Moral Regression in Medical Students and Their Learning Environment

Georg Lind

ABSTRACT
Moral Regression in Medical Students and Their Learning Environment

This study addresses the question whether moral judgment competence develops or regresses during medical education and whether such regression is related to particular features of medical education. It is hypothesized that such regression, when it occurs, is linked to deficiencies in the learning environment provided by medical training. This hypothesis is tested using data from two large scale investigations of German university students, a) in a six-year longitudinal study (N=746), and b) in two cross-sectional studies of medical and other students (total N=4,966). For measuring moral development in the longitudinal study, Lind's Moral judgment Test (MJT) was used. With the MJT, we can measure simultaneously both moral attitudes and moral competencies. The learning environment was assessed through a number of subjective and quasi-objective questions. The data show a) that medical students, like other students, prefer principled moral reasoning over other types of reasoning, b) that this preference remains stable throughout their study, c) but that, at average, medical students show a regression in their moral judgment competence (C-score) while other university students in general show a remarkable increase, and d) that this regression of moral judgment competencies is linked to deficiencies in the learning environment of medical schools. A radical change of the curriculum and organization of medical education is suggested.

RESUMO
Regressao Moral entre Estudantes de Medicina e Seu Ambiente de Aprendizagem

Este estudo aborda a questão relativa ao desenvolvimento ou regressão de competência para julgamento moral durante a educação médica, e, se essa regressão a aspectos particulares da educação médica. Hipotetizou-se que tal regressão, quando ocorre é relacionada a deficiências no ambiente de aprendizagem oferecido pelo treinamento médico. Esta hipótese foi testada utilizando dados de duas investigações em larga escala na com alunos universitários da Alemanha, a) num estudo longitudinal de seis anos (N=746), e b) em dois estudos transversais de alunos de medicina com outros alunos (total N=4,966). Para medir desenvolvimento moral no estudo longitudinal, foi utilizado o Lind Moral Judgment Test (MJT). Com o MJT, nós podemos medir simultaneamente atitude e competência moral. O ambiente de aprendizagem foi avaliado através de um número subjetivo e quasi-objetivo de questões. Os dados mostraram que a) os alunos de medicina, como outros alunos, preferem raciocínio por princípios morais comparando a outros tipos de raciocínio, b) que esta preferência manteve-se estável ao longo do estudo, c) mas que, na média, os alunos de medicina demonstram uma regressão na sua competência para o julgamento moral (C-score) enquanto que outros estudantes universitários em geral demonstraram um aumento notável, e d) que essa regressão na competência para o julgamento moral está relacionada a deficiências no ambiente de aprendizado das escolas médicas. Sugerimos uma mudança radical no currículo e organização da educação médica.
INTRODUCTION

Dr. Paula just started her internship at the department of dermatology, when she already has to make her first big decision. The head of the department asks her to go to the morgue and harvest skin from the corpse of a young person who has died in an accident. He explains that the skin was urgently needed for an emergency surgery of a patient whose skin was burned third degree, and who would not survive without a skin transplant. The hospital has run out of transplants and no other way of obtain transplants seemed feasible. Of course, Dr. Paul knew that harvesting skin from a dead person without his or her consent or the consent of relatives is illegal. Furthermore, Dr. Paula, who is a practicing Catholic, feels that this is a sin. What should she do if she wants to be moral?

She recalls her oath “to help the sick according to my ability and judgment, but never with a view to injury and wrongdoing.” Yet this does not provide sufficient guidance for solving her dilemma. Whatever she does, she does something wrong and transgresses a moral principles (though in favor of another one). If she follows her religious and legal conscience, and decides against harvesting the skin from the dead body, she will be responsible for the dying of the patient with the burning. If she obeys her boss and helps to save this patient, she will have to break the law and disregard her religious duty. Obviously, Dr. Paula has a dilemma.

Anybody who has ever come across a similar dilemma knows that a solution surely is hard to find. The process of finding a solution involves various abilities, in particular, the ability—

- to recognize one’s own complex, conflicting moral feelings,
- to submit those feelings to reflective reasoning, and
- to enter ethical discourse with friends, experts and authorities (Hinman; Lind).

We call such abilities “moral judgment competence,” a term coined by Lawrence Kohlberg to describe 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).

