- Not for citation -[1]

30 Years of the Moral Judgment Test – Support for the Cognitive-Developmental Theory of Moral Development and Education

Georg Lind [2]
University of Konstanz, Germany

Abstract

I have developed the Moral Judgment Test (MJT) on the basis of ideas taken from philosophers (e.g., Habermas, Apel), psychologists (Piaget's notion of affective-cognitive parallelism, Kohlberg's definition of morality as a competence, G.A. Kelly' personal construct theory, H.H. Kelley's idea of subjective variance analysis, Jim Rest's postulate of hierarchical moral preferences), cognitive test theorists (Torgerson's concept of response scaling, N. Anderson's cognitive algebra; Guttman's facette analysis). The idea was not only to measure human attitudes and behaviors but to assess the cognitive structures which organize this behavior. As a result, the MJT uses the N=1 multivariate analysis of variance technology. The empirical criteria for checking the *theoretical validity* of the MJT were drawn from well-established postulates of cognitive-developmental theory: a) morality is a *competence* (Kohlberg, 1964), b) the correlations between different stage-type moral orientations form a *simplex structure* (Kohlberg, 1958), c) *moral orientations are universally preferred in the same order or hierarchy* (Rest, 1973), d) *affective and cognitive aspects of moral behavior are parallel* (Piaget, 1976; 1981).

After 30 years of MJT research, international findings with 29 different language versions not only demonstrate the theoretical and cross-cultural validity of the MJT but also strongly support the basic assumptions of cognitive-developmental theory on which the test was based.

---

[1] Last revision: July 18, 2006

[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-moral/*

Especially, MJT-research corroborates Piaget and Kohlberg's notion that morality has a competence-side (or cognitive aspect) which can be clearly distinguished from, and measured independently of, its affective aspects (like moral orientations or attitudes or values). However, some postulates of cognitive-developmental theory (invariant sequence, age-correlation) have been refuted. In sum, the MJT is well suited for cross-cultural research as well as for the evaluation of educational programs and methods.

The advancement of our knowledge on morality depends strongly on the advancement of measurement in this domain, and vice versa. We need tests in order to validate our theories to the point at which we can rely on them and make them part of our tested knowledge. Tests link our theories to reality and help us determine whether we can rely on our knowledge as a basis for decision-making in every-day life, that is, for designing effective methods of moral education.

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 epistemologist Karl Popper (1968 a, 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/1934, p. 425).[3] Thus, the scope and limits of our knowledge on morality determine what we can measure and how good we can measure it. Kohlberg (1984) has 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 ... abduction." (p. 424)

This spiral process sets in only when our knowledge does not work well anymore, that is, when data are so inconsistent with theory that we can no longer regard them merely as random measurement error and 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 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

[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)

for scientific progress because it lets us “predict some novel, 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, and by rejecting inconsistent data as caused by measurement error or lack of test validity. Lakatos has called such 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 with it. In other words, the hidden psychological assumption 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.

Cognitive-developmental theory of moral behavior and development, I believe, is such a “progressive problem shift” in the domain of moral psychology and education, which can explain the inconsistencies and insufficiencies of the two older and still prevailing paradigms in this domain, behaviorism and emotionalism.

Behaviorism defines the morality of an action purely from outside as its conformity (or non-conformity) with socially given norms or rules (e.g., Hartshorne and May, 1928; Watson, 1970/1924; Skinner, 1971; Moll et al., 2005). Morality is measured only by observing people’s reactions in situations in which they were tempted to transgress social norms, without paying attention to people’s own moral motives and principles.

The limitations of the behaviorist norm-conformity approach are most succinctly pointed out by some of its strongest proponents, Hartshorne and May (1928), after their experiments did not produce the moral trait which they expected, on the very last page of their research report:

> “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)

Emotionalism is the theory that moral action is caused directly by moral emotions and affects. Hence, in this approach tests of moral preferences, moral ideals and value inventories are

used, and morality is seen as belonging to the "affective domain" of human behavior, and excluded from the "cognitive domain" (Bloom et al., 1956; Krathwohl et al., 1964). Many of these tests were designed to measure only people's preferences of moral values set forth by religious or philosophical authorities.

Two of the best know moral attitude tests are the *Defining Issues Test* (DIT) by Jim Rest and his associates and the *Socio-Moral Reasoning Measure* (SRM) by John Gibbs and his colleagues (Rest, 1979; 1986; Gibbs et al., 1992). The DIT measures, according to its authors, "the amount of postconventional 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 preconventional thinking to a preference for conventional to one for postconventional thinking." (p. 14, footnote 1).

Critiques like Emler (1996) have pointed out a severe limitation of confining moral behavior to moral preferences, omitting the cognitive aspect of moral thinking and behavior, which Kohlberg calls "moral judgment competence."

> "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 attitudes (like moral preferences) can be simulated to meet the expectations of the experimenter, 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 kindly comply with these expectations. Indeed, this effect could also be demonstrated in regard to moral judgment behavior. Emler and his associates (1983) tested this hypothesis by having student participants fill out the DIT twice, one time as usual and a second time with the instruction to simulate the responses of students with opposite political attitudes. Their two major findings were fully in line with their social-desirability hypothesis: as one would expect, principled moral statements were preferred much more by political leftists (liberals) than by rightists (conservatives), and, even more important, both groups could easily simulate the moral

preferences of the opposite political camp. Emler's hypothesis also explains why a direct approach to moral teaching is highly effective in regard to moral attitudes or preferences, as Penn (1990) has shown, and why older participants in a dilemma discussion show larger "gains" in moral preferences than younger subjects (Schläfli et al., 1985).

Does Emler's experiment refute the cognitive-developmental approach? In their original article, Emler and his colleagues (1983) actually left this question open. Besides considering their experiment as a refutation of cognitive-developmental theory, they also considered the possibility that measures like the DIT might not produce adequate data for judging the value of cognitive developmental theory. In fact, after a round of debate (Markoulis, 1989; Barnett et al., 1995; Lind, 2002), Rest acknowledges that the DIT can be simulated upward, that is, that the DIT is not a measure of moral competence but of moral preference: "We [...] have eliminated the faking study from our set of the validity criteria." (Rest et al., p. 115)

Cognitive-developmental theories imply new methods of measurement and lets us make new predictions. The most important discovery was the discovery that morality has a very important *competence* aspect, which links moral intentions and preferences on the one hand, with everyday action and decision-making on the other.[4] Up until Kohlberg's discovery, nearly all psychologists and emotivist moral philosophers held the dogmatic belief (and still does so) that morality is merely an attitude or emotional trait which can be assessed only by soft methods like attitudes scales and non-obtrusive interviews, but did not even think that ability has any meaning in the moral domain. Many philosophers in the tradition of Kant believed that once you have committed yourself to moral values you will act morally, but did not see the necessity of building up the abilities to translate moral intentions into behavior.

In cognitive-developmental theory, both terms, performance and competence, are used in an almost interchangeable way. At some time, the idea of moral competencies became confused by an unfortunate distinction between (observable) 'performance' and (unobservable) 'competence.' This distinction was borrowed from Noam Chomsky's who used it to describe the relationship between a small set of 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

[4] Before Kohlberg, already Charles Darwin (1966/1874), in his seminal book *The Descent of Man,* has explicitly talked about *moral abilities*, by which 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).

moral competencies an unobservable entity and, even worse, it is alien to cognitive-developmental theory, as Lourenco 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 “competencies can only be observed through real performances [...] otherwise these competencies could not be measured” (p. 199; my transl.).

The more important distinction is that between moral *ideals, the affective aspects,* and the moral *competencies, the cognitive aspects,* whereas the latter are needed to put the former into action, as Piaget observed:

> “Affectivity constitutes the energetics of behavior patter 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." (Piaget & Inhelder, 1969, p. 158; see also Piaget, 1981/1954; Lourenco & Machado, 1996).
>
> “A systematic general observation of moral behavior, attitudes, or concepts in terms of such 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” (Kohlberg, 1958, p. 16).
>
> “I define stages solely in terms of cognitive structures, or ways of thinking or judging.” (Kohlberg, 1984, p. 398)

While the affective aspect of moral judgment behavior are conceptualized as the preferences of a person for certain moral orientations, which can be assessed by methods of attitude measurement, the cognitive aspect is conceptualized as its organizational property, which needs to be assessed in a new way (Broughton, 1978).

Piaget (1965/1932) and Kohlberg (1958) made a first breakthrough. They used the method of *critical clinical interviews* (Lourenco & Machado, 1996), in which subjects were not just observed but were confronted with a rather difficult moral task, much like in an N=1 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" (p. 146).
>
> “Piaget 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 in Piaget’s methodological steps, used 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)

In clinical-experimental assessment method by Piaget and Kohlberg is based on the assumption that structural properties of a participants responses are real and observable.

> “The fact that some people are more consistent, understandable, involved in our [morality] and theoretically interesting than others.” (Kohlberg, 1958, p. 93).
>
> “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, pretend 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., p. 58).

Structural properties of moral judgment behavior can, however, 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)

Later, under the growing pressure of main-stream psychologists (e.g., Kurtines & Greif, 1974), Kohlberg and his associates unfortunately dropped basic postulates of cognitive-developmental methodology in order to comply with mainstream psychometric standards. The new version of the MJI no longer tries to infer moral judgment competence from a participants interaction with difficult moral *tasks*. Counter-arguments are still brought forth by the interviewer, but the participants' responses to them are counted only one third or are discounted altogether (Colby et al., 1987, p. 161 & 186). "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).

Nor do they look anymore for the form or organization of a participants' pattern of responses but views them as a collection of unrelated responses. Now, they considered structure as "unobservable or hypothetical. [...] The structures themselves can never be observed [...]" (Kohlberg et al., 1984, p. 242)

The *Moral Judgment Test* (MJT)[5]

The Moral Judgment Test (MJT) has been designed to bridge the notorious gap between psychological theory and methodology. Therefore, it implies also a new way of test validation. It is a methodology in the tradition of cognitive-developmental theory and methodology, to be used both for testing new hypotheses about the nature of moral behavior and development as well as for the evaluating educational policies and practices.

In order to use the MJT for testing new assumptions about moral behavior and development, it has been submitted to rigorous validation procedures to make sure that the MJT has the same (semantic and pragmatic[6]) meaning for all participants, even for participants in different cultures, the validity of the MJT was tested not only in regard to its semantic meaning by careful translations but also in regard to its semiotic meaning, using four empirical criteria derived from theory and research: a) *Competence Nature*, no upward simulation possible (Nicolas Emler), b) *Simplex Structure* (Lawrence Kohlberg), c) *Preference Hierarchy* (Jim Rest), and d) *Affective-Cognitive Parallelism* (Jean Piaget).

As we have seen above, classical methods of test construction and test analysis ("validity", "reliability", "consistency" etc.) are far less neutral toward psychological theorizing that most psychological researchers seem to believe. They contain implicit psychological assumptions about the nature of human behavior and development which are at odds with cognitive-developmental theory. For example, classical and modern 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 structure or organization of their behavior (Kohlberg, 1958; 1981; 1984; Lind, 1982, 1989, 2006). With the MJT, so-called response inconsistency or measurement error is believed to contain an important information about a person's moral-cognitive organization or moral judgment competence. In accordance with our

---

[5] Note that the MJT has been constructed only for use in scientific research and evaluation studies (e.g., for evaluating the effects of certain methods of moral or character education, but *not* for diagnosing or selecting individuals or group of individuals. The MJT is not suited for the latter use, and as the author, I do not approve of it. For usage guidelines please visit this web-site: http://www.uni-konstanz.de/ag-moral/ .

[6] In older versions of this paper, we spoke of "semiotic validity" instead of pragmatic validity. Because "semiotic" is the more encompassing term (meaning everything that signs signify), which also includes semantic, we now prefer the term "pragmatic" as more adequate for what we mean.

Dual Aspect Theory of moral behavior and development, the MJT has been developed to assess simultaneously moral attitudes and moral judgment competence.

While there has been a long tradition of measuring *attitudes* in the moral domain and the methodology of measuring attitudes, values and alike is well-developed, measuring *competencies* in the moral domain is relatively new. Up until the work of Piaget and Kohlberg, psychologists had not even be aware of the fact that moral behavior has a competence aspect and, therefore, constrained morality solely to the affective domain of human behavior (Krathwohl et al., 1962). 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).[7] 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 its moral values in a morally consistent manner. This ensures cultural fairness when measuring moral judgment competence.

The measurement of a particular competence it closely tied to the definition of the kind of tasks for which it is needed.[8] Therefore, in order to measure *moral* judgment competence it is necessary to define a *moral* task. In the area of rule-conformity research, several tasks have been developed to this purpose. The ability to obey the rules set up by society (e.g., the school), adolescent participants were confronted with situations in which they had to resist the temptation to cheat or to steal or to submit to some abusive authority (cf., May & Hawthorne, 1928; Milgram, 1974). However, in regard to moral competencies we lacked the specification of appropriate tasks as instruments for measurement.

