

Measuring Moral Judgment: A Review of *'The Measurement of Moral Judgment'* by Anne Colby and Lawrence Kohlberg¹

Volume I: Theoretical Foundations and Research Validation, 397 pp²

Volume II: Standard Issues Scoring Manual, 975 pp³

Georg Lind

Universität Konstanz, FRG

Thirty years ago, in his 1958 dissertation, Lawrence Kohlberg published the outline of a new methodology for measuring moral judgment. He claimed to provide a (valid and reliable) measure of the structure of moral thought and behavior that was to revolutionize psychological assessment in the socio-moral domain.

In the early 1970s, in preparation for a longitudinal cross-cultural study of the process and effects of higher education, we were searching for concepts and instruments for measuring socio-moral competencies in young adults. Most social-psychological measurement dealt solely with attitudes to-

ward moral values, rather than the skills of applying values, or principles, to complex situations and solving conflicts between moral principles. This shortcoming, we felt, accounted for the lack of evidence concerning moral-development effects of (higher) education [Lind, 1985a]. We believed that Kohlberg's structural methodology could fill this void.

For nearly three decades, Kohlberg and his (numerous) associates refined and validated his system for scoring reasoning about moral dilemmas. During this time, the measurement of moral judgment consisted of research in progress, most information being unpublished or scattered in diverse journals. In order to utilize this system, and to keep up with its many revisions, students and reviewers of the scoring system were obliged to travel to workshops held at the Harvard Center of Moral Education, an inconvenience, even annoyance, to many researchers and practitioners.

The publication of the present volumes gratifies the long-suffering audience. On the

¹ Cambridge University Press, New York 1987; ISBN 0-521-32565-X set; ISBN 0-521-24447-1 vol. I; ISBN 0-521-32501-3 vol. II.

² In collaboration with Anat Abrahami, John Gibbs, Ann Higgins, Kelsey Kauffman, Marcus Lieberman, Mordechai Nisan, Joseph Reimer, Dawn Schrader, John Snarey, and Mark Tappan.

³ In collaboration with Betsy Speicher, Alexandra Hewer, Daniel Candee, John Gibbs, and Clark Power.

one hand, it is a very comprehensive, and well-written, manual for scoring moral judgment. It provides a thorough introduction to the interviewing and coding technique (vol. I, chapt. 6, 7). Volume II contains many 'match' samples and illustrative examples of moral 'norms', 'elements,' and 'structural criteria' for scoring moral interviews according to the new Standard Issue Scoring System (SIS). On the other hand, the volumes also contain a summary of the 'psychological and philosophical assumptions underlying the construct and the measure', [p. 1]⁴ and their operational meaning: 'Standard Issues Scoring is offered not only as a measurement and research tool, but also as an operational definition of the stages (of moral development)' [p. vii]. This link between theory and methodology is characteristic of Kohlberg's hypothetico-deductive epistemology: 'The Standard Issue Moral Judgment Interview and Scoring System is a theory-based assessment procedure' [p. 1; see also Broughton, 1978; Lind, 1985b]. This sets it apart from many psychological tests that claim to be less theory-tied, if not theory-free.

Following the outline of volumes, in this review I discuss some of the major historical, theoretical and methodological arguments they contain. For brevity, I refer mostly to Kohlberg as author, in acknowledgement of the fact that the late Lawrence Kohlberg founded, and made the major contribution to, this approach. This is not to neglect the first author, who has been the driving force for many years, and the many collaborators who have made substantial contributions to the work. This multiple authorship might account for both the creativity and some inconsistency in the approach.

History: From Intuition to Science

In the fifties, growth in psychology was 'erratic' and there was 'much casting about for the most crucial problems and the most powerful methods' [Wolfe, 1963, p. v]. A group of scholars, appointed by the American Psychological Association to recapitulate the state of the art, called any theory that addressed inner psychological processes and intervening variables 'sharply into question' [Koch, 1963, p. 735]. Since then, experimental psychology has withdrawn from central themes such as morality and personality, structure and development, in the eyes of some focussing instead on psychologically peripheral phenomena (i.e., physical and physiological aspects of behavior). In empirical studies of personality, structure became operationally defined mostly in terms of interindividual data (e.g., as the factor structure of a set of variables across a set of persons) rather than in terms of individual behavior patterns. As a result, structural aspects of human behavior and its development became invisible in psychological research. The moral domain was separated from the cognitive domain, and morality was narrowly defined as expressed moral knowledge and moral attitude [Krathwohl et al., 1964].

