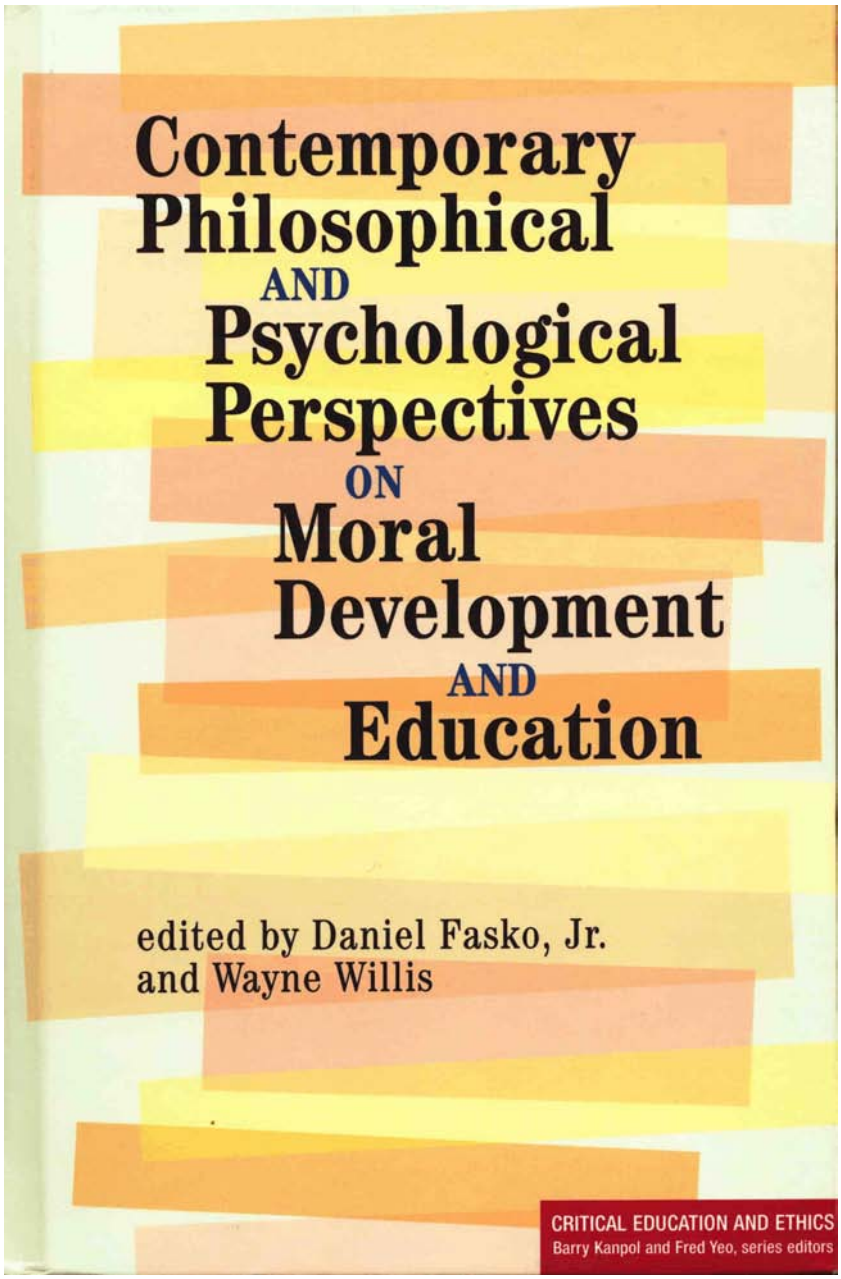
See important note below!

Lind, G. (2008). The meaning and measurement of moral judgment competence. A dual-aspect model. In: Daniel Fasko, Jr. & Wayne Willis, eds.: Contemporary philosophical and psychological perspectives on moral development and education, pp. 185-220.
Creskill: Hampton Press. (The page numbers correspond roughly to the original publication; I made some small corrections afterwards by me; GL 2013)



Important note:
In meanwhile (2013), the Moral Jugment Test (MJT) has been renamed as *Moral Competence Test* (MCT). The name of the test is now aligned to the construct it measures, namely *moral competence* (C-score). Competence is an persisting human trait while judgment is an ephemeral phenomenon.
We also speak now of 'moral competence' instead of 'moral *judgment* competence' to indicate that this competence can be observed only when it shows itself in overt action.

8

THE MEANING AND MEASUREMENT OF MORAL JUDGMENT COMPETENCE

A Dual-Aspect Model

Georg Lind
University of Konstanz, Germany

The capacity to make decisions and judgments which are moral (i.e., based an internal principles) and to act in accordance with such judgments. (Kohlberg, 1964, p. 425)

INTRODUCTION

Bill: Jane, do you think Rick is a moral person?

Jane: Mm, yes, I think so. He said he is against mercy killing.

Bill: Why do you think that this shows he is moral?

Jane: Well, mercy killing is wrong. God wants us to preserve life.

Bill: But Rick is an Atheist. So how do you know whether he has got a moral reason for his opinion?

Jane: Well, yeah, I don't know. I'd just assumed. I also heard him once say, life is something precious.

Bill: Did you know that he voted in favor of capital punishment? He does not seem to regard life as an ultimate moral value.

Jane: No, I didn't know that. He seems to be inconsistent in his moral principles.

Bill: Or he just thinks it is good to get rid both of ill and of bad people.

Jane: I cannot believe this. That would not be very moral. When I'll see him, I will ask him about this, though it is a bit hard to talk with him about these matters. He can get very defensive.

What does this short dialogue tell us about the measurement of morality? If anything, it tells us that it is not easy and that our observation can be easily led astray when we are too quick in making a judgment about someone's morality. Obviously, Bill and Jane try to score at Rick's level of moral development and use various approaches for this end. Jane first uses a statement by Rick to infer his moral attitudes or values, assuming that Rick is against mercy killing because he gives the value of life highest priority and because he believes in God. This inference is invalidated by Bill who quotes Rick's vote in favor of capital punishment. Jane changes her scoring to inconsistent moral judgment. Bill challenges this scoring again by suggesting that Rick might be ill-minded toward sick as well as toward bad people. This could resolve the apparent inconsistency of Rick's stances on mercy killing and capital punishment. However, Jane cannot easily accept this explanation. She feels more information is required to make a valid judgment on Rick's morality, but would Rick be engaged in a moral dialogue? Would he give good reasons or would he just try to rationalize his opinion?

Many generations of psychological researchers have been confronted with similar questions like Jane and Bill when trying to conceptualize and measure people's moral behavior and development. In this chapter, first, I will argue that modern psychological theories of moral behavior and moral development need new, theoretically valid methods of measurement. Unless we have them, we cannot make valid inferences from data on the empirical validity of these theories. The evolution of new and better theories depends an the construction of better research instruments and vice versa.

Therefore, I will first give an overview of these definitions of morality (the *rule-conformity*, the *good-intentions*, and the *competence* definition) and the methods of measurement that they entail. Second, I will describe the Moral judgment Test (MJT), which I developed thirty years ago in order to assess adequately the *competence aspect* of moral judgment behavior. Third, I will report empirical data on the theoretical and cross-cultural validity of the MJT and several translated versions of it. Fourth, I will contrast the MJT with other instruments for measuring moral development, especially Kohlberg`s *Moral Judgment Interview (MJI), Rest's Defining Issues Test (DIT) and Gibbs' Sociomoral Reasoning Measure (SRM).*

Three Definitions of Morality: Conformity, intensity, and competence

One of the earliest approaches to measuring morality is the rule-conformity definition. Here morality is defined through a list of things that should be done and better be avoided: Don't steal, don't murder, don't commit adultery and so forth, and "love your neighbors as yourself," and "love strangers who sojourn with you as yourself" (The Bible, Leviticus 19:17-34). Other cultures and religions know of similar lists. In some cultures, a person is considered moral if his or her rule-conforming deeds outnumber his or her rule transgressions. Morality, then, is measured by the number of instances in which a person exhibited morally correct behavior and avoided morally wrong behavior. In other (sub-)cultures, each and every rule has to be observed. These lists, usually defined by a superhuman instance and interpreted by a special group of people, grew rather long, and eventually it became nearly impossible to conform to each and every rule unless one was a saint. This approach was used by many researchers and is still quite popular. Parts of psychological *Behaviorism* (e.g., Hartshorne & May, 1929) can be seen as a representative of the rule-conformity approach.

Another approach is the good-intentions *definition*, which considers rule-conforming behavior as a poor indicator of morality, but it regards a person's moral intentions as a better, if not the only, indicator of morality. According to this definition, a behavior is morally good if it is based on morally good intentions (or moral values, motives, or principles). Early Christian movements and Stoicism regarded the inner possession and cultivation of ideals as self-sufficient for the essence of morality. Saint Thomas of Aquinas is said to have weighted this definition against the prevailing rule conformity definition of the church of his time. The influential philosopher Immanuel Kant (1948/1785) also maintained that morally good intentions are not only necessary but sufficient conditions of moral goodness: There is nothing good but which results from good will! As a standard for a morally good will he provided his *categorical imperative*: "Act only on that maxim whereby you can at the same time will that it should become a universal law" (pp. 84, 421). The attitude paradigm in moral research seems to be rooted much in this tradition of thought. Morally good behavior, one seems to believe, follows exclusively from morally good attitudes and values, that is, preferences for principled moral thinking (e.g., Rest, 1979; Rest, Narvaez, Bebeau, & Thoma, 1999). Interestingly, Hartshorne and May (1929) have incorporated this approach in their studies, too.

For many centuries, both definitions have been (and still are) theological and philosophical rivals, implying very different theories and methods in psychology and education. Wheras

the behavior-conformity theory resonated mostly in behavioristic moral psychology and reward-and-punishment approaches to moral education, the good-intentions theory resonated in the psychology of moral attitudes, motives and values and in different sorts of value education and moral indoctrination. This rivalry, pointed out by Max Weber (1994), was between *Verantwortungs-Ethik* (being concerned only about the consequences of one's decisions) and *Gesinnungs-Ethik* (being concerned only about the goodness of a person's intentions), or, as John Dewey (1966, p. 349) noted, between Kant's good-intention ethics and Bentham's utilitarianism.

Yet, both definitions were also based on some common beliefs, for example, (1) the belief that morality was not merely inborn, but could be (or even had to be) improved by psychological and educational means; (2) the belief that this improvement was mostly (if not exclusively) the responsibility of special social institutions (parents, teachers, priests, administrators and the like), which means that it is something external to the individual; (3) the belief that any attempt to improve morality in people requires some sort of social power and even force, and (4) the belief that morality is something completely separate from a person's cognitive abilities and competencies. The latter seems to have been a plain truism: a person is to be called moral either (in the case of the rule-conformity theory) if she or he complies perfectly with the behavioral rules set up by society, or (in the case of the good intentions-theory) if she or he has only morally good intentions. This belief was so deeply entrenched in psychological and educational theorizing and research for many decades that, in the 1950s, a joint committee of the American Psychological Association (APA) and the American Educational Research Association (AERA), chaired by Benjamin Bloom, created two separate taxonomies for educational objectives, one for the "cognitive domain" and one for the "affective domain." Moral and democratic behavior was thought to have nothing to do with cognition and competencies; it was put into the affective domain (Bloom, 1994; Bloom, Engelhart, Hill, Furst, & Krathwohl, 1956; Krathwohl, Bloom, & Masia, 1964).