Lind and his colleagues considered several options (Lind, 1978; 1985a; Lind & Wakenhut, 1985). Informed by the theory of communicative ethics (Habermas, 1983; Apel, 1990), by Piaget's use of "counter-suggestions" in his clinical interviews (; see also Orlando, 1986; Inhelder et al., 1974) and by Keasey's (1974) research on adolescents' ability to deal with counter-arguments, they chose as a moral task a communication situation in which the participants had to rate moral arguments pro and contra their own opinion on a specific moral issue. That is, for the participant the

---

[7] Though was not the first to use this concept. Already Charles Darwin spoke of moral competencies in the way we use this term today.

[8] As Kohlberg (1985) notes: "In studying moral behavior we are concerned with studying action in which the subject gives up something or takes risks where not doing so would appear to be to his or her immediate advantage. ... Thus, it is the overcoming of these situational pressures on either a verbal or a physical level that constitutes the test of moral behavior" (p. 522).

main moral task involved in taking the MJT is to engage in a moral discourse by rating arguments speaking *in favor* and *against* her or his opinion on a fundamental moral dilemma. According to Lind's Dual-Aspect Theory of moral behavior, the participant will feel strong moral emotions when he is confronted with a deep moral dilemma situation and has to make a judgment on the protagonist's decision. As much research has shown (e.g., Keasey, 1974; Damasio, 1994; Haidt, 2001), these emotions can get so strong that cognitive processes like moral reasoning and judgment are severely hampered.

The past 30 years of research with the moral judgment test has provided much anecdotal and experimental evidence on the interplay of moral emotion and moral cognition (Lind, 1985, 1985b; 2002). The responses to the probing questions of the MJT illustrate this interplay:

- 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, they show various signs of excitement: straightening their body, touching their head, mumbling, making faces etc.
- At the bottom of their development, they will take a stance on the dilemma but will not, as required, respond to the subsequent arguments. Typical answers go like this: "What is this good for? I did already say what I think about this!" So, in this phase, even the "simple" task to judge arguments seems to be of unsurmountable difficulty.
- Next, they agree to respond to the given arguments but only to the *supporting* arguments (pro-arguments) but *not* to the counter-arguments. "I thought these arguments must be rated only by people who disagree." 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* (Rest, 1979)[9]. For the same reason, an expert colleague advised me to drop counter-arguments from the MJT. Obviously, he did not understand that a test of moral judgment competence had to incorporate a moral task and that counter-arguments could exactly serve this role.

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

- Only in their next phase of development, participants cope with counter-arguments, yet first only by rejecting them indiscriminately. All *supporting* arguments are accepted without sign of any doubt (getting all "+4"-ratings in the MJT), and all *opposing* arguments are definitely rejected (getting all "-4" ratings in the MJT). These participants seem to avoid strongly *cognitive dissonance* (Festinger, 1957) between their stance on a moral issues and their ratings of arguments. This phase of moral judgment development is reached by most participants at least at the age of ten, the lowest age for which the MJT is suited, at least if some special adaptations are made.[10] At this point, the MJT's index of moral judgment competence, the so-called C-score, is nearly zero.
- First signs of a more sophisticated moral judgment show up when the participant begins to discriminate between the moral quality of the given arguments, often starting with the *pro*-arguments. He or she discovers that not all supporting arguments are morally good and should be less accepted or even be rejected because of their moral inadequacy, like not all "friends" are of good character and can be considered as true friends.
- Parallel to this process, or a little later, the participants discover that not all *counter*-arguments are bad, but that some are less bad or even appeal to their own moral ideals. At this point the C-score really starts to climb up on the C-scale from 0 to 100. Some theorists believed, that moral competence indices like the C-score should climb up synchronously on various moral tasks (the assumption of "structural wholeness"). When we conceived the MJT, we rather believed that the C-score would vary according to the kind of the moral task, predicting that the mercy-killing issue would elicit a somewhat higher degree of moral judgment competence than the works dilemma (Lind, 1978). In fact, MJT studies up until some time ago confirmed this expectation (Lind, 1985 a, 1985 b). Studies by Krebs et al. (1991) and Juujärvi (2003) are also in line with this, though their authors they explain those intra-individual differences in moral judgment in a different way. They argue that these differences are due to the difference between (unobservable) competence and (observable) performance. However, because in this theory competence is regarded as "unobservable," this explanation cannot be

[10] 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 like explanations of unfamiliar words.

empirically tested without resorting to unproven assumptions. Like Habermas (1983), we believe that we can measure moral competencies by observing how he or she solves a difficult moral problem.[11]

- More recently, in studies in Latin-America, we discovered a very new phenomenon that has direct impingement on the question of structural wholeness and of cross-cultural validity (Lind, 2000a; Bataglia et al., 2002; Bataglia et al., 2003). In Latin-American countries respondents get much lower C-scores than in European countries. First it was hypothesized that this was due to the workers' dilemma because the respondents were mostly college students with a very affluent background, and that these subjects had no binding to the world of workers and, therefore, would respond to this dilemma with less motivation. Yet, the opposite turned out to be true, respondents in Latin American countries get, at average, very low C-scores on the mercy-killing dilemma. Some evidence suggests that the teaching of the Roman Catholic Church forbids true believers to deliberate about mercy-killing because the church offers a dogmatic solution for this dilemma. This phenomenon reminded us of the findings by Roland Wakenhut and his colleagues in their studies of German soldiers, who showed high C-scores in civilian dilemmas and lower C-score when similar issues were presented in a military context. Wakenhut (1982) touted this phenomenon "moral *segmentation*". Moral segmentation shows that the C-score not only reflects the individual's moral judgment competence and the moral difficulty of a task but also the influence of powerful social agencies like the church, the military and other social institutions.
- A maximum C-score of 100 is reached if the participant solely focuses on the moral quality of the argument when evaluating them, and not on their opinion-agreement and not the situational context. The first condition is necessary given the operationalization of moral competence as the ability to deal adequately with supporting and opposing arguments. Only if people are able to do this they will be able to engage in a moral discourse in order to solve a social problem non-violently by reason rather than by power and violence. The latter condition may be seen dubious in the light of the segmentation

[11] "Da Kompetenzen immer nur an ihren greifbaren Äußerungsformen, also anhand von Performanzphänomenen dingfest gemacht werden können, stehen diese theoretischen Ansätze vor besonderen Messproblemen." (S. 199) [Because competencies can always be assessed only through manifest forms of behavior, that is, through performance, these theoretical approaches [which distinguish competence and performance] are confronted with particular measurement problems.] My translation, GL.

phenomenon. It may require us to replace the C-score by a more refined index like the C-plus index which Lind (1978) had proposed some years ago. He decided against using the C-plus index because both scores were highly correlated, and because it would diminish comparability of MJT data with much existing research. We also believe that social forces refraining the individual from applying his reason to all moral issues are ubiquitous and to overcome them is part of the task that the responded has to solve for showing true moral autonomy. If we would try to compensate for these constraints of autonomy, we would have to give up the notion of moral judgment competence to be a universally valid ability. Yet, alternative indices might still be an option.

The Moral Judgment Test as a Multivariate N=1 Experiment

The experimental design of the MJT is basically the same as the one used in experimental psychology except that the universe of measurement is the *individual person* (N = 1, obviously, the natural unit of research in psychology) and not some group of people. The basic experimental design is this: Before the experiment, the subject is to read a Kohlberg-type moral dilemma story - (Colby et al., 1987), and is to judge the decision of the protagonist: "Was he/she doing right or wrong?" This task is to prime moral feelings in the respondent. Yet, this task sets only the stage for the actual experiment, in which the participant has to rate arguments of different moral quality, six argument 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 stages of moral orientation (Kohlberg, 1984).[12] It is assumed that this situation, in which counter-arguments have to be evaluated, elicits both *self-protective* emotions as well as *moral* emotions, that is, the tendency to protect one's own judgment, and the tendency to seek moral truth as Festinger's (1957) theory of *cognitive dissonance* describes it (see also, Habermas, 1983).

It is the *pattern* of responses to this structure, rather than isolated responses, which lets us see which of the two tendencies in the individual is stronger. Since we are especially interested in

[12] Temporarily, Kohlberg reduced his six stage-model at some point, but returned to it later again (Kohlberg et al., 1990).

the competence too seek moral truth (i.e., moral judgment competence), our focus is on this. In order to facilitate quantitative analysis, we developed a special index of moral judgment *c*ompetence, the so-called *C-score*, the "C" standing for "competence." The C-score reflects the degree to which an individual participant rates the 24 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.[13] Technically, the C-score reflects ration of the response variance accounted for by the experimental factor "moral quality of the argument" by the total response variance.[14] The C-score ranges from zero, meaning that the participant has not attended at all to the moral quality of the arguments, to one hundred, meaning that the he or she has rated the arguments solely for their moral quality .

The claim that the C-score reflects a *competence* and not merely an attitude, was subsequently tested experimentally (Lind, 2002; Wasel, 1994). Above all, it has been shown that the MJT's C-index cannot be simulated upward in the same experimental situations in which other tests could (Emler et al., 1983). Moreover, the accuracy with which participants perceived the moral judgment competence of others was strongly correlated with their own level of moral judgment competence Wasel, 1994).

The MJT's C-score has been designed to be *culturally fair* (Lind, 1995). In contrast to most, if not all other tests of moral development, the MJT's index for moral judgment *competence*, the C-index, reflects solely a participant's ability to apply his or her own moral orientation consistently[15] and is not tied to (though based upon) the participant's moral orientation. This means, in order to get a high C-score, the participant does not need to subscribe to particular moral orientations (as is required by most, if not all other tests of moral development). In theory, a participant could prefer Stage 1 reasoning most and get a high C-score. However, evidence from many MJT

---

[13] As Kohlberg (1985) notes: "In studying moral behavior we are concerned with studying action in which the subject gives up something or takes risks where not doing so would appear to be to his or her immediate advantage. ... Thus, it is the overcoming of these situational pressures on either a verbal or a physical level that constitutes the test of moral behavior" (p. 522).

[14] Note that the meaning of the C-score is derived not only from the numerical calculation of variance components (consistency) but also from the task involved in the MJT. A test, which does not incorporate a moral task does not allow us to measure moral *competencies* even if one calculates an index similar to the C-score like Rest and his colleagues did with the DIT, which assesses only the preference for principled moral reasoning (Rest et al., 1997).

[15] Note that the MJT does not measure an undefined consistency but *moral* consistency, that is consistency *in regard* to the subject's preferred stage of moral orientation. It should be clear that other forms of consistency, like opinion agreement and moral "rigidity" mean the opposite of competence.

studies shows that this is hardly ever the case. As Piaget has hypothesized, the cognitive and the affective aspect of moral behavior are very strongly associated. The higher participants' moral judgment competence is, the more do they prefer principle moral arguments, and the more do they reject low Stage moral arguments (see below).

Finally, it has been shown that the response pattern of the MJT's narrative confirms well with the theoretical assumptions underlying the construction of this test. To check on this claim, we used three empirical criteria for semiotic correspondence, which we will also use here for judging the cross-cultural validity of translated versions of the MJT and of newly developed subtests of the MJT. The same rigorous criteria are being used in addition to more traditional strategies for securing cross-cultural validity (see, e.g., Gielen et al., 1996; Edwards, 1981).

Semantic and Pragmatic Threats to Communicative and Cross-Cultural Validity of Psychological Tests

In an objective test of moral judgment behavior like the MJT, complex moral thoughts must be represented in only one short sentence, because we do not want to overburden the subjects' short term memory with too long and complex statements to remember when making the evaluative response, on which we rely our observations. If we did, we could hardly tell, which parts of the statement she or he was actually responding to. In contrast to an open-ended interview in an objective test, the subject is also deprived of the possibility to make clarifications and explanations to make sure that the test scorer truly understands what she or he wants to communicate. Therefore, objective tests must not only be checked carefully for semantic validity but also for pragmatic validity if we want to be sure that there is maximum communicative validity or, as Campbell (1963) has called it, "conceptual overlap."

By *pragmatic*[16] *validity* we mean that each item of a test and the test as a whole correctly represents what we want it to mean, that is, for example, in the case of the MJT, the accuracy to which each argument included in the test represents one of the six Kohlbergian stages of moral orientation. In the German master version of the MJT we have checked on this by carefully re-rea-

[16] Semantic problems refer to problems of word and sentence meaning in a rather direct sense of "how speakers know that a given utterance represents a given thought or idea. Thus semantics is concerned with a type of translation: the translation from thought into utterance, and vice versa" (Wojcik, 1998).

ding the items and by having six noted experts in the area of moral development research to critically examine each item.[17]

In the process of cross-cultural validation of the MJT, semantic validity was sought to be established though three means. First, the author of the foreign language version, typically herself or himself an expert on Kohlberg's stage model and Lind's *dual aspect theory* of moral behavior and development, not only translates the test item by item but also checks her or his translations against the theory. Second, most authors double-check on semantic validity through backward translations. Third, in cases of uncertainty, experts of the theory who also are native speakers of the target language are asked to critically comment on the items of the MJT.