In this context, Kohlberg proposed a structural theory of moral judgment and an integrating methodology for studying cognitive aspects of moral behavior. He was well aware of the psychological assumptions underlying conventional psychometrics, which is essentially based on the analysis of interindividual correlations between test items. Therefore, Kohlberg initially based his scoring system largely on informed psychological intuition (e.g., global rating) and considered alternative methodological approaches to as-

⁴ If not stated otherwise, quotes refer to vol. I.

sessing validity [Kohlberg, 1981; Oser, 1988]. Intuitions are the major, if not the sole, source of revolutionary developments in science. The next important step however is explication by means of a theoretically valid, objective methodology. Explication is indispensable for both the conveyance and the critical examination of intuitions through rational discourse, and hence for scientific progress. In Kohlberg's case, the theory became alternatively convincing or dismaying before it was altogether understood.

The change from the original scoring system to the new SIS reflects a change from 'a view of interpretation as an art to a view of interpretation as a science' [p. 40]. Many critics have reproached Kohlberg for his alleged neglect of traditional standards of test construction and analysis. The present volumes disprove this reproach. In explicating Kohlberg's basic intuitions, the authors comply fully to the standards of conventional psychometrics. In particular, they rely on the idea of interindividual correlation, as a basis for test analysis: 'Standard Issue Scoring is intended ... to achieve greater objectivity and reliability in scoring by specifying clear and concrete stage criteria and to define the developmental sequence of the specific moral concepts within each stage as well as the sequence of the global or general stage structures' [p. 40].

By the standards of conventional psychometrics, the SIS is highly reliable. This is broadly documented by an unusual array of cross-sectional, longitudinal, and intercultural studies (vol. I, chapt. 2-5). Of course, this evidence does not establish that the SIS is a valid measure of moral judgment. Successful compliance with conventional standards of psychometrics does not suffice to insure congruity between theory and methodology.

Theory: Intuitions about Moral Stage and Behavior

Kohlberg's cognitive-developmental theory of moral judgment rests on grand intuitions. These intuitions were informed by Hobhouse's and G.H. Mead's notions of levels of social perspectives, Spinoza's logic of ethics, Kant's concept of moral cognition, and Piaget's developmental theory of both children's moral behavior (from imitation and heteronomy to autonomy) and logical-physical reasoning (from preoperations to concrete to formal operations). With regard to epistemology, he extracted notions from Pierce's idea of abduction (Kohlberg used the term 'bootstrapping'), Weber's ideal-types, and, more recently, Habermas' communicative hermeneutics. Kohlberg possessed thorough knowledge of these sources, as well as experimental and statistical methodology.

Kohlberg [1984] has condensed these intuitions into four basic assumptions: (1) Morality encompasses cognitive structural aspects; (2) persons at different stages of cognitive-moral development show structurally different modes of moral decision-making (rather than just changes in moral attitudes); each mode of moral thought forms a structural whole; (3) these different modes of thought form an invariant sequence, or order, of moral development; and (4) cognitive-moral stages are hierarchically integrated. Stages form an order of increasingly differentiated and integrated structures to fulfill a common function. The last assumption has two corollaries: (a) individuals usually prefer a solution of a moral problem at the highest level available to them, and (b) the moral stage sequence reflects an order of increasing difficulty, or complexity, and,

thus, it also reflects different levels of judgment competence. [For further discussion of these assumptions, see Lind, 1985b.]

Throughout his writings, Kohlberg spoke of moral judgment in terms of behavior rather than latent moral thoughts or merely linguistic material. 'What we care about is how moral judgments are made when (a moral principle) is actually applied to values in conflict. Morality is a matter of choice and decision. It is not just a matter of using abstract concepts like justice' [p. 58]. Accordingly, the authors 'are concerned with actual moral judgments. We are not trying to measure an individual's abstract philosophical positions, but rather what is going to make a difference when that individual is faced with an actual moral choice' [p. 58].