If Dr. Paula was lucky, medical education at university prepared her well for such a conflict situation. That is, that she did not only learn the technical skills of an dermatologist but also the moral competencies needed to solve various moral dilemmas that will inevitable arise when dealing with human life.

Most likely, however, she was little prepared for such kinds of problems of her profession. In one of my ethics class, for example, in which I used Dr. Paul’s dilemma as starting point for a dilemma discussion (Blatt & Kohlberg; Kuhmerker et al.), fourth year medical students told me that it never occurred to them that, in their profession, they would ever be confronted with such kind of problems. They believed that I made up this story for didactic reasons.

Many studies show that this impression can be generalized. Self and his associates (1994 a, b) have summarized the outcomes of several cross-sectional and longitudinal studies on the effects of medical education on moral development. Most of these studies compared first and fourth year medical students, some first and third year, and one first and fifth year. All studies were conducted in the US, one in Mexico. Moral development was observed using either Kohlberg’s Moral Judgment Interview (MJI; see Colby et al.), Rest’s Defining Issues Test (DIFF), or Gibbs’ Socio-moral Reasoning Measure (SRM). Most studies found no, or only statistically insignificant, upward changes in moral development during medical study colleagues (Sheehan et al; Self et al. 1994 a, 1994 b; 1996 b). Self et al. (1994a, b) conclude from these studies that medical education does not contribute to students’ moral development.

Medical doctors, it seems, get a highly sophisticated technical training but not the moral education needed to cope with the dilemmas that they are confronted with in their professional life. This findings is also supported by a cross-cultural study of physicians in hospitals in Austria and Saudi-Arabia by James DuBois (1997 a, b). Medical doctors in both countries have very low C-scores on the MJT, which we will discuss below in more detail.

The studies on which this conclusion is based, however, show some severe shortcomings: First, the sample sizes of most of these studies were very small so that changes in moral development could not become statistically significant. No measure of effect size are reported. Second, the scores of university students, and especially of medical students, may be so high that no further development can be detected by these scales (ceiling effect). Third, the instruments used for measuring moral development tap mostly moral attitudes, which medical education cannot, and does not want, to change, instead of moral competencies (Lind & Wakenhut; Lind). Fourth, research so far has hardly clarified the question whether the lack of moral development can be clearly attributed to deficiencies in present medical education. For example, one could argue that moral-cognitive development ceases in early childhood and that, therefore, not further increase in moral competencies can be expected (Gross, p. 10). Indeed, none of the above cited studies provide evidence that link the lack of moral development directly to the learning environment in medical schools.

To fill this void and clarify some of these questions, I will present four analyses:

(1) A re-analyze of the studies reported by Self et al. (1994a, b; 1994b) to show that medical students’ gains on mixed moral judgment scales are indeed retarded and that this retardation is unique for medical education.
(2) An analyses of the theoretical validity and cross-cultural validity of the MJT to support our claim that this test provides pure measures of the competence and the attitudinal aspect of moral judgment irrespective of cultural context. For this we will compare validation data from German and Brazilian samples.

(3) An analysis of data from our longitudinal study of medical students' moral judgment competence in Germany (Bargel et al.; Framhain & Langer).

(4) And findings from a large representative study into the learning environments provided by medical education as perceived by students by Bargel & Ramm.

FOUR ANALYSES

Analysis I: Does moral development of medical students stagnate and can this be attributed to medical education?

Recent studies on the relationship between moral development and medical education have reported "non-significant" changes in moral development scores over a three to five year time span (Self et al.; Framhain & Langer). As mentioned above, this finding could be due to the fact that most studies used small samples, which make it hard for changes to become statistically significant, or that in this age no progress of moral development can be expected anymore.

To clarify these two objections, we can estimate the absolute changes of moral development reported in those studies and compare them with changes found in other studies over a similar period of time. Because different instruments have been used in these studies, their findings cannot be compared directly. They need to be standardized on a common scale to make them comparable. The maximum range of WAS scores (MF) is from 100 to 500, of P-scores (DIT) from 0 to 100, and of SRM-scores from 100 to 400. We chose to transform all values to range from 0 to 100. By subtracting pretest scores from posttest scores, we can obtain change scores. The change score is zero when no change is reported. We treat cross-sectional and longitudinal studies alike.

