

Lind, G. (2013). Thirty years of the *Moral Judgment Test* – Support for the Dual-Aspect Theory of Moral Development. In: C. S. Hutz & L. K. de Souza, eds., *Estudos e pesquisas em psicologia do desenvolvimento e da personalidade: uma homenagem a Angela Biaggio*, pp. 143-170. Sao Paulo: Casa do Psicólogo.

35 Years of the *Moral Judgment Test* –
Support for the Dual-Aspect Theory of Moral Development¹

Georg Lind²

University of Konstanz, Germany

Abstract

After the Moral Judgment Test (MJT) has been in use for more than 30 years, I feel it is time to check on its validity and fertility. For developing the MJT, I drew upon ideas from various sources, e.g., from philosophical ideas, e.g., Jürgen Habermas and Karl-Otto Apel's notion of ethical discourse competence, and from psychological concepts, e.g., Jean Piaget's notion of affective-cognitive parallelism, Kohlberg's definition of morality as a competence, Jim Rest's postulate of hierarchical moral preferences), and from various 'cognitive-experimental' approaches to psychological measurement, e.g., Torgerson's concept of response scaling, Norman Anderson's cognitive algebra; Guttman's facet analysis.

My main intention was to assess, beside the affective aspect of moral orientation, also the cognitive aspect of moral behavior, namely *moral judgment competence*. This aspect has

¹ This chapter is based on a presentation at the conference of the Association for Moral Education (AME), Nov. 3.-5. 11. 2005, Cambridge, MA. I like to thank Thomas E. Wren for his invaluable help in revising this paper, James Fearn and Patricia Knoop for proof-reading. Yet, I take all responsibility for remaining mistakes.

² Contact: Prof. Dr. Georg Lind, Department of Psychology, University of Konstanz, 78462 Konstanz, Germany. Georg.Lind@uni-konstanz.de; Home-page: <http://www.uni-konstanz.de/ag-morall>

been hitherto neglected by psychological research or been inadequately addressed as a separate component of behavior. To make sure, that the MJT really measures these constructs, I have rigorously tested its theoretical validity by using four theoretical predictions or criteria found in the research literature: Hierarchy of moral orientations, quasi-simplex-structure of the inter-correlations of these orientations, competence aspect of moral behaviour, and affective-cognitive parallelism.

In more than three decades of research, the MJT has stood the test of time. In numerous experimental, longitudinal and correlational studies in many different countries and with many different participants, not a single study has so far refuted any of these four theoretical predictions. This is a remarkable achievement, because the MJT has been constructed solely on the basis of theoretical considerations and has not been submitted to empirical item selection. No attempts were made to improve its correlation with any of these four criteria or with age. Because of the theoretical validity of the MJT, the accumulated research evidence also demonstrates the empirical validity of the dual-aspect theory of moral behavior. MJT-research, including neuro-imaging studies, especially corroborates the notion that morality requires not only certain moral orientations but also the ability to act accordingly.

Moral Judgment Research: Conjectures and Refutations

The advancement of our knowledge of morality depends strongly on the advancement of measurement in this domain and vice versa. We need tests in order to validate our theories. On the other hand, we need knowledge about the nature of morality in order to design adequate methods of measurement and data analysis. In other words, our measurement tools rest as much on assumptions as they reflect reality. All our observations are “soaked with theory”, wrote the philosopher Karl Popper (1968a, p. 387). “There is no measurement without a theory and no operation which can be satisfactorily described in non-theoretical terms. The attempts to do so are always circular.” (p. 62). Elsewhere he stated, “There is no such thing as 'pure experience,' but only experience interpreted in the light of expectations or theories which are 'transcendent'.” (Popper, 1968 b, p. 425).³ Thus, the scope and limits of our knowledge of morality determine what we can measure and how well we can measure it. The late Lawrence Kohlberg (1984) called this process “bootstrapping” and “saving circularity:” “I have already noted that there is a certain circularity involved in assumptions about truth of a theory and validity of a test. Only a bootstrapping spiral can make this a saving circularity. [...] saving circularity is at the heart of scientific epistemologies of pragmatism of Charles Sanders Pierce and John Dewey.” (p. 424)

This spiral process usually sets in when our knowledge does not work well anymore, that is, when our predictions fail. Then we look out for “bold ideas, unjustified anticipations, and speculative thought” (Popper, 1968 b, p. 280), create new theories, design new measurement tools, test these tools, and gather data to test the empirical validity of the new theories, hoping for a “progressive problem shift” (Lakatos, 1972). A progressive problem shift is a decisive condition for scientific progress because it lets us “predict some novel,

³ “For example, the description of the measurement of length needs a (rudimentary) theory of heat and temperature-measurement; but these, in turn, involve measurements of length.” (Popper, 1968a, p. 62)

hitherto unexpected fact [and] leads us to the actual discovery of some new facts.” (Lakatos, 1972, p. 118)

However, this process can also go wrong in two ways. First, we can immunize our theories against any falsification by creating tests in such a way that they always support our theories, or by rejecting inconsistent data as caused by measurement error or lack of test validity. Lakatos has called this a “degenerating problem shift,” as it immunizes the theory against falsification and hinders any scientific progress.

Second, we may ignore the fact that a particular measurement tool is soaked with theory and that it may be soaked with the wrong theory, and still use it for testing a theory’s empirical validity. In other words, the hidden psychological assumptions about the nature of morality built into some psychometric test may be totally at odds with the assumptions of the theory for which it is used.

The Dual-Aspect Theory

The dual-aspect theory of moral behavior and development, I believe, is such a “progressive problem shift” in the domain of moral psychology and education. This theory states that moral behaviour must be described in terms of affective and cognitive aspects, that is, in terms of the moral orientations which this behavior expresses and the moral competencies which it reveals. In emphasizing the dual nature of moral thinking and behaviour, this theory overcomes the constraints and inconsistencies of the two older and still prevailing paradigms in this domain, namely behaviorism and ‘motivationalism.’

Behaviorism defines the morality of an action which conforms to culturally given

norms or rules (e.g., Hartshorne and May, 1928; Skinner, 1971). Typically, the ‘morality’ of behavior is measured by observing people’s reactions in situations in which they are tempted to transgress social norms. No attention is given to their own moral motives and principles. The limitations of the behaviorist norm-conformity approach are most succinctly pointed out by two outspoken behaviorists, namely Hartshorne and May (1928). On the very last page of their research report, after their experiments did not produce the expected moral trait, they admit that behavior cannot be studied without reference to an actor’s motives: “The essence of the act is its pretense. Hence it can be described and understood only in terms of the human elements in the situation. It is not the act that constitutes the deception, nor the particular intention of the actor, but the relation of this act to his intentions and to the intentions of his associates.” (p. 377)

Motivationalism is the theory that moral action is caused mostly, if not solely, by moral intentions and affects (see, e.g., Emler et al., 1983; Haidt, 2001). Morality is seen as belonging solely to the “affective domain” of human behavior and not to the “cognitive domain” (Bloom et al., 1956; Krathwohl et al., 1964). Hence, this approach uses exclusively tests of moral preferences, and moral value inventories. Many of these tests were designed to measure people’s preferences in regard to moral values set forth by religious or philosophical authorities. One of the best known moral attitude tests is the *Defining Issues Test (DIT)* by James Rest and his associates (Rest, 1979; 1986). The DIT (P-index) measures “the relative importance give[n] to principled moral considerations on making a moral decision.” (Rest, 1979, p. 101), that is, “the amount of post-conventional thinking (in contrast to other kinds of thinking) preferred by the participant.” (Narvaez, 1998, p. 15). Studies using this test of moral preferences, for example, could show that “moral judgment changes [...] from a preference for pre-conventional thinking to a preference for conventional to one for post-conventional

thinking.” (p. 14, footnote 1).

Emler (1996) has pointed out a severe limitation of confining moral behavior to moral preferences, omitting the cognitive aspect of moral thinking and behavior: “With respect to validity, I have argued that scores on the DIT may reflect a person's political attitudes and not only their developmental level (Emler et al., 1983). Debate on this question continues [...], but I believe the possibility remains that the effects of various kinds of educational experience, as revealed by evidence from DIT scores, may in part or in whole be attitude change effects. This possibility is particularly strong with respect to the effects of participation in higher education [...].” (p. 119)

Emler alludes here to the fact that moral preferences and orientations can be simulated to meet the expectations of the experimenter, i.e. they are influenced by the so-called social-desirability effect. If participants in an educational program sense, or are explicitly told, what kind of effects are expected from the program, they often kindly comply with these expectations. Indeed, Emler et al. (1983) could show in their experiment that the P-score of the DIT could be simulated in any direction if the subjects were instructed to do so. This experimental finding may explain why a direct approach to moral teaching seems to be highly effective in regard to moral preferences (Penn, 1990), and why older participants in a dilemma discussion show larger “gains” in moral preferences than younger subjects (Schläfli et al., 1985). After a round of debate (Markoulis, 1989), Rest and his associates acknowledge that the DIT is only a measure of moral preferences: “We [...] have eliminated the faking study from our set of the validity criteria.” (Rest et al., 1999, p. 115)

Moral motives, intentions, and other affects are indispensable aspects of moral behavior. Without moral orientations there would be no moral behavior. Yet, they are not sufficient (Pittel & Mendelsohn, 1966). As Kohlberg (1964) proposed, morality has also a

very important *competence* aspect,⁴ which links moral orientations and preferences on the one hand with everyday action and decision-making on the other.

