase between 5 and 20 points (see Figure 5), which is a much bigger effect than that of a five-year program on college level. This is the good news. It should be noted, however, that this success requires sufficient teacher training. To obtain a Senior KMDD-teacher certificate, about 120 hours of intensive training is required (see Lind, in prep.). This may seem bad news for educational budget officers. However, it has been shown that moral judgment competence is indispensable for good citizenship and professional life (Kohlberg, 1984; Lind, 2008a). The benefits of effective moral education by far outweighs its costs (Bok, 1976; Gutmann, 1999).
- Endeavors to value-add lectures with the KMDD-teaching principles were less successful. Lectures resist the most the changes called for by the KMDD-principles. The continuous talking by the lecturer keeps students from information processing (“thinking”) and prevents them from becoming active learners. As the data from the present study and other studies (Lind, 2000a; in prep.; Schillinger, 2006) show, the effects of teaching on moral development are the stronger the less the teacher talks and the more opportunities for responsibility-taking he or she gives to the students. Yet, as with the seminars, a single dilemma discussion in a lecture course can increase students’ moral judgment competence, too.

In concluding, our findings from this and other studies suggest three major reform efforts in higher education and beyond:

First, we should supply the syllabi of all study programs with sufficient time for self-studies, reflection and taking over responsibilities in and out the university. This does not mean that students engage in jobs to make a living (Pascarella et al., 1998), but that they engage in meaningful activities in which they can try out what they have learned, make competent and responsible decisions, and discuss conflicting courses of action with others (Schillinger, 2006).

Second, the university syllabi should contain regular well-designed dilemma-discussions (Lind, 2008a). This has shown to be highly effective in stimulating not only moral

judgment competence but also an atmosphere of discourse and learning, mutual respect, and better subject learning.

Third, college courses should be redesigned to create an atmosphere of free discourse and mutual respect, to stimulate cognitive growth through challenging moral tasks, and to guide students' ability to reflect and to control their affects.

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.
- Bok, D. (1976). Can ethics be taught? *Change*, 26-30.
- Colby, A. & Kohlberg, L. (1984). Invariant sequence and internal consistency in moral judgment stages. In: W. M. Kurtines & J. L. Gewirtz, Eds., *Morality, moral behavior, and moral development* (pp. 41-51). New York: Wiley.
- Gutmann, A. (1999). *Democratic Education* (2nd edition). Princeton, NJ: Princeton University Press
- Helkama, K., Uutela, A., Pohjanheimo, E., Salminen, S., Koponen, A., & Rantanen-Vantsi, L. (2003). Moral reasoning and values in medical school: a longitudinal study in Finland. *Scandinavian Journal of Educational Research*, 47, 399-411.
- Kohlberg, L. (1984). *Essays on moral development, Vol. II, The psychology of moral development*. San Francisco: Jossey-Bass.
- Kohlberg, L., & Higgins, A. (1984). Continuities and discontinuities in childhood and adult development revisited – Again. In L. Kohlberg, Ed., *Essays on moral development, Vol. II. The psychology of moral development* (pp. 426-497). San Francisco: Harper & Row.
- 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. (1982). Experimental Questionnaires: A new approach to personality research. In: A. Kossakowski & K. Obuchowski, Eds., *Progress in psychology of personality*, pp. 132-144. Amsterdam, NL: North-Holland.

- * Lind, G. (2000a). The importance of role-taking opportunities for self-sustaining moral development. *Journal of Research in Education* 10(1), 9-15.
 - * Lind, G. (2000b). 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? Research findings from modern moral psychology]. Second Edition. Berlin: Logos-Verlag.
 - * Lind, G. (2008a). Teaching students to speak up and to listen to others: Cultivating moral democratic competencies. In: D. E. Lund & P. R. Carr, Eds., *Doing democracy and social justice in education: Political literacy for all students* (pp. 185-220). New York: Peter Lang Publishing.
 - * Lind, G. (2008b). The meaning and measurement of moral judgment competence revisited - A dual-aspect model. In D. Fasko & W. Willis, Eds., *Contemporary philosophical and psychological perspectives on moral development and education* (pp. 185 – 220). Cresskill. NJ: Hampton Press.
- Lind, G. (in prep.). *Morality can be taught. Teacher handbook for moral and democratic education using the Konstanz Method of Dilemma Discussion*.
Spanish edition: Lind, G., 2007. *La moral puede enseñarse. Manual teórico-práctico de la formación moral y democrática*. México: trillas.
- Lupu, I. (2008). *Die Bedeutung von Bildung und Religiosität fuer die Entwicklung moralischer Urteilsfähigkeit*. [The role of education and religiosity in regard to the development of moral judgment competence.] Dissertation, University of Konstanz, Department 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.
- Newcomb, Th.M. (1974). What does college do for a person? Frankly very little. *Psychology Today*, 8(4), 72-80.
- Nowak, E. & Lind, G. (in press). Mis-Educative Martial Law – The Fate of Free Discourse and the Moral Judgment Competence of Polish University Students from 1977 to 1983. In: M. Zirk-Sadowski, B. Wojciechowski, & M. Golecki, Eds., *Between Complexity and Chaos*. Torun, Poland: Adam Marszalek publisher.
- Pascarella, E. (1997). College's influence on principled moral reasoning. *Educational Record*, 47-55.