Findings: In the studies reported by Self et al., the change scores over three to five years of medical education were equal to, or smaller than, 5 points out of 100. In contrast, the size of change reported for other, non-medical students (Rest, 1978), averages 30 points during such an extended period of time. So the change of moral development observed in these studies was less than one sixth of the moral development observed in other college and university students. Hence, these studies contain good news and bad news for medical education. They suggest that medical students' moral development does not completely halt during medical education but that it progresses somewhat during their time at medical school. Yet they also show that this progress of medical students' moral development is much smaller than it could be.

P-scores (DIT) of medical students are usually higher than those of BA students and, therefore, a so-called ceiling-effect may account for the low increase of P-scores over the range of four years, that is, the DIT might be less sensitive to change once a certain level of preference for principled reasoning is reached. However, in an intervention study by Self et al., medical students with initially high P-scores gained more than those with initially low scores. Unfortunately, many authors fail to mean scores (but merely report change scores) so that we cannot test our assumptions against more evidence.

Analysis II: Longitudinal Study of Moral Judgment Competence

The question remains, whether this stagnation of moral development indicates a halt of growth of moral competencies, or of moral attitudes. Most previous studies on the impact of medical education on students' moral development use tests provide only a confounded index of cognitive and affective aspects of students' moral development (see Lind). So they do not allow us to distinguish the change of moral attitudes from the development of moral competencies. To allow this, we have developed a new measurement instrument, the Moral Judgment Test (MJT) by Lind, Lind & Wakenhut). The MJT lets us assess a subject's moral competence independently from his or her moral attitudes. We have designed the MJT as an Experimental Questionnaire (Lind) so that it is possible to calculate an index of pure moral competence (the C-score), unconfounded of subjects' moral attitudes.

The C-score reflects the degree to which a person is able to make or rate judgments on the basis of (universally valid) moral principles in spite of strong counter-tendencies, e.g., the tendency to rate other people's arguments in regard to one's own opinion (opinion agreement). To

*1 I wish to thank James DaBois for pointing out to me the possibility of such an effect.

*2 Funded through a research grant of the German Research Council (DFG) and the University of Konstanz, as part of the Special Research Unit 23 "Educational Research" (Funderforschungsbereich 23 Bildungsforschung). The international extension of this research is usually referred to as the "FORW-project" (Framhain & Langer, 1984; Bargel et al., 1983).

*3 A Kohlberg (1958) argued, discussing arguments that differ not only in regard to their moral quality but also in regard to agreement or disagreement with subject's opinion, provides a good way of probing his or her moral judgment competence. Consequently, in the early version of the Moral Judgment Interview, he used them frequently. Koeser (1974) underscored experimentally the psychological significance of responses to opinion-disagreeing moral arguments. The use of con statements has proven to be essential for using the C-score as an index for the competence aspect of moral reasoning (Lind, 1995). Unfortunately, Kohlberg and his colleagues discontinued this technique (cf. Lind, 1983) and other test authors discarded the use of pro and con statements as artificial (cf. Rest, 1979, p. 69).

*4 I like to thank Tino Bargel for his help in drafting this paper.
arouse that tendency, the MJT confronts the subjects with arguments which oppose his or her opinion on a moral decision. Only if the subject rates the counter-arguments according to moral criteria in the same way as he or she rates supportive arguments, he or she considered to be a morally competent rater and gets a high C score. The C score can range from zero, indicating absence of any moral judgment competence, to 100, indicating perfect judgment competence. That is, the subject is only concerned with the moral quality of the arguments when rating their acceptability.

Validation experiments show that the C score of the MJT is truly an index for a competency and appropriate for evaluation studies. The C score cannot be faked upward in situations in which tests of moral attitudes can (Lind, 33). The MJT is sensitive to educational efforts in situations in which other tests failed to show gains (Lind, 36), and C scores does not imply undecidedness (in the two dilemmas, the correlations between the C score and an index of undecidedness are 0.06 and -0.01, respectively), hence does not reflect a tendency to purely formalistic reasoning.

The medical students of the longitudinal sample have been surveyed four times: in their first, fifth and ninth semester and in their 13th semester when many of them had already graduated and did their internship.** We will report data only on those students who have participated in all four surveys (N=746 out of 1673 students who participated in the first-semester survey). Among these were 104 medical students and 604 students from other fields of study. Altogether, 563 male and 183 female were analyzed. Note that, in Germany, medical school starts immediately after high school graduation (Abitur). That is reached after thirteen years and, in some respects, compares to a college BA.