We speak of *aspects*, not of components. This is an important distinction. Components are separable classes of behavior which are to be measured with different instruments.⁵ In contrast, aspects are distinguishable, though inseparable properties of a pattern of behavior. Piaget and Inhelder (1969) characterized the relationship between affective and cognitive aspects of behavior this way: "Affectivity constitutes the energetics of behavior patterns whose cognitive aspect refers to the structures alone. There is no behavior pattern, however intellectual, which does not involve affective factors as motives; but, reciprocally, there can be no affective states without the intervention of perceptions or comprehension which constitute their cognitive structure. [...]. The two aspects, affective and cognitive, are at the same time inseparable and irreducible." (p. 158; see also Piaget, 1981; Lourenço & Machado, 1996)

In a similar vein, Kohlberg (1958) argues: "A systematic general observation of moral behavior, attitudes, or concepts in terms of such a set of formal criteria of morality ... cross-cuts the usual neat distinctions between moral knowledge or beliefs on the one hand and moral behavior or motivation on the other, since a moral act or attitude cannot be defined either by purely cognitive or by purely motivational criteria." (p. 16)

⁴ Already in his seminal book *The Descent of Man*, Charles Darwin (1966) mentions *moral abilities*, by which he means man's ability to treat others like oneself (p. 169), and the ability to reflect on one's past actions and their motives, accepting some and rejecting others (p. 268).

⁵ As Higgins (1995) notes, it is hardly possible, to assess cognitive components in isolation: "However, one should note that there are cognitive aspects to all of Rest's components, and Kohlberg's idea of a stage as a structured whole or a world view cuts across Rest's componential model. ..." (p. 53)

Measuring Cognitive-Structural Aspects

While the affective aspects of moral judgment behavior are conventionally assessed as the preferences for certain moral orientations, cognitive-structural aspect needs to be assessed in a new way (Broughton, 1978). Piaget (1965) used *clinical interviews*, in which subjects were not just observed but were confronted with a rather difficult moral task, as in an intervention-experiment. “Piaget [...] used judgments plus explanations (instead of judgments only) as criteria for operational competence, and considered counter-suggestions essential to the clinical method” (Lourenço & Machado, 1996, p. 146). He “assessed necessary knowledge by a variety of means – the child's justifications, her or his resistance to a variety of perturbing counter-suggestions or to cues of perceptual seductions are cases in point.” (p. 154)

Kohlberg (1958) followed Piaget's methodological steps, using the clinical method and difficult moral tasks like moral dilemmas and counter-suggestions in order to observe participants' moral judgment *competence*: “We felt that it would be easier to analyze qualitatively a case in which the situation demanded more than a child could respond to than to analyze a case in which a child wanted more challenge than the situation could provide” (p. 76). “On three questions (...) the interviewer disagreed with the child and gave an argument to influence the child to change his mind. This argument was designed to be as 'low level' as possible, and was based on a ten-year-old pre-test response. If the child maintained his previous response, a second 'high-level' argument was offered to the child.” (p. 78)

The clinical-experimental assessment method developed by Piaget and Kohlberg is based on the assumption that structural properties of a participant's responses are real and observable: “Most [developmental] changes are changes in qualitative (structural-organizational) aspects of responses. [. . .] A really new mode of response is one that is different in

its form or organization, not simply in the element or the information it contains." (Kohlberg, 1973, p. 498) "The responses of subjects to the dilemmas and their subsequent responses to clinical probing are taken to reflect, exhibit, or *manifest* the structure. [. . .] There can be no error in the sense of a mistake in inferring from a judgment to some state of affairs concurrent with, prior to, or subsequent to the [scorer's] judgment" (Kohlberg, 1984, p. 407). "We are concerned with actual moral judgment. [. . .] what is going to make a difference when that individual is faced with an actual moral choice [. . .] What we care about is how moral judgments are made when [a moral principle] is actually applied to values in conflict." (Colby et al., 1987, p. 58)

Manifest structures or pattern of behavior can only be observed if the assessment method is properly designed. "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 the 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; it will not emerge through pure factor-analytic responses classified by content." (Kohlberg, 1984, pp. 401-402)

Probably under the growing pressure of main-stream psychology (e.g., Kurtines & Greif, 1974; for an opposing view see Broughton, 1978), Kohlberg and his associates gave up most of their methodological innovations. In contrast to Kohlberg's earlier insights, they later de-emphasized the use of counter-arguments and counter-suggestions (Colby et al., 1987, p. 161 & 186) and considered structure as "unobservable or hypothetical. [...] The structures themselves can never be observed." (Kohlberg et al., 1984, p. 242) Therefore, they no longer assess the structural properties of a *pattern* of responses but try to infer these properties from

the content of the items: “My colleagues and I [...] have required each item in the manual to clearly reflect the structure of the stage to which it is keyed.” (Kohlberg, 1984, p. 403) “Each item must have face validity in representing the stage as defined by the theory.” (p. 410)

At the same time, the concept of moral competence became confused by an unfortunate distinction between (observable) ‘performance’ and (unobservable) ‘competence.’ This distinction was borrowed from Noam Chomsky, who used it to describe the relationship between grammatical rules (competence) and concrete speech (performance). Some Kohlbergians translated it to mean that moral behavior (= performance) follows from moral judgment (= competence) in a similar way as speech follows from grammar. However, this distinction rendered moral competence an unobservable entity. To make things worse, this distinction is alien to cognitive-structural theory, as Lourenço and Machado (1996) have pointed out: “Piaget realized that to oppose competence and performance is to create a false dichotomy” (p. 149). Habermas (1983), who initially sympathized with this distinction, finally conceded, that “competence can only be observed through real performances [...] otherwise these competencies could not be measured.” (p. 199; my transl.)

The *Moral Judgment Test*⁶

See note at the end of this article!

We have designed the Moral Judgment Test (MJT) to assess simultaneously both affective and cognitive aspects of moral judgment behavior, i.e., moral orientations and moral judgment competence. It has been designed to bridge the notorious gap between psychological

⁶ Note that the MJT has been constructed only for use in scientific research and evaluation studies but *not* for diagnosing or selecting individuals or group of individuals or for high-stakes testing. For usage guidelines please see: <http://www.uni-konstanz.de/ag-moral/> .

I wish to thank the many colleagues who have helped to construct, examine and revise the MJT, especially those who have been involved in an early phase of test validation: Tino Bargel, Rainer Döbdert, Thomas Krämer-Badoni, Gertrud Nuner-Winkler, Gerhard Portele (†), and Roland Wakenhut (†).

theory and methodology by picking up Kohlberg's original methodological ideas as I have outlined them above and at other places (Lind, 1978; 1982; 1989; 2008; Lind & Wakenhut, 2010). It is not a psychological test in the traditional sense but a N=1 clinical experiment. Because the MJT is constructed on the basis of a well-developed theory and a large body of research, clear and rigorous criteria exist for its validation. Because the construction of the MJT did not involve item selection in order to increase the correlation of its index with external validity criteria like age and stage sequence, it is an unbiased tool for theory testing.

As we have seen above, classical methods of test construction and test analysis ("validity", "reliability", "consistency" etc.) are far less neutral towards psychological theories than most psychological researchers seem to believe. They contain implicit psychological assumptions about the nature of human behavior and development which are at odds with modern psychological insights: "Test theory that dominates educational measurement today [is] the application of 20th century statistics to 19th century psychology. Sophisticated estimation procedures [...] applied within psychological models that explain problem-solving ability in terms of a single, continuous variable [...]. Educational measurement faces today a crisis today that would appear to threaten its very foundations." (Mislevy in Frederiksen et al., 1993, p. 19)

For example, conventional methods of test construction are based on the implicit assumption that "inconsistency" in participants' responses reflects nothing but measurement error, ignoring the fact that people's traits can be quite different in regard to the structure or organization of their behavior (see above; also Lind, 1982; 1989; in press). Yet, so-called response inconsistency or measurement error is an important source of information about a person's moral-cognitive organization or moral judgment competence. Hence, the MJT has been designed to use the consistency of rating of arguments in regard to their moral quality as

an index of moral judgment competence (the C-index).

While there is a long tradition of measuring *attitudes* in the moral domain and the methodology of measuring attitudes, values and the like is well-developed, defining and measuring *competencies* in the moral domain is new. Kohlberg (1964) was the first to explicitly define *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" (p. 425; emphasis added). Note that this definition refrains from imposing specific moral values on each and every individual but requires only that each individual – regardless of culture – pursues her or his moral values in a morally consistent manner.