By *pragmatic validation* we mean the degree to which the subjects, who probably have no knowledge of the theory, understand the arguments they are to rate, in the same way as we, the experts, do. Our instructions and test items may be hundred percent semantically correct and still the subjects may understand them in a different way and thus their responses may be easily misinterpreted and mis-scored. As Campbell (1963) argues this problem exists even in situations where the instructions and stimuli are very simple and hardly prone to semantic ambiguities, as in situations in which animals are involved as experimental subjects. The more we must be on alert when we deal with complex matters like morality. As Kohlberg (1958, 1981) maintained, a single argument, taken by itself can never be a reliable sign of a specific moral orientation or moral judgment competence, it must be always interpreted in context. This is why traditional ways of checking the "reliability" and "validity" of test items are insufficient if not just wrong. A reliable and valid interpretation of test data cannot be achieved just by looking at isolated arguments. We need to look at *relationships* and *structures*.

We have checked the *semantic validity* of the master version of the MJT in two ways. First, we asked a small sample of subjects to talk aloud when filling out the MJT and write down any comments they wanted to make. From this material we could detect many misunderstandings which were provoked by the wording of the MJT. Subsequently we revised the test items and resubmitted them to the same procedure again. Second, we submitted the responses of the subjects to four types of relational analysis, which will be described below in more detail. Looking at the

---

[17] I wish to thank all experts involved in this: Tino Bargel, Rainer Döbert, Thomas Krämer-Badoni, Gertrud Nunner-Winkler, Gerhard Portele, and Roland Wakenhut.

relations and inter-correlations between the responses to the MJT, let us detect further instances of pragmatic invalidity, which the subjects did not need to be aware of.

The latter method is also used to check on the pragmatic equivalence of translated versions of the MJT. Because it does *not* require to analyze foreign language material, it seemed well suited as a basis for cross-cultural validation. I will now describe the criteria which we used for this validation process in more detail and present the findings for the validation of the translated versions of the MJT.

Improper Criteria of Validity for Structural Measures of Moral Competence

In psychological research, little has changed since the Fifties, when the philosopher Ludwig Wittgenstein (1971) observed that the "the problems and the methods miss each other."[18] Conventionally, the validity of measures of moral development is inappropriately judged on the basis of *sample* statistics (like the degree of internal consistency) rather than intra-individual analysis, and on the basis of stability of rank orders ('reliability') and the correlation with age rather than the goodness of fit with psychological theory (see, e.g., Colby et Colby et al., 1987; Gibbs et al., 1992; Rest, 1979). In doing so, the research method used in moral psychological research is at odds with modern moral psychology (Kohlberg, 1981; Lourenco & Machedo, 1996; Lind, 1989, in press).

Whereas modern psychology, especially structural models of human functioning imply that each person has a unique structure of cognitive-affective functioning, most if not all statistical analysis employed in psychological research rests on the hidden assumption that all people do not differ structurally but only gradually in regard to some universally shared trait dimensions. As far as people do not into this universal trait structure model, their responses are perceived to be "inconsistent" (in regard to that model!) and are treated as measurement error. Thus our measurements do not allow s to distinguish between amoral, immoral and highly differentiated moral responses because of the overly simplicistic personality models on which most prevailing statistical

[18] My translation. Full quote: "Die Verwirrung und Öde der Psychologie ist nicht damit zu erklären, daß sie eine 'junge Wissenschaft' sei; ihr Zustand ist mit dem der Physik z.B. in ihrer Frühzeit nicht zu vergleichen ... Es besteht nämlich, in der Psychologie, experimentelle Methode und Begriffsverwirrung ... Das Bestehen der experimentellen Methode läßt uns glauben, wir hätten die Mittel, die Probleme, die uns beunruhigen, loszuwerden; obgleich Probleme und Methode windschief aneinander vorbeilaufen" (p. 370).

methods are based (Pittel & Mendelsohn, 1966). In contrast, a few methodologies were developed to match psychological concepts like, for example, Kohlberg's (1958) initial, individual-focused method of clinical interview for assessing a person's structure of moral judgment behavior (see also Kohlberg, 1981), George A. Kelly's (1955) grid technique for assessing "personal constructs," Norman Anderson's (1991) cognitive algebra paradigm for assessing moral information processing. Building on these methods and on the concept of N=1 experimental design, we developed the method of *Experimental Questionnaire* (Lind, 1982), on which the MJT is based.

Whereas modern moral psychology is interested in the structural organization of moral reasoning and feeling and in manifest pattern of judgment behavior (Kohlberg, 1981; Burton & Kunce, 1995[19]), high "internal consistency" is viewed as invariantly high for all people and, therefore, as one of the principal criteria of measurement validity (Colby et al., 1987, p. 71), and thus data indicating structural differences are discarded as "measurement error." As we have seen above, there are different forms of consistency of judgment on different levels of moral-cognitive development. At an earlier stage, participants show consistency of judgment in regard to their own opinion on an issue (Keasey, 1974), and behave mostly inconsistently in regard to behavioral norms. Later, their judgments and behavior become more and more consistent in regard to their own moral orientations ((Burton & Kunce, 1995; Lind, 2002).

Whereas modern moral psychology finds a strong correlation of moral development with the amount and quality of education, and only spurious correlations with chronological age, that is, correlations which disappears once education is held constant (Lind, 2002; Lourenço et al., 1996; Rest, 1979; Rest & Thomas, 1985),[20] a high correlation of age with moral development ("invariant sequence") is called the second important criterion for measurement validity: ""Nothing is more crucial to a cognitive-developmental construct than evidence of change over time from less

---

[19] "[In Hartshorne & May's study] consistency [of behavior] across situations was positively correlated with age, with some children becoming more consistently honest and others more dishonest as they matured." (Burton & Kunce, 1995, p. 148).

[20] Kohlberg (1984) remarks: A correlation with age "is not 'validating.' Many adults are morally immature, so that a test which maximized correlation with age would ecologically relate to age but have little relation to moral development" (p. 194). Similarly, Rest (1979) reports: "Coder (1975) found a slightly negative correlation of P with age, r(86) = -.10, while the P score was positively and significantly related to education, r = .25. [...] Crowder (1976) found that age correlated with P , r = -.05, whereas education correlated r = .25.[...] G. Rest (1977) found a correlation of 45 of education with P in a sample of 43 adults randomly selected from the Minneapolis telephone book.[...] Negative age trend in adult s can be completely accounted for by differences in education. [...] Therefore, consistently in adults, moral judgment is more positively related to education than to age. Cognitive restructuring of one's moral thinking seems to be more related to formal education than to the passage of years" (p. 112).

advanced forms of thinking [. . .] to more advanced forms of thinking." (Rest, 1986, p. 106; see also Kohlberg, 1958, pp. 17, 88, and 101; Colby et al., 1987, p. 71; Rest, 1979, p. 98 and 143). Many studies have found undisputable signs of regression and, therefore, *age* and *invariant sequence* cannot be regarded as a valid criterion for validating moral development tests. It is clear by now that moral development is not fueled by biological age but by quality and quantity of education (Lind, 2002; Rest, 1979; 1991). The illusion of age-correlation seems to have been caused by the fact that in many studies age and level of education are confounded because most longitudinal and cross-sectional studies have involved only youth who later attended college or university. Only when we studied youth who did not enter an academic career but entered the workforce at the age of 14 or 15, we found clear evidence of moral regression once they did not receive adequate education (Lind 2000b; Helkama et al., 2003). Age by itself does *not* foster moral judgment competence. Up until a certain critical stage of development, it seems, educational input is a necessary condition for moral growth. In many studies, the correlations between level of education and moral development have been higher than all other variables studied (Lind, 2002; Rest, 1979; 1991). For these reasons, we reject the notion that age could be a criterion for test validity.

Would years of education than be a good criterion for the MJT's validity? For some time it appeared that level of education could be a good criterion for validating moral development tests, but we no longer suggest to do so. Recently, studies have shown that some institutions of education do not foster moral development but hamper it. If education is low quality in regard to moral development, more education can mean less moral judgment competence as is the case with medical education (Lind 2000b; Helkama et al., 2003). Unpublished data from Colombia also show a lowering of C-scores from grade 8 to grade 12.

Although, we do not use education as a validation criterion anymore, this variable seems helpful in drawing a proper validation sample with large enough variance of moral judgment competence. The size of variance of the C-score in the validation sample restricts the correlation that can be found. If there is no or only very little variance, no correlations can be found at all. To make sure that the validation sample is appropriate for testing the cross-cultural validity of the MJT, we suggest that the validation sample consists of three participants representing three different levels of education, with the level about two years of education apart. In most cases, it seems, this sampling rule helps to maximize the variation of moral judgment competence in any sample so that

a meaningful validation analysis is possible. Note that there can be not correlation and not validity proof if there is no or little variation in the measurement values.

Proper Criteria for the Validity of the MJT: Three Invariant Universals as Signs of Pragmatic Equivalence [21]

In contrast, we judge the construct validity and cross-cultural equivalence of the MJT by checking the goodness of fit of the measures to three well-established facts regarding the nature of moral judgment behavior:[22] Any deviation of the data from these facts would indicate a lack of validity of the measurement:

- Moral judgment *Competence* is a genuine ability, and not merely an *orientation, ideology* or *attitude* (Kohlberg, 1963). Thus it should not be possible to simulate measures of moral judgment competence upward (for confirmatory findings, see Lind, 2002).
- Moral *orientations* which are perceived as adjacent (e.g., stage 2 and 3 orientations in Kohlberg's stage model) are accepted or rejected in a more similar way than moral orientations which are seen to be more distant (e.g., stage 2 and stage 5 orientation). Thus moral stage orientations should correlated the higher the closer they are on Kohlberg's stage model. They should form a *Simplex Structure* (see Kohlberg, 1958).
- Stage-typical moral orientations form a universal *Hierarchical Preference Order* (Kohlberg 1958, 1984). Thus, regardless of cultural and ideological background, social class, age or gender, people should prefer (or reject) them in the same way (see Rests, 1969).

- Hence, moral orientations and moral competencies can be clearly distinguished as different properties or aspects of moral behavior, but they cannot be *separated* as

[21] In older publications Lind also used correlation of moral judgment competence with level of education as a fourth criterion. Because there is a certain circularity involved in this criterion, it is dropped from this list.

[22] This does not mean that this knowledge has been fancied out of the blue but that it is research-based knowledge which has been distilled into a concise and coherent theory (Lind, 1985; Lind & Wakenhut, 1985) rather than merely "piled up" as is so often the case in "empirical" validation studies.

behavioral components (Piaget, 1976; 1981; Higgins, 1995).[23] According to Piaget's theory of *Affective-Cognitive Parallelism*, measure of moral orientation and moral competence should correlate strongly with each another (Lind 1978; 2002). However, this correlation may break down if something is at stake for the participant because he or she may then simulate the orientation measures towards greater 'social desirability.'

In the following, I will discuss only these last three validation criteria because the first one has been well-established in two laboratory experiments (Lind, 2002; Wasel, 1994) and have not been made a criterion for cross-cultural validation for economical reasons. I will now discuss these technical meaning of these criteria in turn in more detail.

Criterion # 1: Competence Nature of Morality and No Upwarf Simulation

(To be supplemented)

Criterion # 2: Simplex Structure of Moral Orientations

The validation criterion of "Simplex Structure" refers to a special way that ordered variables may inter-correlate: more adjacent or similar variables correlate more highly with one another than more distant or dissimilar variables. This validity criterion for scales of moral development is rooted in Kohlberg's (1958) dissertation study, in which he states:

> "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

[23] "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. ..." (Higgins, 1995, p. 53)

between them, . . . Such a pattern in quantitative data is called a 'simplex' by Guttman" (pp. 82-83).[24]

"The implication is that each level would correlate most highly with its neighbor(s) and its correlations with other levels would decrease steadily as those levels were increasingly distant from its position in the order of levels. If the matrix of these correlations were arranged in this developmental order, the correlations would decrease in any direction moving away from the main diagonal." (p. 84)

Kohlberg employed a graphical inspection method to evaluate the degree of fit of this prediction with observed data. So, 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. In contrast, use two statistical methods to assist this visual inspection. First, 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 rather crude and insensitive.