At the beginning of the twentieth century, some psychologists started to see the need for overcoming this separation of affect and cognition. Eventually, this need led to the formulation of the *dual-aspect theory* of morality and moral development, which defines moral behavior as consisting of two inseparable, yet distinguishable aspects: (1) a person's affection for certain moral ideals or principles, and (2) his or her ability to reason and act according to these ideals and principles (e.g., Lind, 1985a; 2002).

Some early formulations of a dual-aspect theory can be found in the nineteenth and early twentieth centuries. In 1912, the German psychiatrist Max Levy-Suhl found that moral attitudes or values (that is, a lack thereof) cannot account for juvenile delinquency. When we interviewed delinquent and nondelinquent youth about stealing, he found that both groups

rejected stealing in a similar manner. However, in the reasons given by the two groups, Levy-Suhl found signs of different levels of moral maturity. Juvenile delinquents often said stealing was wrong because it causes personal damage, and they mentioned less frequently sociomoral reasons or autonomous ethical reasons than did nondelinquent youth (Heidbrink, 1989; Levy-Suhl, 1912). Another example is altruistic attitudes or values, which, since August Comte's times, have been regarded as indicators of moral maturity. As early as 1892, Georg Simmel (1989) rejected this definition because altruism and egoism are always intermingled (one can be altruistic for very egoistic reasons and vice versa). Later it was empirically shown, that altruistic attitudes or intentions were hardly related to altruistic behavior (Darley & Latane, 1968; Krebs, 1982; McNamee, 1977). Krebs (1982) echoes and specifies Simmel's objections: "In my view, altruistic behavior is not necessarily moral or just behavior. In fact, inasmuch as the idea of altruism means giving more than one's share, or giving more than one should, it entails a violation of the balance of reciprocity that defines justice [. . .] behaving justly involves much more than living by a simple rule of generosity" (p. 73).

This need for a new point of view was reinforced by Hartshorne and May (1928), who based their study into the nature of deceit on the behavior conformity definition of morality. For them, moral behavior must be observed, and measured, "without any reference [. . .] to its motives or its rightness or wrongness. The first question to ask is: What did the subject do? Unless this question is answered in quantitative terms, so that what he did is clearly known, there is little use in going on to ask why he did it, and still less use in speculating whether he is to be blamed or praised" (p. 11). As a result of their extensive empirical study of children, they found that honesty or dishonesty is not a unified craft in children of the ages studied, but a series of responses to specific situations (Hartshorne & May, 1930, Book Two, p. 243). Therefore, as a conclusion, Hartshorne and May (1928) questioned their own definition of morality: "It is not the quality of the isolated act which distinguishes the good man from the bad, but the quality of the man as an organized and socially functioning self" (p. 413). Reviewing a large bulk of behavioristic rule-conformity studies on moral development, Pittel and Mendelsohn (1966) concluded that the problems of measurement

> seem to be the result of an insufficient effort to *conceptualize the nature of moral values and their relation to behavior.* Perhaps the greatest single shortcoming underlying each of the specific criticisms discussed is the failure to view evaluative attitudes as subjective phenomena whose measurement is best achieved [...] without reference to the relationship of these attitudes to conventional and normative standards of moral valuation. It is important to assess at an individual level the content, strength and patterning of subjective attitudes or evaluation per se. Whether these attitudes would be approved or disapproved by society is a subsequent question which need not he considered in the construction of

measures of evaluative attitudes. (p. 32; emphasis added)

The need for a new definition of moral character and development was felt by many, but only a few actually tried to bridge the gap between moral ideals (or attitudes or values or principles) on the one side and moral behavior on the other. Even fewer suggested methods for measuring concepts that could fill this void. In his famous studies of children's play, Piaget (1965/1932) made the gap between moral rules and actual behavior his explicit aim of research, discerning stages or phases of moral-cognitive development similar to the ones Levy-Suhl (1912) found. He used combinations of methods of direct observation and interviews for his studies. Pittel and Mendelsohn (1966) described succinctly the criteria for a more adequate method of measurement: "It is important to assess at *an individual* level the content, strength and *patterning* of subjective attitudes or evaluation per se" (p. 32; emphasis added).

It was Kohlberg (1958, 1964, 1984) who tried to integrate various attempts to measure moral judgment competence and found a new, promising paradigm of moral psychological research. He picked up Piaget's assessment methods, Pittel and Mendelsohn's criteria for measurement, and, most importantly, he provided a clear conceptualization of the relationship between moral ideals and moral behavior through his definition of moral judgment competence, which 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"* (Kohlberg, 1964, p. 425; see also Kohlberg, 1984, p. 523).

Kohlberg's definition of moral judgment signified a true paradigm shift in a Kuhnian sense (Kuhn, 1962) as it was revolutionary in three ways:

1. For the first time, morality is defined in terms of a competence (or ability or capacity) rather than merely as an attitude or value, and, through this, the unfortunate separation between the cognitive and the affective domain of behavior is overcome.
2. Moral behavior is defined in reference to a subject's internal, accepted moral principles, rather than to external social norms and standards (as in the rule-conformity definition of morality).
3. The enactment of one's judgment is an integral part of the definition, making all three *aspects* (the affective, cognitive, and behavioral aspect) part of a proper definition rather than viewing them as separate components, which can be observed or measured in isolation from one another.

The relationship of these three aspects of moral behavior is summarized in Figure 8.1. To be moral, a behavior needs to be guided by moral ideals or

principles, yet in order to be morally mature a behavior must also be informed by developed reasoning competencies. However, note that moral principles and competencies are not, as this model may suggest, separate things, but just different aspects of behavior (See the following). Subsequently, Kohlberg (1984) elaborated his concept of moral judgment competence in more detail and suggested theoretical criteria for an adequate measurement approach:

> 1. The test constructor must postulate structure from the Start, as opposed to inductively finding structure in content after the test is made. "[...] If a test is to yield stage structure, a concept of that structure must be built into the initial act of observation, test construction, and scoring" (pp. 401-402).
>
> 2. Hence, measurement must be directed at assessing "parameters of organizational wholes or Systems of internal relations" (p. 8), rather than merely focusing on isolated items of behavior as in classical psychometric testing.
>
> 3. These "cognitive structures are always structures (Schemata) of action" (p. 8). More precisely, Kohlberg (1984) stated after, that structure "is a construct rather than an inference, and is warranted only on the grounds of intelligible ordering of the manifest items" (p. 408). Hence, "the responses of subjects to the dilemmas and their subsequent responses to clinical probing are taken to reflect, exhibit, or manifest the structure.. . . There can be no error in the sense of a mistake in inferring from a judgment to some state of affairs concurrent with, precedent to, or subsequent to the [scorer's] judgment" (1984, p. 407). Kohlberg's manifest behavior pattern approach contrasts sharply with approaches like classical test theory and Rasch scaling, which regard each response item as an indicator of some unobservable, hypothetical variable or latent entity, and the structure of the individual response pattern only as a sign of measurement error.

FIGURE 8.1. The Dual Aspect Model

4. The measurement must use a complex moral situation to elicit moral judgment competence. "The organization of these modes [of cognitive activities] is always an organization of action upon objects" (Kohlberg, 1984, p. 8). Hence, decontextualized interview questions or test items, which are constructed on the basis of classical test theory, do not allow us to assess moral judgment competence, though they may be valuable instruments for assessing a person's moral attitudes and values.

5. "Affective development and functioning and cognitive development and functioning are not distinct realms. Affective and cognitive development are parallel; they represent different perspectives and contexts in defining structural change" (Kohlberg, 1984, p. 8), that is, they must be seen as aspects of the very same pattern of behavior and must be measured using the same instrument rather than seen as separable behavioral components which can be measured with separate instruments (see also DeVries, 1997, p. 6; Piaget & Inhelder, 1969, pp. 114, 117).

6. The measurement of a competence always requires the construction of proper moral tasks. Dilemmas and counter-suggestions seem to provide such a task. The way actors deal with dilemmas and counter-suggestions is a very good indicator for the actors' ability to solve a conflict by engaging in a peaceful, nonviolent moral discourse rather than using their status and power to coerce others into accepting their convictions (Habermas 1990). Levy-Suhl (1912) already used counter-suggestions to elicit all reasons that the children were capable of. Kohlberg (1958) used dilemmas and counter-suggestions in his interview method. He used dilemmas because their "solution must do justice both to what the self believes and yet meet the situation. Thus the choice is *difficult* in the sense ... of doing justice to all the values which the self believes are true and important" (Kohlberg 1958, pp. 128-129; emphasis added).

He used counter-suggestions "for two purposes: first, the assessment of independence of judgment, and secondly, the validation of developmental levels" (p. 78). Similarly, Piaget “considered counter-suggestions essential to the clinical method” (Lourenzo & Machado, 1996, pp. 146, 154). In order “to facilitate the interpretive analysis and to clarify the subjects' real understanding,” as Blasi (1987) noted, it has become a general standard for cognitive developmentalists to resort "to probes, alternative hypothetical situations, and counter suggestions" (p. 89). In fact, experimental studies showed that the participants' response to counter-suggestions is indicative of their (lack of) moral-judgment competence (Keasey, 1974) and may, therefore, be used as a task in a moral judgment competence test.

Besides dilemmas and counter-suggestions, Kohlberg also recommended using an open interview format, rather than closed questions, to make it difficult for an individual to get a high score just by guess work. Although many researchers (e.g., Gibbs, Widaman, & Colby, 1982; Kohlberg, 1979, p. xiv) regard the production of arguments as the most distinct feature of Kohlberg's MJI, contrasting it to preference tests like the DIT and the MJT, this distinction does not seem to be as important as the difference between competence tests, which contain *a moral task* like counter-suggestions, and attitude questionnaires, which do not. As Lind (2002) has shown, MJI scores seem to be always higher than DIT and MJT scores, speaking against the hypothesis that the production of arguments is a more difficult task than meeting the criteria set forth by those preference tests. Moreover, although dilemmas and counter-suggestions possess a prima facie relevance for moral problem solving, the question of whether the reasons used for a deliberation process have come to mind spontaneously or have been suggested by the interviewer seems of less moral relevance. The production of reasons may just add a morally irrelevant linguistic difficulty to the task (Darley & Shultz, 1990).

THE DUAL-ASPECT THEORY OF MORAL-JUDGMENT COMPETENCE

Based on a review of Kohlberg's definition of moral-judgment competence and subsequent attempts to measure it, I proposed the *dual-aspect theory of moral behavior and development* and a new method for measuring this construct, the *Moral Judgment Test* (Lind, 1978, 1982, 1985a, 1985c, 2000a, 2002).