If we wish to assess a particular ability, we must first define an adequate task. Therefore, in order to measure *moral judgment competence* it is necessary to define a *moral* task. When we constructed the MJT in the 1970ties, we considered several options (Lind, 1978; 2008; Lind & Wakenhut, 2010). Informed by the theory of communicative ethics (Habermas, 1983; Apel, 1990), by Piaget's use of "counter-suggestions" in his clinical interviews (see also Inhelder et al., 1974) and by Keasey's (1974) research on adolescents' ability to deal with counter-arguments, we envisioned a communication situation in which the participants is confronted with arguments supporting and opposing his or her stance on a moral dilemma. We believed that dealing with arguments and especially with counter-argument would be a challenging task for more respondents and that moral feelings would be aroused with which he or she had to cope with. As research has shown, sometimes these emotions can get so strong that cognitive processes like moral reasoning and judgment are severely hampered (e.g., Keasey, 1974; Damasio, 1994).

How does the Moral Judgment Test work? At the beginning of each dilemma-experiment, the subject is requested to read a short moral dilemma story, and to judge the

decision of the protagonist: “Was he/she doing right or wrong?” This task is to prime moral feelings in the respondent. It sets the stage for the actual experiment, in which the participant has to rate six arguments in favor and six against the decision of the protagonist, and, therefore, also for and against his or her own evaluative judgment. Each argument represents one of Kohlberg’s six types (stages) of moral orientation (Kohlberg, 1984).⁷ This makes it possible to measure to which extent participants attend to the moral quality of the arguments. Since each argument is related to either of two dilemmas, to either of two sets of arguments (pro and contra), and to one of six type of moral orientations, the whole set of arguments in the MJT form a fully crossed 2 x 2 x 6 experimental design for each participant.

Thus, the *pattern* of responses to the 24 arguments of the MJT shows to which degree a participant takes the moral quality of the arguments into account when judging them. To obtain a quantitative measure for this degree of moral considerations, we calculate the proportion of response variance accounted for by moral concerns and multiply this number by 100 so that the resulting “C-score” can range from 0 to 100.⁸ The C-score reflects the degree to which an individual participant rates the arguments of the MJT in regard to their moral quality rather than in regard to their opinion agreement or other aspects of the situation like dilemma-context. A C-score of zero means that the participant has not attended at all to the moral quality of the arguments; a C-score of one hundred means that the he or she has rated the arguments solely for their moral quality.

⁷ Note that we use Kohlberg’s six categories only for describing the affective aspect of moral judgment (i.e., moral orientations) but not for describing the cognitive side of development. Temporarily, Kohlberg reduced his six stage-types to five, but returned to the six again later (Kohlberg et al., 1990).

⁸ Note that the psychological meaning of the C-score is based not only on the numerical calculation of variance components (consistency) but also on the task involved in the MJT. A test which does not incorporate a moral task does not allow us to measure moral *competence* even if one calculates an index similar to the C-score.

How does the MJT, and especially its C-score, reflect moral development? An answer to this question emerges when we assemble various experimental and anecdotal evidence from MJT research (Lind, 1985; 2002; in press):

- When responding to the MJT as part of a battery of tests and questions, participants get noticeably excited. While they are usually very calm when filling in survey questionnaires, here they show various signs of excitement: straightening their body, touching their head, mumbling, making faces etc. Obviously, the MJT triggers some (moral?) affects in the participants.
- In a very low phase of their development, participants take a stance on the dilemma but not, as required, respond to the subsequent arguments. Typical answers go like this: "Why should I go through all this? I did already say what I think about the issue!" It seems that at this stage, respondents do not understand the function of arguments or reasons. The seemingly simple task of judging arguments cannot be coped with.
- On the next level, respondents respond to the given arguments but only to the *supporting* arguments (pro-arguments) and *not* to the counter-arguments. "I skipped these arguments because I disagree," said one respondent. Maybe because of such responses, Kohlberg and his colleagues largely dropped counter-suggestions from their interviewing schedule (Colby et al., 1987; Lind, 1989), and Rest decided against them when constructing his *Defining-Issues-Test*.⁹

⁹ "The artificiality of the [con] statement interfered with its usefulness in studying modes of reasoning. For the most part, information from these statements was useless and had to be eliminated from the analysis." (Rest, 1979, p. 89)

- However, at the age of ten, the lowest age for which the MJT is suited,¹⁰ most participants can deal with all arguments, even with counter-arguments, yet in a peculiar way. At this age hardly anyone attends to the *moral quality* of the arguments but only to their *opinion-agreement* or disagreement (Keasey, 1974). All *supporting* arguments are accepted without any sign of doubt (all getting "+4"-ratings in the MJT), and all *opposing* arguments are definitely rejected (all getting "-4" ratings in the MJT). For such a pattern of responses, the C-score is zero. A C-score of zero indicates that a respondent's moral orientations may exist but are too weak to overrule other factors in this situation like opinion-agreement.
- First signs of a more sophisticated moral judgment show up when participants begin to discriminate between the moral qualities of an argument, often starting with the *pro*-arguments. They discover that not all supporting arguments are morally good. Some supporting argument may be inadequate or even bad, just as not all "friends" are of good character and should be considered a true friend.
- At some later point of development children discover that not all *counter*-arguments are bad, but that some may even appeal to moral ideals which they hold themselves. At this point the C-score really starts to climb up on the C-scale, which ranges from 0 to 100.
- However, moral judgment and discourse competence does not develop evenly in all life domains. It seems that people may show a highly developed moral judgment competence in one domain and a low one in another. We came across this phenomenon of 'moral segmentation' in MJT studies in Latin-American countries like

¹⁰ For children of this age and for participants with low educational training, some special adaptations need to be made, like, e.g., larger font, more spacing between the lines, shorter response scales (from -2 to +2, instead of -4 to +4), and some technical assistance needs to be given like explanations of unfamiliar terminology.

Brazil, Colombia, and Mexico. These college students got much lower C-scores than their European counterparts (Lind, 2000a; Bataglia et al., 2002; Bataglia et al., 2003; Moreno, 2005). Because these studies mostly involved college students with a very affluent background, some hypothesized that they lacked motivation to respond to the so-called workers dilemma. However, a separate analysis of the two dilemmas revealed that not the moral judgment competence shown in this dilemma was suppressed but in the mercy-killing dilemma. Thus the phenomenon of moral segmentation may be related to the religious culture in these countries. This phenomenon resembles findings by Roland Wakenhut (1982) and his colleagues in their studies of German soldiers. Soldiers showed higher C-scores in civilian dilemmas than when the dilemma was placed in a military context. These and other findings on moral segmentation make me believe that these differences of level of moral development in different life domains do not, as Piaget maintained, reflect an intra-psychic process of developmental *decalage* but that it reflects the influence of powerful cultural agencies like the church, the military and other social institutions.

- A high C-score (of about 40 and higher) is reached if the participant, when evaluating an argument, focuses predominantly on its moral quality but hardly on the arguments' opinion-agreement or on cultural dogmas. On this level of moral-democratic maturity, the participants are able to enter a moral discourse about controversial issues and maintain such a discourse even with opponents.

Four Research-Based Criteria for MJT's Validity

Does the MJT measure what it is supposed to measure? We have checked on this question of validity on the basis of four theory-based criteria concerning both aspects of moral behaviour: the hierarchical preference order of moral orientations (Kohlberg, 1958; 1976; Rest, 1973; Lind, 1978), the simplex structure of the inter-correlations of moral orientations (Kohlberg, 1958), the competence nature of moral behavior (Kohlberg, 1964; Lind, 2002), and the parallelism of moral orientations and moral competencies (Piaget, 1965; 1976; 1981):

Criterion # 1: Hierarchical Preference Order of Moral Orientations

In the MJT, *moral orientation* is defined as the subject's mean *acceptability ratings* of all arguments that represent a particular stage. Because the standard MJT has two dilemmas, and in each dilemma there are two arguments – one *in favor* and one *against* the respondent's decision on the dilemma – for each of the six Kohlbergian moral orientation is represented by four items. The respondent can choose a number from “-4” (“I completely *reject* it”) to “+4” (“I completely *accept* it”).

Kohlberg (1958; 1984) and Rest (Rest, 1973) have compiled ample evidence suggesting that the highest moral orientations are preferred by most people regardless of social background, education, age and gender. The psychiatrist Max Levy-Suhl (1912) found that even juvenile delinquents valued universal moral principles more highly than conventional or pre-conventional reasons. In many MJT studies, this finding is clearly corroborated (see Wischka, 1982; Scheurer, 1993; Lind, 2002). University students as well as

delinquents of the same age revealed the identical hierarchical order of moral preferences or attitudes (Figure 1).