Second, we are using Principle Component Analysis (which is similar to factor analysis, except that the diagonal of the correlation matrix to be analyzed contains ones rather than estimates of reliability) with simple *varimax* rotation. The criterion predicts that the analysis (with *eigenvalues* set as 1.0) should produce *two* factors and that the *factor loadings* of each stage should lie on a circle, being perfectly ordered from stage one to six. Because research findings suggest that the order between stages 1 and 2, as well as the order 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 *it's a priori* probability is very low. Theoretically, there are 4! or 4 x 3 x 2 x 1 = 24 ways in which the stages can be ordered. Hence, the probability that they are ordered in the predicted way by chance is p = 1/24 = 0,042.

[24] 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 is 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.

In his original study of 83 boys age 10 to 16, Kohlberg (1958, pp. 100 & 104) found this pattern of correlations:

	St1	St2	St3	St4	St5	St6
St1	–					
St2	.55	–				
St3	-.41	-.19	–			
St4	-.52	-.41	.18	–		
St5	-.52	-.58	.09	.00	–	
St6	-.37	-.43	-.29	-.07	.23	–
$\bar{x}$	17.2	15.6	20.8	19.0	10.5	4.3
s	16.8	12.6	10.5	12.6	11.6	11.0

Note that Kohlberg used *relative* frequency of stage usage in interviews as an indicator of stage preference (we call them *ipsative* because they must add up to 100 percent), implying that some indices must correlate negatively with one another (as some percentages get high, others must go down by definition). I submitted this correlation matrix to principle component analysis, getting the graph depicted in Figure 2. Comparing Kohlberg's data with an ideal Simplex-Structure from fictitious data (Figure 1) shows that they fit well though not perfectly. Studies using the MJT show even a better fit to this criterion (for an example, see Figure 3).

This finding may sound trivial but it is not. Firstly, because the intercorrelations may have resulted in a completely different stage order, and secondly, because other tests have indeed shown other stage orderings.

DIT (Sprinthall & Stewart, 1995)

Intercorrelations between Stages:

The entries in italics clearly refute the hypothesis of stage ordering

Stage	4	3	1 & 2	
5	*-.47*	-.19	-.19	
4		*-.30*	-.19	
3			.14	

Criterion # 3: Hierarchical Preference Order of Moral Orientations

Moral ideals and principles, which we also call moral preferences or moral attitudes, are usually seen as separate components of behavior. However, according to Piaget (1981) they are *aspects* of behavior which must be *distinguished* from cognitive aspects but must *not* be *separated* from them. Accordingly, the MJT, though it is mainly made for assessing subjects' *competence* to make moral judgments, it provides also indices for a person's *moral attitudes* or *moral preferences*. (I will come back to this below.)

To categorize a person's moral attitudes/preferences, we use Kohlberg's original six stages of moral reasoning, which he once reduced to five stages, but later reconfirmed.[25] The subject is given arguments that resemble each of theses stages, one argument pro and one speaking against the particular decision made in each of the two dilemma situation, which the MJT contains, and is to express his or her degree of acceptance or rejection to each of them. Up until the year 2001, the test asked to rate the arguments' "degree of acceptability." To emphasize more the subjectivity of this rating task, the subject is now instructed to express how much the subject *accepts* or *rejects* each argument.[26]

In the MJT, *moral attitude* toward the stages is defined as the subject's mean *acceptability ratings* of all arguments in the MJT that represent a particular stage. Because the standard MJT has two dilemmas, and in each dilemma two arguments – one *in favor* and one *against* the respondent's decision on the dilemma – for each of the six original Kohlbergian stages, each stage is represented by four items. The respondent can choose a number from "-4" ("I completely *reject* it") to "+4" ("I completely *accept* it"). Thus, attitudes toward each stage are represented either by an index ranging from -16 to +16 or, if means are calculated, by an index from -4 to +4.

Kohlberg (1958; 1984) and Rest (Rest, 1973; Rest et al., 1969) have compiled ample evidence suggesting that the highest stages of moral orientation are not only preferred as the ideal level of reasoning by philosophers but also by most ordinary people. Before them, the psychiatrist Max Levy-Suhl (1912) had found that even juvenile delinquents valued universal moral principles higher than conventional or pre-conventional reasons. In many MJT studies, these findings were

---

[25] See Kohlberg, Boyed & Levine, 1990.

[26] I wish to thank Dr. Michael Hauan, University of Missouri, for convincing me on this issue.

clearly corroborated (see Lind, 2002b). University students as well as delinquents of the same age revealed the identical hierarchical order of moral preferences or attitudes (Figure 4).

We use this well-established finding of a Hierarchical Preference Order as our second criterion for testing the cross/cultural validity of translated versions of the MJT. In order to be valid, the attitudes toward the six Kohlbergian stages must be ordered according to their stage numbers, with the highest stage 6 preferred the most and stage 1 preferred the least.

Criterion # 4: Affective-Cognitive Parallelism

As noted above, Piaget (1951; 1976; 1981; Piaget & Inhelder, 1969) speaks at many places about the intriguing problem of inseparability of affective and cognitive aspects of human behavior, which, however, can be clearly distinguished. For a long time, this idea could hardly be tested empirically, because there was no method of measurement available which allowed us to assess both aspects of behavior simultaneously as 'distinct" and yet 'inseparable.'

> "Affective and cognitive mechanisms are inseparable, although distinct: the former depend on energy, and the latter depend on structure." (Piaget, 1976, S. 71)

The Moral Judgment Test, it seems, is the first and hitherto the only instrument which makes possible a simultaneous measurement of affective and cognitive aspects of moral judgment behavior. Piaget's parallelism hypotheses has been very well supported by MJT studies. In Figure 5, the findings from a study of German 1st semester university students are depicted as an example (for more examples, see also Lind, 1985; 1985a).

The MJT's index for moral judgment competence, the C-score, correlates systematically with the subjects' attitudes towards each of the six Kohlbergian stages of moral orientation: highly negative with the attitudes towards the low stages and highly positive in the case of high stages, with the other correlations stage-ordered 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 as adequate stages of reasoning and discourse for solving 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 (Lind, 2002a; 2002b). In such situations, the subjects may simulate socially desirable moral attitudes or may "underachieve" by showing less moral judgment competence than they are capable of, or do both, thus blurring the picture we get. While indices of moral attitudes are susceptible to simulation either "up" or "down" (Emler et al., 1983), moral competencies may not fully show if the situation is aversive (as, e.g., under time pressure).

Findings from Cross-Cultural Validation Studies

For certification, first each newly translated version of the MJT had to be double-translated or back-translated to make sure that the new language version was semantically equivalent. Second they had to submit the new version to an empirical validation study using the validation criteria #2 to #4 (see the guidelines on the internet: http://www.uni-konstanz.de/ag-moral/). Most importantly, the raw data had to be submitted to the principle author of the MJT for validity checks. In instances, when the findings did not meet one or several of the validation criteria, the reviewer recommended revisions of the translations and gave specific hints on the basis of the validation analysis. These hints mostly helped the author(s) of the foreign language versions to identify the source of invalidity and to improve the new test version. In some cases, technical flaws (like mistakes regarding data entry and processing or scoring) were detected which accounted for a lack of validity. Such technical errors included among other things confusing variable numbers, keying in wrong data, making mistakes when *hand-sorting* the data to arrange the variables according to the stage numbers (I strongly recommend using a computer program to do this!), and errors in the scoring program. Only two (or three) of twenty validations studies produced a perfect or near to valid test version on the "first shot." All others had to be corrected and to be submitted to an empirical validation study a second time. None had to be corrected after this again. Only when the findings showed a (near-to-)perfect fit, new versions were certified as cross-culturally valid.

There are now 29 different language versions of the MJT which have undergone this rigorous validation and certification process (see also Appendix). In all cases, the validity criteria were met, either on the first shot (in two cases) or after some revisions concerning technical and pragmatical problems detected in the process of certification. Although the C-score varied largely between the 29 countries and showed different correlations with level of education, the three crite-

ria were corroborated. In each country a) the moral-stage-orientations correlation in the predicted way and formed a Simplex, b) the higher stages were preferred over the lower stages of moral reasoning, and moral judgment competence was strongly correlated with moral orientations. In the appendix, some sample findings are depicted.[27]

A copy of the English version of the Moral Judgment Test is attached in the appendix of this paper. Information on the rational of the test, the condition of use, and its scoring can be also assessed in the internet. So can other language versions.

Taken together, the three criteria used for validating the Moral Judgment Test and for certifying new language-versions (Simplex, preference hierarchy, and affective-cognitive parallelism) are extremely unlikely to be met by chance alone. Because they address different aspects of the pragmatic meaning structure of moral judgment behavior, they are very sensitive to any lack of construct validity, much more sensitive than most conventional psychometric criteria. Moreover, they made it possible for independent reviewers without knowledge of the target language, to closely analyze the cross-cultural validity of translated versions of the MJT. Note also, that this validation procedure is very arduous and costly. Yet, all authors of new versions agree that the procedure was worth-while because it built up trust in the validity of the data gained with a certified version.

All certified versions of the MJT meet the three demanding validation criteria and can be considered cross-culturally valid or equivalent. Note that this does not imply that the level of moral judgment competence is the same in all countries. On the contrary, the C-index varies greatly between these cultures. Yet, only because the meaning structure of the test has shown to be invariant across these cultures, these variations can be regarded as valid signs of different levels of moral judgment competence. This means that the Moral Judgment Test offers a good way of cross-cultural research and research into the effectiveness of educational policies and practices (however, not for selection and sanctioning of people or institutions!)

This validation strategy contrasts sharply with other more conventional methods of test and item selection, which select and revise items and scoring techniques in order to maximize the correlation of test scores with certain criteria like stage invariance, age, education or political attitude. Such validation methods produce a tautology, which makes an instrument unsuitable for

[27] More can be found on the authors's web-site (http://www.uni-konstanz/ag-moral/) in the certification reports which are available there for many of the studies or can be requested from the respective authors

testing those hypothesis anymore. Even Kohlberg was not immune against such 'conventional wisdom.'[28] The items of the MJT were constructed strictly on the basis of three criteria that have shown to be universally true. These universally supported criteria helped to validate the instrument in what Kohlberg (1981) called a "bootstrapping process.". The fact that it was possible to certify so many translated versions as cross-culturally valid means that Kohlberg's (1964) construct of "moral judgment competence" is cross-culturally applicable.

In this context it should be noted that the author dropped not only age but also "level of education" from the list of validation criteria because quality and quantity of education varies greatly between the cultures that have been studied (cf. Colesante & Biggs, 2003; Lind, 1986; Schillinger-Agati & Lind, 2003). As the data shows (see appendix, Figures 8 through 14), the correlations vary strongly from country to country, indicating that this criterion is only of limited value as validations criterion. Only in Germany and some other European countries there are strong linear relationships between education and the development of moral judgment competence. In most other countries, schools do not seem to have such a positive impact on moral development. This finding deserves further research. Therefore, any direct comparison of the C-scores across cultures and countries is likely to be misleading, unless differences in level and especially in quality of education in each country are taken into account. Even when we compare C-scores across levels of education within one country, we cannot be sure if these levels really correspond to different quantities of education if the quality of education is not considered. For instance in one country the validation process brought about a higher mean C-score in 12th graders than in college students. On my inquiry I found out that the 12th graders were taken from a private high school in an affluent neighborhood, whereas the college was located in a very poor area and was badly financed.

---

[28] For example, Kohlberg (1976), the main author of the Moral Judgment Interview (MJI) writes: "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) second by stating "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). Hence, the Moral Judgment Interview, which has been thoroughly revised to meet these criteria, cannot be used to test the hypothesis of invariant sequence because it is biased towards supporting this hypothesis by virtue of construction. For a detailed analysis of the MJI, see Lind (1989).

Similarly, Rest (1979) regards "the demonstration of age trends as crucial" (p. 143) for the validity of his Defining Issues Test (DIT). Hence, the DIT cannot be used to test the hypothesis of age-correlation because it is engineered to support this hypothesis.

Conclusion and Outlook

Good measurement and progress of 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 both fields is not linear but is 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 construction of the Moral Judgment Test has been based on what we know about the nature of moral judgment behavior for sure, in order to explore contested assumptions and hypotheses. So far, three properties of human moral judgment behavior seem to be well-confirmed facts: Simplex structure of moral orientations, hierarchical preference order of moral orientations, and affective-cognitive parallelism. These three facts served as rigorous criteria for validating and certifying the Moral Judgment Test. In this study it was shown that the scores produced with the MJT and with its various language versions confirm very well to these three facts and thus the MJT can be considered as highly valid. Differences in C-scores across various cultures reflect true difference in moral judgment competence and cannot be discounted as lack of pragmatic equivalence of the different test versions.

Hardly any nonscorable test data have been reported, and all data sets (which have been made available to me in many cases) show only a very few missing data. This signifies that the MJT can be used in all this countries to produce meaningful data. This contrasts favorably with tests which have produced many nonscorable data in various cultures (up to 50 percent; see, e.g., Gielen et al., 1986).

