Last revision: 2006, Jan. 16

The *Moral Judgment Test*: Comments on Villegas' Critique

Georg Lind

University of Konstanz, Germany

Contact information:

Address correspondence to Prof. Dr. Georg Lind, Department of Psychology, University of Konstanz, 78457 Konstanz, Germany, or e-mail (georg.lind@uni-konstanz.de).

Erschienen in: Psychological Reports, 2006, 98, 580-584

Summary. — The MJT is an operationalization of Kohlberg's notion of moral judgment *competence* and Piaget's two-aspect model of behavior, and thus differs from tests of preferred moral orientations, and Kohlberg's own Stage measurement. As the MJT presents counter-arguments to be rated in regard to their moral quality, it is a difficult *task* for many, and the C-score reflects the respondents' competence to cope with this task. Validation studies of 27 language versions and thirty years of use in research and evaluation studies have proven that the MJT is valid and useful for assessing affective and cognitive aspects of moral behavior for research and evaluation purposes.

In the past, moral and democracy education programs have mostly been evaluated through tests of moral and democratic preferences or tests of declarative knowledge, that is, of moral and democratic concepts. No measure was available to study the effect of education on moral and democratic *competencies*. Therefore, thirty years ago, Lind has designed the *Moral Judgment Test* (MJT) to fill this void (Lind, 1978; Lind & Wakenhut, 1985; Lind, in press), and submitted it to many validation studies. These validation studies were unusually rigorous because the predictions that were used as criteria possess an unusually low *a priori* probability (see below). Meanwhile, the MJT has helped to evaluate and improve methods of moral education such as dilemma discussion (Lerkiatbundit et al., 2004; Lind & Althof, 1992: Lind, 2002) and teacher education (Lind, 2003), and also to enhance our understanding of the nature and development of moral development (Gross, 1997; Lind, 2002).

Without taking any of this research into account, Villegas' (2005) criticizes that the MJT presents "some problems" related to the trait measured, reliability, and validity of its main score (C-index). Further she maintains that the C-index reflects consistency in moral judgment, but this construct would be different from moral development as stated by Kohlberg.

The MJT never claimed to measure Kohlbergian *Stages*, but *moral judgment competence*, the concept which Kohlberg (1958; 1964) originally had in mind. An elaborate discussion of the Stage concept and its measurement can be found in Lind (1989; in press). The MJT follows closely Kohlberg's theoretical definition of moral judgment competence and the methodological principles he sets up for measuring it. It agrees with Kohlberg's original intentions even better than the Stage measurement because the MJT allows us to measure moral *competencies* and moral *preferences* or *attitudes* in a distinct and non-confounded way,

whereas Kohlberg's Stage model conflates both aspects. Kohlberg's Stage scores reflect a mixture of judgment consistency and stage preference (Colby et al., 1987). Yet, moral preferences can be easily simulated upward (Emler et al., 1983). Therefore, the ubiquitous moral preference measures are inappropriate for evaluating programs of moral education. Mixed scores like Kohlberg's Stages are better but still not optimal for that. In contrast, the MJT measures moral judgment competence in a pure way, independently but not separately from moral preferences, as affective and cognitive aspects of moral judgment behavior are not to be conceived of as separate components but as aspects or properties of the same behavior (Piaget, 1976). Hence, the MJT provides a 'hard' measure for evaluating the effects of moral or character education.

Initially, Kohlberg (1964, p. 425) 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." The MJT renders this competence observable by confronting the participant with a *moral task*, namely with the task to apply one's own moral orientations to judging arguments in favor and against the solution of a moral dilemma. For most people this is a very difficult task. Many people find it even hard to cope with arguments at all, while others respond to arguments but judge them only in regard to their opinion agreement yet not in regard to their moral quality (Keasey, 1974). No other test of moral development defines or incorporates a difficult moral *task*, which would be a prerequisite to call it a moral *competence* test (Emler, Renwick, & Malone, 1983).