The preference for post-conventional moral orientations seems to be a universal phenomenon (Gielen et al., 1993). Indeed, MJT studies in various cultures show that the Hierarchical Preference Order of the six Kohlbergian moral orientations is ubiquitous: everywhere, type 6 orientations preferred the most and type 1 preferred the least (Lind, 1986; 2002; 2005; Schillinger, 2006). As an example the findings from our five-country study of 1st semester university students is depicted in Figure 2.

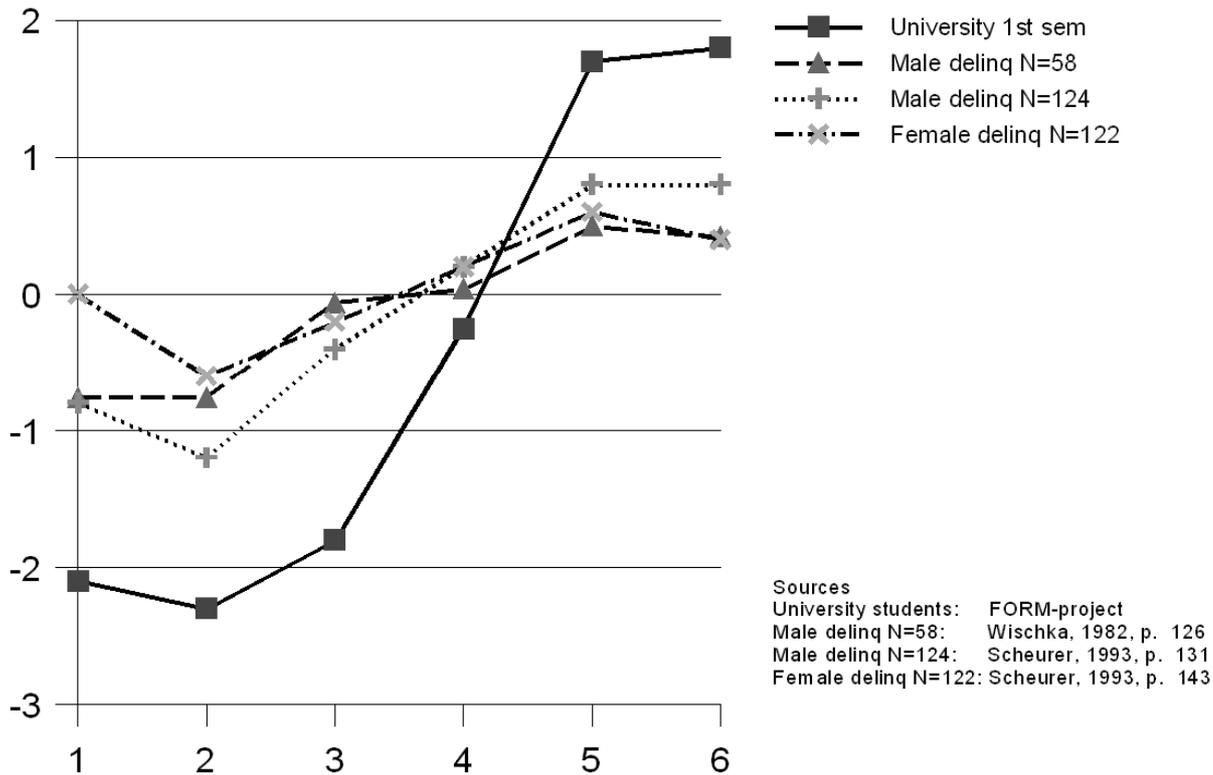


Figure 1. The preferences for the six Kohlbergian orientations (stages) are hierarchically

ordered. This is true for non-delinquent and for delinquent young adults. Source: Lind (2002).

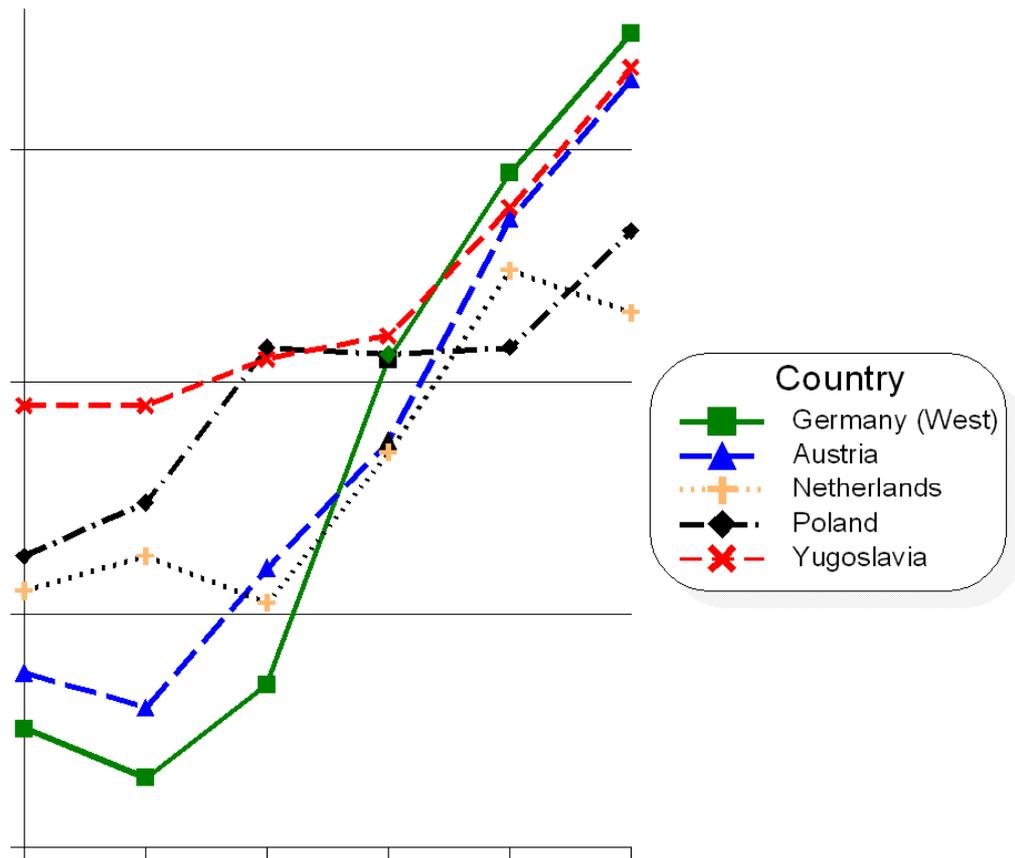


Figure 2. The hierarchical preference order of the six Kohlbergian moral orientations by 1st semester university students in five European countries in the late 1970s, while Poland and the former Yugoslavia were under communist rule. Source: FORM-project (see Lind, 2002).

Criterion # 2: Quasi-Simplex Structure of the Inter-correlations of Moral Orientations

In his dissertation study, Kohlberg (1958) hypothesized that his six moral orientations (or stages) should be inter-correlated in such a way that they form a quasi-simplex:¹¹ “The relevant rationale seemed to be suggested by the thinking of L. Guttman. This thinking specifies some necessary, though not sufficient conditions for inferring a developmental sequence. If certain tests or items or dimensions stand in a developmental sequence, with regard to one another, then a certain pattern of associations should hold between them, . . . Such a pattern in quantitative data is called a 'simplex' by Guttman.” (pp. 82-83)

Kohlberg employed a graphical inspection method to evaluate the goodness of fit of his data. Like a medical doctor who looks at an X-ray picture for signs of some disease, he judged the goodness of fit of his data by visual inspection of the correlation matrices obtained in his study.

We have used two statistical methods to assist this visual inspection. Formerly we used a method for reordering the correlations in order to maximize the simplex structure (Nagl et al., 1986). If the application of this method suggests an ordering of the six stages different from the one Kohlberg postulated we would regard this as a violation of the validity criterion. Yet this method is still rather crude and insensitive.

Today we use principle component analysis (which is similar to factor analysis, except that the diagonal of the correlation matrix contains ones rather than estimates of reliability) with *simple varimax* rotation. The criterion predicts that the analysis should produce *two*

¹¹ Often, in literature the terms simplex and quasi-simplex are used in an interchangeable way. The term quasi-simplex is used if sizable error measurements are allowed. A perfect simplex is reasonable only if measurement errors are negligible. I prefer the weaker assumption, i.e., quasi-simplex, though one could also argue otherwise. I wish to thank Debbie D. Reese for this clarification.

factors and that the *factor loadings* of each orientation should lie on a circle, being perfectly ordered from orientation type one to six. Because research findings suggest that the rank order between stages 1 and 2 as well as between stages 5 and 6 are not as clear as between the other stages, we allow for small deviations from this prediction. Note that even with this tolerance for deviations, this prediction is very risky as *its a priori* probability is very low.