The confirmation of cross-cultural validity of the MJT in 29 different countries also supports core assumption of cognitive-developmental theory (except the assumption of non-regression) and of Lind's Dual-Aspect Theory of moral judgment behavior (Lind, 2002):

- Morality has a strong competence aspect, which can be demonstrated in experimental settings by objective data.

- Regardless of cultural background (and regardless of age, gender, and education), people prefer moral orientations (as defined by Kohlberg's stages) in the predicted order.
- These moral orientations also constitute some developmental order as neighboring stages are more correlated than more distant stages, and
- the higher people's moral judgment competence, the more they prefer principle moral reasoning over lower stage reasoning.

However, MJT research refutes the postulate that the development of moral competencies is invariant and cannot regress. We found ample evidence that peoples' moral competencies can erode if they have not yet reached a critical level of autonomous moral development, and if at the point the educative support by the social environment has been withdrawn, as is the case, for example, with graduated from German middle school who enter a vocational training program or directly the workforce (Lind, 2002). Another example are medical school students, who loose their moral judgment competence during study, as has been shown in the longitudinal study by Lind (2001b). 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.

The findings accumulated in 30 years of MJT research have two important practical implications in education and therapy. First, the MJT is a valid method for comparing the effects of educational and therapeutic methods and educational systems across different cultures, thus opening up a whole lot of interesting research questions which have a bearing on educational policy making, like the question, whether different learning environments have a differential impact on moral learning.

Second, these findings show 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" values, but rather a need to teach children (and adults) how to apply their own moral values to specific decision-making and to resolve inevitable conflicts, when they try to do so. So for moral or character education, methods should be used which are appropriate for this objective like Blatt and Kohlberg's method of dilemma discussion or our revised Konstanz version of the dilemma discussion, which shows to be highly effective and also well manageable by teachers (Lind, 2003).

Finally, the phenomenon of 'moral segmentation' needs attention in further studies because it is of immediate relevance to the topic of cross-cultural validity (see Lind, 2000a; Schillinger-Agati & Lind, 2003). This phenomenon does not invalidate the MJT nor does it necessarily prevent cross-cultural comparison. Rather it reveals another strength of the MJT, namely that it allows us to detect and study the segmentation phenomenon in detail. Therefore, we strongly recommend for research and evaluation studies not only to look at the overall C-score but also to analyze the C-scores for each dilemma in order to check on the possibility of 'segmentation,' especially when the score appear unusually low.

References

Apel, K.-O. (1990). *Diskurs und Verantwortung.[Discourse and responsibility.] Das Problem des Übergangs zur postkonventionellen Moral* (Original 1988). Frankfurt: Suhrkamp.

Anderson, N.H. (1991). Moral-social development. In: N.H. Anderson, *Information Integration Theory*, pp. 137-187. Hillsdale, NJ: L. Erlbaum.

Barnett, R., Evens, J., & Rest, J. (1995). Faking moral judgment on the Defining Issues Test. *British Journal of Social Psychology, 34*, 267-278.

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.

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.

Campbell, D.T. (1963). Social attitudes and other acquired behavioral dispositions. In: S. Koch, Ed., *Psychology: A study of a science*, pp. 94-172. New York: McGraw-Hill.

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.

Colesante, R. & Biggs, D. (2003). *Cross-cultural comparison of moral competencies and moral attitudes.* Paper presented at the AERA conference in Chicago, April 21st - 25th, 2003.

Culler, J. (1998). Semiotics. In: *Grolier Multimedia Encyclopedia*. CD-Edition.

Damasio, A. (1994). *Descartes' error: emotion, reason and the human brain*. New York: G. P. Putnam's Son.

Dana, R.H. (2000). Multicultural assessment of child and adolescent personality and psychopathology. In: A.L. Comunian & U. Gielen, Eds., *International perspectives on human development*, pp. 233- 258. Berlin: Pabst Science Publishers.

Darwin, Ch. (1966). *Die Abstammung des Menschen*. (Original, The Descent of Man. 2nd Edition, 1874). Stuttgart: Kröner Verlag.

Edwards, C.P. (1981). *The comparative study of the development of moral judgment and reasoning*. In R.H. Munroe, R.L. Munroe & B.B. Whiting, eds., Handbook of cross-cultural human development, pp. 501-527. New York: Garland Publishing.

Edwards, C.P. (1986). *Cross-cultural research on Kohlberg's stages. A basis for consensus*. In S. Modgil & C. Modgil, eds., Lawrence Kohlberg: Consensus and controversy. Brighton GB: Palmer Press.

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.

Festinger, L. (1957). *A theory of cognitive dissonance.* New York: Harper & Row, Publishers, Inc.

Gibbs, J.C., Basinger, K.S., & Fuller, R. (1992). *Moral Maturity: Measuring the Development of Sociomoral Reflection*. Hillsdale, NJ: Erlbaum

Gielen, U., Cruckshank, H., Johnston, A., Swanzey, B. & Avellani, J. (1986). The development of moral reasoning in Belize, Trinidad-Tobago, and the U.S.A. *Behaviour Science Research, 20* (1-4), 178-207.

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., Saliminen, 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.

Juujärvi, S. (2004). *The ethics of care and its development. A longitudinal study among practical nursing, bachelor-degree social work and law enforcement students*. Helsinki, Finnland: Helsinki 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.

Keller, M. & Edelstein, W. (1991). The development of socio-moral meaning making: Domains, categories, and perspective-taking, in W. Kurtines & J. Gewirtz (eds.), *Handbook of moral behavior and development,* Vol. 2, pp. 89-114. Hillsdale, NJ: L. Erlbaum.

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

Kohlberg, L. (1958). *The development of modes of moral thinking and choice in the years 10 to 16*. University of Chicago: Unpublished doctoral dissertation.

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*. Fan Francisco, CA: Harper & Row.

Kohlberg, L. & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development, 12*, 93-120.

Kohlberg, L., Levine, C., & Hewer, A. (1984). The current formulation of the theory. In: L. Kohlberg, *Essays on moral development, Vol. II, The psychology of moral development*, pp. 312-319 Fan 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.

Krebs, D.L., Vermeulen, S.C.A., Carpendale, J.I. & Denton, K. (1991). Structural and situational influences on judgment: The interaction between stage and dilemma. In: W.M. Kurtines & J.L. Gewirtz, eds., *Handbook of moral behavior and development.* Volume 2: Research, pp. 139-169. Hillsdale, NJ: Erlbaum.

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago: The University of Chicago 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. (1985a). *Inhalt und Struktur des moralischen Urteilens*. [Content and structure of moral judgment.] University of Konstanz: Unpublished doctoral dissertation.

Lind, G. (1985b). 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 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*. Cresskill. NJ: Hampton Press, in press.

Lind, G. (2006). The Moral Judgment Test: Comments on Villegas de Posada's Critique. *Psychological Reports, 98*, 580-584.

Lind, G. & Wakenhut, R. (1985). Testing for moral judgment competence. In: Lind, G., Hartmann, H.A. & Wakenhut, R., Eds., *Moral Development and the Social Environment. Studies in the Philosophy and Psychology of Moral Judgment and Education* , pp. 79 - 105. Chicago: Precedent Publishing Inc.

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 persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Milgram, S. (1974). *Obedience to authority. An experimental view*. New York: Tavistock.

Miller, G.A. (1969). *Psychology, the science of mental life*. Harmondsworth: Penguin (Original 1962).

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.

Nunner-Winkler, G. (1996). Moral development. In E.D. Corte & F.E. Weinert (eds.), *International encyclopedia of developmental and instructional psychology*, pp. 222-228.

Penn, W. (1990). Teaching ethics – A direct approach. *Journal of Moral Education, 19*, 124-138.

Piaget, J. (1965). *The moral judgment of the child*. New York: The Free Press. (Originally published 1932).

Piaget, J. (1951). *Play, dreams and imitation in childhood*. New York: Norton.

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). *Intelligenz 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).

Power, F.C., Higgins, C. & Kohlberg, L. (1989). *Lawrence Kohlberg's approach to moral education*. New York: Columbia University Press.

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. (1991). Research on moral development in college students. In: A. Garrod, Ed., *Emerging themes in moral development and moral education*, New York: Columbia University Press.

Rest, J.R., Thoma, S.J., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preferences and stage consistency approaches. *Journal of Educational Psychology, 89*(1), 5-28.

Rest, J.R. & Thoma, S.J. (1985). Relation of moral judgment development to formal education. *Developmental Psychology, 21*, 709-714.

Schillinger-Agati, M. & Lind, G. (12003). *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 judgment? 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.

Snarey, J. (1985). The cross-cultural universality of social-moral development: A critical review of Kohlbergian research, *Psychological Bulletin, 97*, 202-232).

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

Walker, L.J. (1995). Sexism in Kohlberg's moral psychology? W.M. Kurtines & J.L. Gewirtz, eds., *Moral development: An introduction*, pp. 83-107. Needham, MA: Allyn & Bacon.

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

Watson, J.B. (1970). *Behaviorism*. New York: Norton (Orginally published 1924).

Wittgenstein, L. (2002). *Philosophical investigations*. Blackwell Publishers (Originally published 1952)

Wojcik, (1998). Semantics. In: *Grolier Multimedia Encyclopedia*. CD-Edition.

The *Moral Judgment Test* (MJT)

(Last revision of this text: March 4, 2004)

The MJT is designed for research and for the evaluation of programs and policies only. It is **not** designed as an instrument for evaluating people or groups or the use as a high-stakes test. The MJT has been constructed on the basis of Lind´s Dual Aspect Theory of moral judgment and development to assess subjects' moral judgment competence as defined by Lawrence Kohlberg (e.g., 1964, p. 425). More details on the MJT and guidelines to establish cross-cultural validity of translated versions are available from this web-site: *http//www.uni-konstanz.de/ag-moral/.*

In pretest-posttest-studies, test weariness may be a problem, resulting in an unusual lowering of the C-score on the retest. This or a similar instruction helps to avoid this problem: "Some of the questions will be the same as you have been given the first time. We want to know whether your thoughts have changed. Please fill them out as sincerely as you did the first time."

This version of the MJT has been in use since 1977, slightly revised in Dec. 2001, replacing "acceptability" judgments by "acceptance" and "rejection" judgments.

Note: Do <u>not</u> publish this test without written consent by the author.

[29] Author's address: Prof. Dr. Georg Lind, University of Konstanz, Department of Psychology, D-78457 Konstanz, Germany. Fax: +49-7531 882899, Phone: +49-7531 882895. E-mail: *Georg.Lind@uni-konstanz.de*. Web site: *http://www.uni-konstanz.de/ag-moral/*

Only for the project administrator:

1.	Date:	
2.	Name of the project administrator:	
3.	Name of the project :	
4.	Number of the Questionnaire:	

Questionnaire
on Ethical Problems

(The following identification code is only needed in follow-up studies)

This questionnaire study will be repeated. To be able to match parts from the same person, we need an identification. Please do not enter your name anywhere on the questionnaire.
For *identification*, please fill in the following questions instead:

The first two letters of your mother's first name:	___ | ___	(e.g., Jessica: -> "J | E")
The first two letters of your father's first name:	___ | ___	
The last two digits of your house number:	___ | ___	(if it has only one digit, please write a leading "O", e.g. "0 | 2")
The day of your birth (e.g., "0|5" for fifth of...):	___ | ___	

5. Please copy the above numbers to here: ___ | ___ ___ | ___ ___ | ___ ___ | ___
(Personal identification number)

6. If you responded to these questions previously:
Has your address changed in meantime? If YES please give your previous house number: ___ | ___ (last two digits)

1. Workers´ Dilemma

Due to some seemingly unfounded dismissals, some factory workers suspect the managers of eavesdropping on their employees through an intercom and using this information against them. The managers officially and emphatically deny this accusation. The union declares that it will only take steps against the company when proof has been found that confirms these suspicions. Two workers then break into the administrative offices and take tape transcripts that prove the allegation of eavesdropping.

	I strongly *dis*agree / **I strongly agree**
7. Would you disagree or agree with the workers´ behavior?	-3 -2 -1 0 +1 +2 +3

How acceptable do you find the following arguments *in favor* of the two workers´ behavior? Suppose someone argued they were *right* . . .	**I strongly reject** / **I strongly accept**
8. because they didn't cause much damage to the company.	-4 -3 -2 -1 0 +1 +2 +3 +4
9. because due to the company's disregard for the law, the means used by the two workers were permissible to restore law and order.	-4 -3 -2 -1 0 +1 +2 +3 +4
10. because most of the workers would approve of their deed and many of them would be happy about it.	-4 -3 -2 -1 0 +1 +2 +3 +4
11. because trust between people and individual dignity count more than the firm's internal regulations	-4 -3 -2 -1 0 +1 +2 +3 +4
12. because the company had committed an injustice first, the two workers were justified in breaking into the offices	-4 -3 -2 -1 0 +1 +2 +3 +4
13. because the two workers saw no legal means of revealing the company's misuse of confidence, and therefore chose what they considered the lesser evil.	-4 -3 -2 -1 0 +1 +2 +3 +4

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

2. Doctor's Dilemma

A woman had cancer and she had no hope of being saved. She was in terrible pain and so weakened that a large dose of a painkiller such as morphine would have caused her death. During a temporary period of improvement, she begged the doctor to give her enough morphine to kill her. She said she could no longer endure the pain and would be dead in a few weeks anyway. The doctor complied with her wish.