- Pascarella, E. T., Edison, M. I., Nora, A., Hagedorn, L. S., & Terenzini, P. (1998). Does work inhibit cognitive development during college? *Educational Evaluation and Policy Analysis*, 20, 75-93.
- Penn, W. (1990). Teaching ethics – A direct approach. *Journal of Moral Education*, 19, 124-138.
- Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years*. New York: Holt.
- 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.
- Prehn, K., Wartenburger, I., Mériaux, K., Scheibe, C., Goodenough, O. R., Villringer, A., van der Meer, E., & Heekeren, H. R. (2008). Influence of individual differences in moral judgment competence on neural correlates of socio-normative judgments. *Social Cognitive and Affective Neuroscience*, 3(1), 33-46.
- Rest, J. R. (1969). Level of moral development as a determinant of preference and comprehension of moral judgments made by others. *Journal of Personality*, 37(1), 220-228.
- Rest, J. R. (1986). Moral development. Advances in research and theory. New York: Praeger.
- Rest, J. R. (1988). Why does college promote development in moral judgment? *Journal of Moral Education*, 17, 183-194.
- Rest, J. R., Narvaez, D., Bebeau, M. J. & Thoma, S. J. (1999). *Postconventional moral thinking. A Neo-Kohlbergian approach*. Mahwah, NJ: Erlbaum.
- Rosenthal, R. & Rosnow, R. L. (1984). *Essentials of Behavioral Research*. New York: McGraw-Hill
- Schillinger, M. (2006). *Learning environments 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.
- Self, D. J., Baldwin, D. C., & Wolinsky, F. D. (1996). Further exploration of the relationship between medical education and moral development. *Cambridge Quarterly of Healthcare Ethics*, 1996, 444-449.

- Sprinthall, N. A., Reiman, A. J., & Thies-Sprinthall, L. (1993). Role-taking and reflection: promoting the conceptual and moral development of teachers. *Learning and Individual Differences, 5*, 283-299.
- Trow, M. (1976). Higher education and moral development. *AAUP Bulletin 4*, 227-238.
- Waal, F. de (2006). *Primates and philosophers: How morality evolved*. Princeton, NJ: Princeton University Press.

* These papers by Lind can be downloaded from the author's web-site:

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

Note

The dual-aspect theory of moral behavior and the role of moral and democratic competencies will be discussed extensively at the international conference "Can morality be taught? Is it a competence?", July 27th to 31st, 2009, in Konstanz, Germany. For information and registration see:

