Abstract

According to modern psychology, morality is a matter not only of attitudes towards moral principles but also of people's competence to utilize those principles. In a series of longitudinal and cross-cultural studies, comprising several thousand subjects, both aspects have been assessed simultaneously by using the Moral Judgment Test, based on the methodology of Experimental Questionnaire (EQ). Similar to Kohlberg's methodology, this test contains a discourse about moral dilemmas. However, with the MJT, the discourse between the subjects and the interviewer is designed as a multivariate task. The design of the MJT includes three factors:

(a) Types of dilemma,
(b) stages of reasons discussed,
(c) opinion agreement or disagreement about the solution.

Its main function is to assess individuals' competence of making consistently moral judgments in a dispute situation (rather than merely defending one's own opinion).

In this presentation, major findings of the studies mentioned will be reviewed, and their implications for both the validity of cognitive theory of moral development and for educational research in the moral domain will be discussed.
Theory and Method

Piaget and Kohlberg have revolutionized the way we are looking at moral thought and behavior. Essentially, they have shown that morality is not, as so many had assumed, purely a matter of having right moral attitudes or knowing moral terminology, but that morality is a competence which needs to be developed, and measured, much the same way as other competences are developed and measured. This revolution in theorizing has many implications for psychological and educational practice, which are not fully understood.

As I shall argue, their cognitive-developmental theory is completely at odds with the psychological assumptions underlying prevailing research methodologies, i.e., attitude measurement and test theory. The key issues is the definition and operationalization of “structure,” as opposed to content, which plays a pivotal role in cognitive-developmental theory - but its methodological meaning still riddles opponents as well as proponents of the theory (see Kurtines & Greif 1974, Flavell 1977, Rest 1979, Kohlberg 1979).

The mainstream psychometrician would hold that structure can only be conceived of as a “latent variable,” or “underlying thought organization” that is not directly observable but may, at best, be only inferred through a set of probabilistic indications. If they speak of measuring structure, they mean a set of relations among variables in a group of persons rather than within an individual person. Some cognitive-developmentalists, among them Piaget and Kohlberg, would hold that, through the methodology of clinical interviews, it is possible to penetrate the “surface of behavior” (“content”) to get at the deeper layers of a person's behavior, i.e., its “structure” (see Colby & Kohlberg, 1987). But we cannot be sure that asking many “why” questions really purges content from structure (Rest 1979, pp. 42-43; Lickona 1976, p. 13), nor do we know what's left when we squeeze out situation-bound behavior to get at “pure thought” (see Lind 1989). Though disagreeing in some aspects, both main-stream and cognitive-developmental methodologies have one conjecture in common: they believe that structure is a component that is separable, if not separated, from behavior.

In my view, this notion of separate components is the very cause of the methodological riddles. It is at square with the explicit statements of cognitive-developmental theory which speaks of structure as the organization of behavior, e.g., its degree of consistency, differentiation and integration (Kohlberg 1984), and which speaks of structure of behavior. Moral judgment competence is defined as “the capacity to make decisions and judgments which are moral (i.e., based on internal principals) and to act in accordance with such judgments” (Kohlberg 1964, p. 425; emphasis added). Hence, on the outset, psychometricians ought to “care about ... how moral judgments are made when [a moral principle] is actually applied to values in conflict” (Colby & Kohlberg 1989, p. 58; emphasis added). In order to close this gap between psychological theory and psychological and educational measurement, we have developed the Moralisches
Urteil Test (MJT), which represents a new concept of measurement, the concept of Experimental Questionnaire (Lind 1982). This concept rests on the notion that structure cannot be separated from behavior, though it can be distinguished from latter. Structure is a property of a pattern of behavior exhibited by a person. Behavioral pattern may be differently structured, e.g., high or low consistent, differentiated or integrated, but structure can be separated neither from behavior nor from situation, as wetness cannot be separated from water and from the one who feels that water is wet. Accordingly, with Experimental Questionnaires, moral structure is defined as a particular pattern of judgment behavior in a moral situation, in which two or moral values conflict.