		I strongly *dis*agree						**I strongly agree**
20.	Do you disagree or agree with the doctor's behavior?	-3	-2	-1	0	+1	+2	+3

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

		I strongly reject								**I strongly accept**
21.	because the doctor had to act according to his conscience. The woman's condition justified an exception to the moral obligation to preserve life.	-4	-3	-2	-1	0	+1	+2	+3	+4
22.	because the doctor was the only one who could fulfill the woman's wish; respect for her wish made him act as he did.	-4	-3	-2	-1	0	+1	+2	+3	+4
23.	because the doctor only did what the woman talked him into doing. He need not worry about unpleasant consequences.	-4	-3	-2	-1	0	+1	+2	+3	+4
24.	because the woman would have died anyway and it didn't take much effort for him to give her an overdose of a painkiller.	-4	-3	-2	-1	0	+1	+2	+3	+4
25.	because the doctor didn't really break a law. Nobody could have saved the woman and he only wanted to shorten her suffering.	-4	-3	-2	-1	0	+1	+2	+3	+4
26.	because most of his fellow doctors would presumably have done the same in a similar situation.	-4	-3	-2	-1	0	+1	+2	+3	+4

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

		I strongly reject								**I strongly accept**
27.	because he acted contrary to his colleagues´ convictions. If they are against mercy-killing the doctor shouldn't do it.	-4	-3	-2	-1	0	+1	+2	+3	+4
28.	because one should be able to have complete faith in a doctor's devotion to preserving life even if someone with great pain would rather die.	-4	-3	-2	-1	0	+1	+2	+3	+4
29.	because the protection of life is everyone's highest moral obligation. We have no clear moral criteria for distinguishing between mercy-killing and murder.	-4	-3	-2	-1	0	+1	+2	+3	+4
30.	because the doctor could get himself into much trouble. They have already punished others for doing the same thing.	-4	-3	-2	-1	0	+1	+2	+3	+4
31.	because he could have had it much easier if he had waited and not interfered with the woman's dying.	-4	-3	-2	-1	0	+1	+2	+3	+4
32.	because the doctor broke the law. If one thinks that mercy-killing is illegal, then one should refuse such requests.	-4	-3	-2	-1	0	+1	+2	+3	+4

Thank you!

Appendix: Validated and Certified Foreign Language Versions of the MJT

	Language	(Co-)Authors
1.	Deutsch (Master copy)	Dr. Georg Lind; Item-Reviewers: Tino Bargel, Dr. Rainer Döbert, Michael Hauan Ph.D., Dr. Thomas Krämer-Badoni, Dr. Gertrud Nunner-Winkler, Dr. Roland Wakenhut, Dr. Thomas E.Wren et al. (1977-2002)
2.	Armenian	Mery Avetisyan (in progress)
3.	Basque	Prof. Santiago Palacios Navarro (1982)
4.	Czechian	Dr. Birgita Slovácková (1999)
5.	Chinese	Zhao Zhanqiang M.A. (2004).
6.	Chinese (Taiwan)	Dr. Chi-Ming Lee (2004)
7.	English	Dr. Georg Lind (1984)
	English, additional subtest: Judge Steinberg	Dr. Patricia Bataglia, Marcia Schillinger-Agati & Dr. Georg Lind (2003)
8.	Finnish	Prof. Matti Ýlen (1999)
9.	Flemish (Belgium)	Dr. Bart Duriez & Pieter-Jan De Marez, Catholic University Leuven, Belgium
	Flemisch (Netherlands)	Dr. Michael Gross (1992)
10.	French	Dr. Michael Gross (1992)
11.	Greek	Dr.. Katerina Mouratidou (2002). (provisionally certified)
12.	Hebrew	Dr. Michael Gross (1992)
13.	Hungarian	Dr. Varine Szilagyi Ibolya (1994)
14.	Iranian	Soudabeh Saeidi-Parvaneh, M.A. (2003)
15.	Italian	Prof. Dr. Anna Laura Comunian (1995)
16.	Latvia	Gints Malzubris, M.A. (2002)
17.	Macedonian	Marijana Handziska, M.A. (2001)
18.	Moroccan (Arabic)	Dr. Ahmed Aghbal (2003) (provisionally certified)
19.	Philippine	Jasmine Tuboro, M.A. (2001)
20.	Polish	Aleksandra Cislak, M.A. (2005)
21.	Portuguese (Brazilian)	Dr. Patricia Bataglia (1998)
	Portuguese, additional subtest: Judge Steinberg	Dr. Patricia Bataglia, Marcia Schillinger-Agati, M.A. & Dr. Georg Lind (2003)
22.	Romanian	Tatiana Chicu, M.A., Beatrice Popescu, M.A. & Stefania Puschila, M.A. (2004) (provisionally certified)
23.	Russian	Ilya Krumer, M.A. (2000)
24.	Sinhala (Sri Lanka)	Sanjee Perera, M.A. (2002)
25.	Slovakian	Mgr. Petra Lajciakova (in progress)
26.	Spanish	Dr. José Luis Trechera (1996), revision: Cristina Moreno, R. Hernández (1999)
27.	Tamil (Sri Lanka)	Sanjee Perera, M.A. (2002)
28.	Thai	Prof. Sanguan Lerkiatbundit (2003)
29.	Turkish	Dr. Nermin Ciftci (1996)

Criterion #1: Simplex Structure

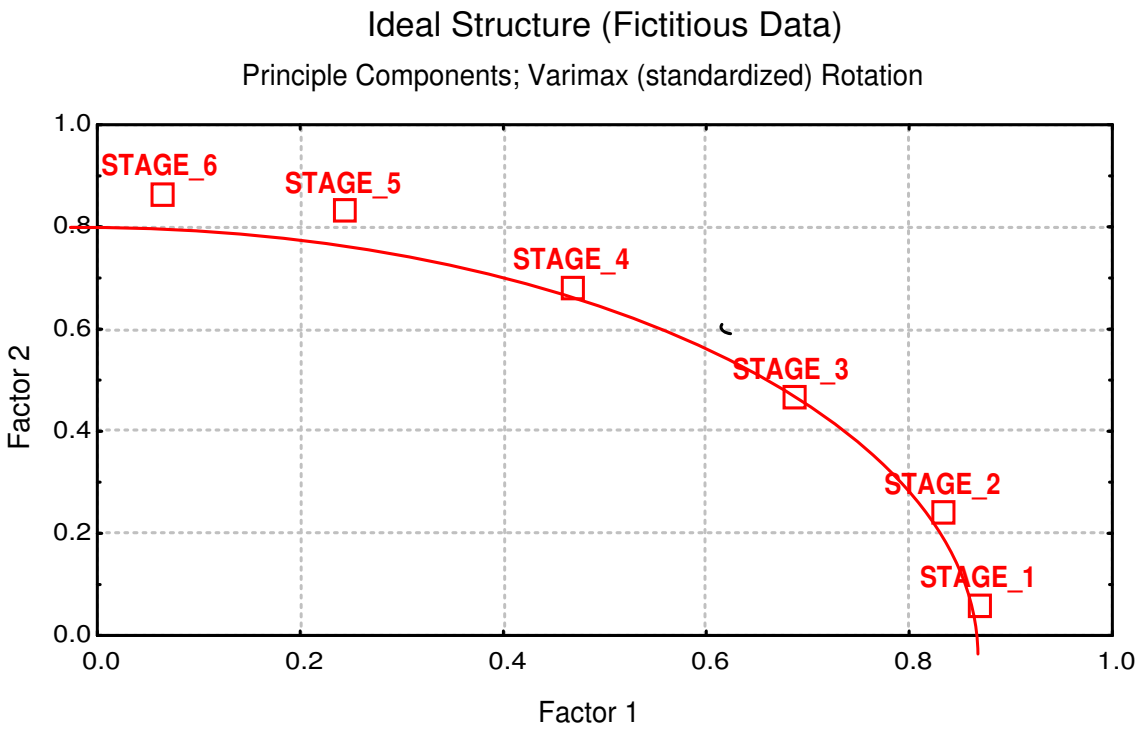


Figure 3

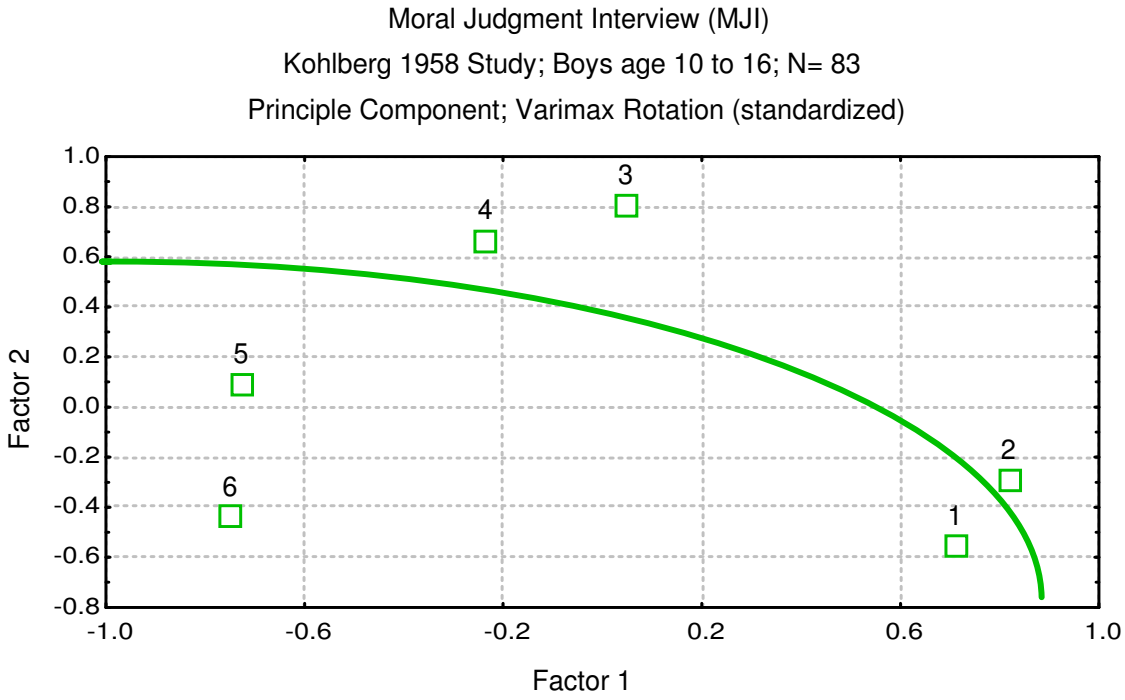
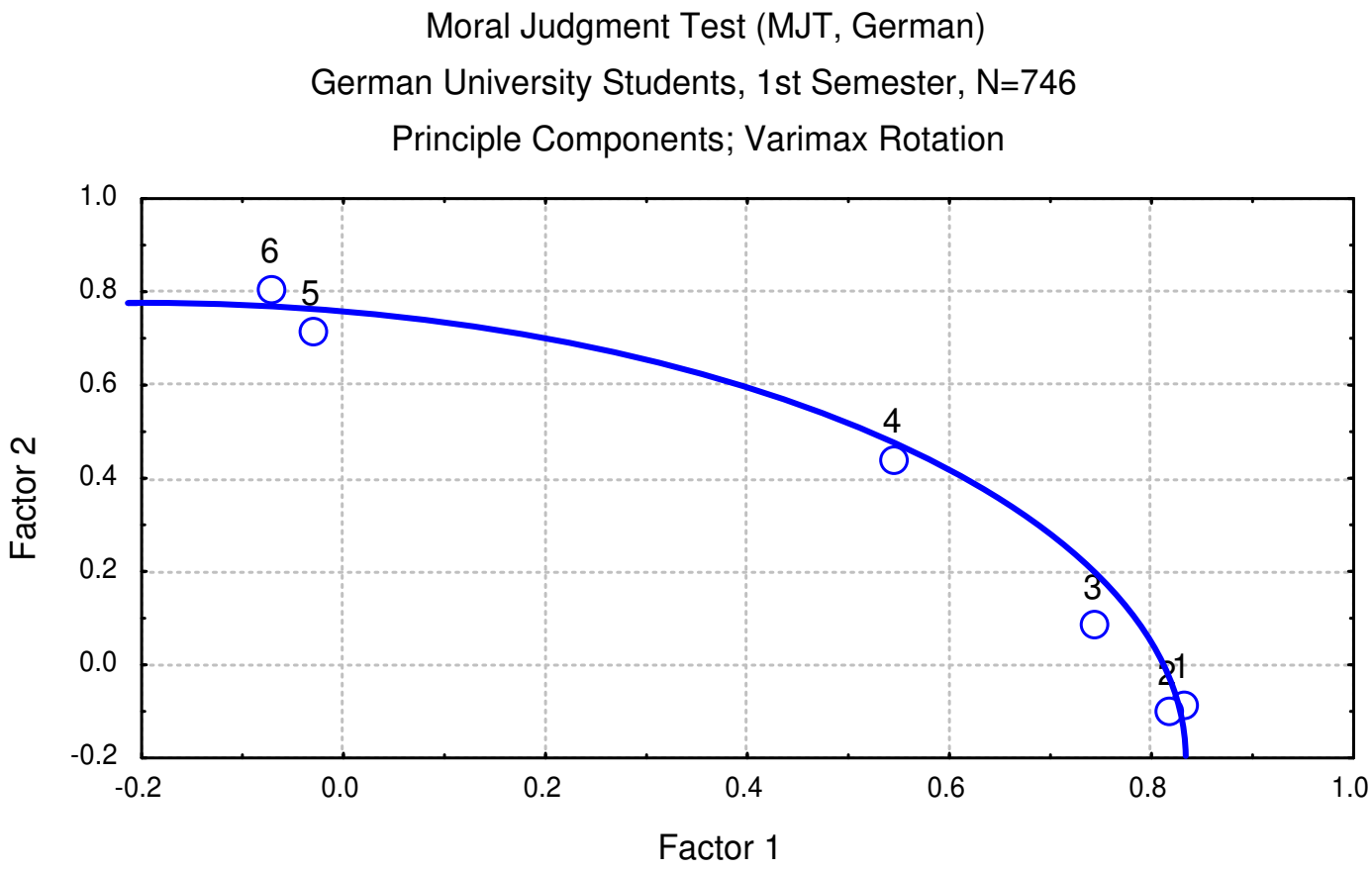