The dual-aspect theory incorporates many postulates from cognitive developmental theory by Piaget and Kohlberg, but makes also some clarifications and modifications to make the theory more coherent and consistent with empirical data. The main postulates that are relevant for measuring moral judgment competence are:

1. *Inseparability:* Affective and cognitive mechanisms are inseparable, although distinct. Moral affects (values, ideals) are exhibited in moral behavior in various ways depending on the individuals' cognitive structures and competencies. Thus, in order to measure them properly, we must also study them closely. Moral competencies, on the other hand, cannot be defined and measured without reference to an individual's moral ideals or principles. Therefore, an adequate measurement must be designed to assess both aspects of a person's judgment behavior as distinct aspects of the same pattern of behavior (Piaget, 1976, p. 71; Lind, 1985a, 2000a).

2. *Moral task:* To measure moral judgment competence, the instrument must contain a moral task, which requires this competence for its solution. A proper task seems to be to require the subject to deliberate on moral dilemmas and to rate counter-suggestions or arguments against one's own stance on a dilemma along with arguments agreeing with one's opinion (Keasey, 1974; Kohlberg, 1958; Lind, 1978, 1985a; Lind & Wakenhut, 1985).

3. *Nonfakeability*: To be a reliable measure of moral judgment competence, individuals should not be able to fake their competence scores upward (Emler, Renwick, & Malone, 1983; Lind, 2002).

4. *Sensitivity to change*: Although the competence score should not be upward fakeable, it should be sensitive to real changes over a wide range of the scale, either to upward changes, as a function of moral learning and intervention, or to downward changes, as a function of competence erosion. This postulate contrasts with the position of Kohlberg and his colleagues, who tuned their measurement to prevent the occurrence of such erosions: "Several years of reflection," Kohlberg (1984) wrote, "led me to decide that the data called into question the construct validity of my measure rather than the truth of the Piagetian sequence hypothesis" (p. 411). When we drop this measurement bias, we can observe cases of competence erosion (Lind, 20004, 2002).

5. *Internal moral principles*: The score for moral-judgment competence should take the individual's own moral principles into account and not impose on him or her external moral expectations (e.g., the test constructor's moral preferences) (Kohlberg, 1964; Pittel & Mendelsohn, 1966).

6. *Quasi-simplex:* lf the test dilemmas demand principled moral judgment, the acceptability ratings of each stage should support the notion of an ordered sequence, that is, the correlations among the stage ratings should form a quasi-simplex structure (Kohlberg, 1958, pp. 82-85; Lind, 2000a).

7. *Parallelism*: Although, the affective and the cognitive aspects of moral judgment behavior are distinct and independently scored, the two aspects should be parallel, that is, they should correlate highly with each other (Piaget & Inhelder, 1969, pp. 114, 117; Lind, 1985a, 2002).

8. *Equivalence of pro and con-arguments:* To be able to measure participants' moral competencies irrespective of their particular

stance on the dilemmas presented, they must be confronted with pro- and contra-arguments that are equivalent for both stances. For example, the arguments in the test in favor of mercy killing should altogether be just as acceptable or defensible as the arguments in the test against mercy killing.

THE MORAL JUDGMENT TEST (MJT)

In the early 1970s, we looked for an instrument to measure individuals' moral-judgment competence besides assessing their moral *attitudes* (see also Lind, 1985a, 2000a, 2002). As there was no adequate instrument available at that time that allowed us to measure both aspects simultaneously and yet in a clearly distinguished way, we developed a new instrument, the *Moral Judgment Test (MJT)*, which should make it possible to assess the ability of people to judge arguments pro and con a controversial moral problem on the basis of their own moral principles, that is, irrespective of their opinion on the particular problem. Besides this, it should provide measures of the participants' attitudes toward the six Kohlbergian stages of moral reasoning, as many other instruments do, too.

The design of the MJT reflects our understanding of the psychological nature of moral-Judgment behavior and the testing situation. We believe that the quality of moral judgment cannot be reduced to single responses or dimensions, but must be seen in context. Moral virtues like trust, reliability, sincerity, fairness cannot be judged on the basis of a single act or one kind of behavior of a person, but only on the basis of a whole pattern or structure of acts. For example, in the dialogue presented at the beginning of the chapter, Bill and Jane try to assess Rick's kind of morality. As it turns out, their first attempt at inferring Rick's morality from a single quote turned out as invalid. They realized that much more information about Rick is needed to make a good judgment. Or take Robert (a fictitious boy of seventeen), who invokes a certain moral reason for his decision to help a friend cheat on an academic test (to help a friend in need is a much higher value than obeying some abstract rules set by the teacher), we would expect him to be consistent when this moral reason is invoked in another context or by other persons. Let us imagine that Robert learns that many have cheated on this test and, therefore, his ranking fell below a critical mark so he cannot apply for his favorite college. Now he argues that the other students should have obeyed the rule. Obviously, there is something wrong with this kind of reasoning because it is not consistent. Either he should have thought of this consequence in the first instance, or he should have stuck to his argument the second time. This and the following example are kept very simple for the reason of clarity. Of course, there can be good reasons for changing

one’s mind on the quality of a reason. Yet certain changes of the context are thought to be inadmissible justifications. For example, we accept that someone getting older gains new insights and, therefore, rejects reasons he or she once invoked. Yet when a certain reason is invalidated merely through the fact that another person or an opponent uses them, we do not regard this judgment behavior as *morally* consistent. Another example for such an immature level of moral judgment is represented by a man who goes out to kill doctors who do abortions because he wants to preserve *life* as an absolute principle that must never be broken.

In accordance with this notion of moral-judgment competence as moral consistency across well-defined changes in the context, the MJT was designed as a complex situation to which the participants were to respond.[1]

The MJT confronts individuals with a situation that puts a high demand on them. They have to decide on a difficult moral dilemma and then to rate arguments in favor and against this decision on a scale from "I strongly reject (-4) to I strongly accept (+4)." The two sets of arguments (pro and con) are matched to represent the same qualities or levels of moral reasoning though with opposing implications. What we want to find out with this arrangement is: Do participants base their ratings on the different moral qualities of the arguments, and thus demonstrate some moral-judgment competence, or do they base their judgment rather on the fact whether the arguments speak in favor of or against their own opinions? Obviously, this question cannot be answered by looking merely at isolated answers, but only by looking at the whole pattern of answers given by a person. Neither can this question be answered by looking at aggregated "structures" across individuals (like correlation patterns or results of factor analysis), but only by conceiving the individual as a unique person acting on the basis of unique cognitive-affective structures.

Like Piaget and Kohlberg in their clinical interview method, the MJT uses moral dilemmas, which is a situation in which a person cannot make a decision without transgressing an important moral rule or principle. In the standard MJT, two dilemmas are used, the mercy-killing dilemma from Kohlberg's MJI (Colby et al., 1987b) and the worker's dilemma taken from Max von der Grün's (1975) theater play and novel *Stellenweise Glatteis*. In the Doctor's Dilemma, the woman's request for mercy killing brings highly valued moral principles into conflict with each other (e.g., the value of preserving life, of quality of life, of helping a person who is in great distress). In the Workers' Dilemma, several controversial issues are invoked, such as the issue of arbitrary law, of solidarity, and of the rights of employers and employees. It is required that the individual gives his or her opinion on the doctor's and the workers' behavior in a scale from -3 to +3 (see Box 1 and Box 2).

BOX 1: DOCTOR' S DILEMMA
A woman had cancer and she had no hope of being saved. She was in terrible pain and so weakened that a large dose of a painkiller such as morphine would have caused her death. During a temporary period of improvement, she begged the doctor to give her enough morphine to kill her. She said she could no longer endure the pain and would be dead in a few weeks anyway. The doctor complied with her wish. Do you disagree or agree with the doctor's behavior?

BOX 2: WORKER'S DILEMMA
Due to some seemingly unfounded dismissals, some factory workers suspect the managers of eavesdropping on their employees through an intercom and using this information against them. The managers officially and emphatically deny this accusation. The union declares that it will only take steps against the company when proof has been found that confirms these suspicions. Two workers then break into the administrative offices and take tape transcripts that prove the allegation of eavesdropping. Would you disagree or agree with the workers' behavior?

The two dilemmas for the standard version of the MJT were chosen because they both confront the participants with highly demanding *moral principles*. Although we expected that the mercy-killing dilemma (taken from Kohlberg's *Moral Judgment Interview*) would require a moral discourse on the highest level of moral reasoning on Kohlberg's six stages, we expected the workers' dilemma at least at stage 5 reasoning (Lind, 1985a). Research shows that, in fact, individuals typically prefer stage 6 moral discourse most in the doctor's Dilemma and stage 5 reasoning most in the Workers' Dilemma (Lind, 2000a).

Findings from European studies also show that the mercy-killing dilemma pulls slightly higher moral-judgment competence scores. Recent studies in Latin America and the United States indicate that this relationship can be reversed when people submit to religious authorities, who discourage them from reasoning about issues on which the authority is dogmatic (Lind, 2003b).

Although the dilemmas are believed to be difficult to solve, and most respondents indeed show some signs of emotional arousal when being pressed to make a decision, this part of the test is not used for scoring, rather, it sets the stage for the subsequent task. The dilemma may also be used for indexing decision-making competence, but for this, the MJT needs to be administered electronically. In a laboratory study, Mansbart (2001) demonstrated that the time participants needed to make a decision on the presented dilemma varied widely among participants and was correlated substantially with the C-score, the main index of moral-judgment competence.

The moral task of the MJT is contained in the arguments that are subsequently presented to the participants. For each dilemma, the participants have to say how much they accept or reject a set of twelve arguments. Six of these arguments are *in favor* and six are *against* the doctor and the workers, respectively. These arguments are to represent the different moral qualities of reasoning as described by Kohlberg's six stages of reasoning.[2] To make sure that they are valid representatives of these stages, they were first subjected to an expert rating[3] and then empirically tested according to the validity criteria described earlier. In addition, eight participants were asked to respond to these arguments aloud, and were encouraged to comment on them critically. Sample arguments (taken from the workers' dilemma) are given in Box 3.

Participants should say how much they accept or reject each argument on a nine-point scale (see Scale in Box 4).

BOX 3: SAMPLE ARGUMENTS

How acceptable do you find the following arguments in favor of the behavior of the two workers? Suppose someone argued they were right ...

- because they ain`t cause much damage to the company.
- because due to the company's disregard for the law, the means used by the two workers were permissible to restore law and order.
- because most of the workers would approve of their deed and many of them would be happy about it.
- because trust between people and individual dignity count more than the firm's internal regulations.
- because the company had committed an injustice at first, the two workers were justified in breaking into the offices.
- because the two workers saw no legal means of revealing the company's misuse of confidence, and therefore chose what they considered the lesser evil.

BOX 4: SCALE									
I strongly reject									I strongly accept
	-4	-3	-2	-1	0	+1	+2	+3	+4

TEST ADMINISTRATION

The MJT is usually administered without a time restriction. lt takes approximately 10 to 20 minutes to complete it. Mostly, participants like to respond to the test and to my knowledge, nearly all participants completed the test. Usually, only a very few, if any, tests needed to be relegated from scoring and analysis. The MJT can be used with children as young as ten years of age, if they have no learning deficits. When applied to young children or persons with cognitive deficits, the MJT should be administered individually; the font of the text should also be increased and the response scale should be reduced from 9 points to 5 points (ranging from -2 to +2 rather than from -4 to +4). However, the wording of the dilemmas and the arguments should not be changed. If changed, the new version must be submitted to a validation study again (explained later).