The “Moral Judgement Test” (MJT)

In the MJT, structure is precisely and objectively defined as the degree to which a subject's judgment pattern is determined by the moral features of the task set before him or her. The MJT provides a scale with which we can measure the degree to which a respondent's judgments are morally consistent (as opposed to the degree to which his or her judgments are consistent in regard to non-moral aspects of the situation).

The MJT has borrowed many features from Piaget's and Kohlberg's clinical interview methodology though it differs from them in some decisive aspects. The test contains stories telling about a moral dilemma. The respondent is first asked to state his or her opinion about the way the protagonist solves the dilemma. Thereafter, the subject is to evaluate a number of Stage-typical reasons. These reasons represent pro and con views, and they advance several (six) levels of moral reasoning, conceived according to Kohlberg's Stage theory of moral development (Kohlberg 1984). The subject is asked to indicate his or her preferences for Stage-typical moral reasons. The subject does not need to be self-consciously aware of his or her own stage structure in order to exhibit structures of judgment behavior, nor does the research need to rely on the subject's own abstractions of structure in the process of measurement. Structural features can be abstracted from preference judgments as well as from spontaneously produced reasoning. Hence, Kohlberg's (1979, p. xiv) global critique of preference tests is unwarranted. The real problem is how we ought to “abstract” structure from behavioral pattern.

The MJT differs from any other moral judgment test in that the task of solving moral dilemmas is regarded as central and, therefore, is conceptualized as a multi-factorial, orthoiiogonally designed experiment with each factor being selected on the basis of psychological theory rather than as a psychometric test. According to Torgerson the distinction between the MJT and other in-struments such as Kohlberg's MJI or Rest's DIT it one between the “subject-centered approach” on the one side and the “response approach” on the other (Torgerson 1967, p. 46-48). Former approach has also been labeled as “measurement by definition.” This term points to the fact that this type of measurement requires a well developed theory on which definitions can be
based. Unfortunately, in psychology “definitions” are lacking and, for this reason, psychometricians have hardly ever dealt with this branch of measurement (which is so eminent in the natural sciences). It is not at all, as Torgerson and others seem to believe, confined to “the one-way classification in the analysis of variance” (Torgerson 1968, p. 46). To researchers who are interested in the structure of judgment and in the interplay of several dispositions of an individual person in dilemma situations, multifactorial measurement by definition is a very powerful, if not the only tool. All other traditional methods rely on groups of individuals (thus confounding individual structure with inter-individual differences) and preclude the study of interacting dispositions. The MJT provides unconfounded measures of individual judgment structure as well as of moral content, i.e., subject's attitudes toward particular moral orientations (see Lind & Wakenhut 1983). It thus allows to assess simultaneously cognitive and affective aspects of moral judgment behavior (Lind 1986).

The individual scores for the cognitive-structural aspect are computed by multivariate analysis of variance components for each individual pattern of judgment behavior. The resulting coefficients of determination are taken to indicate the degree to which moral criterions are cognitively anchored in the individuals' judgment. A scale value of 'zero' means that the respondent's judgment behavior is not determined by moral concerns. A value of 'one hundred' indicates that his or her judgment is completely determined by moral considerations. Since this structural measure is conceptually different from Kohlberg's we have refrained from assigning Stage numbers. Nevertheless, it claims theoretical validity. So far, studies with the MJT have well corroborated its developmental properties, in particular the hypotheses of sequentiality and of age-relatedness (Heidbrink 1983; Lind 1983 b). The affective aspect is defined, alike classical attitudes, as the direction and the strength of the respondents' affective commitment to Stage-typical moral concerns. For each of the six Kohlberg Stages of moral judgment, we have assessed the degree to which the subject accepts, or rejects, Stage-typical reasons for and against a particular solution of a behavioral dilemma. The test contains two dilemmas, the euthanasia dilemma by Kohlberg and the working men's dilemma which is taken from a German prose (for the standard version of the MJT see the appendix). We claim that the psychometric concept of Experimental Questionnaire agrees better with cognitive-developmental theory, and therefore, more adequate for measuring moral judgment, than classical test theory prevailing in education and psychology. Moreover, with the MJT it is possible to measure structure of moral judgment directly, and independently from affective aspects, and hence it is for the first time possible to test the hypothesis of affective cognitive parallelism (Lind 1986). Let me discuss both claims in turn.
Methodological Implications of Cognitive-Developmental Theory

