

Georg Lind

**Review and Appraisal of the
*Moral Judgment Test (MJT)***

2000
Revised 2003

Contact:

Prof. Georg Lind
University of Konstanz
FB Psychologie
78457 Konstanz
E-Mail: Georg.Lind@uni-konstanz.de

For further information and publications on this topic see www.uni-konstanz.de/ag-moral/b-publik.htm

**Psychology of Morality &
Democracy and Education**

© Georg Lind

Review and Appraisal of the *Moral Judgment Test* (MJT)

Georg Lind
University of Konstanz¹

Lawrence Kohlberg defined moral judgment competence 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; emphasis added).

The *Moral Judgment Test* (MJT) has been constructed in 1975-77 to assess subjects' moral judgment competence in accordance with this definition. It is a *behavioral* test of a subject's ability to judge controversial arguments in a discussion about a moral problem on the basis of moral principles and orientations rather than on the basis of other criteria like opinion-agreement or opinion domain. The *counter-arguments* are the central feature of the MJT. They represent the *moral task* that the subjects has to cope with. In the standard version of the MJT, the subject is confronted with two moral dilemmas and with arguments pro and contra the subject's opinion on solving each of them. This competence is indexed by the so-called *C-score* or *C-index*, which ranges from zero (lowest score) to one hundred (highest score).¹

Another unique feature of the MJT is that it is possible to measure not only cognitive aspects of moral judgment behavior but also affective aspects simultaneously. It provides measures of subjects' moral attitudes: To which degree does a subject prefer or reject each of the six Kohlbergian stages of moral reasoning? (See Kohlberg et al., 1990); and gives several other measures of the cognitive and the affective aspect of moral judgment behavior. For a discussion of these and other indices, see Lind (1978; 2000a), and Lind and Wakenhut (1985) and Heidbrink (1985).

Because of the orthogonal, multi-factorial (experimental) design of the MJT, most of these measures are logically independent from each other and, therefore, allow us to study their *empirical* relationship. So it is possible to study Piaget's hypothesis of cognitive-affective parallelism directly and without any logical circularity (Lind, 2000a; 2000b).

The MJT is designed for research purposes and for use in evaluation studies, not for individual diagnostics.

Since its construction, the MJT has been used in many cross-sectional, longitudinal, experimental, cross-cultural and educational intervention projects involving more than 30.000 subjects of all ages, gender, social classes, educational backgrounds and cultures. Adaptations of the MJT to other languages and cultures undergo an arduous validation-process. Besides standard procedures like translation and back-translation, each new version is submitted to an empirical examination based on three highly sensitive criteria: 1. Monotonous preference hierarchy, 2. Quasi-Simplex-Structure of Stage Correlations, and 3. Cognitive-Affective Parallelism.

The findings show that the MJT is useful, theoretically valid and the new versions certified by the original author are cross-culturally valid: The MJT is much shorter than most other tests of moral development (the standard version has 26 items), it is easily to administer, there are hardly any non-scorable cases (as compared to some fifty percent unscorable cases in DIT-studies; see Gielen et al., 1994; Self et al., 1992) and fully scorable by a computer. Like the Kohlberg's Moral Judgment Interview, the MJT can already be used with 5th graders (the DIT has a testing floor of 9th graders; Rest et al., 1999).

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 (see also last paragraph on using the MJT in repeated testing projects).

The *dual aspect theory* of moral behavior and development

A core ground for constructing the MJT, was the dual-aspect theory of moral behavior and development as outlined by Piaget, Kohlberg and Lind. For Piaget (1976) “affective and cognitive mechanisms are inseparable, although distinct: the former depends on energy, and the latter depend on structure” (p. 71). Accordingly, Kohlberg meant his stage model of moral development to be a description of both the affective and the cognitive aspect of moral behavior (Kohlberg, 1958; 1984). Lind further explicated this theory and analyzed its implications for the measurement and stimulation of moral-cognitive development (Lind 1985a; 1985b; 1985c; 2000; 2000a; 2000b).

The dual aspect theory states that for a comprehensive description of moral behavior both affective as well as cognitive properties need to be considered. A full description of a person's moral behavior involves a) the moral ideals and princi-

¹ Author's address: Prof. Dr. Georg Lind, Dept of Psychology, University, 78457 Konstanz, Germany, e-mail: Georg.lind@uni-konstanz.de; URL www.uni-konstanz.de/ag-moral/

ples that informs it and b) the cognitive capacities that a person has when applying these ideals and principles in his or her decision making processes.

In contrast, other theories state that affect and cognition represent separate *components* of the human mind, separated also from moral behavior. They state that there is an affective domain of moral behavior and a cognitive domain, which can be dealt with separately. These theories imply that there are purely affective, cognitive and behavioral responses that can also be assessed separately, for example by using different tests for both components, eliciting the respective type of behavior. “However,” Higgins (1995) succinctly states, “there are cognitive aspects to all [. . .] components, and Kohlberg’s idea of a stage as a structured whole or a world view, cuts across [. . .] componential models” (p. 53).