The C-index describes the degree to which a respondent judges arguments given by other people on the basis of their moral quality rather on the basis of the arguments' agreement with his or her own opinion. By doing so, the MJT follows Kohlberg's requirements for

structural measurement. He defined moral judgment competence in terms of cognitive *structures*, or ways of thinking or judging (Kohlberg, 1984, p. 398). An index for judgment competence, he says, should be abstractable from responses as their form or quality (p. 402). Kohlberg argues that the measurement of structure or competence is warranted only on the grounds of "intelligible ordering of the manifest items," which reveals the principle of *organization* of an individual's tacit moral reasoning (p. 408). All these requirements imply that the measurement of moral judgment competence must use some kind of *consistency* information, namely moral consistency, rather than summated ratings as in Likert scales.

Thus, from a cognitive science point of view, "inconsistent" responses are not a characteristic of a test ("reliability"), but of participants' moral judgment competence. We can no longer interpret consistency as "test reliability" or its lack as "measurement error" (Mischel & Shoda, 1995; Lind & Wakenhut, 1985; Lind, in press). Without the arbitrary attribution of structural properties of behavior to the test situation, we can now assess the competence aspect of morality directly without relying on self-assessment. Because for many people it is difficult to deal with arguments at all, and even very difficult to deal with counter-arguments (Keasey, 1973; Lind, in press), the MJT confronts the respondents with arguments *supporting* and *opposing* their stance on moral dilemmas. Thus the design of the measurement is a N=1, three-factorial, multivariate design, the arguments representing three design-factors: opinion-agreement, quality or stage of moral orientation, and dilemma context. This kind of experimental measurement borrows ideas from Brunswik's (1955) *diacritical method,* personal construct theory (Kelly, 1955), information integration theory and cognitive algebra (Anderson, 1990), as well as facet analyses and structural theory (Guttman, 1971).

Using moral consistency as a source of information about a participant's "structure,"

"form of response," or "intelligible ordering of the manifest items" implies, however, that we must define the quality of a test in a new way. The MJT is objective, unbiased toward a certain psychological theory, is sensitive to the effects of educational interventions, and fulfills four rigorous criteria of construct validity. *Objectivity*: No subjective rating intervenes with the measurement process that could bias the test scores toward some end. *Unbiasedness*: The MJT was never submitted to "item analysis," that is, no items were introduced or omitted to artificially maximize stability and variability of test scores and test-retest correlations. *Sensitivity*: The MJT showed to be highly sensitive to educationally induced moral development of adolescents ($r > 0.70$; Lind, 2002), while tests of moral preferences showed no or only little effects ($r = 0.13$; e.g., Schläfli et al., 1985). *Construct validity*: To certify MJT versions as valid, scores must meet four rigorous criteria grounded in cognitive-developmental theory: a) Rest's (1969) "preference hierarchy" postulates that Kohlbergian Stage one is preferred least and the higher Stages are preferred increasingly according to their number. Note that six stage preferences can be ordered in 720 different ways and thus the probability of the predicted order is only $p = 1/720 = 0.001$, making it high unlikely that this prediction is confirmed by chance. b) According to Kohlberg's (1958) "quasi-simplex" theory, the preference for neighboring stages should be highest and lower for more distant stages so that the correlations are highest along the diagonal of an inter-correlation matrix and fall off toward the corners. c) According to Piaget's (1976) "affective-cognitive parallelism" theory, when nothing is at stake, preferences for each Kohlbergian stage should be systematically correlated with the index for moral judgment competence in such a way that the higher the judgment competence is the more should lower stages be rejected and at the same time upper stages be accepted. d) Finally, participants should not be able to simulate the index

of moral judgment *competence* upward through simple instruction (Lind, 2002; Lind, in press), as they can when instructed to simulate tests of moral *preferences* (Emler et al., 1983; Rest et al., 1999, p. 115). All 27 language versions of the MJT meet all four criteria (see, e.g., Gross, 1997; Bataglia et al., 2003; Schillinger et al., 2003; Lind, 2005; Lind, in press). Because of these findings and because the MJT is the only pure test of moral judgment competence, it is highly suitable for research and evaluation purposes. Copies of the test are available for free from the author. The use for individual diagnosis or selection purposes is not permitted.