Theoretically, there are 6! (that is, $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$) ways in which the stages can be ordered. Hence, the probability that they are ordered in the predicted way only by chance is $p = 1/720 = 0,0013$ for one individual!

Ideally, the factor loadings should look like that in Figure 3, if the data confirm to the hypothesis of a quasi-simplex. I submitted Kohlberg's correlation matrix to principle component analysis, getting the graph depicted in Figure 4. Comparing Kohlberg's data with an ideal simplex-structure from fictitious data shows that they fit well though not perfectly. Note that Kohlberg's (1958) used *relative* frequency of stage usage in interviews as an indicator of moral orientations (we call them *ipsative* because they always add up to 100 percent), implying that most indices must correlate negatively with one another (as some percentages rise, others must go down by definition).

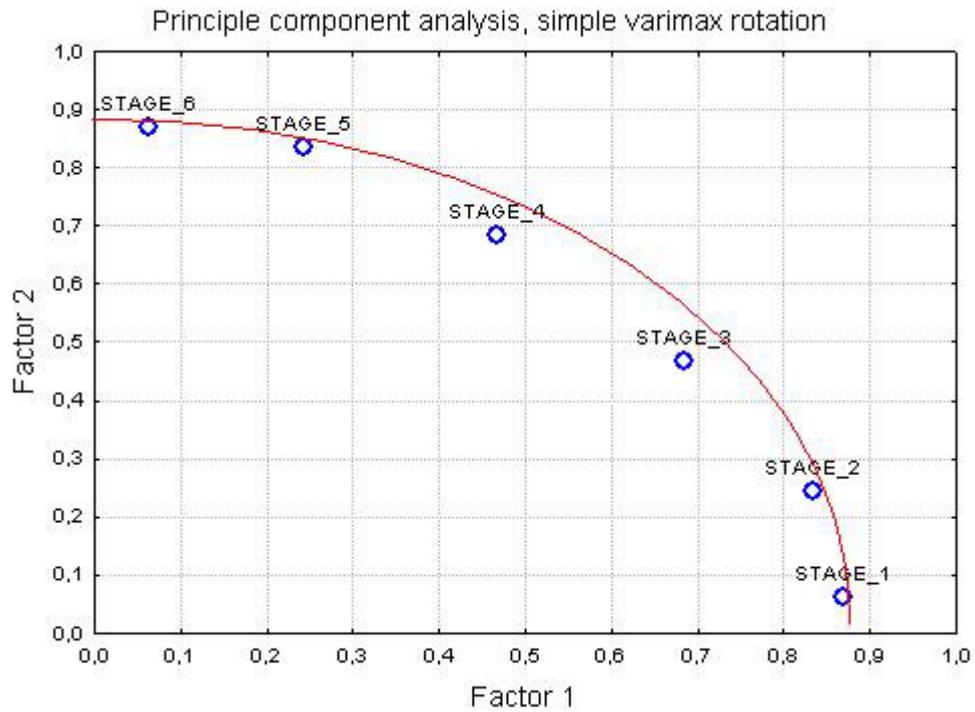


Figure 3. Ideally, the correlations between the six Kohlbergian orientations (stages) form a quasi-simplex structure.

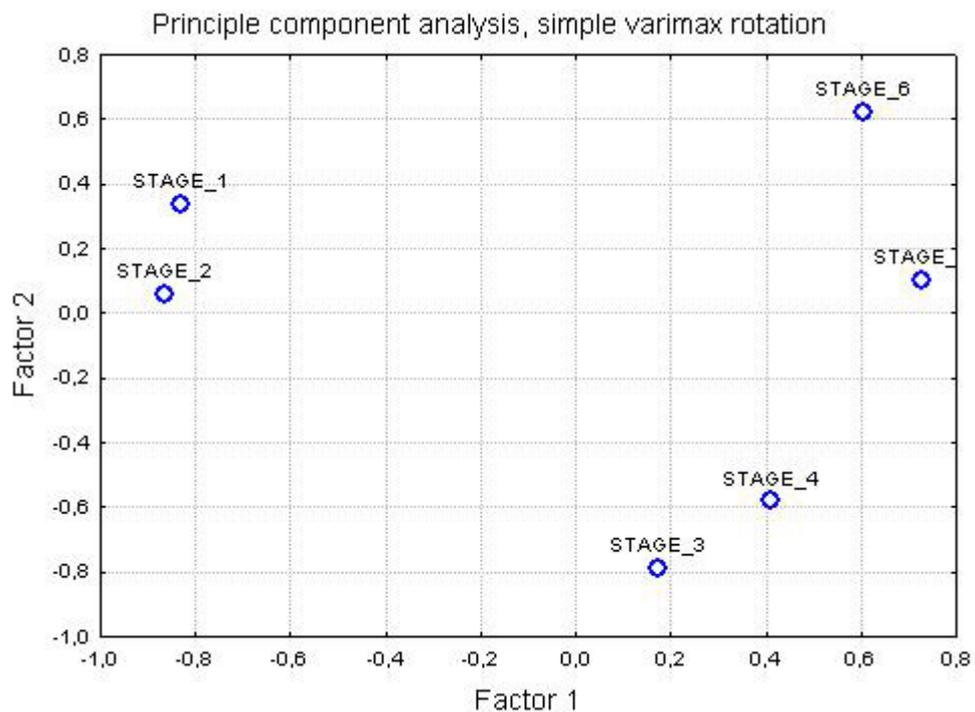


Figure 4. The correlations between the frequencies of “stage-use” in Kohlberg’s Moral Judgment Interview resemble a quasi-simplex structure. Source: My principled component analysis of the correlations reported by Kohlberg (1958, p. 84). The simple varimax rotation resulted in an unusual pattern of factor loadings because Kohlberg’s index of stage usage is an ipsative measure, forcing many correlations to be negative.

Studies using the MJT show an even better fit to the simplex criterion than MJI-studies.

Figure 5 shows the findings from German first semester university students in the 1970ties.

This finding has been replicated in each and every study which uses a validated version the MJT. All studies show also a good or very good fit of response pattern with the simplex structure predicted by Kohlberg (1958).

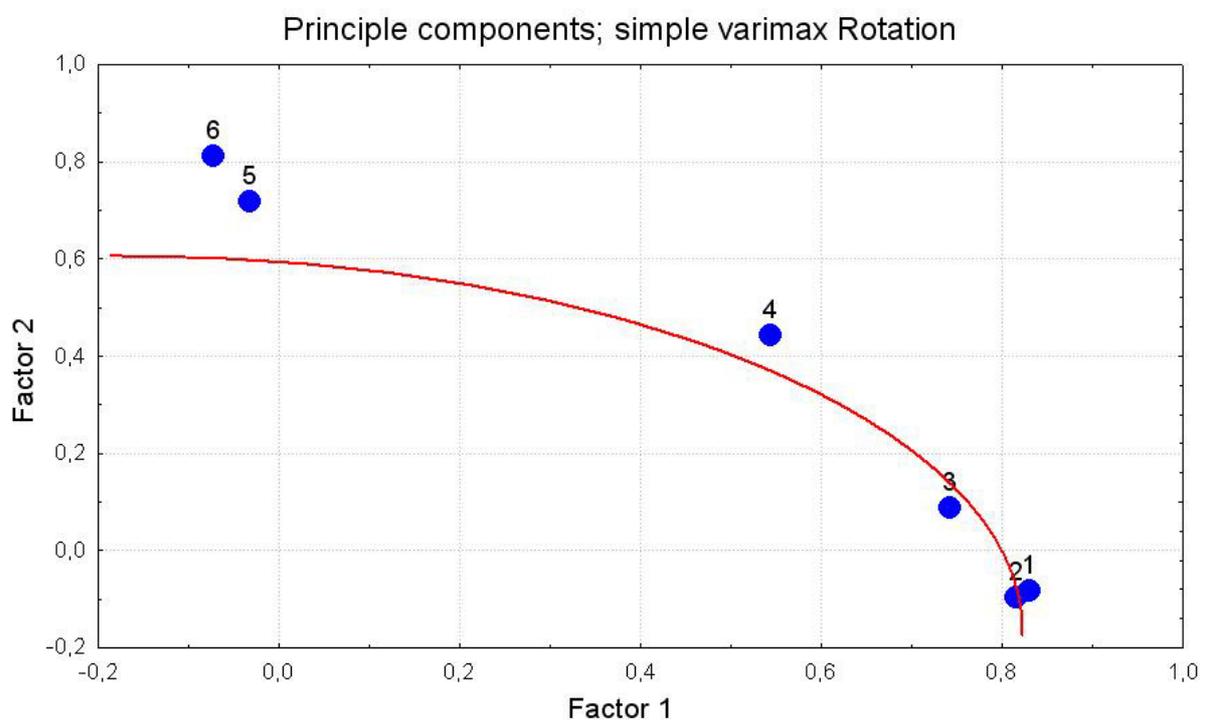


Figure 5. The factor loadings of the preferences for the six orientations form a quasi-