While this theory assumes the cognitive and affective aspects are inseparable properties of human behavior, it also insists that they are distinct and not reducible to one another. In other words, we assume that each aspect provides us with important information about the nature of a person’s behavior. The affective aspect informs us about the direction or orientation of human behavior, and the cognitive aspect about the organisation and structure of it. Hence, both aspects are needed to achieve a comprehensive description of human action and each contributes in a unique way to the prediction of external criteria. In fact, studies indicate that the preference for a particular stage of moral reasoning (affective aspect) varies only little between subjects, and, therefore, correlates little with behaviors like helping, being honest, avoiding violence etc., but (cf. Rest et al., 1999; Lind, 2000c), but that moral judgment competence is consistently related to various kinds of moral behavior (Lind et al., 1985). In a recent study, Biggs et al. (1999) have demonstrated that both aspects of moral judgment behavior contribute differently to the correlation with various behaviors and attitudes of college students. Lind (1985b) used the combined analysis of both aspects to show that the moral regressions found by Kohlberg and his associates (Kohlberg & Higgins, 1984) may have been pseudo-regressions, that is, an artifact of their method of assessing both aspects through one and the same index.

Being able now both aspects of to assess subjects’ moral judgment behavior simultaneously but in a distinct manner, opens up the opportunity to design adequate studies to test and to prove Piaget’s hypothesis of parallelism between the affective and the cognitive side of morality. In fact, both aspects correlate highly, but this correlation can break down in a predictable way (Lind, 2000a; Lind, 2000b). This and some other findings have shown to be so clear-cut and replicable in many studies that we can use them now as criteria for testing the validity of new versions of the MJT (see below).

The *MJT* as a Multi-Variate N=1 Experiment

The *MJT* rests on modern, cognitive-structural approaches to psychological measurement (see, amongst others: Anderson, 1991; Broughton, 1978; Brunswik, 1955; Burisch, 1984; Cronbach & Meehl, 1955; Kohlberg, 1984; Lind, 2000a; 2000b; Loewinger, 1957; Lourenço & Machado, 1996; Mischel & Shoda, 1995; Pittel & Mendelsohn, 1966; Travers, 1951).

The basic approach we started out with coincides with Kohlberg’s: „In order to arrive at the underlying structure of a response, one must construct a test, [. . .] so that the questions and the responses to them allow for an unambiguous inference to be drawn as to underlying structure. [. . .] 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” (1984, p. 401-402).

We felt that the best ways of fulfilling this postulate was to design the *MJT* as a multi-variate $N = 1$ experiment because that way we can make sure that all relevant aspects of a moral task are present in the test and that these aspects are uncorrelated and thus can be clearly identified. As modern psychology reveals, individuals do not only differ in regard to certain moral preferences, attitudes or values but are *structurally* different. Therefore, we must base the measurement of psychological properties (such as moral judgment competence), on the assessment of *individual* pattern of behavior rather than on the behavioral pattern of a *sample* of persons (as is usually done). Otherwise we would commit an *ecological fallacy*, that is, we would falsely hypothesize that the structure of behavioral data in a sample of individuals is identical with that of each individual. Such a hypothesis, however, is hardly tenable (see Mischel & Shoda, 1995).

Because the function of this experiment is not to test the effects of some treatment but to describe the nature and development of behavioral properties, we call it an *ideographic* experiment. This special function entails some special kinds of experimental analysis. The *independent* variables (or *factors*) are varied in order to study the functioning of the individual’s mind but not to assess ‘general’ effects of these factors. Modern cognitive-structural research found that these effects differ much from one person to another depending on their level of development (Lind, 1978; 1985a; 1985c; 2000a; 2000b; Krebs et al., 1990).

The *dependent* variable is represented by the subject's judgment behavior, that is, by his or her rating of the arguments on a scale from -4 to +4. (Note that this scale may also range from -2 to +2 for subjects that have difficulties with such a fine-graded scale; see a sample of test items in the appendix).

The moral factor determining subjects' judgment behavior is represented by the moral quality of the arguments. With the *MJT*, moral quality was defined using Kohlberg's six stages of moral reasoning (Kohlberg, 1958; 1984).

The task factor, opinion agreement or disagreement, is represented by the implication of the argument pro or contra the subject's opinion about the decision of the story's protagonist. The pro-arguments indicate which ideal level of moral discourse the subject prefers; the contra arguments indicate how much the subject let this moral ideal determine his or her judgment of arguments in the presence of other powerful psychological forces.

Finally, the different dilemmas contained in the *MJT* represent different *moral demand structures*. In the standard version these differences are small yet noticeable. While the mercy-killing dilemma (taken from Kohlberg's *Moral Judgment Interview*) is thought to *pull* the highest level of moral reasoning on Kohlberg's stage six, the worker's dilemma (adapted from the novel "Stellenweise Glatteis" by Max von der Grün) is thought to pull more Stage 5 reasoning (Lind, 1985a). In fact, subjects typically prefer stage 6 reasoning most in the mercy killing dilemma and stage 5 reasoning most in the workers' dilemma (Lind, 2000b).

Because of its rationale and its design, the *MJT* is a behavioral experiment rather than a classical psychometric test (Lumsden, 1976). Hence, response consistency and inconsistency indicate properties of a person's moral-cognitive structure, rather than properties of the instrument like "measurement error" or "unreliability" (see Lind, 2000b).

Therefore, the *MJT* was not submitted to traditional „item analysis.“ Hence, no items were selected to increase the correlation of the *C-index* with an empirical criteria like age, political attitudes, or higher education. This fact guarantees that the *MJT* is not biased in favor or against certain predictions like stability of rank orders among people, age-correlation, or invariant sequence. 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 use of experimental terminology may seem unusual in the context of measurement. Yet, as Brunswik (1955) and Anderson (1991) had shown, the structure of dispositions can only reliably assessed through some kind of "dispositional experiment" or, as Brunswik (1955) put it, through "diacritical method" (see also Lind, 1982).