Figure 4

Criterion #1: Simplex Structure (cont'd)



Source: Lind, 2002

Criterion #2: Preference Hierarchy of Kohlbergian Stages of Moral Orientation

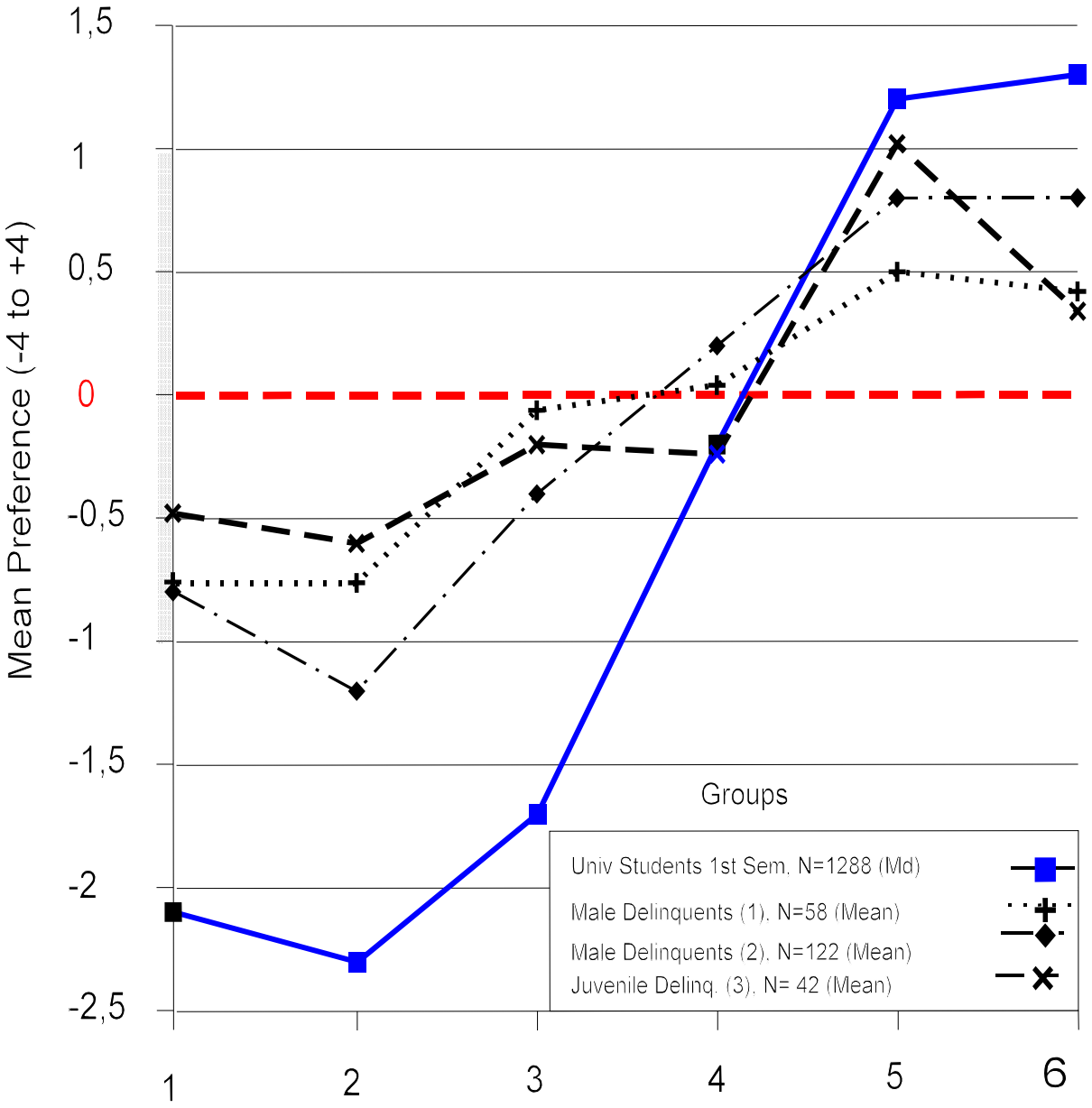
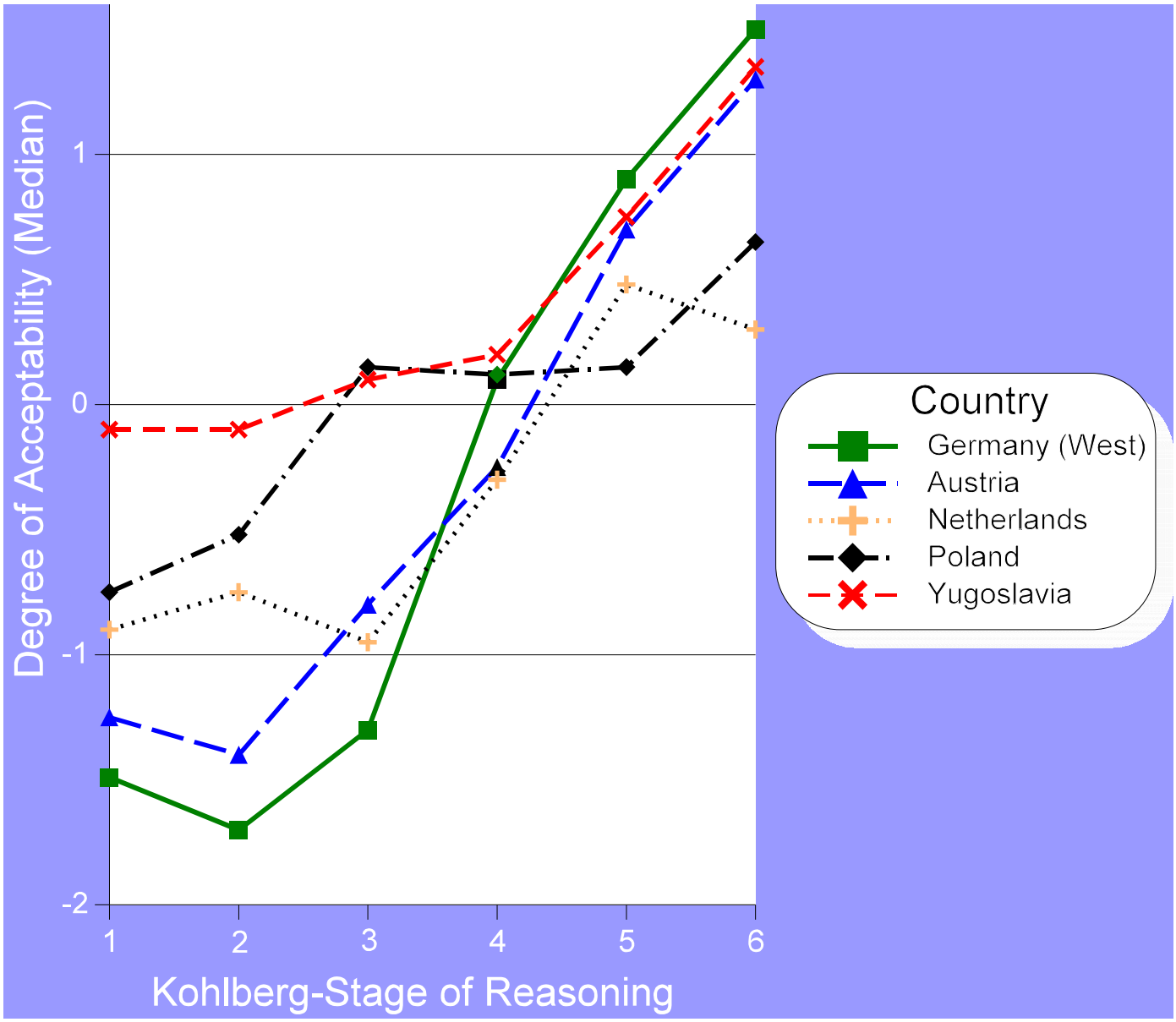