There is some ambiguity about the term 'behavior' in this context, which has caused much confusion between Kohlberg and his critics. In a subject's reasoning about moral dilemmas, three different kinds of behavior need to be distinguished. (1) The endorsement of a particular decision is too ambiguous to serve as a basis for inferring moral stages. For example, in the renowned Heinz dilemma the subject is asked to make a decision whether or not Heinz should steal a very expensive drug, which he cannot afford to buy, in order to save his ill wife's life. This decision is situation-bound. Modifications of the situation (e.g., if the drug is paid for by health insurance, or the respondent takes the perspective of either Heinz or his ill wife) will cause the subject to respond quite differently, without affecting his or her moral structure. (2) If certain norms and values can be linguistically classified as 'moral' (e.g., honesty, justice), their use signifies that the subject values, and knows, moral words. But in moral terms, the endorsement of moral

words is ambiguous. For example, the subject may say, 'It's wrong for Heinz to steal a drug in order to save his wife's life, *because it's against the law.*' Taken by itself, this statement may be interpreted as stage 1/2 [p. 229]. However, other stage interpretations would be just as valid. The 'norm of law is found to be used at every stage' [Kohlberg, 1985, p. xvi]. Furthermore, mentioning moral norms does not show whether or not a person is willing, and able, to apply moral norms and values in concrete decision-making. When pushed to give reasons, a person may mention moral principles merely to explain a choice. (In psychoanalytic terms, such post hoc explanations are called 'rationalization'.) (3) According to cognitive-developmental theory, the constitutive component of a subject's moral behavior is the degree to which a person *applies* moral principles consistently and differentiatedly to decision-making. Hence, 'the measurement of moral judgment involves the analysis of observable patterns of thought revealed in the subject's responses to the moral interview' [p. 4].

Methodology: The Design of Probing and Scoring

Since, from a cognitive-developmental point of view, neither tests of moral decision-making nor tests of moral attitudes can provide insight into moral judgment, a structural methodology is needed that combines, and extends, the two. These volumes provide, or aim at, such a methodology, while at the same time maintaining the standards of conventional psychological test construction and analysis. Kohlberg's methodology facilitates the measurement of the structural com-

ponent of a person's judgment behavior in three interrelated ways. First, in a Piagetian manner, the subject is confronted with a moral dilemma, that is, a short story in which two or more moral principles oppose each other. He or she is asked to make a choice. Second, the interviewer uses intensive probing, that is, *why* questions, and questions that stimulate the respondent to consider varying situational contexts. Third, stage scoring of the interview is based on well-conceived and meaningful measurement units.

Through the confrontation with moral dilemmas, the subject is stimulated to consider moral norms rather than merely technical knowledge involved in solving a problem. (Most people suggest a technical solution first, which seems an appropriate strategy in most everyday decision-making.) With the moral dilemma, the stage is set for a moral discussion between the subject and the interviewer. According to the design of the SIS interview, the choice between two alternative courses of action implies a choice between two moral values (or issues) and a particular cluster of moral norms [pp. 45-49]. The authors maintain that, neither logically nor empirically, would this choice determine the stage of moral reasoning the subject uses.

Issue and stage are logically independent by virtue of the scoring system's design. Unfortunately, the authors provide insufficient data to support their empirical hypothesis. Issue scores, based on a selected sample, are presented only in one figure [p. 111]. No information is provided about sample sizes and dilemmas (and issues) involved. In regard to a former version of the Kohlberg interview, Eckensberger [1983] observed a correlation between moral issue, or norms, and stages. In the studies reported in these

volumes, the different pulls of certain material toward certain stage classifications seem to be largely neutralized by intensive probing. Other studies (in particular those relying on written responses) may not be as successful. But even in the studies reported, 'absolute differences between the forms' (containing different content) have been found [p. 66].

Probing and high quality interviews are 'of critical importance for assessing moral judgment stage' [p. 151]. Probing is necessary (a) to achieve good scorability of answers [p. 169]; (b) to elicit the most advanced stage of moral reasoning of which a subject is capable [pp. 5, 61]; and (c) to test the degree to which the subject is capable of applying moral principles to decision-making [p. 58].

In contrast to its great importance, interviewing is not treated systematically in these volumes. Interviewing is still largely founded on intuition and common sense [pp. 153-158]. In some instances, instruction remains vague. For example, the researcher is instructed to probe intensively, but at the same time he or she should not ask 'why' too many times [p. 54]. For each dilemma, the standard interview questions systematically touch upon the two main issues involved in the dilemma. Yet, the authors leave it to the interviewer to probe into the norm conflicts involved in a dilemma [vol. II, pp. 4, 56]. While, in his dissertation, Kohlberg [1958] used probing to test the firmness of moral convictions, now probing is used to produce only a sufficient amount of linguistic material for scoring [p. 151]. The authors play down the significance of probing nonselected issues.