Computing the *MJT's C-index*

With the *MJT*, several indices for cognitive and for affective aspects of moral judgment behavior can be computed. The main index is the *C-index* of moral judgment competence. This index takes into account the whole *pattern* of a subject's responses rather than merely isolated acts. The *C-score* is not based on the mere summation of responses but mirrors the relationship or structures among the responses of an individual.

The *C-index* is computed by a multivariate analysis of variance components, similar to a multi-variate analysis of variance (MANOVA). In most cases some programming is inevitable because most commercial packages do not provide ready methods for analyzing data individually. Most packages (like SAS, SPSS, STATISTICA) have a programming language module included that lets one quickly write a program for scoring the *MJT*. The *coding scheme* for the standard version and a sample program for STATISTICA, version 5.x can be requested from the author. A scoring service is also available.²

The *C* index can range from 1 to 100. It indicates the percentage of an individual's total response variation due to a person's concern for the moral quality of given arguments or behavior. Following a proposal by Cohen (1988), the *C* is sometimes graded *low* (1-9), *medium* (10-29), *high* (30-49) and *very high* (above 50). For typical mean *C* of various groups of people, see Lind (2000a). How *C* is computed is explained in more detail on this document: www.uni-konstanz.de/ag-moral/pdf/mjt-scoring.pdf.

Note that only because the *MJT* contains a *moral task* (see below), a person's moral judgment competence can be scored by the *C-index*. If used with a test that does not contain a moral task (for example with the *DIT*) the *C-index* is meaningless.

² In former publications, another name for the *C-index* were used: "DetStufe" (German) or "DetStage" (English), which is an abbreviation for the degree of determination of the individual's judgment behavior by the experimental factor „stage of reasoning.“

As second set of indices, 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, whether the preferences for the six stages form a hierarchical order as Kohlberg assumed, that is, which stage of moral reasoning a person prefers 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.

There is no logical connection between the cognitive and affective dimensions of moral reasoning. Although many individuals prefer higher stage moral arguments, only those with more cognitive structures exhibit consistency or reversibility, i.e., the capacity to recognize the moral merit of opposing viewpoints. Invariably, most subjects prefer sophisticated moral arguments when assessing factors favorable to their own position, an outcome stemming from successful socialization to the language of democracy: civic responsibility, civil rights and justice. It is only when asked to evaluate a position contrary to one's own that the importance of cognitive structures emerges. One may prefer universal norms of justice (a high affective score) but be unable to use them consistently, particularly when evaluating the moral position of an adversary (a low competence score). Or, one may exhibit a preference for parochial moral norms (a low affective score) but use them consistently to judge competing moral positions (a high cognitive score).

Research Projects Using the MJT

The MJT has been used in many studies, comprising more than 40.000 subjects. In the following table, a rough overview is given on studies using the MJT or translated and validated versions of it together with main author(s), topics of study and approximate sample sizes. The figures are conservative estimations, giving the lower limits. Information on ongoing international research using the MJT is provided on the web: <http://www.uni-konstanz.de/ag-moral/>.

| Author(s), Year of Study | Study Topic | N |
|--|---|------|
| Tino Bargel, Barbara Dippelhofer-Stiem, Gerhild Framhein, Georg Lind, Hansgert Peisert, Johann-Ulrich Sandberger, Hans-Gerhard Walter, 1976 - 1977 | High School Graduate Follow-up Study (2 waves) | 1200 |
| dto. 1982 - 1985 | German University Graduates Follow-up Study (2 waves) | 800 |
| dto. 1977 - 1984 | German University Students Panel Study (4 waves), with initially 2000 first semester students | 4500 |
| Hansgert Peisert, Markiewicz, Tino Bargel, Georg Lind et al. 1977 - 1984 | In addition: Five National Panel Study (three waves) in Austria, Netherlands, Poland and Yugoslavia | 8000 |
| Rainer Senger, 1979 | Study of German soldiers and officers on the segmentation phenomenon | 130 |
| Georg Lind & Leonore Link 1980 | Study of German school students age 13-20 | 212 |
| Georg Lind, Fritz Oser et al., 1987-1991 | DES, project (Democracy and Education in Schools) evaluation | 500 |
| Michael Gross, 1990-1999 | International studies on ethics and political activism in the US, France, Netherlands and Israel | 500 |
| Klaus Beck, 1991-1993 | Study of German apprentices and students of vocational schools | 9000 |
| Horst Heidbrink, 1980 | Study of German vocational school students | 140 |
| Georg Lind with EMNID 1990 | Cross-sectional study of German apprentices and vocational students | 1600 |