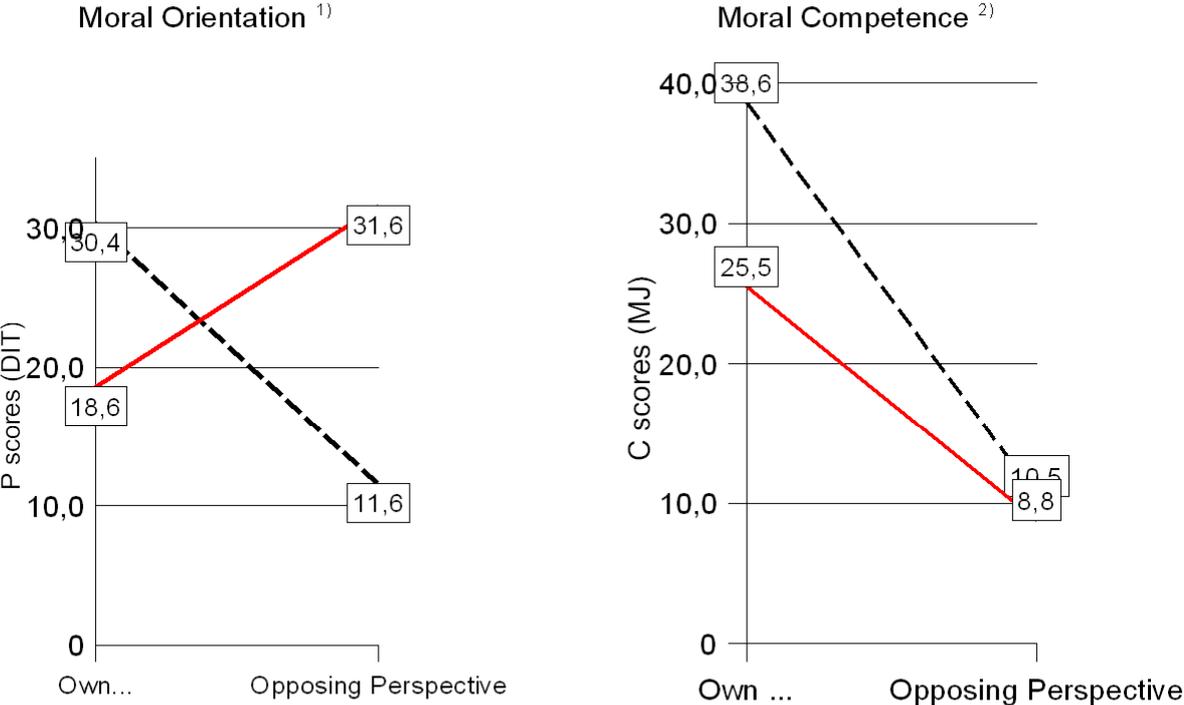
simplex. Data source: FORM-project, German 1st semester university students, N = 746 (for more details see Lind, 2002).

Criterion # 3: The Competence Nature of Moral Behavior

Moral judgment *Competence* is a genuine ability, and not merely an *orientation, ideology* or *attitude*. While Kohlberg (1964) acknowledged the affective nature of moral judgment, i.e., moral orientations and preferences, he maintained that moral behavior is also determined by the “capacity to make decisions and judgments which are moral (i.e., based on internal principles) and to act in accordance with such judgments.” (p. 425) Because the C-score has been constructed to index this competence aspect, it should not be possible to fake it upwards.

MJT research corroborates this most important postulate of modern moral psychology. There is a strong competence aspect involved in moral behavior. In their experimental study, Emler et al. (1983) showed that participants could successfully be instructed to fake their preferences for postconventional moral reasoning (low scoring participants were able to push their P-score of the Defining Issues Test upward after being instructed to simulate the responses of high-scoring subjects (Figure 6, graph on the left). However, in a replication of this experiment with the MJT (see Lind, 2002), participants were not able *to* simulate the C-index of the MJT upwards but only downwards (Figure 6, graph on the right). In Figure 6 the *solid* line shows the data of the critical group, namely the initially ‘low-scoring’ group, which scored low on the first administration of the DIT and the MJT. Before taking their test a second time, the participants were instructed to simulate the responses of the other, high-scoring group. In both experiments the low-scoring group were students who indicated that

they were politically conservative or rightist, and the high-scoring group were students who described themselves as being progressive or leftist. Wasel (1994) supported the competence hypothesis in a modified experiment with the MJT. In addition, he found that subjects with high C-scores were better able to diagnose other people's moral judgment competence than subjects with low C-scores.



Source
 1) Emler, N., Renwick, S. & Malone, B. (1983).
 2) Lind, G. (2002).

Figure 6. Emler et al. (1983) was able to instruct participants with low moral development index to simulate the moral preferences of participants with a high index (P-score of the DIT, left graph). In the second experiment by Lind (2002), the participants with an initially low moral judgment competence were *not* able to fake their

moral judgment competence upwards (C-score of the MJT, right graph).

The role of moral judgment competence in moral decision-making has recently been supported also by brain research. Prehn et al. (2008) have shown through a fMRI study that moral judgment competence (as measured with the MJT) is highly correlated ($r = 0.45$) with brain activities in the dorsolateral prefrontal cortex. The lower the participant's moral judgment competence (as measured with the MJT), and the longer he or she needs to solve a moral conflict and the longer the neurons in the right frontal part of the brain are active. In this part of the brains moral feelings and moral consciousness are coordinated (Damasio, 1994; Robertson et al., 2007).

Criterion # 4: Affective-Cognitive Parallelism

Piaget (1951; 1976; 1981; Piaget & Inhelder, 1969) holds that affective and cognitive aspects of behavior are “parallel.” This assumption of parallelism concerns the development of both aspects and their empirical correlation. For a long time this hypothesis could not be tested empirically, because there no method of measurement was available which let us assess affect and cognition simultaneously as ‘distinct’ and yet ‘inseparable’ aspects of behavior (Piaget, 1976, p. 71). The Moral Judgment Test, it seems, is the first instrument which allows this.

MJT studies support very well Piaget's hypothesis of correlational parallelism. In Figure 7, the findings from a study of German 1st semester university students are depicted as

an example (for more examples, see also Lind, in press). The C-score correlates systematically with the six Kohlbergian stages of moral orientation: it correlates highly negative with the preference for low stage reasoning, and highly positive with the preference for high stage reasoning, with the other correlations in between. In other words, the higher the moral judgment competence of people, the more clearly they reject low stage moral reasoning as inadequate, and the more clearly they prefer stages 5 and 6 orientations as an adequate basis for a moral discourse about a moral dilemma. However, note that this is true only for observations in 'regular' situations. Affective-cognitive parallelism seems to become unobservable if something is at stake for the subject. In such situations, the subjects may simulate socially desirable moral orientations 'upward' or may, because of time pressure or social control, exhibit a lower level of moral judgment competence than they are actually capable of.

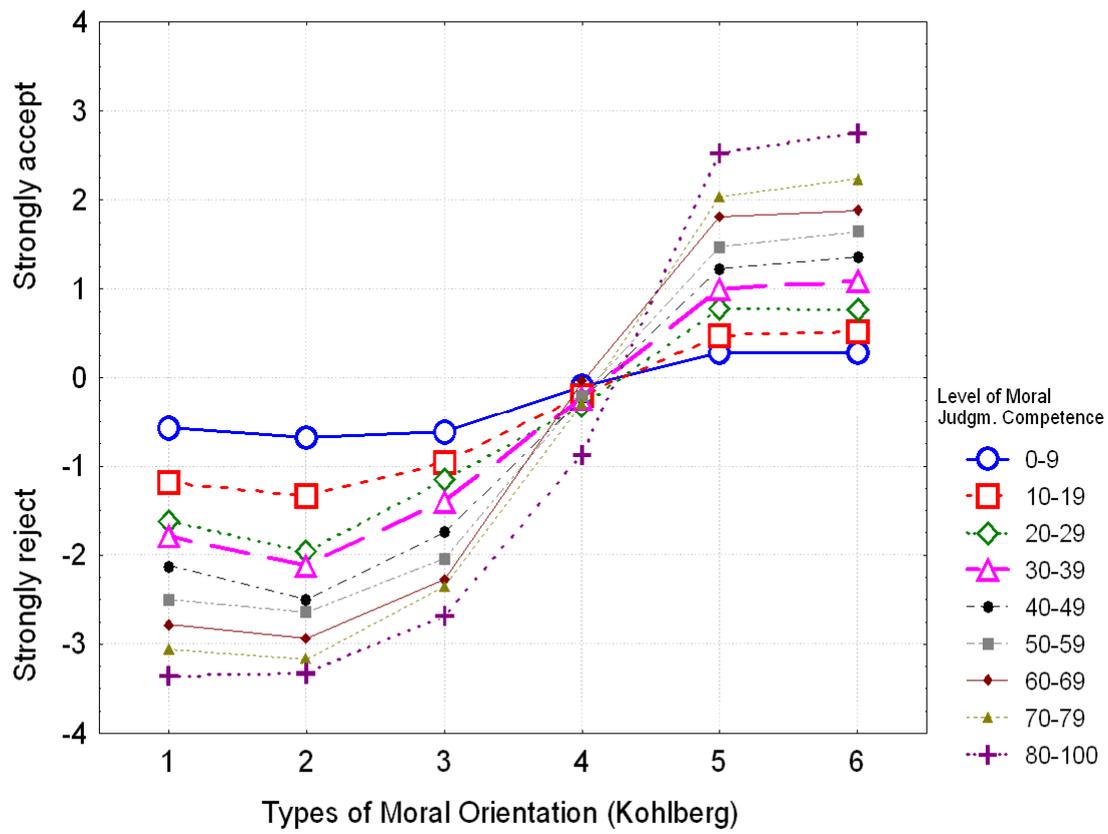


Figure 7. Affective-cognitive parallelism: The higher the moral judgment competence, the more participants prefer higher moral orientations and reject lower orientations.