When the MJT is used repeatedly, as in evaluation or intervention studies, special attention should be given to the phenomenon of retesting fatigue. As our own unpublished studies show, some subjects may become weary of responding twice to the test within a short time period. The participants should be prepared for this by proper instructions, like: "Now, we present again the dilemmas that you had previously seen. Please read the questions again as carefully as you did the last time." Usually, this reduces test weariness considerably.

Note that the MJT is designed for research and program evaluation, but not for individual diagnosis or for the evaluation of individual persons (e.g., grading). A person's moral judgment behavior depends considerably on situational factors like fatigue, involvement, prior experience. Therefore, an instrument for assessing an individual's degree of moral judgment competence must have built-in safeguards against misinterpretations, which the MJT does not have. When doing basic research or evaluation studies with *groups* of people, such situational factors mostly cancel out so that the average C-scores can be reliably interpreted as a "true" level of moral Judgment competence.

SCORING THE MJT

According to the dual-aspect theory of moral behavior, the MJT produces two sets of scores, one for its cognitive aspect and one for its affective aspect. The most important cognitive score is the C-score, which indexes the subjects' moral-judgment competence. With the MJT, moral-judgment competence is operationally defined *as the ability of a subject to accept or reject arguments on a particular moral issue consistently in regard to their moral quality even though they oppose the subject's stance on that issue*. Only the fact that they have to cope with an emotionally difficult task to accept moral arguments that oppose their opinions makes this a competence index. Mere preference for certain stages of moral reasoning would only indicate their moral attitudes.

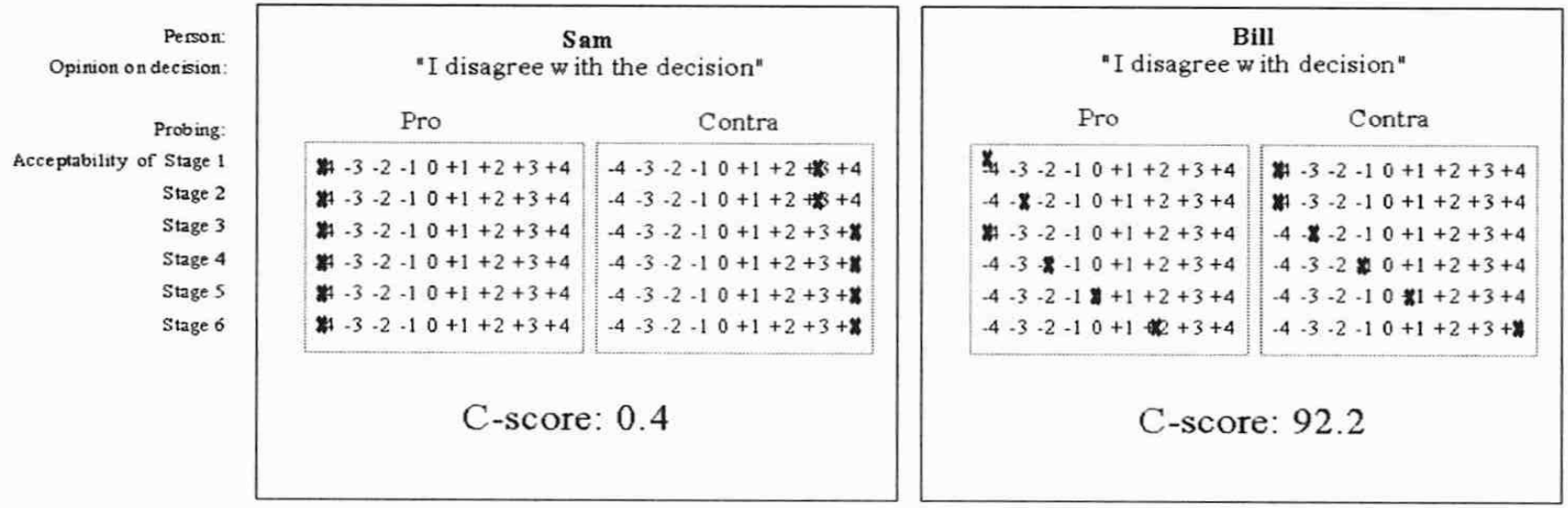
Accordingly, the C-score reflects the degree to which an individual accepts or rejects arguments in a discussion on a moral issue *with regard to their moral quality* rather than with regard to their agreement with the individual's opinion (or other nonmoral properties). Note that it does not index consistency *per se*, as some researchers (e.g., Rest, Thoma, & Edwards, 1997) seem to assume. Only if the participant's judgment is consistent with his or her moral principles is it scored as moral competence. Other forms of judgment consistency, for example, with regard to an individual's opinion, are seen as lack of moral competence (for a more refined statement, see Lind, 1978, 2000a). Of course, the C-Index indexes a competence only when calculated for a test that contains *a moral task* (like the MJT's counter arguments). This is not the case when the C-Index is calculated for a moral attitude test like the DIT (Rest, Thoma, & Edwards, 1997), which does not contain a task. (How the C-Index is computed is explained in more detail at http://www.uni-konstanz.de/ag-moral/mut/mjt-engl.htm; and Lind [2000a]).

The C-Index can range from 1 to 100. The C is sometimes categorized as *very low* (1-9), *low* (10-19), *medium* (20-29), high (30-39), *very high* (40-49) and *extraordinary high* (above 50). Note that this categorization is very rough and applies only to the standard version of the MJT. If the C-score is calculated only for each dilemma separately, it will be higher because variance due to dilemma context is omitted.

In Figure 8.2 the response patterns of two fictitious participants to the mercy-killing dilemma are presented for Illustration. Both Sam and Bill oppose mercy killing, but both respond very differently to the arguments given for and against this decision. Sam accepts all con arguments, because they agree with his opinion on mercy killing, and he rejects all pro arguments, which for him are counter-suggestions. Only in regard to stages 1 and 2 does he seem to sense the differential moral quality of the supporting arguments. In contrast, Bill demonstrates his capa-

FIGURE 8.2. The Experimental Design of the MJT

The C-Scores for the Patterns of Moral Judgment Behavior of Two Fictitious Participants



Design of factors: 6 x 2 multivariate, orthogonal, N=1 (Standard MJT, two dilemmas: 6 x 2 x 2)

Independent variable:
1. Moral stage (or quality) of reasoning
2. Probing (pro and contra arguments)
(3. Dilemma context; standard version, not depicted here)

Dependent Variable: Rating of 12 arguments as to their acceptability (standard version: 24 arguments)

Note: In the MJT the arguments are presented in random order and not according to the stage they represent.

bility to make judgments on the basis of moral principles by rating the arguments consistently with regard to their moral quality. Even though he disagrees with the decision and accepts the con arguments slightly more than the pro arguments, he clearly differentiates between the moral qualities of the arguments. The C-scores for their response patterns express numerically this visual interpretation. Sam's C-score for the Doctor's Dilemma is C = 0.04, Bill's is C = 92.2 (of 100 points). In other words, whereas Sam's response pattern to the arguments indicates a low degree of moral-judgment competence, Bill's pattern indicates an extremely high level of competence, one that is rarely found in real life. Bill judges consistently with regard to the *moral quality* of the arguments, whereas Sam is consistent, but in regard to *his own stance on the issue at stake.*

If strong differences of C-Scores occur within the same participant, we call it "segmentation phenomenon" (Lind, 2000c; see also Schillinger, 2006). A more refined score was proposed by Lind (1978, 2000a) to take into account dilemma type and other aspects of a person's structure of moral judgment behavior. Because, in the past, this refined index correlated almost perfectly with the C-score, this line of scoring was put aside. In the light of new findings regarding segmentation, it is now being reconsidered. With the electronic version of the MJT, additional measures are available, like response latency and decision-making time, which give valuable insights into the cognitive processes underlying moral reasoning. Note that the C-score and other indices of a subject's moral-judgment competence would not be meaningful unless the test contained a moral task, namely, the task to use one's own moral principles persistently in the face of opposition to one's own position on a moral issue.

As indices for the affective aspects of moral behavior, the MJT produces scores for a person's *attitudes* toward each of the six levels of moral reasoning that Kohlberg identified. An inspection of these attitudes tells us, for example, which stage of moral reasoning people prefer most and which least, and for which type of dilemma people prefer a reasoning on the highest level and for which dilemmas they believe a lower stage to be most adequate. Moral attitudes are indexed by summary ratings over all four arguments that represent a Kohlbergian stage. (For a discussion of these and several other indices, see Heidbrink [1985]; Lind [1978, 2000a], Lind & Wakenhut [1985]). The six attitudes are usually depicted as profiles of acceptability. One could also calculate which stage of reasoning is mostly preferred, but this measure would be of little use in research because hardly any variance is expected. Most, if not all, persons prefer principled stages (5 and 6) most.

THEORETICAL AND EMPIRICAL VALIDATION OF THE MJT

Validity refers to whether a test measures what it is supposed to measure (Messick, 1994). Does the MJT measure what it claims to measure? Do MJT data allow us to make inferences on the empirical fittingness of our hypotheses? Can we compare data from studies using different language versions of the MJT? These questions must be answered by theoretical as well as by empirical analysis: Is the design and content of the test in line with what we want to measure, and, if this is the case, do the data produced with this test agree with our theoretically founded expectations? To analyze the original MJT's validity (or, to be more precise, the validity of our interpretation of the C-score), we assessed the original version of the MJT and all certified translated versions of the test.[4]

To be called valid, the MJT had to meet five *empirical* criteria derived from the cognitive-developmental theory and the dual-aspect theory of moral behavior (e.g., Kohlberg, 1958, 1964, 1984; Lind, 1978, 2000a, 2002):

1. *The preferences for the six Kohlbergian stages of moral reasoning* (affective *aspect) are ordered in a predictable way*: Moral reasoning on high Kohlbergian stages is preferred over reasoning on lower stages. (To my knowledge, this preference hierarchy has been found in all Moral Judgment Test-studies including studies in many different cultures (Lind, 1986, 2000a, 2002; Gross, 1997).

2. The correlations between the stage-preferences form a quasi-simplex structure, that is, the correlation between the preferences of neighboring stages (like four and five) should be higher than the correlation between more distant stages like four and six (Kohlberg, 1958, pp. 82-84). (This hypothesis was tested conducting a principle component analysis with varimax rotation. If the data form a quasi-simplex, as predicted, a two-factor solution should result with the factor loadings of stages 1 to 6 being located on a simplex arc nicely ordered. This hypothesis is well supported by all empirical studies.)

3. *Cognitive and affective aspects are parallel*: The higher a participant's moral Judgment competence is, the more clearly does he or she accept higher stage arguments and reject lower stage arguments. Piaget (1976) assumed that cognitive and affective aspects of human behavior are systematically correlated. One way to translate this general postulate is to assume that participants' moral attitudes should be systematically correlated with their moral judgment competence. We expect high negative correlations between the

C-score on the one hand, and attitude scores for stages 1 and 2 on the other, and moderate correlations between C and attitudes to stages 3 and 4, and substantial positive correlations between C and attitudes to stages 5 and 6. Nearly all MJT-studies found such patterns of correlations corroborating the *parallelism theory*.