Our first research question was whether medical students' moral attitudes change during their study. The data show that students' attitudes toward each of the six Kohlbergian stages of reasoning remained very stable over the five years in which they were observed. Neither medical nor other students changed their preferences for stages 6 and 5 over all the other stages. These preferences for moral orientations even remained stable in absolute terms. This finding is fully in line with the findings from the previous studies reported by Self et al. 37.

However, when we look at students' moral judgment competence, we can see remarkable downward changes in medical students. Medical students exhibit a sizable regression of moral judgment competence during the first two years of their education. This regression is only partly made up by a small increase at the end of their study 0. So the stagnation of moral development in medical students is not attributable to a fixation of moral attitudes but pertains indeed to a stagnation, or even regression, of moral judgment competence.

We may object that this phenomenon is due to a so-called ceiling effect, that is, due to the fact that the competence scores and the attitude scores of medical students are already so high when enter medical school that they cannot get any higher. This argument is only partly supported by our data. Medical students moral exhibit indeed an ideal profile of moral attitudes, clearly preferring stage 6 and 5 reasoning over any other stage of reasoning, and rejecting stage 1 and 2 reasoning most. They share this ideal attitude profile with many other groups, even with 5th graders and with juvenile delinquents (Lind, 38). So it would make no sense to improve their moral attitudes through teaching ethical values.

With regard to their moral competence scores, however, the stagnation of scores cannot be explained by a ceiling effect. Medical students could develop further. Due to strict admission policies, medical students are a highly selective group of students. Accordingly, the C scores of first semester medical students are higher (average C = 46.4) than the average C scores of non-medical university students (see ). Yet, higher scores can be still obtained. Theoretically, C scores can go up to 100 and non-medical university students indeed reach C scores considerably higher than medical students 0. In a study of ninth semester students, psychology majors obtained an average C score of 50.0 (Herberich, 39). Herberich explains the differential increases in moral judgment competence in different fields of study as caused by different opportunities for role-taking and guided reflection (see also Lind, 38; Sprinthall, 14).

**The standard version of the MJT has a multi-factorial 2 x 2 x 6 design. It consists of two sub-tests (factors STORY, each containing a dilemma-story and questions regarding this story. The first question is, whether the subject feels the decision made in the dilemma to be right or wrong. Subsequently, the subject is to rate arguments for their acceptability both in favor and against the solution presented in the dilemma story (factor OPINION AGREEMENT). All arguments are also constructed to represent one of Kohlberg's six stages of moral development (factor STAGE). The C score is computed as the proportion of variance of the subject's ratings accounted for by his or her concern for the moral quality of the arguments (STAGE factor) rather than, for example, for their opinion-agreement (OPINION AGREEMENT factor) or other factors and combinations thereof.

**It should be noted that the German educational system differs markedly from the US American and other systems. As I already noted, there are no fees for studying at a university. There are no colleges in Germany though in English some institutions of higher education (Fachhochschulen) are called colleges. They do not offer a liberal arts curriculum as in the US and, therefore, compare rather to institutes of technology. The undergraduate curriculum of American colleges is largely integrated in German high schools, which prepare for university and are called "Gymnasium" (for further information on the German educational system, see Peisert & Frankheim, 1994; the study design is reported in more detail in Frankheim & Langer, 1984).

ANALYSIS III: REPRESENTATIVE STUDIES OF LEARNING ENVIRONMENT

To answer the question whether the hypothesis of deficient learning environments holds true for medical students (not part of Herberich's 42

*For a description of these stage of moral development, see Kohlberg (1984). Although Kohlberg and his associates dropped the sixth stage from the scoring manual, he nevertheless adhered to the notion of stage 6 reasoning throughout his theoretical writings.
In this climate of learning, students of medicine report much less contacts with their peers, much less working in groups and considerably more competition, than other students do (p. 71). Their relationship to their professors, they report, is very bad, worth than the student-professor-relationship reported by students from other fields of study (only the relationship between students and professors in law school seems to be worth). Remember that these data stem from German universities. I do not have data available from the United States, where they may look better.

Finally medical students were asked how much they feel their knowledge about subject matter and general abilities and competencies have been fostered by university study (rather than by other factors). Compared to other students, they report, consistent with the overall picture they drew of their study, that they have learned much factual knowledge through their study. However, when it gets to general education, students of medicine report much less fostering than other students and, on top of this, this lack of fostering seems to decrease with the year of study. So it is clear that medical education, more than most other field of professional education, is puts one-sided emphasis on learning factual knowledge but neglects all other areas of human development.