Cognitive-developmental theory (Kohlberg 1958, 1964, 1969, 1984), which has greatly contributed to our rethinking of the nature of moral behavior and moral development is basically at odds with the behavioristic assumption underlying mainstream educational measurement:

a. Individuality. The basic unit of cognitive-developmental inquiry is the moral judgment of individual persons, whereas classical measurement methodology, as well as most alternatives that have been suggested, involve only the judgment of samples of persons.

b. Structural Uniqueness. The core assumption of cognitive developmental theory is that, due to developmental processes, there are basic structural differences in moral behavior between persons (apart from differences on single behavioral dimensions), whereas mainstream test theory overlooks them or regards them indiscriminately as “measurement error,” rather than as unknown variance.

c. Unity of Affect and Cognition. The cognitive-developmental approach postulates a “methodological” unity of affect and cognition, whereas the “hidden psychology” of behavioristic educational measurement states that both aspects belong to completely different domains and, therefore must also be assessed through different methodologies (cf. Krathwohl et al., 1964).

d. Pattern of Behavior. Cognitive-developmental theory maintains that morality is real, and can be measured, only if it affects behavior, whereas classical measurement theory speaks of morality in terms of “hypothetical constructs,” or “latent traits.”

This issue cannot be settled solely on the ground of empirical data, since the quarrel is about the quality of data. For example, since classical test theory overlooks structural differences, is does not produce data on a persons' personality structure, and, hence, there is no basis for reconciling opposite assumptions. However, this is not to mean, that we ought to dismiss behavioral data altogether in favor of introspection and interpretative speculations. Rather I am asking for a new type of methodological behaviorism that takes psychological and educational theory seriously, and thus produces higher quality, i.e. structural, data. In accordance with cognitive-developmental theory, enlightened behavioristic methodology should allow us

- to study individual moral behavior and moral dispositions without involving data from a randomly selected sample of individuals,
- to assess moral structure, that is, sets of relations among elements of moral behavior, and
- to measure moral affect and cognition as aspects of moral judgment pattern simultaneously, rather than as separate behavioral domains or components.
The last “should” needs a qualification. While affective and cognitive aspects should be measured simultaneously, the assessment of moral behavior and development should produce distinct scores for either aspect rather than combining them into one. Otherwise it would be impossible to study the interplay of these two aspects.

A Task Conception of Test Design

Morality is adequately understood only if we regard it both as an attitude syndrome and as a competence structure (rather than either as an attitude or as a cognition). The discovery of this unity of the affective and the cognitive side of morality certainly is Piaget’s and Kohlberg’s major achievement. Indeed, morality implies an affective attachment to some moral rules or principles, and knowledge of specific rules of conduct. Yet, it seems that neither is sufficient for achieving stable and differentiated solutions in a moral conflict situation. A person must (a) sensitive for the moral issues involved in a conflict, and (b) he or she must be able to apply the preferred principles to particular problems. In other words, morality implies moral judgment competence, that is, in Kohlberg’s terms,

“the capacity to make decisions and judgments which are moral (i.e., based on moral principles) and to act in accordance with such judgments.” (Kohlberg 1964, p. 425)\(^3\)

This definition does not mean that a person must always act in accordance with one particular moral rule (such behavior would be rightly called “rigid” rather than moral) but that he or she acts integratedly and differentiatedly in accordance with several moral principles. As Kohlberg once put it,

“The solution [of a moral problem] must do justice both to what the self believes and yet meet the situation” (Kohlberg 1958, S. 118).