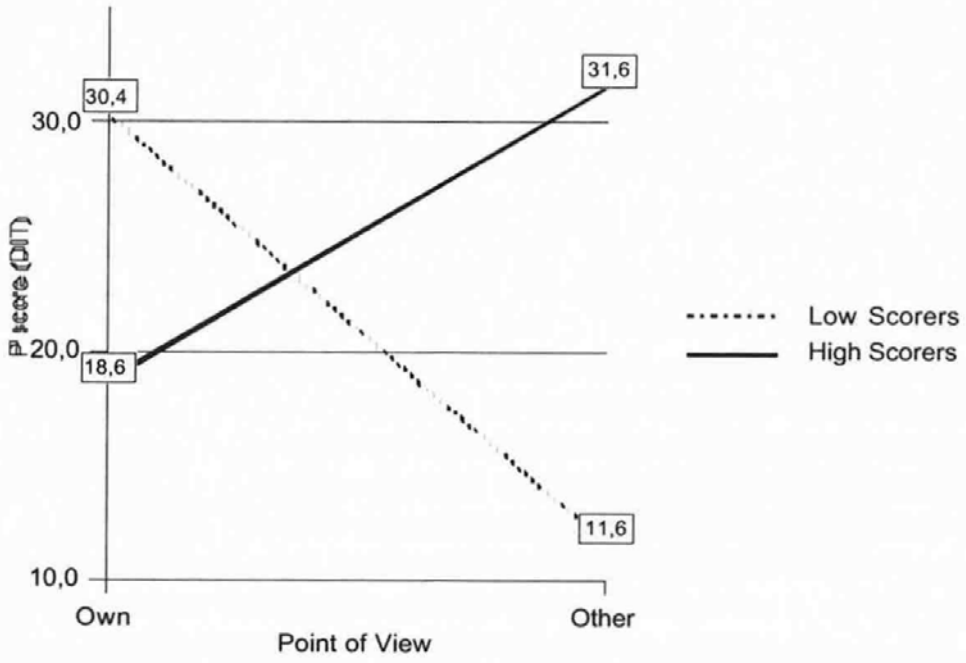
4. *Pro and con arguments are equivalent*: Indeed, the profile of preferences for pro arguments by the pro subjects was almost identical with the preferences of con arguments by con subjects. The same holds true for the preferences of opposing arguments by the two groups (Lind, 2000a).

5. *The MJT is a difficult moral task and, hence, the C-index is an index of moral competencies* (rather than of moral *attitudes*). Whereas in Emler et al.'s (1983) experimental study using the DIT the participants Could simulate a higher *P-score* when they were instructed to do so (see also Markoulis, 1989), Lind (2002) showed with an identical experiment, that participants could not simulate the *C-index* upward (see Figure 8.3). In Figure 8.3 the solid line shows the data of the critical group, namely, the group *(rightists)* that scored low on the first administration of the DIT and the MJT before taking their test a second time, the participants were instructed to simulate the responses of a group *(leftists)* with average scores above their own ones. Although low scorers were able to elevate their preference for principled reasons *(P-scores)*, they could not simulate their moral-judgment competence (C-score) upward. Wasel (1994) replicated this finding in another experiment with the MJT In addition, he found that subjects with high C-scores were better able to correctly diagnose other people's moral-judgment competence than subjects with low *C-scores*.

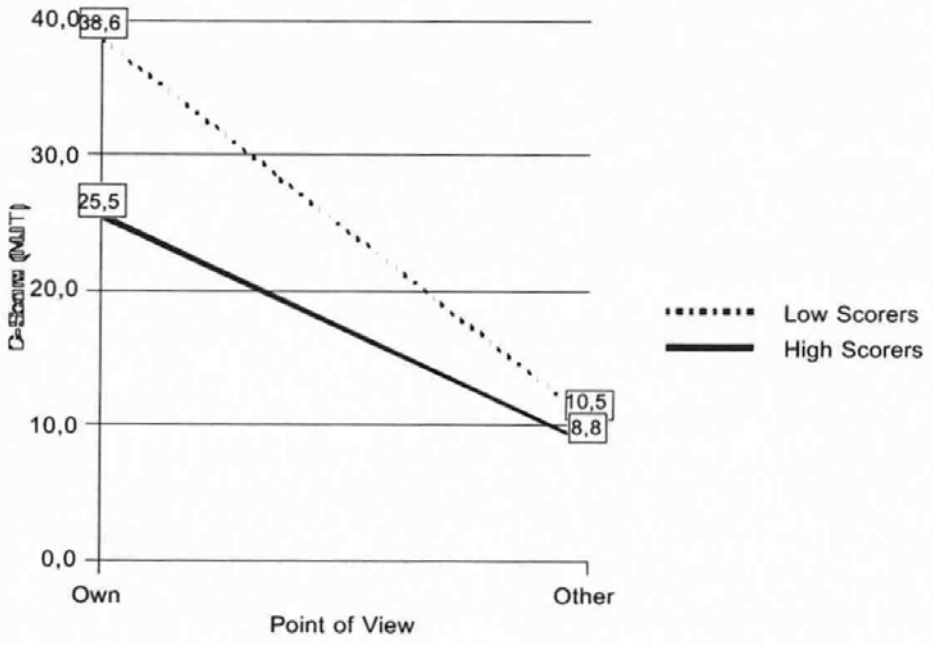
The competence nature of the *C-index* is also supported by the fact that upward and downward changes seem to happen gradually rather than abruptly (Lind, 2000a, 2002). Gradual changes are typical for the acquisition of abilities, but not for the change of attitudes, which may sometimes be very abrupt and dramatic when people change their social context. Moral competence also erodes slowly. The "forgetting curve" of subjects' *C-score* is negatively accelerated, that is, the longer participants do not practice their moral abilities, the faster they lose them.

In order to avoid tautologies or circularities in theory testing, we have not used hypotheses for validation like the hypotheses of invariant sequence, age-correlation and internal consistency of judgment behavior, which we

FIGURE 8.3. The Experimental Design of the MJT.
When being asked to simulate the moral judgment behavior, "Low Scorers" can simulate the moral *attitude* of "High Scorers" but not their moral judgment *competence*.



Emler et al. (1983) Experiment: The *P*-Score (DIT) *Can* be Simulated Upward.



Lind (2002) Experiment: The *C*-Score (MJT) *Cannot* be Simulated Upward.

wanted to test empirically. This is a problem with empirically validated tests of moral development. For example, Kohlberg and his colleagues revised their MJI several times to maximize the internal consistency of subjects' judgements and the correlation of MJI-scores with age: "The appropriate question is whether the interview and scoring system provides a valid assessment of moral judgment stage. . . . The Standard Issue Scoring . . . yields scores that agree very closely with the theoretical predictions of invariant sequence and internal consistency" (Colby et al., 1987a, p. 71). Hence, with MJI data, the two core hypotheses of cognitive-developmental theory, invariant sequence and structural wholeness, cannot be tested anymore without circularity. Of course, this restriction applies also to other tests, which have been optimized for internal consistency or correlation with age In sum, the standard version of the MJT is a valid measure of moral judgment competence and moral attitudes by virtue of test design and by five empirical criteria derived from the dual-aspect theory of moral behavior. This is not only true for the original German version but also for all twenty-nine certified translations of the MJT, which have been submitted to the same rigorous set of validity tests (For a List of certified versions of the MJT and recommendations regarding the validation process, see: http://www.unikonstanz.de/ag-mroral/.)

Further, the MJT has fared well in cross-national and cross-cultural settings, even though our validation criteria are demanding (Lind, 2005). I believe that this unlikely career of the instrument is not the least due to the universal validity of Kohlberg's core notions about the nature of morality and moral development, which was utilized when constructing the MJT in the early 1970s. Since then, only minor editorial changes have been made; none of them were necessitated by cross-cultural validation studies.

EMPIRICAL DISCOVERIES FACILITATED BY THE MJT

The MJT has facilitated research on hypotheses, which otherwise would not have been possible. Because the MJT allows us measure cognitive and affective aspects of moral judgment behavior simultaneously, differential hypotheses about the correlation between moral development and social behavior can now be tested empirically (Colesante & Biggs, 2001; Lind, 2002). Piaget's hypothesis of cognitive-affective parallelism can be studied directly without resorting to indirect indicators. MJT studies support this hypothesis so well that we can even use this parallelism as a criterion for evaluating the validity of new test versions (Lind, 2000a, 2002).

Because both aspects can be assessed and the MJT is not biased against regressions, it can also help us to distinguish between real regressions of

moral Judgment competence and pseudo-regressions. In MJT studies, only participants with the highest moral judgment competence showed signs of regression at the transition from high school to college/university (Lind, 2002). In addition, it was shown that only the competence aspect regressed, but not the affective aspect. That is, the regressors did not, as the Kohlberg and Kramer (1969) data seem to imply, suddenly prefer preconventional moral reasoning over principled reasoning. This supports the reinterpretation by Kohlberg and Higgins (1984) without making additional assumptions. However, with the MJT, we also discovered real competence regressions in youth, who had only nine to ten years of schooling (Lind, 2002), and also in medical students (Lind, 2000d; Schillinger, 2006).

The electronic version of the eMJT, called MJT, opens up a plethora of possibilities for testing psychologically important hypotheses about how an individual's moral-judgment behavior is affected by certain changes of context. The experimental design of the test can be altered and supplemented in many ways to study *the effect* of, for example, (1) changes of the response scale from a nine-point scale to shorter or longer scales; (2) different order of presenting the arguments (e.g., from a random, as it is, to a systematic ordering from Stage-1 to Stage-6 arguments, or vice versa); and (3) changing the order and number of dilemmas to be discussed.

Moreover, with the *eMJT*, the researcher can also assess the whole judgment process (e.g., the way a participant works on the test) and, most important, the latency times for each element of the *eMJT*. Response latency for each argument can be used as an indicator for the difficulty of the different stages of reasoning. In a laboratory experiment, Mansbart (2001) studied the time the subjects needed for deciding on the dilemma as a function of their moral-judgment competence (C-score) as compared to so-called motivators. He found that moral-Judgment competence was the best predictor of the decision-making time. Finally, two experimental studies have linked moral judgment competence to learning ability. Heidbrink (1985) found that students with high C-Scores learned more facts from a short film clip than students with low C-Scores. Lind and Knoop (2001) reported that several indicators for learning and applying newly learned things in adulthood correlate positively with moral-judgment competence. These findings add experimental evidence to the body of research showing that moral-judgment competence is highly relevant for everyday behavior (Kohlberg & Candee, 1984; McNamee, 1977).

Recently, in a fMRI study, neuro-psychologists have found a high correlation between low moral judgment competence (the C-score) and extended brain activities in the right dorso-lateral prefrontal cortex (DLPFC), indicating that people with low moral judgment competence need more time and energy for solving moral dilemmas (Prehn et al., 2008; added GL).