| | | |
|--|---|--------------|
| Fritz Oser et al., 1980-1982 | HASMU cross-sectional study of Swiss vocational students | 200 |
| Fritz Oser, André Schläfli et al, 1980-1982 | HASMU educational intervention study with Swiss vocational students and apprentices | 50 |
| Don Biggs & Robert Colesante, 1998 | Study on US college students | 200 |
| Patricia Bataglia, 1998-2000 | Validation study in Brazil of Portuguese MJT | 60 |
| Nermin Ciftci, 1998-1999 | Validation study of Turkish MJT | 600 |
| Louis Trechera, 1992 | Validation study of Spanish MJT | 101 |
| Cristina Moreno / Roberto Herandez/ Monica Rangel / A. Quiroga et al., 1996-2000 | Validation studies in Mexico of Spanish MJT version | 1200 |
| Bart Duriez, 1999 | Validation study of the Flemish MJT in Belgium | 180 |
| Jame Dubois, 1998 | Study of medical personnel in Austria and Saudi Arabia | 257 |
| Anna Laura Comunian et al., 1995 | Validation of the Italian MJT | 465 |
| Anna Laura Comunian et al., 1997 | Study of university's learning environment and students' moral development | 928 |
| Gabriele Klewitz, 1995 | Validation of the French MJT | 77 |
| Gaby Plitzko, 1999 | Study on the moral development of female professionals and housewives | 84 |
| Dale Richesin, 1999 | Study of drug therapists in the US | 88 |
| Ibolia Vári-Szilágyi et al. 1994 | Validation of Hungarian MJT and evaluation of school programs | 301 |
| Manfred Scholz, 1995 | Evaluation of dilemma discussions | 58 |
| Georg Lind, 1999-2000 | Various evaluations of dilemma discussions | 100 |
| Wolfgang Wasel, 1994 | Simulation experiment with college students | 105 |
| Beatriz Peña, 2000 | Study of students in Colombia | 140 |
| B. Slovackova, 1999-2000 | Czech Republic | 327 |
| M. Ylén, 1999 | Validation study of the Finish MJT | 60 |
| Diverse studies with no information on date and sample size | Thailand, Indonesia (W. Marlind; Y. Santi), Uganda, Uruquay, Peru, Argentina, Costa Rica, England, Germany, Jordania (R. Amman), Kuwait, Hongkong | |
| Secretaría de Bogotá | Secondary school students, Bogotá, Colombia | 6600 |
| Total number of measurements using the MJT = | | 38763 |

Theoretical Validity

Because the MJT was intended for research (for testing theoretical propositions and hypotheses) and program evaluation, it was of paramount importance to secure its theoretical validity. That is, the question “Does the MJT really measure what it claims to measure” should be answered positively.

The MJT claims to measure moral judgment *competence* rather than only moral orientations or attitudes.³ It is the only test which makes such a claim and which is explicitly designed to contain a moral task (see above). In that respect, the MJT is theoretically valid by virtue of construction.

Moreover, this claim has received strong empirical support by many validation studies. These studies employ five *empirical* criteria for theoretical validity which we derived from cognitive-developmental theory and research (e.g., by Kohlberg, 1958; Rest, 1979; Lind, 2000b; Walker, 1986).

1. *The preferences for the six Kohlbergian stages of moral reasoning (affective aspect) are ordered in a predictable way:* In a truly moral dilemma, subjects should prefer the stages of moral reasoning in the order of their number, with highest preference for stage-six-reasoning and lowest preference for stage-one-reasoning. To my knowledge, all *MJT*-studies have found such a preference order (Lind, 2000a; see also below the paragraph on cross-cultural validity).
2. *The preferences form a quasi-simplex structure:* According to Kohlberg (1958), 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). On other words, in the correlation matrix of all stages, the coefficients should decrease monotonously from the diagonal toward the corners of the matrix. This can be tested by submitting the correlations to a *principle component analysis* (this is a factor analysis that makes use of the total variance rather than only of the common variance) with *varimax rotation* and *Eigen values* greater than 1.0. 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 (Lind, 2000a; 2000b; see also below).
3. *Cognitive and affective aspects are parallel:* The higher a subject’s moral judgment competence, the more clearly does he or she accept higher stage arguments and rejects lower stage arguments. Piaget assumed that cognitive and affective aspects of human behavior are systematically correlated. One way to translate this general postulate is to assume that subjects’ moral attitudes should be systematically correlated with their moral judgment competence (Lind, 1985a, 2000a; 2000b). 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. Most, if not all, *MJT*-studies found such pattern of correlations corroborating the *parallelism* theorem (Lind, 1985a; 2000a; 2000b).
4. *Pro and con arguments are equivalent:* The arguments in favor of a certain solution of a moral dilemma should be equivalent to the arguments of against it, that is, subjects agreeing with the given solution of a dilemma should be confronted with arguments of approximately the same quality as subjects disagreeing with this solution. Lind (2000b) found this hypothesis to be clearly supported. 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.
5. *The MJT contains a real, difficult moral task and, hence, the C-index is a indexes of moral competencies* (rather than of moral *attitude*). This task consists in the fact that subjects are asked to judge moral arguments which *oppose* their opinion on a certain moral issue. This task has shown to be very difficult for many subjects (Keasey, 1974). Subjects typically get much lower *C*-scores than *P*-scores (Lind, 2000a). While in Emler et al’s (1983) experimental study using the *Defining-Issues-Test* the *P*-index the subjects could simulate a higher score than they had (see also Markoulis, 1989; Barnett et al., 1996), Lind (2000b) showed with an identical experiment, using the Moral Judgment Test instead, that subjects could not simulate the *C*-index (Lind, 2000b). Wasel (1994) replicated this finding in a similar experiment. He also found a substantial correlation between subjects’ *C*-scores and the precision of their perception of other people’s moral judgment competence. The competence nature of *C-index* is also supported by the fact that, in longitudinal studies, upward changes were always gradual rather than abrupt (Lind, 2000a; 2000b). 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 subjects do not practice their moral abilities, the faster they lose. So the *C-index* of the *MJT* meets all criteria of a competence-index: moral task by virtue of construction, non-fakability, gradual learning curve and negatively accelerated, smooth forgetting curve (Lind, 2000a).