Scoring interview judgments according to moral stages involves three major steps:

(1) identify the proper unit of measurement; (2) assign a stage score to it, and (3) calculate global and average scores for the individual. In the first step, the scorer is to proceed as follows: Start with an issue (e.g., the issue 'life' in the Heinz dilemma), and classify all responses by moral norm and by moral element (the manual gives ample explanation for both). The resulting *issue* \times *norm* \times *element* combination is called the 'measurement unit'. This labelling is misleading, however, because, in the SIS, this unit is the basis of hermeneutic analysis rather than of measurement operation. The authors' alternative term 'interview judgment' (IJ) seems more appropriate. IJs are classified according to the issues to which they are related. The 'rules for issue classification' [pp. 163–165] provide the actual unit of measurement. The authors use this unit as a basis for scoring and analysis of test reliability and validity.

In the SIS, the aim of hermeneutic analysis is to understand the socio-moral perspective, or stage, of the subject's reasoning with respect to a particular issue. This analysis involves [pp. 165–177]: (1) tentative rating ('reread all the responses classified under the issue you are scoring and think about what they mean'); (2) deciding about scorability (Do the IJs provide reasons, are they considered valid by the subject, and are they prescriptive in nature?); (3) guess scoring and transposition ('constructing meaning out of garbled words and mangled sentences'); (4) matching IJs with the Criterion Judgments (CJs) of proper stages (vol. II); and (5) 'structural evaluation' of proposed matches to 'weed out or "veto" those literal matches for which the stage of the criterion judgment is not a true reflection of the structural significance of the interview material'. False stage matches between IJs and CJs mainly result when 'lower stage ideas are

included within a higher stage'. According to the rules, 'only the most mature expressed version of a particular moral idea is scored' [p. 5].

In hermeneutic analysis, which is to be clearly demarcated from conventional test analysis, clusters of information must not be disjoined into independent items, but are used as a whole to assess the meaning of a person's reasoning. Its major tool is the concept of consistency. It would be pointless to use consistency information for calculating conventional coefficients of reliability and internal consistency.

The authors contribute to the misunderstanding by sometimes using an inadequate terminology, as for example in this statement: 'A scoring system based on a very literal, "strict interpretation" approach might yield greater reliability, but responses could be misscored from the structural point of view. The increase in reliability would be outweighed by a decrease in validity' [p. 172f.]. The terms 'reliability' and 'validity', which are clearly defined only within conventional test analysis, do not apply to hermeneutic, or structural, analysis. In the latter, meaningless individual pieces of information are drawn together to constitute meaningful units of information. In the former, meaningful units of information are assumed and taken as a basis for evaluating the quality of the measurement tool. Thus, confusing the struggle for reliability and for meaningfulness would indeed compromise the overall validity of the measurement. The new SIS avoids this fallacy. Rest's [1979] relevant critique of earlier versions does not apply anymore.

Reliability is exceptionally high. The authors report that raters agree almost perfectly within the range of one third of a stage.

Raters were university graduates with psychology background who had worked with the SIS for some time, or who at least had participated in a 4-day scoring seminar. Scores are quite stable within a 3- to 6-week period (test-retest reliability is in the high 0.90s). Alternate form reliability is reported to be 0.95 [pp. 63–65]. However, the three forms are not absolutely equivalent, since scores for these forms differ up to one fifth of a stage [p. 68].

More subtle problems linger on. The scoring rules favor moral ideas at the highest stage of which the subject is capable. Both 'transposition' of garbled answers and 'structural evaluation' ensure that the researcher assigns the highest stage. Yet these rules do not guarantee that the assigned stage is a subject's true stage of moral judgment competence. The SIS does not unambiguously reflect the subject's ability to apply moral principles to action choice. It fails to evaluate structural information, like the degree of differentiation and integration of the subject's moral judgment *behavior*. Relevant differences of reasoning between issues, e.g., the fact that some subjects produce fewer reasons, and lower stage scores, for nonchosen than for chosen issues [p. 161], are excluded from hermeneutic analysis by the authors' issue classification rule. Differences within an issue are excluded by the lower-stage inclusion rule.