COMMENTS ON OTHER INSTRUMENTS

Three other instruments also are thought to give indices for the development of moral-judgment competence, the *Moral Judgment Interview (MJI), the Defining Issues Test (DIT) and the Sociomoral Reflection Measure (SRM).*

The Moral Judgment Interview (MJI) by Lawrence Kohlberg and his associates (Kohlberg, 1958; Colby et al., 1987) may be seen as a moral-competence test, but the design of the MJI and its scoring agrees only partially with the criteria of the dual-aspect theory of moral-judgment behavior. The original MJI incorporated the two central features of a moral task– dilemmas, and counter-suggestions (Kohlberg, 1958) – and was scored by looking at the whole structure of an individual's response pattern. After several revisions, the interview guide does not mention the use of counter-suggestions anymore (Colby et al., 1987a, 1987b), and the scoring has regressed to classical measurement theory, which strips off all relational or structural properties from the participants' responses. The authors now require "each item in the manual to clearly reflect the structure of the stage to which it is keyed" (Colby et al., p. 403; see also p. 410) rather than define the stage or quality of reasoning through the pattern or structure of responses of an individual. Finally, the MJI gives only mixed indices (stage score and MMS) for the affective and cognitive aspects of moral-judgment behavior and, thus does not allow us to identify the two aspects independently. (See Lind [1989] for an extensive discussion of the MJI.)

Nonetheless, the MJI may be regarded as a useful method for indexing moral-judgment competence if these shortcomings are considered (Lind, 1989). It seems that the MJI cannot be faked upward and thus its gain scores give a true picture of the effectiveness of educational interventions (Lind, 2002). When used in intervention studies, the MJI seems sensitive to education-induced change over a wide age range (Lind, 2002), whereas other instruments, like the DIT, seem to be sensitive only to changes in the adult age from conventional to postconventional stages (Schläfli, Rest, & Thoma, 1985).

If properly administered, the MJI seems to indicate the highest (rather than the typical) level of moral reasoning of which a person is capable. However, if not properly instructed, individuals may tend to reason below this level and get scores below their actual level. In a study of university students Schuhler (1977) found that the instruction to do their best elevated considerably the participants' level of moral reasoning. Kohlberg and Kramer (1969) reported that in late adolescence people tend to regress to lower stages of reasoning. Although Kohlberg and Kramer believed that this indicated a change in moral orientation, Kohlberg and Higgins (1984) argued that this regression is rather due to errors of the old scoring system. With MJT data, we found that regression is confined to the competence aspect and does not affect the subjects' moral

orienta tions, and that it is confined to about 20 percent of the participants with the highest C-Scores (Lind, 1985b). Finally, regression Scores may also indicate that the interview Situation does not challenge subjects enough to produce arguments at their highest level of moral reasoning. Although these findings seem to indicate that some forms of regression are only artifacts of the measurement method, Lind (2002) demonstrated that real regression of moral judgment competence can take place lf children are not provided with sufficient education to reach a critical level at which self-education takes up the lead in the process of development.

Other measures of moral development, like Rest's (1979) *Defining Issues Test* (DIT) and Gibbs, Basinger, and Fuller's (1992) *Socio-Moral Reasoning Measure* (SRM), seem to contain only few, if any, moral tasks.

Thus, they are valid instruments only for assessing people's moral *attitudes*. They do not provide indices for moral-judgment *competence* though under certain conditions moral attitudes correlate highly with moral competencies, and, therefore, may be used as an indirect indicator when direct indicators are not available.

With the DIT, the respondents are asked to give their judgment of the importance for several moral statements in relationship to a number of dilemmas. Each statement represents one of the Kohlbergian stages (from stage 1 to 5). The DIT contains dilemmas, but, as in the other tests, these play no role for scoring. The DIT does not contain Counter-suggestions; Rest (1979) even brings this feature of competence tests like the MJT and the (early) MJI into ridicule: "Ironically, [...] some researchers still regard such orthogonally balanced, pro-and-con Statements as a mark of design craftsmanship" (p. 89).

Most studies use as an index the P-score, which "provides a percent score that indicates the amount of postconventional thinking (in contrast to other kinds of thinking) preferred by the participant" (Narvaez, 1998, p. 15).[5] Other indices have been suggested, which basically give the same information though they seem to have some better statistical properties (Rest et al., 1999). If the research participants have no reason to fake their score upward, the P-score can be regarded as an (indirect) indicator of the level of moral development in many cases, because, as has been shown, moral attitudes toward postconventional moral reasoning correlate very consistently with moral judgment competence (Lind, 1985a, 2000a, 2002). If, however, the DIT is used as a test for evaluating educational interventions or admission to institutions of higher education, this correlation can break down. lf the participants desire to meet external expectations (e.g., to please an instructor or to meet the requirements of an institution), they may change their moral attitudes upward and thus let us overestimate their level of moral development or the effect of an intervention. "[Some] intervention studies that directly taught Kohlberg stages as part of the

learning experience . . . had pretest to posttest gains around 10 points, nearly twice as much as in the Panowitsch-Balcum study, which . . . seems to me to be the most concentrated, powerful interventions to facilitate development in moral thinking.. . . Thus these interventions may be only successful in instructing subjects how to fake high on the DIT posttest without invalidating the questionnaire by the *M-score*" (Rest, 1979, p. 218). The DIT is also of limited value for evaluating intervention studies because the DIT measures only preferences for postconventional (stage 5) reasoning and, therefore is insensitive to changes on lower stages and age ranges (Schläfli et al., 1985; Kim, 2006). Unfortunately, this methodological limitation of the DIT has misguided some textbook authors to conclude that effective moral education is only possible with adults.

With a competence index, it should not be possible to get higher scores just because someone has given the instruction to do so, though, of course, a competence index should be sensitive to real improvements due to practice and education. Recently, Rest et al. "have eliminated the faking study from our set of the validity criteria" (p. 115). Hence, they no longer seem to assume that the P-score or similar scores of the DIT are competence indices. The DIT is a valid test of moral attitudes and may be used as an indirect indicator of moral-judgment competence if it is reasonable to believe that the subjects had no desire to fake their scores upward. What about alternative indices to the DIT, like the *U-score* (U for utilizer; see Rest et al., 1999, p. 104) and the DIT-score, adapted from our C-score (Rest et al., 1997), which can be found in the literature? The so-called utilizer score *(U-score)* assesses the degree to which general moral attitudes are linked to specific moral choices. However, the U-score seems to have been incorporated only in a few studies (Thoma, 1994). The DIT C-score has been discarded because it was allegedly outperformed by the *P-score* on statistical criteria (Rest et al., 1997). However, the C-score, as we have defined it, is meaningful only in conjunction with a moral task as contained in the MJT. Because the DIT does not contain such tasks, the DIT-score is meaningless. [6]

The *Sociomoral Reflection Measure* (SRM; Gibbs, 1995; Gibbs, Basinger, & Fuller, 1992; see also Gielen, Comunian, & Antoni, 1994; Krettenauer & Becker, 2001) has been constructed as an economic replacement of the MJI, but with some unique theoretical positions in mind. The main (if not only) criteria for validity are (1) a high correlation with Kohlberg's MJI, and (2) a high correlation with age (Gibbs, 1995, p. 35; Gibbs et al., 1992; Krettenauer & Becker, 2001). The SRM does not contain a moral task; it has no counter-suggestions and even eliminates dilemmas. A sample item is: "How important is it for people to tell the truth?" Thus, among the three alternative tests discussed here, the SRM is the purest moral-attitude measure. Like the MJI and the DIT, the SRM may be used to give indirect indices of moral-competence development

only if the research participants have no reason to simulate their scores upward. To my knowledge, the SRM was never tested for fakeability, though one should expect that among the tests discussed here, the SRM can be most easily faked upward.[1]

In sum, while there are many good tests of moral *attitudes*, there seems to be no test, besides the MJT, which contains a difficult moral task and thus measures moral judgment *competence* in a theoretically valid way. Furthermore, the MJT is shorter, provides more indices, can be used with a wider age range (age ten and older) than most of the other tests, and is sensitive to education-induced change.

SUMMARY APPRAISAL OF THE MJT

The MJT is a theoretically valid, change-sensitive, short, computer-scorable instrument for the measurement of moral judgment competence and moral attitudes. It is designed for use in research and evaluation studies, but not for individual assessment or selection.

Our structural measurement approach contrasts to classical measurement theory and its modern derivatives like Rasch scaling, which sees the target trait to be represented by each response test item, and which analyzes structure on the basis of between-subject variance and item-correlations in a sample. Because the classical approach misinterprets response characteristics (like consistency and inconsistency) as characteristics of the test (reliability and measurement error), these approaches (e.g., Rasch scaling) are inadequate for assessing structural properties of an individual's thought and behavior, and, therefore, for testing the dual-aspect theory of moral development (Broughton, 1978; Lind, 1985c, 2000a).

The MJT is designed as an intra-individual behavioral experiment to measure some theoretically well-defined constructs. Confronted with a systematically constructed set of moral arguments, the participant's pattern of reactions tells us something about that individual's structure of moral reasoning. Rigorous testing of several hypotheses on the nature of moral reasoning structure derived from theory has confirmed an unusual degree of theoretical and empirical validity of the test. Hence, because the MJT is an experimental questionnaire rather than a test based on classical measurement, we interpret response consistency and inconsistency of a person's

[1] In the first edition of this paper (Lind, 2008), I claimed that the SRM was not sensitive to change in the study by Oser & Althof (1992, p. 162). I do not uphold this claim. Unfortunately, it is based on a mis-reading of the findings. Actually, in that study the Kohlberg-Interview was used, not the SRM.

response pattern as properties of that person's moral-cognitive structure rather than properties of the instrument, like measurement error or unreliability (see Lind, 2000a). Because the focus of measurement is on patterns and structures rather than on itemized information, it would be meaningless to submit the MJT to traditional item analysis or to calculate test reliability. Interestingly though, the MJT was not optimized for "reliability." Lerkiatbundit, Utaipan, Laohawiriyanon, & Teo (2004) report a test-retest correlation of r = 0.90.

With the MJT no items were selected to increase the correlation of the C-Index with empirical criteria like age, political attitudes, or higher education. This fact guarantees that the MJT, unlike other instruments, is not biased in favor of or against certain predictions like stability of rank orders among people, age-correlation, invariant sequence, or correlation with education. Most important, the items were not screened either to maximize stability of scores (reliability) at the expense of the test's sensitivity for education-induced change, nor were they selected to maximize sensitivity for change at the expense of theoretical validity.

The MJT is much shorter than most other tests of moral development, (the standard version has twenty-six items), it is easy to administer and fully scorable by a computer. Like Kohlberg's Moral Judgment Interview, the MJT can already be used with fifth graders (the DIT has a testing floor of ninth graders; Rest et al., 1999). There are hardly any non-scorable cases (as compared to up to 50 percent non-scorable cases in DIT studies; see Gielen, Comunian, & Antoni, 1994).

Although the MJT cannot be faked upward in experimental test situations in which other tests of moral development have shown that they can be faked upward (Lind, 2000, 2000b), it is sensitive to educational changes. As with any ability test, the problem of retesting weariness can depress gain scores and thus can lead to an underestimation of treatment effects. Proper instructions seem to solve this problem. In recent intervention studies, very high effect sizes r > 0.70) have been reported indicating that the MJT is sensitive to real increase in moral judgment competence (Lind, 2008, Lerkiatbundit et al., 2006).