Conceiving morality as a competence implies a task conception of psychometric test design. By the way a subject solves this task we may make inferences on the degree to which he or she is able to make integrated and differentiated moral judgments, that is, which dispositions are at work when a particular subject solve a certain moral problem. In recent research on moral-cognitive development three major dispositions or dispositional syndromes have been identified as influencing moral judgment (Lind 1989):

\(^3\) Judgment competence, the cognitive aspect of morality, is acquired in the process of development (Kohlberg 1982), and is especially fostered through general education at schools and universities (Lind 1987, Rest 1979) and moral educations programs such as dilemma discussion and just community (see Power, Higgins & Kohlberg 1989; Oser 1984, 1987).
a. *Moral Opinions*, that are emotionally toned beliefs usually aroused first when a moral conflict occurs. Almost every person has an opinion on an issue, although people gain their opinion from different sources, and in some persons they are malleable, and in others quite persistent (cf. Keasey 1973).

b. *Moral Concern*, also referred to as rational reflection on the basis of general moral concepts. In some persons, moral concerns are too weak to have an impact on his or her moral opinion. At most, they serve as a “rationalization” for a solution. Only as a consequence of a developmental process of conflict resolution, the individual acquires the competence to change his or her moral opinions on the basis of moral principles, i.e. to make moral judgments (Walker 1986).

c. *Situational Judgment*, also referred to as prudential (or meta-ethical) judgment on the situational adequacy of particular levels of moral reasoning. In everyday life only a few (though very important) situations “ask” for a solution on the highest level of moral reasoning (Lind 1985). Many conflicts may be adequately solved on medium or low levels of moral reasoning.

Applying this psychological model of moral judgment to measurement implies an experimental, multifactorial test design including the following three factors:

1. Type of Moral Situation or Dilemma (in short, Dilemma)
2. Agreement or Disagreement with one's Opinion (Pro/Con)
3. Stage of Moral Reasoning (Stage)

These three “factors” constitute the design of the “Moralisches Urteil Test” (MJT). The standard version of the MJT consists of

1. Two moral stories (mercy killing and theft) in which two or more moral principles conflict with one another (=Dilemma Factor),
2. Two opposing opinions on the solution of the dilemma suggested in the story (= Pro/Con Factor), and
3. Six stages of moral reasoning (pro and con) about the dilemmas.

Hence the MJT forms a 2 x 2 x 6 experimental design. The subject is asked to judge the acceptability of each of the reasons presented (24 altogether) on a scale reaching from -4 (completely unacceptable) to +4 (completely acceptable). From the methodological point of view, the subject's judgment of acceptability may be labelled the dependent variable. The whole judgment pattern (in this case consisting of 24 single ratings) is considered as the subject's behavior, i.e., as the proper basis for inferences on his or her moral dispositions. This behavioral unit is definitely more than the sum of the subject's discrete acts. Its pattern comprises relational
information through which we may infer the structure of the subject's moral judgment. This relational information would be lost if we, like mainstream psychometrics, would break individual judgment pattern into pieces. Classic behaviorism has failed because it reduced psychological reality to atomized acts rather than studying whole pattern of behavior. (For this reason, we can still believe in behaviorism and yet do not have to buy classical test theory.)

**Types of Scores**

Basically two sorts of scores are computed from MJT data:

(a) Measures of moral attitudes, e.g., a subject's attitudes toward one of the six Kohlbergian stages of reasoning. This score reflects the agreement or disagreement with a particular level of moral reasoning. It is calculated like Likert scales through summated ratings. With the MJT, the sum scores are divided by the number of items to reflect an “average attitude,” across situations.

(b) Measures of moral structure, e.g., a subject's ability to evaluate reasons consistently from a moral point of view (as opposed to judging reasons on the basis of opinion-agreement). This score reflects the degree to which the subject's judgment is determined by moral concerns. It is a percentage score calculated through multivariate analysis of variance of individual judgment pattern. This type of analysis is unique for the MJT, as it is only feasible in experimentally designed, multifactorial tests.