³ Rest and his colleagues (1999) “have eliminated the faking study from our set of the validity criteria” (p. 115) and thus dropped the claim that the DIT would measure a moral competence.

Note that it would be pointless to validate the *MJT* by correlating it with other tests. It was constructed on the basis of founded theories in order to replace other tests which seemed to be not sufficiently valid. So, if the correlation with another test was low, we would think that this other test rather than the *MJT* is not valid.

Validating new and translated versions of the *MJT*

Because any inference that we base on data depends heavily on their quality and the validity of the measurement process, careful validation procedures of a measurement instrument are very important. In cross-cultural research they are even more important as we may falsely interpret methodological differences between cultures as substantial ones. Only if the translation of the test has been carefully done (including backward translation) and if the empirical data agree fairly well with the first four indicators, one can assume that a newly created version of the *MJT* is fairly valid and equivalent to the original (German) version, which is a necessary condition for comparing findings across countries. If cultural validity is low or unknown, we have no way to attribute difference in data in any substantial way.

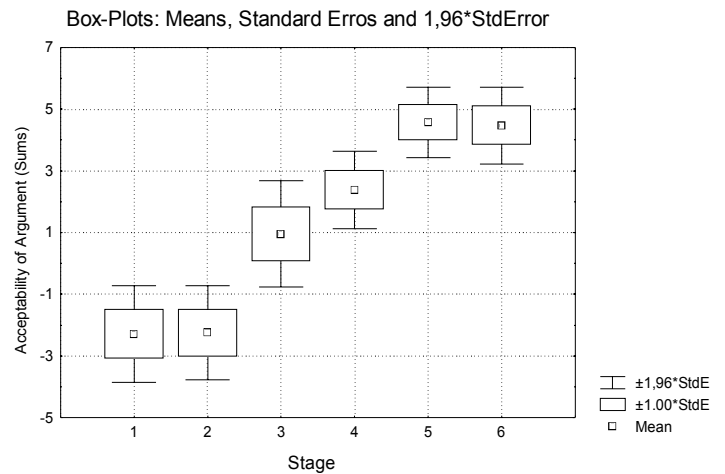
This theory-based empirical validation procedure requires more time and funding than is usually thought necessary. However, it pays well in terms of more trustworthy, meaningful and comparable data. The procedure provides validity criteria independent from the research for which it will be subsequently used, making circular conclusions and tautologies less likely and findings more credible. Studies that omit this validation process because it is expensive, usually pay for this omission by producing ambiguous, if not useless data. They warrant no substantial interpretation other than that the data are not valid.

In the following we summarize a procedure for securing the validity of new sub-tests and foreign language versions of the *MJT*, which we have been found very effective for detecting and curing flaws in new or translated versions. It employs the first four empirical criteria discussed above and one additional criterion.

The first step of validating any foreign version of a test, is to translate that version backward into the original language. There are a German and an English master version of the *MJT* that can be used for backward translation and checking the equivalence of the new version. Many flaws can be already detected this way.

The second step in validating new *MJT* versions involves a small empirical validation study (N = 60) with a group of subjects representing different levels of education, e.g., 9th graders, 12th graders, undergraduate college students, and graduate students. Such a study provides all validity checks that we have discussed above. I will illustrate this by the findings of the first validation study of the Brazilian *MJT* conducted by Patricia U. Bataglia, University of Sao Paulo, Brazil.

- **Preference hierarchy:** The preferences for the six Kohlbergian stages of moral reasoning should be ordered as theoretically predicted, with stage 6 preferred most, stage 5 second most etc. (see the figure on the left). Some small inversions of stage preferences (especially between stages 1 and 2, as well as between stages 5 and 6) may occur, and do not invalidate the new test version. Cross-cultural research supports this hypothesis quite well (Lind, 1986; Gross, 1996).



- **Quasi-Simplex structure** of stage preference inter-correlations: Neighboring stages (for example, stages 5 and 6) should correlate higher than more distant stages (for example, stages 4 and 6), that is, the correlations should monotonously decrease from the diagonal to the left lower corner of the correlation matrix. This criterion can be tested in two ways: a) in a main-component factor analysis with varimax-rotation, two factors should be produced and the factor loadings for the preference scores for the six stages should be orderly located on a simplex curve between the two (see graph); b) attempts to bring the matrix of the inter-correlations between the preferences for the six stages into a more simplex-like order should not result in a re-ordering of the six stages.

- **Affective-cognitive parallelism**: The stage preferences should correlate in a predicted manner with the *MJT's* C-index of moral judgment competence, i.e., while the preference for the highest stages should correlate highly positively with the competence score, the preferences for the lowest stages should correlate highly negatively with that score, and the other *MJT* preference indices should show correlations in between these extremes.

- **Correlation with education**. Given the above described sample, *C* should correlate highly positive ($r > 0.40$) with amount and quality of education of the subjects. Correlation of *C* with Ss' age should be small or close to zero when level of education is hold constant. In the first validation study of the Brazilian *MJT*, the correlation between *C*-score and level of education was $r = 0.40$.