Source: FORM-Project, University of Konstanz (see Lind, 2002). Data of 1st semester university students, N=746, 1977, Germany.

Conclusion

Good measurement and progress in scientific understanding depend on each other. Our understanding of the process and conditions of moral development depends as much on good

measurement, as good measurement depends on our knowledge of this object (see, e.g., Messick, 1995). Obviously, we first need to know what we intend to assess before we can say how valid our measurement is and what the scores mean. Therefore, progress in the two fields is not linear but mutually dependent. It is, as Kohlberg (1981) called it, a “bootstrapping” process in which we base the construction of new measurement methods on the facts which we already know, in order to explore further yet unknown grounds.

The validation procedure chosen for the MJT was theory-based. This contrasts with other more conventional methods of test validation. These tests are made by selecting and revising items in order to maximize the correlation of test scores with certain validity criteria like stage invariance, age, education or political attitudes. In contrast, the MJT has been constructed without item selection in order to avoid a tautology. If a test is constructed to maximize its correlation with age, it should not be surprising that its scores highly correlated with age. Such a validation method produces a tautology.¹²

The main purpose of the MJT is to assess what Kohlberg (1973) called the structural-organizational aspects of responses: “A really new mode of response,” he states, “is one that is different in its form or organization, not simply in the element or the information it contains.” (Kohlberg, 1973, p. 498)

We designed the MJT on the basis of Kohlberg’s notion that “the responses of subjects to the dilemmas and their subsequent responses to clinical probing are taken to reflect,

¹² Even Kohlberg (1976) was not immune against such tautology: “Our conception of construct validity implies assignment of individuals to stages in such a way that the criterion of sequential movement is met” (p. 47). Colby et al. (1987) also state “the appropriate question is whether the interview and scoring system provides a valid assessment of moral judgment stage [...] the Standard Issue Scoring [...] yields scores that agree very closely with the theoretical predictions of invariant sequence and internal consistency” (p. 71). Similarly, Rest (1979) regards “the demonstration of age trends as crucial” (p. 143) for the validity of his Defining Issues Test (DIT). Hence, both tests are biased in favor of the assumption that moral development is correlated with age.

exhibit, or *manifest* the structure" (Kohlberg, 1984, p. 407). Like he we were "concerned with actual moral judgment. [. . .] what is going to make a difference when that individual is faced with an actual moral choice. [. . .] What we care about is how moral judgments are made when [a moral principle] is actually applied to values in conflict." (Colby et al., 1987, p. 58)

So the findings produced by MJT-studies tell us also something about the empirical validity of the dual-aspect-theory of moral behavior. They clearly support its core assumptions:

- Morality has a strong competence aspect, which could be demonstrated by a series of experiments, including a fMRI imaging study of the brain.
- Regardless of cultural background (and regardless of age, gender, and education), people prefer moral orientations (as defined by Kohlberg) in the predicted stage order;
- Moral orientations exhibit a developmental order as neighboring orientations are stronger correlated with one another than more distant orientations.
- The affective and the cognitive aspect of moral behavior are parallel, that is, the higher people's moral judgment competence, the more they prefer principled moral reasoning over lower stage reasoning (if the test situation does not involve high stakes).
- Aside from this, MJT research revealed a new phenomenon, namely the phenomenon of "moral segmentation" (see Wakenhut, 1982; Lind, 2000a; Schillinger-Agati & Lind, 2003). This phenomenon needs more attention in further studies.

However, the data produced by MJT research let us refute Kohlberg's assumption of invariant sequence of moral development. There is much evidence that moral judgment competence can regress. For example, regression has been found in graduates from German middle school

who enter a vocational training program or the workforce directly (Lind, 2002), and in medical school students, whose moral judgment competence erodes during study (Lind, 2000b). This moral regression in medical students has also been found in a longitudinal study by Helkama et al. (2003), using Kohlberg's *Moral Judgment Interview* method.

Three of the theoretical criteria (preference hierarchy, quasi-simplex-structure, and affective-cognitive parallelism) for validating the German master version of the MJT, have been also used for test validation of 30 foreign language versions of the MJT (Lind, 1986; 1995; in press). They helped to test the cross-cultural validity of the MJT. They make sure that the different language versions are semantically and pragmatically equivalent. Moreover, because the C-score does not require the participants to rank the six moral orientations in the same way (though actually they mostly do), this index is not biased by a certain moral value system. This made it possible to study moral judgment competence in many different cultures (Lind, 2005).

The MJT research has important practical implications, especially for education. On the one hand, these findings clearly demonstrate that the preference for post-conventional moral reasoning seems to be a universal phenomenon. Hence there seems to be no need to "teach" or "instill" moral values, but rather a need to teach moral competencies, that is, the ability to apply one's own moral values to specific decision-making (Lind, 2007). On the other hand, these findings also show that favorable educational environments which offer at least some opportunities for responsibility-taking do foster moral development (Schillinger, 2006). The MJT has also been instrumental for developing the *Konstanz Method of Dilemma Discussion* (KMDD), which has been developed on the basis of Blatt and Kohlberg's (1975) method of dilemma discussion. Continuous evaluation studies using the MJT have helped to make the KMDD a very effective method of teaching (Lind, 2007).

References

- Apel, K.-O. (1990). *Diskurs und Verantwortung.[Discourse and responsibility.] Das Problem des Übergangs zur postkonventionellen Moral* (Original 1988). Frankfurt: Suhrkamp.
- Bataglia, P., Agati, M.M., Siva Torres, S., Zanoni Crivelaro, D.B., Oliviera, D.D., & Lencione Quevedo, T. (2002). *The development of moral competence and religious commitment in Brazil*. Paper presented at the meeting of the Association for Moral Education, Chicago, Nov. 2002.
- 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.
- Blatt, M. & Kohlberg, L. (1975). The effect of classroom moral discussion upon children's level of moral judgment. *Journal of Moral Education*, 4, 129-161.
- Bloom, B.S., Engelhart, M.D., Hill, W.H., Furst, E.J. & Krathwohl (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York: McKay.
- Broughton, J.M. (1978). The cognitive-developmental approach to morality: A reply to Kurtines and Greif. *Journal of Moral Education*, 7, 81-86.
- 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 foundations and research validation*. New York: Columbia University Press.
- Damasio, A. (1994). *Descartes' error: emotion, reason and the human brain*. New York: G. P. Putnam's Son.
- Darwin, C. (1966). *The descent of man, and selection in relation to sex. The Works of Charles*

- Darwin*, ed. Paul H. Barrett & R. B. Freeman. London: Pickering (Originally published 1874).
- Emler, N. (1996). How can we decide whether moral education works? *Journal of Moral Education*, 25, 117-126.
- Emler, N., Renwick, S. & Malone, B. (1983). The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology*, 45, 1073-80.
- Frederiksen, N., Mislevy, R. J., & Bejar, I. I., eds. (1993). *Test theory for a new generation of tests*. Hillsdale, NJ: L. Erlbaum.
- Gielen, U.P. & Markoulis, D. (1993). Preferences for principled moral reasoning: A developmental and cross-cultural perspective. In: L.L. Adler & U.P. Gielen, eds., *Cross-cultural topics in psychology*, pp. 73-87. Westport, Connecticut: Praeger.
- Habermas, J. (1983). *Moralbewußtsein und kommunikatives Handeln*. [Moral consciousness and communicative action.] Frankfurt: Suhrkamp.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 4, 814-834.
- Hartshorne, H. & May, M.A. (1928). *Studies in the nature of character*. Vol. I: Studies in deceit, Book one and two. New York: Macmillan.
- Helkama, K., Uutela, A., Pohjanheimo, E., Salminen, S., Koponen, A., & Rantanen, Vantsi, L. (2003). Moral reasoning and values in medical school: a longitudinal study in Finland. *Scandinavian Journal of Educational Research*, 47, 4, 399-411.
- Higgins, A. (1995). Educating for justice and community: Lawrence Kohlberg's vision of moral education. In: W.M. Kurtines & J.L. Gewirtz, eds., *Moral development: An introduction*, pp. 49-81. Boston: Allyn & Bacon.
- Inhelder, B., Sinclair, H. & Bovet, M. (1974). *Learning and the development of cognition*.