The MJT can be applied in a wide developmental range. Originally, the MJT was constructed for use in studies with university students. Meanwhile, its application has been successfully widened to encompass studies with children as young as 10 years of age, high school students, working-class youth, delinquents, professionals, political activists, senior people, and so on, in many countries. The test has been translated into, and validated for, twenty-nine languages. Data from the Standard MJT are now available from many cross-sectional, longitudinal, experimental, and Intervention studies. The MJT is less value loaded than other tests, but it is not value free. The moral ideal on which the test is based can be justified on a universal basis and, hence, I believe,

makes it acceptable in most cultures. With this test, we acknowledge that questions like "Which solution is the right one in a particular conflict situation?" and "On which level of moral reasoning should it be discussed?" are matters of cultural definition, and we believe that answers to such questions should not be made the base of *moral* evaluation. Therefore, the central index of the MJT, the *C-score*, does *not* depend on such questions. However, the test is based on the belief that we should expect from everyone moral-judgment competence, that is, the ability to make judgments in accord with one's *own* moral principles and to *behave* consistently in regard to these principles even though other feelings oppose them.

In some sense, the MJT builds upon and summarizes what we know now about moral behavior and its core aspects: moral *ideals* and moral *competencies*. However, it makes the starting point rather than the endpoint of a new research paradigm, helping to generate new questions about the nature, conditions, and effects of moral action and development: What are the causes of moral segmentation (Lind, 2000c)? Can the influence of role-taking opportunities on moral competencies be generalized (Lind, 2000b)? How is moral judgment competence related to human learning (Heidbrink, 1985; Lind & Knoop, 2001) and decision making (Mansbart, 2001; Prehn et al., 2008)? Eventually we will need new tests tailored to special research questions and even new design features. However, so far it seems that the possibilities of the standard MJT for answering these and other questions have not been exhausted yet (see Bataglia et al., 2002, 2008; Lind, 2003b).

NOTES

1. The design and construction of the MET rest both on the dual-aspect theory of moral judgment behavior, as outlined earlier and on modern, cognitive-structural and experimental approaches to psychological measurement. We used pertinent ideas from Egon Brunswick (1955; the diacritical method), Norman H. Anderson (1991; cognitive algebra), George A. Kelly (1955; personal constructs), Leon Goodman (1961; facet analysis) and other authors (Lohman & Ippel, 1993; Mischel & Shoda, 1995; for the concept of experimental questionnaire, see Lind, 1982). Our experimental questionnaire approach, we believe, is an adequate translation of the clinical method advocated by Piaget and Kohlberg into a more formalized, objective measurement methodology.
   Within these methods, the basic unit of measurement of the MET is the total response pattern of an individual subject and the within-subject response variance. The subject's moral-judgment competence and the structural organization of his moral judgment is to reveal itself in his or her whole response pattern rather than in each and every item as is postulated by classical measurement theory

and its modern derivatives like the Rasch scaling method. In the Moral Judgment Test, the dependent variable is the subjects' judgment of acceptability on a scale from -4 to +4 (reject/accept). The independent variables or design factors are: (1) moral quality (or stage) of reasoning, (2) the task factor (opinion-agreement), and (3) the dilemma type. Hence, the MET represents a 6 x 2 x 2 factorial design. In combination, they are to elicit a response pattern that lets us infer the moral-judgment competence of the subject. In contrast, classical test designs are based only on one design factor. If other factors are reported, they typically represent variance between subjects rather than intra-individual variance. Moreover, classical approaches regard intra-individual variance as error variance rather than as a manifestation of an individual's cognitive structure.

2. We use the six-stages model, which Kohlberg (1958) had originally proposed and finally revised (Kohlberg, Boyd, & Levine, 1990), after he had abandoned Stage 6 for a while.
3. My gratitude for those colleagues who critically examined the first versions of the MJT has grown over the years: Tino Bargel, Rainer Döbert, Thomas Krämer-Badoni, Ekkehard Lippert, Gertrud Nunner-Winkler, Gerhard Portele, Roland Wakenhut, and Hans-Gerhard Walter.
4. Our validation procedure is more rigorous than that of classical measurement theory. The validation criteria, which are derived from accumulated research and theorizing, set highly specified standards, which cannot he easily met by chance alone. In a similar unprecedented, rigorous way, all foreign language versions of the MJT are empirically analyzed to establish their cross-cultural validity before they are certified. These cross-sectional, longitudinal, experimental and intervention studies comprise participants ranging from 10 to over 90years of age, of both genders, and of widely differing educational and social backgrounds and cultures (Lind, 1978, 2000a/1984). Validated versions of the standard version of the M,/"I' are available in German and in twenty-nine additional languages, e.g., in English, Spanish, Italian, Portuguese, French, Russian, Dutch, Hebrew, Finnish, Polish, Czech, Macedonian, Philippine (for updates on the MJT see http://democracy-education.net).
5. Rest (1979): “The P index [. . .) is interpreted as the relative importance give(n) to principle moral considerations on making a moral decision” (p. 101); see also Rest (1986).
6. Yet it is possible to define a *P-score* for the Moral judgment Test because it is also a moral attitude test (Colesante & Biggs, 2001).
7. The asterisked articles (*) are also available on the Internet: http//www.uni-konstanz.de/ag-moral/

REFERENCES [7]

Anderson, N. H. (1991). Moral-social development. In N. H. Anderson (Ed.), Information integration theory, Volume III: Developmental (pp. 137-187). Hillsdale, NJ: Erlbaum.

Bataglia, R, Agati, M., Siva Torres, S., Zanoni Crivelaro, D. B. et al. (2002, November). The development of moral competence and religious commitment in Brazil. Paper presented at the Meeting of the Association for Moral Education, Chicago.

Bataglia, P, Schillinger-Agati, M. M., Lind, G., & Quevedo, T. L. (2003, July). Testing the segmentation hypothesis with an extended version of the MJT. Poster presented at the meeting of the Association for Moral Education, Krakow.

Blasi, A. (1987). Comment. The psychological definitions of morality. In J. Kagan & S. Lamb (Eds.), The emergence of morality in young children (pp. 83-90). Chicago: University of Chicago Press.

Bloom, B. S., Engelhart, M. D., Hill, W H., Furst, E. J., & Krathwohl, D. R. (1956). Taxonomy of Educational objectives. Handbook 1: Cognitive domain. New York: McKay.

Broughton, J. (1978). The cognitive-developmental approach to morality: A reply to Kurtines and Greif. Journal of Moral Education 1, 81-96.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. Psychological Review, 62, 193-217.

Colby, A., Kohlberg, L., Abrahami, A., Gibbs, J., Higgins, A., Kauffman, K., Lieberman, M., Nisan, M., Reimer, J., Schrader, D., Snarey, J., & Tappan, M. (1987a). The measurement of moral judgment. Volume I, Theoretical foundation and research validation. New York: Columbia University Press.

Colby, A., Kohlberg, L., Speicher, B., Hewer, A., Candee, O., Gibbs, J., & Power, C. (1987b). The measurement of moral Judgment. Volume II, Standard issue scoring. New York: Columbia U Press.

Colesante, R. J., & Biggs, D. A. (2001, October). A controversy about the measurement of moral development: Stage preference or moral consistency? Paper presented at the Association for Moral Education Conference, Vancouver, Canada.

Comte, A. (1975). Cours de philosophie positive. 2 vols. Paris, Hermann.

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibilities. Journal of Personality and Social Psychology, 8, 377-383. Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. Annual Review of Psychology, 41, 525-556.

DeVries, R. (1997). Piaget's social theory. Educational Researcher, 26(2), 4-17.

Dewey, J. (1966). Democracy and education. An introduction to the philosophy of education. New York: The Free Press.

Emler, N., Renwick, S., & Malone, B. (1983). The relationship between moral reasoning and political orientation. Journal of Personality and Social Psychology, 45,1073-80.

Gibbs, J. C. (1977). Kohlberg's stages of moral judgment. A constructive critique. Harvard Educational Review, 47, 43-61.

Gibbs, J. C. (1995). The cognitive developmental perspective. In W M. Kurtines & J. L. Gewirtz (Eds.), Moral development: An introduction (pp. 27-48). Boston: Allyn & Bacon.

Gibbs, J. C., Basinger, K. S., & Fuller, R. (1992). Moral maturity: Measuring the development of socio-moral reflection. Hillsdale, NJ: Erlbaum.

Gibbs, J. C., Widaman, K. F., & Colby, A. (1982). Construction and validation for a simplified, group-administrable equivalent to the Moral Judgment Interview. Child Development, 53, 895-910.

Gielen, U. P., Comunian, A. I.., & Antoni, G. (1994). An Italian Cross-sectional study of Gibbs' Socio-moral Reflection Measure-short form. In U. P. Gielen & A. I. Comunian (Eds.), Advancing psychology and its application: International perspectives (pp. 125-134). Milan, Italy: Franco Angeli.

Gigerenzer, G., & Murray, D. J. (1987). Cognition as intuitive statistics. Hillsdale, NJ: Erlbaum.

Gross, M. (1997). Ethics and activism. The theory and practice of political morality. New York: Cambridge University Press.

Grün, M.v.d. (1975). Stellenweise Glatteis. Neuwied: Luchterhand.

Goodman, L. (1961). Measurement as structural theory. Psychometrika, 36, 329-347.

Habermas, L. (1985). Philosophical notes an moral judgment theory. 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. 3-20). Chicago: Precedent Publishing.

Habermas, J. (1990). Moral consciousness and communicative action (C. Lenhard & S. W Nicholson, Trans.). Cambridge: MIT Press.

Hartshorne, H, & May, M. A. (1928). Studies in the nature of character. Vol. I: Studies in deceit, Boole one and two. New York: Macmillan.

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 (pl). 259-271). Chicago: Precedent Publishing.

Heidbrink, H. (1989). Moralpsychologie: Die Reanalyse einer Untersuchung von 1912 [Moral psychology. The reanalyses an experiment of 1912]. Geschichte der Psychologie, 17, 4-9.

Kant, I. (1948/1785). Groundwork of the metaphysic of morals (H. Paton, Ed.). London: Hutchinson (Original 1785).

Keasey, C. B. (1974). The influence of opinion-agreement and qualitative supportive reasoning in die evaluation of moral judgments. Journal of Personality & Social Psychology, 30, 477-482.

Kelly, G.A. (1955). The psychology of personal constructs. New York: Norton.

Kim, S. (2006, April). A comparative study an measurement instruments for moral judgment: Selecting an appropriate instrument. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Kohlberg, L. (1958). The development of modes of moral thinking and choice in the years 10 to 16. Unpublished doctoral dissertation, University of Chicago.

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. (1979). Foreword. In J. R. Rest, Development of judging moral issues (pp. vii-xvi). Minneapolis: University of Minnesota Press.

Kohlberg, L. (1984). The psychology of moral development. San Francisco: Harper & Row.

Kohlberg, L., Boyd, D., & Levine, C. (1990). The return of Stage 6: Its principle and moral point of view. In T. E. Wren (Ed.), The moral domain: Essays in the ongoing discussion between philosophy and the social sciences (pp. 151-181). Cambridge, MA: MIT Press.