In sum, the MJT is an operationalization of Kohlberg's notion of moral judgment *competence* and Piaget's two-aspect model of behavior, and thus differs from tests of preferred moral orientations, or even Kohlberg's own measurement methodology. Insofar, it "differs," as Villegas (2005) notes, from the Stage model. The MJT is as much or as little a "preference" test as scholastic achievement tests, in which the participant must "prefer" the right answer, is one. The MJT is a competence test because to rate counter-arguments in regard to their moral quality is a difficult *task* for many, and the C-score reflects the respondents' competence to cope with this task. As the MJT has been successfully submitted to sophisticated and rigorous validation studies, the C-score should be considered a valid index of moral judgment *competence* as Kohlberg (1964) has defined it.

References

Anderson, N. H. (1990) *Contributions to information integration theory*. Hillsdale, NJ: L. Erlbaum.

Bataglia, P., Schillinger-Agati, M., Lind, G. & , Quevedo, T.L. (2003) *Testing the segmentation hypothesis with an Extended Version of the MJT*. Poster presented at the meeting of the Association for Moral Education, Krakow, July 19, 2003.

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., & Speicher, B. (1987). *The measurement of moral judgment. Volume I+II*. New York: Columbia University Press.

Emler, N., Renwick, S., & Malone, B. (1983) The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology*, 45, 1073-1080.

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

Guttman, L. (1971) Measurement as structural theory. *Psychometrika,* 36, 329-347.

Keasey, C.B. (1973). The influence of opinion-agreement and qualitative supportive reasoning in the evaluation of moral judgments. *Journal of Personality and Social Psychology, 30*, 477-482.

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

Kohlberg, L. (1958) *The development of modes of moral thinking and choice in the years 10 to 16.* Unpublished 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. New York: Russel

Sage Foundation, pp. 381-431.

Kohlberg, L. (1984) *Essays on moral development, Vol. I. The psychology of moral development*. San Francisco, CA: Harper & Row.

Lerkiatbundit, S., Utaipan, P., Laohawiriyanon, C., & Teo, A. (2004) *Randomized controlled study of the impact of the Konstanz method of dilemma discussion on moral judgement.* Unpublished report, University of Sonkla, Thailand.

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 possibilities of measuring a complex construct]. In G. Portele, (Ed.), *Sozialisation und Moral*. Weinheim: Beltz, pp. 171-201.

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

Lind, G. (2002) *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung* [Can morality be taught? Research findings from modern moral psychology.] Berlin: Logos (3rd ed.).

Lind, G. (2003) *Moral ist lehrbar. Handbuch zur Theorie und Praxis moralischer und demokratischer Bildung*. [Morality can be taught. Handbook on theory and practice of moral and democratic education]. München: Oldenbourg.

Lind, G. (2005) Validation and certification procedure for the Moral Judgment Test. http://www.uni-konstanz.de/ag-moral/mut/mjt-certification.htm

Lind, G. (in press) The meaning and measurement of moral judgment competence – A dual aspect theory. In D. Fasko & W. Willis (Eds.), *Contemporary philosophical and psychological perspectives on moral development and education*. Creskill, NJ: Hampton.

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*. Chicago: Precedent Publishing Inc., pp. 79-105.

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.

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

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

Rest, J. R. (1969) Level of moral development as a determinant of preference and comprehension of moral judgments made by others. *Journal of Personality,* 37, 220-228.

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

Schillinger-Agati, M. & Lind, G. (2003) *Moral judgement competence in Brazilian and German university students*. Paper presented at the Annual Meeting of the American Education Research Association, Chicago, April 2003.

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.

Villegas de Posada, C. (2005) Some problems in use of the Moral Judgment Test. *Psychological Reports, 96,* 698-700.