Cambridge, MA: Harvard University Press.

- Keasey, C.B. (1974). The influence of opinion-agreement and qualitative supportive reasoning in the evaluation of moral judgments. *Journal of Personality and 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, pp. 381-431. New York: Russel Sage Foundation.
- Kohlberg, L. (1973). Stages and aging in moral development -- Some speculations. *The Gerontologist*, 13, 497-502.
- Kohlberg, L. (1976). Moral stages and moralization: The cognitive- developmental approach. T. Lickona, Ed., *Moral development and behavior: Theory, research and social issues*, pp. 31-53. New York: Holt, Rinehart & Winston.
- Kohlberg, L. (1981). *The meaning and measurement of moral development*. The Heinz Werner Lecture Series, vol. 13. Worcester, MA: Clark University Press.
- Kohlberg, L. (1984). *Essays on moral development, Vol. II, The psychology of moral development*. San Francisco, CA: 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.
- Krathwohl, D.R., Bloom, B.S. & Masia, B.B. (1964). *Taxonomy of educational objectives. Handbook II: Affective domain*. New York: McKay.
- Kurtines, W.M. & Greif, E.B. (1974). The development of moral thought: Review and evalu-

- ation of Kohlberg's approach. *Psychological Bulletin*, 81, 453-470,
- Lakatos, I. (1972). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave, eds., *Criticism and the growth of knowledge*, pp. 91-196. Cambridge: University Press.
- Levy-Suhl, M. (1912). Die Prüfung der sittlichen Reife jugendlicher Angeklagter und die Reformvorschläge zum § 56 des deutschen Strafgesetzbuches. *Zeitschrift für Psychotherapie*, 232-254.
- Lind, G. (1978). Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts. 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 psychology of personality*, pp. 132-144. Amsterdam, NL: North-Holland.
- Lind, G. (1985). The theory of moral-cognitive judgment: A socio-psychological assessment. In: G. Lind, H.A. Hartmann & R. Wakenhut, eds., *Moral development and the social environment. Studies in the philosophy and psychology of moral judgment and education*, pp. 21-53. Chicago: Precedent 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. (1989). Measuring moral judgment: A review of 'The measurement of moral judgment' by Anne Colby, Lawrence Kohlberg et al. *Human Development*, 32, 388-397.
- Lind, G. (1995). *Cultural fairness and the measurement of morality*. Paper presented at the Conference of the Association for Moral Education (AME), New York, November 16-18, 1995.

- Lind, G. (2000 a). *Off limits. A cross-cultural study on possible causes of segmentation of moral judgment competence*. Presentations at the AERA Conference. New Orleans.
- Lind, G. (2000 b). Moral regression in medical students and their learning environment. *Revista Brasileira de Educacao Médica*, 24(3), 24-33.
- Lind, G. (2002). *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung* (Can morality be taught? Research findings from modern moral psychology). Berlin: Logos-Verlag.
- Lind, G. (2005). *The cross-cultural validity of the Moral Judgment Test: Findings from 30 cross-cultural studies*. Paper prepared for the American Psychological Association APA, Aug. 18-22, 2005.
- Lind, G. (2007). *La moral puede enseñarse. Manual teórico-práctico de la formación moral y democrática*. México: trillas.
- Lind, G. (2008). The meaning and measurement of moral judgment competence revisited - A dual- aspect model. A revised version of this paper is to appear in: D. Fasko & W. Willis, eds., *Contemporary philosophical and psychological perspectives on moral development and education*. pp. 185 - 220. Cresskill. NJ: Hampton Press.
- Lind, G. & Wakenhut, R. (2010). *Moral judgment and social education. Studies in the psychology and philosophy of moral judgment and education*. Edison, NJ: Transaction Publisher.
- Lourenço, O. & Machado, A. (1996). In defense of Piaget's theory: a reply to 10 common criticisms. *Psychological Review*, 103, 143-164.
- Markoulis, D. (1989). Political involvement and socio-moral reasoning: Testing Emler's interpretation. *British Journal of Social Psychology*, 28, 203-212.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from per-

- sons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Moreno, C. (2005). *Moral education in higher education and the transformation of a concern: a historical account*. Paper presented at the 31st annual meeting of the Association for Moral Education, Cambridge, Mass. November 3-5, 2005.
- Nagl, W. & Walter, H.G., eds. (1986). *Konstanzer Statistisches Analyse System KOSTAS. Bericht 37*. Zentrum I Bildungsforschung. University of Konstanz, Germany.
- Narvaez, D. (1998). The influence of moral schemas on the reconstruction of moral narratives in eighth graders and college students. *Journal of Educational Psychology*, 90, 13-24.
- Penn, W. (1990). Teaching ethics – A direct approach. *Journal of Moral Education*, 19, 124-138.
- Piaget, J. (1951). *Play, dreams and imitation in childhood*. New York: Norton.
- Piaget, J. (1965). *The moral judgment of the child* (Original 1932). New York: The Free Press.
- Piaget, J. (1976). The affective unconscious and the cognitive unconscious. In: B. Inhelder & H.H. Chipman, Eds., *Piaget and his school*, pp. 63-71. New York: Springer.
- Piaget, J. (1981). *Intelligence and affectivity: Their relation during child development*. Palo Alto, CA: Annual Reviews. (Originally published 1954).
- Piaget, J. & Inhelder, B. (1969). *The psychology of the child*. Basic Books (Originally published 1966).
- Pittel, S.M. & Mendelsohn, G.A. (1966). Measurement of moral values: a review and critique. *Psychological Bulletin*, 66, 22-35.
- Popper, K.R. (1968a). *Conjectures and Refutations*. New York: Harper & Row.
- Popper, K.R. (1968b). *The Logic of Scientific Discovery*. London: Hutchinson (Originally

- published 1934).
- Prehn, K., Wartenburger, I., Mériaux, K., Scheibe, C., Goodenough, O.R., Villringer, A., van der Meer, E., & Heekeren, H.R. (2008). Influence of individual differences in moral judgment competence on neural correlates of socio-normative judgments. *Social Cognitive and Affective Neuroscience*, 3(1), 33-46.
- Rest, J.R. (1973). The hierarchical pattern of moral judgment: A study of patterns of comprehension and preference of moral stages. *Journal of Personality*, 41, 86-109.
- Rest, J.R. (1979). *Development in judging moral issues*. Minneapolis, MI: University of Minnesota Press.
- Rest, J.R. (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.
- Robertson, D., Snarey, J., Ousley, O., Bowman, D., Harenski, K., & Kilts, C. (2007). The neural processing of moral sensitivity to issues of justice and care: An fMRI study. *Neuropsychologia*, 45(4), 755-766.
- Scheurer, H. (1993). *Persönlichkeit und Kriminalität. Eine theoretische und empirische Analyse*. Regensburg: S. Roderer Verlag.
- Schillinger, M. (2006). *Learning environment and moral development: How university education fosters moral judgment competence in Brazil and two German-speaking countries*. Aachen: Shaker Verlag.
- Schillinger-Agati, M. & Lind, G. (2003). *Moral judgment competence in Brazilian and German university students*. Paper presented at the AERA conference in Chicago, April 21st - 25th, 2003.
- Schläfli, A., Rest, J.R. & Thoma, S.J. (1985). Does moral education improve moral judg-

ment? A meta-analysis of intervention studies using the Defining Issues Test. *Review of Educational Research*, 55, 319-352.

Skinner, B.F. (1971). *Beyond freedom and dignity*. New York: Bantam/Vintage.

Wakenhut, R. (1982). Lebensweltliche Moral und moralische Segmentierung bei Unteroffizieren der Bundeswehr. In: P. Klein, ed., *Das strapazierte Rückgrat. Unteroffiziere der Bundeswehr*, pp. 309-327. Baden-Baden: Nomos Verlagsgesellschaft.

Wasel, W. (1994). *Simulation moralischer Urteilsfähigkeit. Moralentwicklung: eine kognitiv-strukturelle Veränderung oder ein affektives Phänomen?* Konstanz: Unveröffentlichte Diplomarbeit, Fachgruppe Psychologie, Universität Konstanz.

Important note:

In meanwhile (2013), the Moral Judgment Test (MJT) has been renamed as *Moral Competence Test (MCT)*. The name of the test is now aligned to the construct it measures, namely *moral competence (C-score)*. Competence is an persisting human trait while judgment is an ephemeral phenomenon.

We also speak now of 'moral competence' rather moral judgment competence to indicate that this competence can be observed only when it shows itself in overt action.