Correlations

| Stage | 1 | 2 | 3 | 4 | 5 |
|-------|-----|-----|-----|-----|-----|
| 2 | ,45 | | | | |
| 3 | ,51 | ,47 | | | |
| 4 | ,11 | ,41 | ,31 | | |
| 5 | ,21 | ,27 | ,22 | ,12 | |
| 6 | ,01 | ,14 | ,24 | ,14 | ,16 |

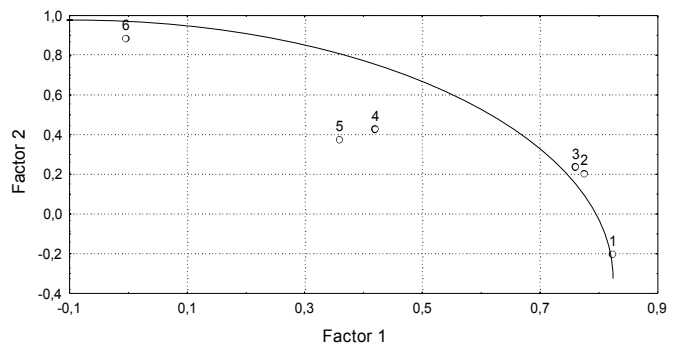
Paired exclusion of missing data
 First Validation Study with the Brazilian *MJT*, N=60
 Author: Patricia U. Bataglia

Additional (optional) criterion:

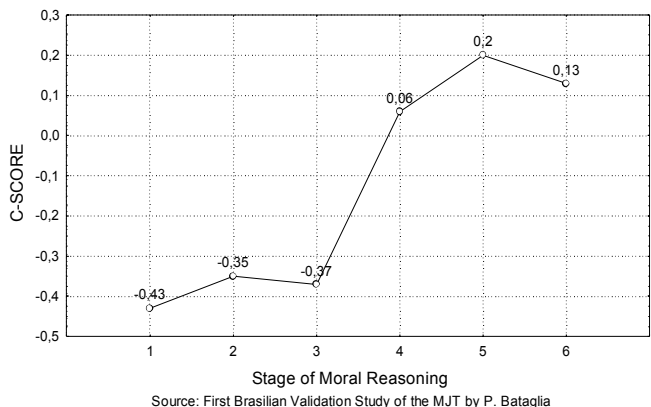
- **No upward simulation of the C-index**. As shown above, the *MJT* has been constructed to assess the cognitive, or competence aspect of moral judgment behavior rather than merely moral attitudes. When submitted to experimental setting like that used by Emler et al. (1983), the *C-score* of a newly constructed *MJT* version should also show no upward change. This criterion has been made optional for the validation process because the other four criteria have proven to be very powerful for detecting faulty items.

Note that these findings and findings from other validation studies do not only support the cross-cultural validity of the new version of the *MJT*, but also the universal validity of core assumptions of cognitive-developmental theory of moral behavior and development. Especially, they prove the validity of Piaget's and Kohlberg's assumption that morality has a strong *cognitive* or *competence* aspect. Only one core assumption as (clearly) proven as invalid, namely the assumption of invariant sequence of moral development. Moral competencies can regress and do regress when educational support ceases before some critical threshold of development has been reached (Lind, 2000a).

Factor Loadings of the Preference Scores for Six Stages
 First Validation Study with the Brazilian *MJT* (Director: Patricia U. Bataglia)
 Sample: Lower and Upper Secondary School and College Students; N = 60
 Rotation: Simple Varimax; Extraction: Main Components



Correlations between C-score and Preferences for the Six Stages



Source: First Brazilian Validation Study of the *MJT* by P. Bataglia

Caveats

The *MJT* was designed for research and program evaluation. It was not designed and should not be used for individual diagnosis or for making judgments about individual persons. 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 build-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 "true" level of moral judgment competence.

Recent evaluation studies also indicate that the *MJT* can underestimate some subjects' moral judgment competencies when used in pretest posttest studies, as is usually the case in evaluation studies. Unpublished studies indicate that some subjects lack motivation to fill in the test a second time in exactly the same manner as before. They sometimes even get zero scores thus lowering the mean scores considerably and creating the impression that the educational intervention had a negative effect.

This problem is generic for competence tests and has not been properly solved yet for the *MJT*. It seems not present when sufficient time intervals are chosen as in some longitudinal studies, or when the *MJT* is administered individually and a motivating instruction is given. When administered in groups, subjects are more likely to express their dissatisfaction with the second testing, stimulating others to feel alike. Currently, we pursue several ideas for coping with this problem and discuss it on our web-site.

Finally, the *MJT* is not, as stated in some textbooks, another version of the *DIT*. Nor is it a psychometric test in the classical sense. It is a multi-variate, $N=1$ experiment, or behavioral test which contains a moral task. For tests that do not contain a moral task, the *C*-score is not meaningful.