After each issue has been assigned a stage score, these stage scores are added to yield two overall scores for a subject: global stage score (ranging from 1 to 5, including pure, transitional, and minor stages), and weighted average score (ranging from 100 to 500) on the basis of three or more dilemmas (depending on how many of the three forms of the interview are administered). The use

of mixed stage scores rather than pure stage scores creates an interpretation dilemma. [Notably, in the one study that reports the distributions of stage usage, only 29 of 188 subjects are assigned to a pure stage of moral judgment; pp. 136–139]⁵. On the one hand, we may argue that mixed stage classifications are due to unreliability (i.e., measurement error). If stages of moral judgment represent structural wholes, a subject should receive the same score on each issue or dilemma. The unreliability assumption saves the structural whole hypothesis, but it also renders both fractions of a stage, and differences smaller than one stage, meaningless. On the other hand, we may construe mixed stages as stage transition. In this case, Kohlberg's structural whole assumption is no longer tenable.

This dilemma points to a fundamental problem in conventional test construction, and analysis [Lumsden, 1976]. The proverbial cake cannot be had and eaten too. The consistency, or variance, observed in research data cannot be attributed to both the subject's disposition and the measurement instrument at the same time. This problem also affects the question of measurement validity.

Validity: Does the SIS Measure Moral Judgment?

'The appropriate question is whether the interview and scoring system provides a valid assessment of moral judgment stage' [p. 71]. How can we answer this question? The authors give a straightforward answer.

⁵ This is my calculation. There is no information given about the basis of the 'Stage Usage Percentage'. I presume it is based on the issue unit.

They argue that SIS is valid if 'it yields scores that agree closely with the theoretical predictions of invariant sequence and internal consistency' [p. 71]. If validity is understood this way, the longitudinal data presented do provide substantial support for the validity of SIS. This support is unusually strong and based on unusually thorough investigations. The main longitudinal study by Kohlberg included six test occasions over a period of 22 years [pp. 80–82]. The results agree well with the developmental sequence hypothesis. Additionally, the cognitive-developmental nature of the SIS scores is supported by the fact that persons' moral judgment scores can be increased by exposing them to conflicting opinions [Walker, 1986].

However, these findings are merely necessary, not sufficient, for establishing the validity of a structural measure of moral judgment. There are many measures that fulfill these two criteria but are *not* valid measures of moral judgment (e.g., weight and height, mathematical skill). Yet another criterion must be satisfied in order to consider the system valid. This criterion pertains to the nature, or quality, of the task presented to the subject. The validity question must be extended to ask: How well does the Kohlberg interview challenge the subject's ability to apply moral principles to decision-making in a moral dilemma task?

Although this question is most crucial for validity, it is not raised in these volumes (nor is it raised in most literature on psychological tests). The solution, and discussion, of a truly moral dilemma is a difficult task. Subjects are asked to coordinate multiple expectations and values while trying to understand the task. First, of course, a commitment must be made to a particular course of action: Do you think Heinz should steal the

drug? Yes, the subject may respond. Are you sure? Yes, I would do it if I were in his situation. Second, a justification must be given. Why do you think Heinz should steal the drug? Because his wife needs it or will die without it. Third, this justification must be generalized. Is it important for people to do everything they can to save another's life? Fourth, the subject is asked to consider moral norms that oppose the course of action that he or she has advocated. Is it against the law for Heinz to steal? Does that make it morally wrong?

There is ample evidence that these questions form an order of increasing difficulty [Keasey, 1974; Lind, 1985b]. With less developed subjects, the interview hardly goes beyond the first or second question. Even many adult subjects find it difficult, if not impossible, to participate in rational discourse regarding opinions opposed to their own.

Thus, the moral judgment interview can be considered a valid test of moral judgment competence if it is construed as a moral task. The authors, however, did not intend this, and, maybe for this reason, did not discuss questions of proper interview design and scoring methodology.

Conclusion

This review has addressed the historical and theoretical foundations and the methodological design of the Standard Issue Interview and Scoring System. Because of limited space, the types of moral judgment have been excluded [for a general discussion, see Lind, 1985b].

In sum, these volumes present an unprecedented methodology for understanding

moral-cognitive development. Developmental, school and clinical psychologists will find it very useful. The methodology, grounded in a grand theory of moral judgment and its development, is supported by a massive body of empirical research. Yet, as we have seen, the meaning and measurement of structure are still the Achilles' heel of the system.

Transforming informed intuitions about the structure of moral judgment into an objective methodology is a formidable task. Kohlberg dedicated most of his research career to it. To some of us, it seemed like an obsession paralyzing his other work [Oser, 1988]. However, this dedication has proven most fruitful. His mission proved to be a great stimulation and an invaluable source for others' efforts, for example Rest's [1979] Defining Issues Test and our 'Moralisches Urteil-Test' [Lind and Wakenhut, 1985]. The Standard Scoring System is unique for its clearness, level of standardization, and serious treatment of its subject. By the standards of classical test theory, it is a highly reliable and valid measure. Last but not least, the methodology (and its descendants) made it possible for researchers to demonstrate convincingly the immense effects of education on cognitive-moral development [Lind, 1985c]. The importance of this achievement can hardly be overestimated.