**CONCLUSION**

Medical education is out of balance. Although the medical profession has probably higher demands for morally competent professional than any other profession, students of medicine are trained only to handle the technical aspects of this profession but not the moral. Through the ever faster growth of medical technologies, medical doctors are more and more in the position to make highly consequential decisions about our well-being and our lives. Yet, in regard to the moral implications of these decisions they are ill-prepared. Reviewers of recent studies have pointed to some methodological limitations of recent studies that showed stagnation of moral development in medical students (Self et al.46). However, the present study, which include a large longitudinal sample followed up four times, and which is based on a pure measurement of moral judgment competence, shows that these findings are valid. While most other students show in increase of moral competencies, medical students do not, or even regress.

This imbalance of technical and moral competencies seems to characteristic not only for present medical education but seems to have been with us as long as there is a medical profession. Though most medical students love their work and are well-intended and have high moral ideals, the tone in moral education is set by one-sided, technological, and highly competitive standards. So when, in their later work, they are confronted with concrete decision making processes, medical doctors seem to be little prepared to cope with the moral aspect of their profession. It seems, most solve a moral problem either by denying the problem or by misinterpreting it as a technological issue.

In his book on The Nazi Doctors, Lifton47, gives many example of physicians who were extremely cruel to prisoners of concentration camps, killing hundreds to pursue their scientific interests, but insisted that

---

* Funded by a research grant of the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Wissenschaft). The second authors has conceived and co-directed these surveys from the beginning (see Bargel & Ramm, 1994).
they were not aware that they broke Hippocrates oath, and that they even had no doubts about the ethical goodness of their doing. Many of these doctors got excellent reviews of their character by their teachers and bosses. Yet, obviously they could not, or did not want to, rigorously evaluate their behavior in the light of their moral consciousness, less that they could let their action be discussed by peers and patients. Lifton\textsuperscript{48} reports how one of these doctors broke down when his uncle engaged him in a night-long discussion of the moral dimension of his doing.

What can be done about the obvious imbalance of technical and moral education in medicine? Would special courses could be effective? Do medical students want this kind of moral training? Would the learning of medical knowledge suffer if we would introduce more course in moral or ethical education in medical schools?

Self et al.\textsuperscript{49,50} report a number of studies that show that adequately designed courses do foster moral development in medical students. Especially moral dilemma discussions (Blatt & Kohlberg\textsuperscript{51}) effectively foster growth of moral competencies (Lind\textsuperscript{52}). In contrast, as Michael Gross\textsuperscript{54} as shown, ethics lectures, in which only communications skills and ethical theories are taught, do not seem to be effective in fostering moral judgment competence. If anything, they foster ethical rhetoric that is of little help when it comes to solving severe moral dilemmas that arise out of everyday professional work.

Some have objected that, moral development halts in early adulthood and that, therefore, any teaching of moral competencies would be in vain. Research clearly shows that this objection is not tenable (Lind\textsuperscript{55}; Sself et al.\textsuperscript{56}; Kohlberg\textsuperscript{57}). Adults clearly do develop beyond their twenties if provided with opportunities of role-taking and guided reflection (Lind\textsuperscript{58}; Self et al.\textsuperscript{59}).

As the extensive experiences and surveys by Sponholz et al.\textsuperscript{60} demonstrate, medical students are highly interested in courses in bio-ethics. Dr. Sponholz and Dr. Baisch, who run small, case-oriented ethics courses at the University of Ulm, Germany, say that they cannot provide so many courses as are demanded by students, although students do not get any credits for these courses. My own experience with dilemma discussion courses in classes of medical students, is fully in line with this experience.

Finally, contrary to conventional wisdom, small, case-oriented ethics courses and dilemma discussions also foster the development of factual knowledge (Self et al.\textsuperscript{61}) and do not hamper or diminish it. It seems that the solution of basic moral problems of their profession sets free a additional learning motivation in the students.

It is high time for a change in medical education in Germany and, as it seems, also in the United States and in many countries that have used these systems of medical educational as a blueprint. There is no doubt anymore that present medical education fails to foster, or even obstructs, the development of moral competencies that medical doctors will need to become good physicians.
REFERENCES