References

- Anderson, N.H. (1991). Moral-social development. In: N.H. Anderson, ed., *Information Integration Theory*, Volume III: Developmental, pp. 137-187. Hillsdale, NJ: L. Erlbaum.
- Arbuthnot, J. (1979). Error in self-assessment of moral judgment stages. *The Journal of Social Psychology*, 107, 289-290.
- Barnett, R., Evens, J., & Rest, J. (1996). Faking moral judgment on the Defining Issues Test. *British Journal of Social Psychology*, 34, 267-278.
- Beck, K. & Krumm, V. (1999). *Wirtschaftskundlicher Bildungs-Test (WBT)*. Göttingen: Hogrefe.
- Biggs, D., Colesante, R., Agafonov, A., Felming, J., & Campbell, K. (1999). *Education about controversial issues: the role of tolerance, moral judgment, and course involvement*. Symposium presentation at the AME conference in Minneapolis, Nov. 1999.
- 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. (1987). *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 University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cronbach, L. & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin* 52, 281-301.
- 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., Basinger, K.S., & Fuller, R. (1992). *Moral Maturity: Measuring the Development of Sociomoral Reflection*. Hillsdale, NJ: Erlbaum.
- Gibbs, J. & Schnell, S. (1985). Moral development 'versus' socialization: A critique. *American Psychologist*, 40, 1071-80.
- Gielen, U.P., Comunian, A.L. & Antoni, G. (1994). An Italian cross-sectional study of Gibbs' Sociomoral 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.
- Gross, M. (1997). *Ethics and activism. The theory and practice of political morality*. New York: Cambridge University Press.
- Habermas, J. (1985). Philosophical notes on 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 Inc.
- Habermas, J. (1990). *Moral consciousness and communicative action*. (C. Lenhard & S. W. Nicholson, Trans.). Cambridge: MIT Press.
- Heidbrink, H. (1985). Moral judgment competence and political learning. In: G. Lind, H.A. Hartmann & R. Wakenhut, eds., *Moral Development and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education* (pp. 259-271). Chicago: Precedent Publishing Inc.
- Keasey, Ch. 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.
- Kohlberg, L. (1958). *The development of modes of moral thinking and choice in the years 10 to 16*. University of Chicago: Unpublished doctoral dissertation.
- 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. New York: Russel Sage Foundation, S. 381-431.
- Kohlberg, L. (1984). *The psychology of moral development*. San Francisco: Harper & Row.
- 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., 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.
- Krebs, D., Denton, K.L., Vermeulen, S.C., Carpendale, J.I. & Bush, A. (1990). The structural flexibility of moral judgment. *Journal of Personality and Social Psychology*.
- Kurtines, W.M. & Gewirtz, J., eds. (1995). *Moral development: An introduction*. Boston: Allyn and Bacon.
- Levy-Suhl, M. (1912). The examination of the moral maturity of juvenile delinquents. [in German: Die Prüfung der sittlichen Reife jugendlicher Angeklagter und die Reformvorschläge zum § 56 des deutschen Strafgesetzbuches.] *Zeitschrift für Psychotherapie*, 232-254.
- Lind, G. (1978). How does one measure moral judgment? Problems and alternative ways of measuring a complex construct. [German: Wie mißt man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts.] 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: Precedents Publishing Inc.
- 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 Publishing Inc.
- 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 Publishing Inc.
- 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. (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 SIG MDE, ERA meeting, New York, April 1996.
- Lind, G. (2000a). *Can ethics be taught? Research findings from modern moral psychology. [in German: Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung]*. 2nd Edition.
- Lind, G. (2000b). *Content and structure of moral judgment*. Second corrected edition. University of Konstanz: Doctoral Dissertation (Original 1984). Electronic edition: <http://www.ub.uni-konstanz.de/kops/volltexte/2000/504>
- Lind, G. (2000c). The importance of role-taking opportunities for self-sustaining moral development. *Journal of Research in Education*, 10, 9-15.
- Lind, G. & Althof, W. (1992). Does the Just Community program make a difference? Measuring and evaluating the effect of the DES project. *Moral Education Forum*, 17, 19-28.
- Lind, G., Sandberger, J.-U. & Bargel, T. (1985). Moral competence and democratic personality. 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 Publ. Inc.
- 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 Publishing Inc.
- Loevinger, J., 1957: Objective tests as instruments of psychological theory. *Psychological Reports*, 9, 635-694.
- Lourenço, O. & Machado, A. (1996). In defense of Piaget's theory: a reply to 10 common criticisms. *Psychological Review*, 103, 143-164.
- Lumsden, J., 1976: Test theory. *Annual Review of Psychology*, 27, 251-280.
- Markoulis, D. (1989). Political involvement and socio-moral reasoning: Testing Emler's interpretation. *British Journal of Social Psychology*, 28, 203-212.
- 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.
- Moody, E.E. (1999). *Using pair counseling to reduce violence in youth offenders*. Paper presented at the AME meeting in Minneapolis, 1999.
- Oser, F. (1986). Moral education and values education: The discourse perspective. In: M.C. Wittrock, ed., *Handbook of research on teaching*, pp. 917-941. New York: Macmillan.
- Piaget, J. (1965/1932). *The moral judgment of the child* (first publ. 1932). New York: The Free Press.
- Pittel, S.M. & Mendelsohn, G.A. (1966). Measurement of moral values: a review and critique. *Psychological Bulletin*, 66, 22-35.
- Popper, K. (1979). *Objective Knowledge. An evolutionary approach*. Oxford: At the Clarendon Press (Revised Edition).
- Power, C., Higgins, A., & Kohlberg, L. (1989). *Lawrence Kohlberg's approach to moral education*. New York: Columbia University Press.
- Rest, J.R. (1979). *Development in Judging Moral Issues*. Minneapolis, Minnesota: University of Minnesota Press.
- Rest, J. (1986). *Moral development: Advances in research and theory*. New York: Praeger.
- Rest, J.R., Narvaez, D., Bebeau, M.J. & Thoma, S.J. (1999). *Postconventional moral thinking. A Neo-Kohlbergian approach*. Mahwah, NJ: Erlbaum.
- Self, D.J., Baldwin, D.C., Jr. & Wolinsky, F.D. (1992). Evaluation of teaching medical ethics by an assessment of moral reasoning. *Medical Education*, 26, 178-184.
- Travers, R.M., 1951: Rational hypotheses in the construction of tests. *Educational and Psychological Measurement*, 11, 128-137.
- 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. [German: Simulation moralischer Urteilsfähigkeit.]* University of Constance, unpublished thesis in Psychology.