http://www.uni-konstanz.de/ag-moral/kurse/conference_2009.htm

Appendix: Figures

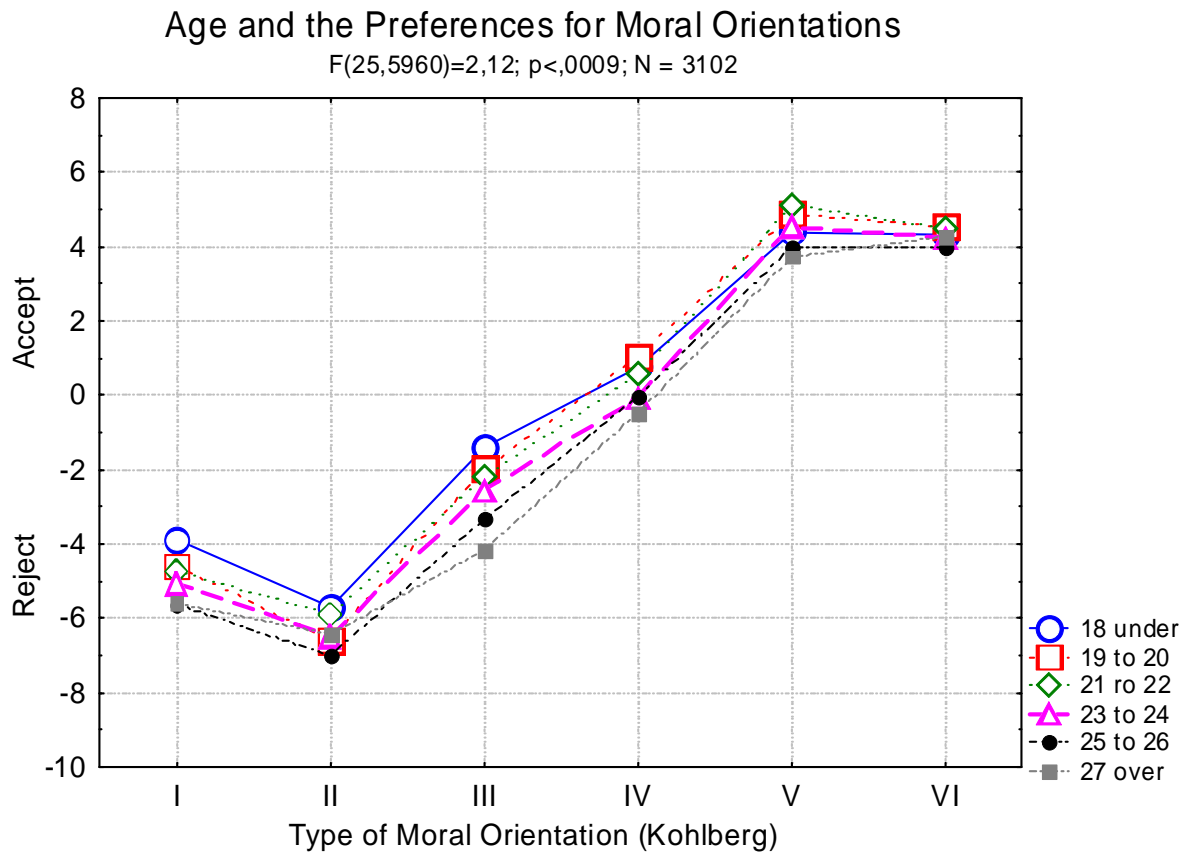


Figure 1

Impact of Education (one Term) on Preference for Moral Reasoning

$F(5,5960)=7,68; p<,0000; N = 3102$

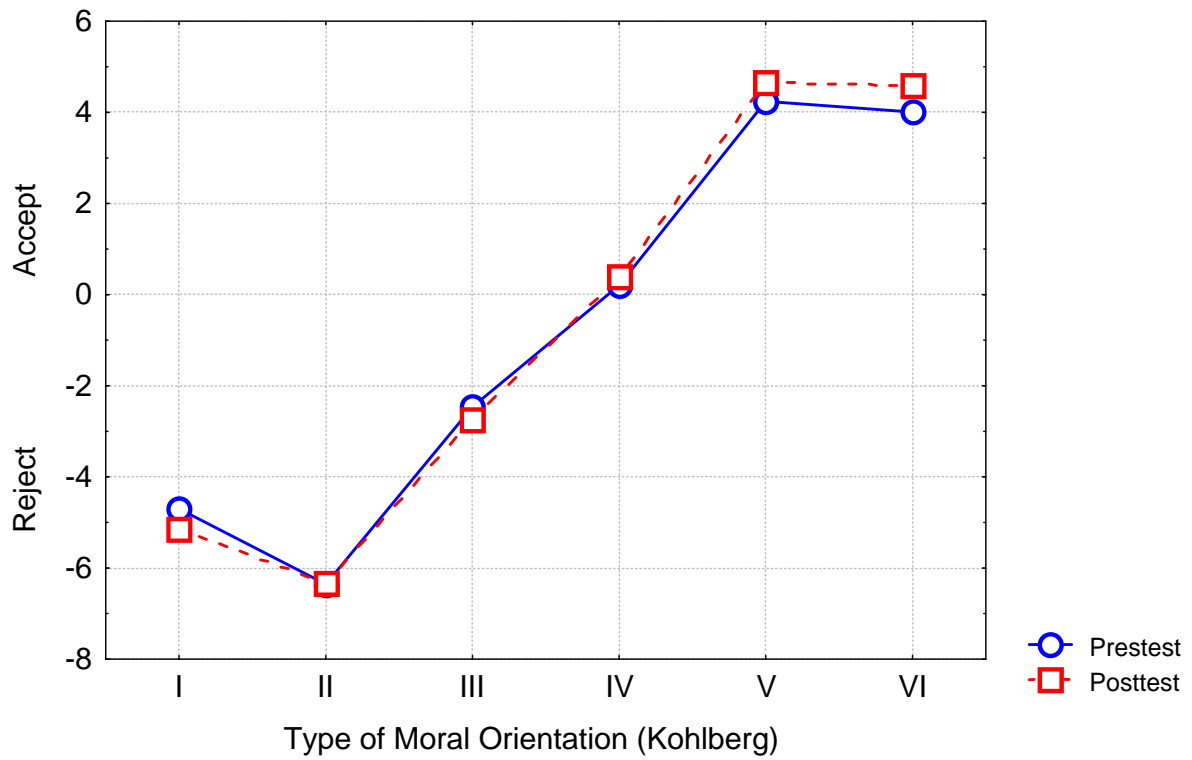


Figure 2