Gielen, U. P., Comunian, A. L.., & Antoni, G. (1994). An Italian cross-sectional study of Gibbs' Socio-moral Reflection Measure-short form. In U. P. Gielen & A. L. Comunian (Eds.), Advancing psychology and its application: International perspectives (pp. 125-134). Milan, Italy: Franco Angeli.

Gigerenzer, G., & Murray, D. J. (1987). Cognition as intuitive statistics. Hillsdale, NJ: Erlbaum.

Gross, M. (1997). Ethics and activism. The theory and practice of political morality. New York: Cambridge University Press.

Grün, M.v.d. (1975). Stellenweise Glatteis. Neuwied: Luchterhand.

Goodman, L. (1961). Measurement as structural theory. Psychometrika, 36, 329-347.

Habermas, J. (1985). Philosophical notes an moral judgment theory. 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. 3-20). Chicago: Precedent Publishing.

Habermas, J. (1990). Moral consciousness and communicative action (C. Lenhard & S. W Nicholson, Trans.). Cambridge: MIT Press.

Hartshorne, H., & May, M. A. (1928). Studies in the nature of character. Vol. I: Studies in deceit, Book one and two. New York: Macmillan.

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.

Heidbrink, H. (1989). Moralpsychologie: Die Reanalyse einer Untersuchung von 1912 [Moral psychology. The reanalyses an experiment of 1912]. Geschichte der Psychologie, 17, 4-9.

Kant, I. (1948/1785). Groundwork of the metaphysic of morals (H. Paton, Ed.). London: Hutchinson (Original 1785).

Keasey, C. B. (1974). The influence of opinion-agreement and qualitative supportive reasoning in the evaluation of moral judgments. journal of Personality & Social Psychology, 30, 477-482.

Kelly, G.A. (1955). The psychology of personal constructs. New York: Norton.

Kim, S. (2006, April). A comparative study an measurement instruments for moral judgment: Selecting an appropriate instrument. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Kohlberg, L. (1958). The development of modes of moral thinking and choice in the years 10 to 16. Unpublished doctoral dissertation, University of Chicago.

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. (1979). Foreword. In J. R. Rest, Development of judging moral issues (pp. vii-xvi). Minneapolis: University of Minnesota Press.

Kohlberg, I.. (1984). The psychology of moral development. San Francisco: Harper & Row.

Kohlberg, L., Boyd, D., & Levine, C. (1990). The return of Stage 6: Its principle and moral point of view. In T. E. Wren (Ed.), The moral domain: Essays in the ongoing discussion between philosophy and the social sciences (pp. 151-181). Cambridge, MA: MIT Press.

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.

Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. Human Development, 12, 93-120.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of educational objectives. Handbook II: Affective domain. New York: McKay.

Krebs, D. (1982). Altruism: A rational approach. In N. Heisenberg (Ed.), The development of prosocial behavior (pp. 53-76). New York: Academic.

Krebs, D. L., Vermeulen, S. C. A., Carpendale, J. L., & Denton, K. (1991). Structural and situational influences an judgment: The interaction between stage and dilemma. In W. M. Kurtines & J. L. Gewirtz (Eds.), Handbook of moral behavior and development. Volume 2: Research (pp. 139-169). Hillsdale, NJ: Erlbaum.

Krettenauer, T, & Becker, G. (2001). Entwicklungsniveaus sozio-moralischen Denkens [Developmental levels of socio-moral thinking]. Diagnostica, 47,188-195.

Kuhn, T (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kurtines, W M., & Gewirtz, J. Eds.. (1995). Moral development: An introduction. Boston: Allyn & Bacon.

Lerkiatbundit, S., Utaipan, P., Laohawiriyanon, C., & Teo, A. (2006). Randomized controlled study of the impact of the Konstanz method of dilemma discussion an moral judgment. Journal of Allied Health, 35(2), 101-108.

Levy-Suhl, M. (1912). The examination of the moral maturity of juvenile delinquents [Die Prüfung der sittlichen Reife jugendlicher Angeklagter und die Reformvorschläge zum §6 des deutschen Strafgesetzbuches]. Zeitschrift für Psychotherapie, 232-254.

Lind, G. (1978). Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts. [How does one measure moral judgment? Problems and alternative ways of measuring a complex construct]. In G. Portele (Ed.), Sozialisation und Moral (pp. 171-201). Weinheim: Beltz.*

Lind, G. (1982). Experimental questionnaires: A new approach to personality research. In A. Kossakowski & K. Obuchowski (Eds.), Progress in the psychology of personality (pp. 132-144). Amsterdam: North-Holland.*

Lind, G. (1985a). The theory of moral-cognitive judgment: A socio-psychological assessment. 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. 21-53). Chicago: Precedent.*

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

Lind, G. (1985c). Attitude change or cognitive-moral development? How to conceive of socialization at the university. 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. 173-192). Chicago: Precedent.

Lind, G. (1986). Cultural differences in moral judgment? A study of West and East European university students. Behavioral Science Research, 20, 208-225.*

Lind, G. (1987). Moral competence and education in democratic society. In G. Zecha & P. Weingartner (Eds.), Conscience: An interdisciplinary approach (pp. 37-43). Dordrecht: Reidel.

Lind, G. (1989). Measuring moral judgment: A review of The Measurement of Moral judgment by Anne Colby and Lawrence Kohlberg. Human Development, 32, 388-397.

Lind, G. (1996). The optimal age of moral education. A review of intervention studies and an experimental test of the dual-aspect-theory of moral development and education. Paper presented at the annual meeting of the American Educational Research Association, New York.*

Lind, G. (2000a). Content and structure of moral judgment (2nd, corrected ed.). Doctoral dissertation, University of Konstanz. (Original work published 1984).*

Lind, G. (2000h). The importance of role-taking opportunities for self-sustaining moral development. Journal of Research in Education, 10, 9-15.*

Lind, G. (2000c, April). Off Limits: A cross-cultural study on possible causes of segmentation of moral judgment competence. Round-table paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA..*

Lind, G. (2000d). Moral regression in medical students and their learning environment. Revista Brasileira de Educacao Medica, 24(3), 24-33.*

Lind, G. (2002). Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung [Can ethics be taught? Research findings from modern moral psychology] (2nd ed.). Berlin: Logos-Verlag.

Lind, G. (2003b, July). Does religion foster or hamper morality and democracy? Invited lecture at the annual meeting of the Association for Moral Education, Krakow, Poland.*

Lind, G. (2005, August). The cross-cultural validity of the Moral Judgment Test: Findings from 27 cross-cultural studies. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

Lind, G. (2008). Moral ist lehrbar. Ein Handbuch zu Theorie und Praxis moralischer und demokratischer Bildung. [Morality can be taught. A handbook on theory and practice of moral and democratic education.] München: Oldenhourg.(2nd enlarged edition).

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

Lind, G., & Knoop, P. (2001). Weiterbildungsseminar: Mentorik für Gymnasiallehrer [Continuing education program: Mentoring for secondary school teachers.] University of Konstanz.*

Lind, G., Sandberger, J.-U., & Bargel, T. (1985). Moral competence and democratic personality. 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. 55-78). Chicago: Precedent.*

Lind, G., & Wakenhut, R. (1985). Testing for moral judgment competence. 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. 79-105). Chicago: Precedent.*

Lohman, D. F., & Ippel, M. J. (1993). Cognitive diagnosis: From statistically based assessment toward theory-based assessment. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), Test theory for a new generation of tests (pp. 41-71). Hillsdale, NJ: Erlbaum.

Lourenco, O., & Macchiato, A. (1996). In defense of Piaget's theory: A reply to 10 common criticisms. Psychological Review, 103, 143-164.

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, Dept of Psychology.

Markoulis, D. (1989). Political involvement and socio-moral reasoning: Testing Emler's interpretation. British Journal of Social Psychology, 28, 203-212.

McDougall, W (1908). An introduction to social psychology. London: Methuen.

McNamee, S. (1977). Moral behavior, moral development and motivation. Journal of Moral Education, 7(1), 27-31.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23(2), 13-23.

Mischel, W, & Shoda, Y. (1995). A cognitive-affective system theory of personality: Re-conceptualizing situations, dispositions, dynamics, and invariance in personality structure. Psychological Review, 102, 246-268.

Narvaez, D. (1998). The influence of moral Schemas an the reconstruction of moral narratives in eighth graders and college students. Journal of Educational Psychology, 90, 13-24.

Piaget, J. (1965). The moral judgment of the child. New York: The Free Press. (Original work published 1932)

Piaget, J. (1976). The affective unconscious and the cognitive unconscious. In B. Inhelder & H.H. Chipuran (Eds.), Piaget and his school (pp. 63-71). New York: Springer.

Piaget, J., & Inhelder, B. (1969). The psychology of the child. New York: Basic Books (Original work published 1966).

Pittel, S. M., & Mendelsohn, G. A. (1966). Measurement of moral values: A review and critique. Psychological Bulletin, 66, 22-35.

Prehn, K., Wartenburger, I., Meriau, K., Scheibe, C., Goodenough, O., Villringer, A., van der Meer, E., & Heekeren, H. (2008). Influence of individual differences in moral judgment competence on neural correlates of socio-normative judgments. Social Cognitive and Affective Neurosience. 3(1), 33-46.

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

Rest, J. R. (1986). Moral development: Advances in research and theory. New York: Praeger.

Rest, J. R., Thoma, S. J., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preferences and stage consistency approaches. Journal of Educational Psychology, 89, 5-28.

Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S.J. (1999). Postconventional moral thinking. A neo-Kohlbergian approach. Mahwah, NJ: Erlbaum.

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.

Schuhler, P. (1977). Anregungsbedingungen zur Förderung der moralischen Urteilsreife [Stimulus conditions to foster moral judgment competence]. Unpublished master thesis (Diplomarbeit), University of Saarbrücken.

Simmel, G. (1989). Einleitung in die Moralwissenschaft [Introduction into the moral sciences]. Frankfurt: Suhrkamp. (Original work published 1892)

Thoma, S. J. (1994). Moral judgments and moral action. In J. R. Rest & D. F. Narvaez (Eds.), Moral development in the professions: Psychology and applied ethics (pp. 199-211). Hillsdale, NJ: Erlbaum.

Walker, L. J. (1983). Sources of cognitive conflict for stage transition in moral development. Developmental Psychology 19, 103-110.

Walker, L. J. (1986). Cognitive processes in moral development. In G. L. Sapp (Ed.), Handbook of moral development (pp. 109-145). Birmingham, AL: Religious Education Press.

Wasel, W. (1994). Simulation of moral judgment competence [Simulation moralischer Urteilsfähigkeit]. Unpublished thesis in Psychology. University of Constance.

Weber, M. (1994). Political writings. New York: Cambridge University Press.

For more references on the *Moral Judgment Test* (MJT) see:
http://www.uni-konstanz.de/ag-moral/
or
http://democracy-education.net

Research studies founding and using the MJT can be found on this web-site:
http://www.uni-konstanz.de/ag-moral/mut/mjt-references.htm