Nevertheless, the authors' struggle to accommodate their scoring system to conventional psychometrics may have obstructed the development of a genuinely structural methodology. Kohlberg's [1981] criticism of the psychological and epistemological assumptions involved in conventional test analysis failed to impress his research group. In my view, in order to improve construct validity, future revisions of the system must

take into consideration the dynamics of moral judgment behavior in the interview process, rather than merely its linguistic outcomes [Lind and Wakenhut, 1985]. A task conception is necessary to address the problem of proper design in regard to (a) dilemmas and issues, (b) interview probing, and (c) structural scoring. Although it insinuates an iconoclastic digression from conventional psychological methodology, a task conception of interviewing will help to fulfill the main goal of cognitive-developmental psychology – the assessment of structure.

References

- Broughton, J. (1978). The cognitive-developmental approach to morality: A reply to Kurtines and Greif. *Journal of Moral Education*, 1, 81–96.
- Eckensberger, L. (1983). Research on moral development in Germany. *The German Journal of Psychology*, 7, 195–244.
- Keasey, C.B. (1974). The influence of opinion-agreement and quality of supportive reasoning in the evaluation of moral judgments. *Journal of Personality and Social Psychology*, 30, 477–482.
- Koch, S. (1963). Epilogue. In S. Koch (Ed.), *Psychology: A study of a science: Vol. 3. Formulations of the person and the social context* (pp. 729–788). New York: McGraw-Hill.
- Kohlberg, L. (1958). The development of modes of moral thinking and choice in the years 10 to 16. Unpublished doctoral dissertation, University of Chicago.
- Kohlberg, L. (1981). *The meaning and measurement of moral development (The Heinz Werner Lecture Series, Vol. XIII)*. Worcester MA: Clark University Press.
- Kohlberg, L. (1984). *Essays in moral development. Vol. II. The psychology of moral development*. San Francisco: Harper & Row.
- Kohlberg, L. (1985). Foreword. 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.

- Kohlberg, L., Colby, A., Gibbs, J., Speicher-Dubin, B., & Power, C. (1977). Assessing moral stages: A Manual. Part I: Introduction. Unpublished ms., Harvard University.
- Krathwohl, D.R., Bloom, B.S., & Masia, B.B. (1964). *Taxonomy of educational objectives. Handbook II: affective domain*. New York: MacKay Company.
- Lind, G. (1985a). 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.
- Lind, G. (1985b). The theory of moral-cognitive development: A socio-psychological assessment. In G. Lind, H.A. Hartmann, & R. Wakenhut (Eds.), *Moral development and the social environment: studies in the philosophy and psychology of moral judgment and education* (pp. 21–53). Chicago: Precedent Publishing.
- Lind, G. (1985c). Moral judgment competency and education in democratic society. In P. Wein-gartner & G. Zecha (Eds.), *Conscience: An interdisciplinary approach* (pp. 91–121). Dordrecht: Reidel.
- 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.
- Lumsden, J. (1976). Test theory. *Annual Review of Psychology*, 27, 251–280.
- Oser, F. (1988). Verlust im Gewinn: Biographie und Determination des Entwicklungsmodells von Lawrence Kohlberg. *Schweizerische Zeitschrift für Psychologie*, 47, 109–119.
- Rest, J. (1979). *Development in judging moral issues*. Minneapolis MN: University of Minnesota Press.
- Walker, L. (1986). Cognitive processes in moral development. In G.L. Sapp (Ed.), *Handbook of moral development* (pp. 109–145). Birmingham AL: Religious Education Press.
- Wolfe, (1963). Preface. In S. Koch (Ed.), *Psychology: A study of a science: Vol. 3. Formulations of the person and the social context*. New York: McGraw-Hill.

Georg Lind
Universität Konstanz
Postfach 5560
D-7750 Konstanz (FRG)

Note: The *Moral Judgment Test*, mentioned above, has been renamed into *Moral Competence Test* (MCT). See: Lind, G. (2016). How to teach morality. Berlin: Logos.

<http://www.uni-konstanz.de/ag-moral/>