I. Workers' 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.

| | I strongly disagree | | | | | | | I strongly agree |
|--|----------------------------|----|----|---|----|----|----|-------------------------|
| 1. Would you disagree or agree with the workers' behavior? | -3 | -2 | -1 | 0 | +1 | +2 | +3 | |

| | I find the argument ... | | | | | | | | |
|---|--------------------------------|----|----|----|---|----|----|-----------------------|----|
| How acceptable do you find the following arguments <i>in favor</i> of the two workers' behavior? Suppose someone argued they were <i>right</i> ... | completely unacceptable | | | | | | | completely acceptable | |
| 2. because they didn't cause much damage to the company. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 3. because due to the company's disregard for the law, the means used by the two workers were permissible to restore law and order. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 4. because most of the workers would approve of their deed and many of them would be happy about it. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 5. because trust between people and individual dignity count more than the firm's best | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 6. because since the company had committed an injustice first, the two workers were justified in breaking into the offices | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 7. 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. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |

| | I find the argument ... | | | | | | | | |
|--|--------------------------------|----|----|----|---|----|----|-----------------------|----|
| How acceptable do you find the following arguments <i>against</i> the two workers' behavior? Suppose someone argued they were <i>wrong</i> ... | completely unacceptable | | | | | | | completely acceptable | |
| 8. because we would endanger law and order in society if everyone acted as the two workers did. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 9. because one must not violate such a basic right as the right of property ownership and to take the law into one's own hands, unless some universal moral principle justifies doing so. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 10. because risking dismissal from the company is unwise because of other people. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 11. because the two should have run through the legal channels at their disposal and not committed a serious violation of the law. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 12. because one doesn't steal and commit burglary if one wants to be considered a decent and honest person. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |
| 13. because the dismissals of the other employees did not affect them and thus they had no reason to steal the transcripts. | -4 | -3 | -2 | -1 | 0 | +1 | +2 | +3 | +4 |

II. Doctor's Dilemma

A woman had cancer and she had no hope 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.

14. Do you disagree or agree with the doctor's behavior?

| | |
|----------------------------|-------------------------|
| I strongly disagree | I strongly agree |
| -3 -2 -1 0 +1 +2 +3 | |

How acceptable do you find the following arguments *in favor* of the doctor? Suppose someone said he acted *rightly* . . .

| | |
|--|------------------------------|
| I find the argument . . . completely unacceptable | completely acceptable |
|--|------------------------------|

15. because the doctor had to act according to his conscience. The woman's condition justified an exception to the moral obligation to preserve life. . .

-4 -3 -2 -1 0 +1 +2 +3 +4

16. because the doctor was the only one who could fulfill the woman's wish; respect for her wish made him act as he did.

-4 -3 -2 -1 0 +1 +2 +3 +4

17. because the doctor only did what the woman talked him into doing. He need not worry about unpleasant consequences.

-4 -3 -2 -1 0 +1 +2 +3 +4

18. because the woman would have died anyway and it didn't take much effort for him to give her an overdose of a painkiller.

-4 -3 -2 -1 0 +1 +2 +3 +4

19. because the doctor didn't really break a law. Nobody could have saved the woman and he only wanted to shorten her suffering.

-4 -3 -2 -1 0 +1 +2 +3 +4

20. because most of his fellow doctors would presumably have done the same in a similar situation.

-4 -3 -2 -1 0 +1 +2 +3 +4

How acceptable do you find the following arguments *against* the doctor? Suppose someone said that he acted *wrongly* . . .

| | |
|--|------------------------------|
| I find the argument . . . completely unacceptable | completely acceptable |
|--|------------------------------|

21. because he acted contrary to his colleagues' convictions. If they are against mercy-killing the doctor shouldn't do it.

-4 -3 -2 -1 0 +1 +2 +3 +4

22. because one should be able to have complete faith in a doctor's devotion to preserving life even if someone with great pain would rather die.

-4 -3 -2 -1 0 +1 +2 +3 +4

23. because the protection of life is everyone's highest moral obligation. We have no clear moral criteria for distinguishing between mercy-killing and murder.

-4 -3 -2 -1 0 +1 +2 +3 +4

24. because the doctor could get himself into much trouble. They have already punished others for doing the same thing.

-4 -3 -2 -1 0 +1 +2 +3 +4

25. because he could have had it much easier if he had waited and not interfered with the woman's dying.

-4 -3 -2 -1 0 +1 +2 +3 +4

26. because the doctor broke the law. If one does not think that mercy-killing is legal, then one should not comply with such requests

-4 -3 -2 -1 0 +1 +2 +3 +4

Thank you!

Note that the items are not ordered according to the Kohlberg stages. The coding scheme is available from the author.