Note that both sets of measures derived from the MJT are logically independent from one another. The attitudinal scores do not depend on the structural scores, nor do the scores within each set depend on one another other. This fact does not preclude an empirical relationship among some of these measures, as those relationships that cognitive-developmental theory predicts. On the contrary, the logical independence between the measures of the affective and the cognitive aspects of moral judgment is essential for testing hypotheses.

**Findings**


Yet people differ largely in regard to their moral judgment competence. [Findings are reported in the presentation more extensively.]
Conclusion

Cognitive-developmental theory has given birth to several methodologies of measuring moral judgment which are, by standards of mainstream educational and psychological measurement, highly objective and reliable, and possess a great amount of predictive power. To mention just a few: Larry Kohlberg's Moral Judgment Interview (Colby, Kohlberg et al., 1987), Jim Rest's Defining Issues Test (Rest 1979, 1986), and John Gibbs' Sociomoral Reflection Measure (Gibbs et al. 1980). However, their attempt to meet conventional standards casts doubt on their theoretical validity. These standards are fundamentally at odds with cognitive-developmental theory to which these methodologies refer.

Experimental questionnaires are a radical alternative to classical testing methodology. They represent, as Torgerson said (1958, pp. 46 f.) a “subject-centered approach to scaling.” Unfortunately, this approach has been confined to “measurement by definition,” that is, to mental testing “where the individual's score is a simple or adjusted sum of the number of items answered correctly” (pp. 47-48). In the realm of morality we cannot define “correct” answers without referring to conventional standards of morality. If we want avoid imposing conventional moral standards on our subjects we need to assess a person's moral judgment competence in regard to his or her own standards of morality. These are found in the individual's moral attitudes.

Hence our approach to the measurement of moral judgment, underlying the MJT, is formal rather than conventional. The MJT measures the degree to which the subject's judgments are determined by his or her own moral principles regardless of his or her stage of reasoning. Only through this approach we could uptake the task of studying these two questions: (a) To which extent do individuals agree or disagree in their moral preferences, and (b) how does moral preferences and moral cognition correlate?

As our findings show, both questions get a positive answer. First, people, across many cultures and across gender differences, strongly agree about the preference for the level of moral reasoning when moral conflicts are at stake. Second, a pronounced preference for higher order moral reasoning (and a rejection of low stage reasoning) strongly correlates with the development of a high moral judgment competence.
References


[Later publications: http://www.uni-konstanz.de/ag-moral/ ]
<table>
<thead>
<tr>
<th>Dilemma a:</th>
<th>Workers (Theft)</th>
<th>Physician (Mercy killing)</th>
<th>SUM of Pro Items</th>
<th>SUM of Con Items</th>
<th>SUM of Stage Items</th>
<th>Average Acceptability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro/Con:</td>
<td>Pro</td>
<td>Con</td>
<td>SUM</td>
<td>Pro</td>
<td>Con</td>
<td>SUM</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
<td>-2</td>
<td>-4</td>
<td>-6</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>4</td>
<td>-2</td>
<td>0</td>
<td>-2</td>
<td>2</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>SUM:</td>
<td>6</td>
<td>-6</td>
<td>0</td>
<td>4</td>
<td>-11</td>
<td>-7</td>
</tr>
</tbody>
</table>

Cognitive Aspects of Moral Judgment: Structure

<table>
<thead>
<tr>
<th>Experimental Factor</th>
<th>Disposition Involved Determination $r^2$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAGE</td>
<td>S's judgment is determined by moral concerns</td>
<td>19 %</td>
</tr>
<tr>
<td>STAGE x DILEMMA</td>
<td>S's judgment is moral depending but situationally differentiated</td>
<td>4 %</td>
</tr>
<tr>
<td>STAGE x PRO-CON</td>
<td>S's judgment is partially moral, depending on opinion agreement</td>
<td>7 %</td>
</tr>
<tr>
<td>PRO-CON</td>
<td>S's judgment depends solely on S's own opinion</td>
<td>45 %</td>
</tr>
</tbody>
</table>