Figure 6

Criterion #2: Preference Hierarchy of Kohlbergian Stages of Moral Orientation



Criterion #3: Affective-Cognitive Parallelism

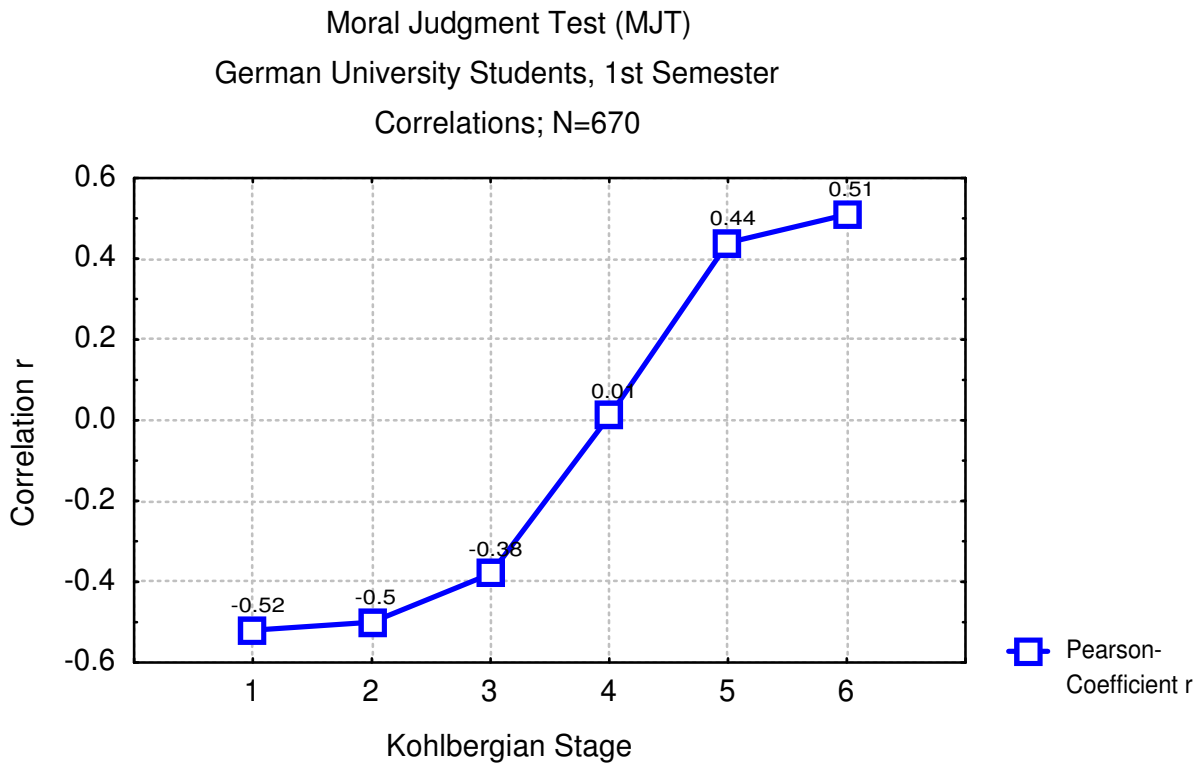


Figure 8

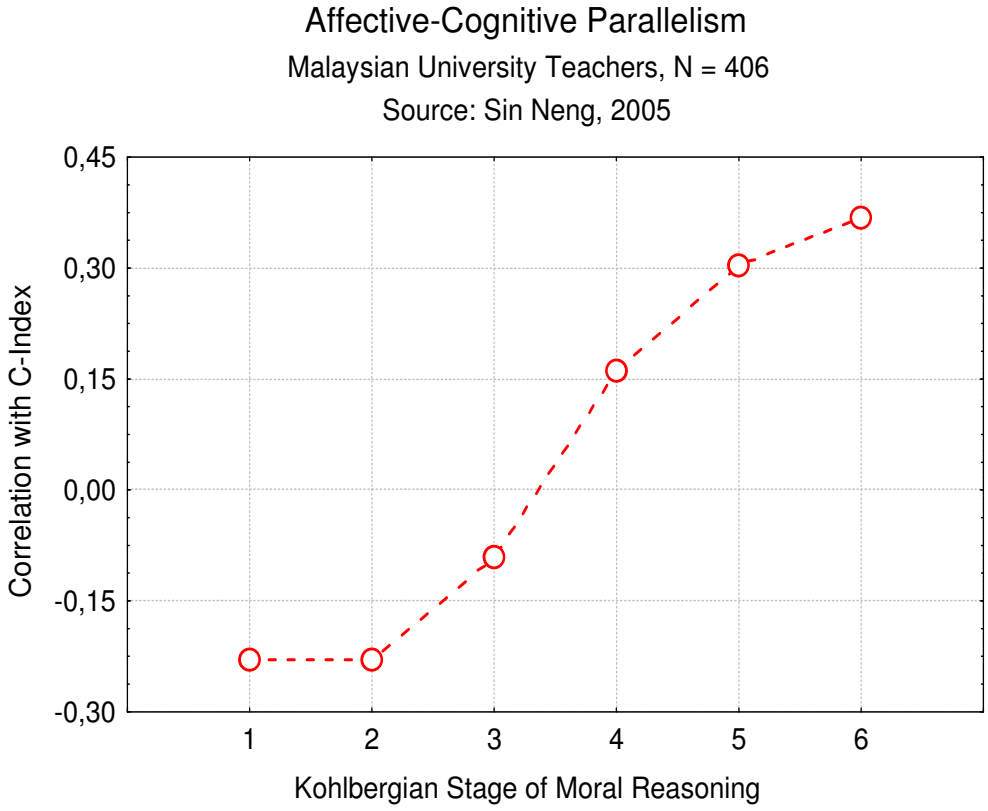


Figure 9

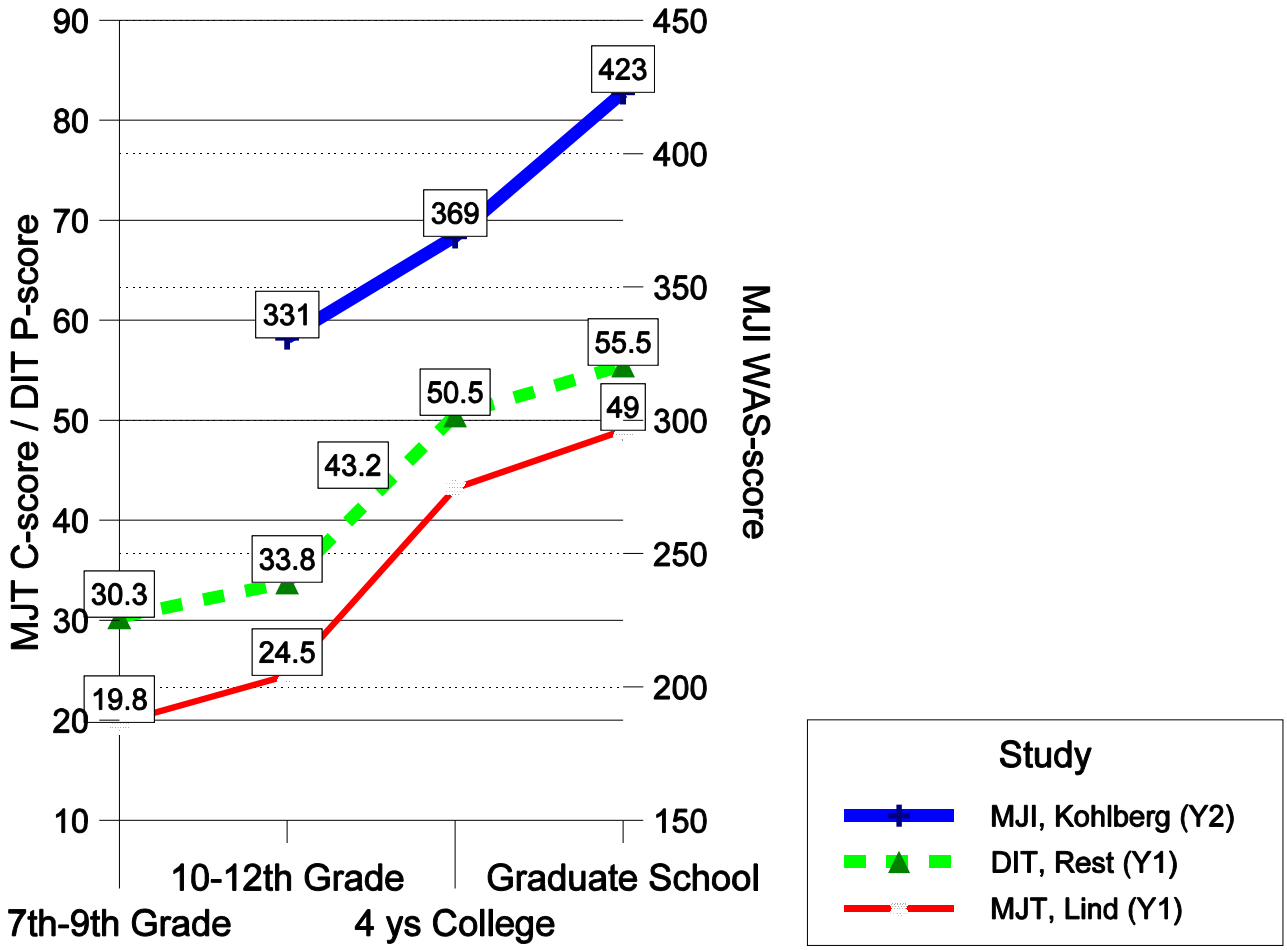

90
80
70
60
50
40
30
20
10
450
400
350
300
250
200
150
MJT C-score / DIT P-score
MJI WAS-score
423
369
331
55.5
50.5
49
43.2
33.8
30.3
24.5
19.8
10-12th Grade
Graduate School
7th-9th Grade
4 ys College
Study
MJI, Kohlberg (Y2)
DIT, Rest (Y1)
MJT, Lind (Y1)

Figure 10

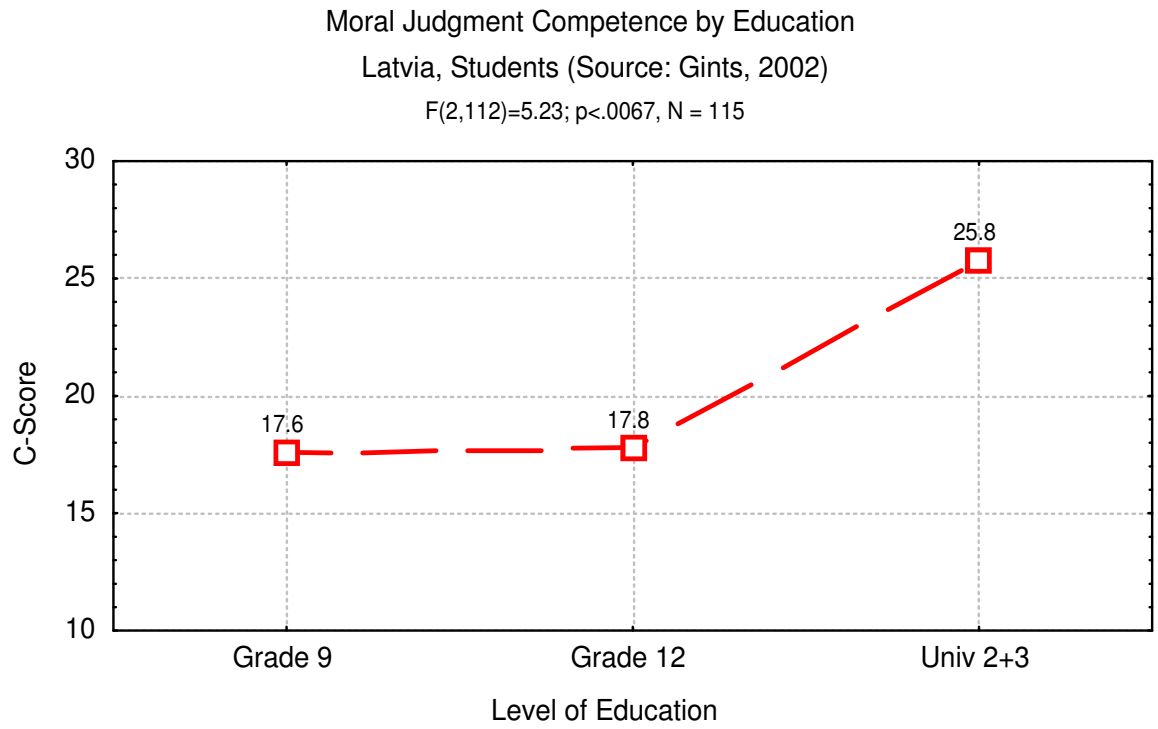


Figure 11

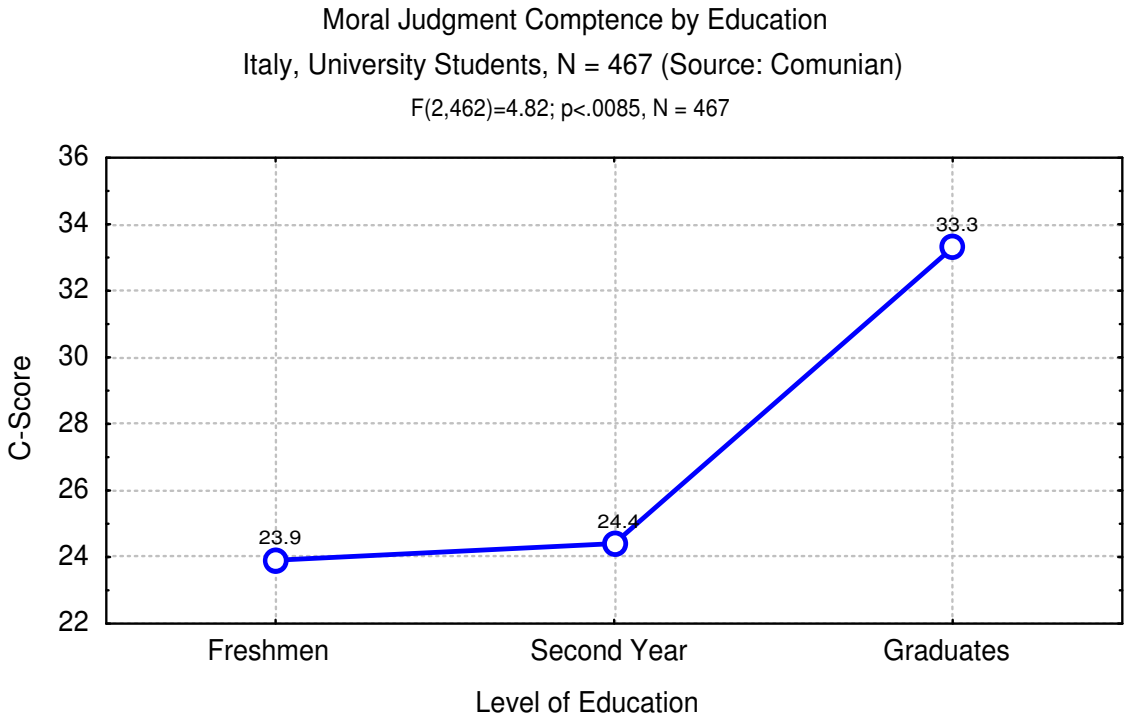


Figure 12

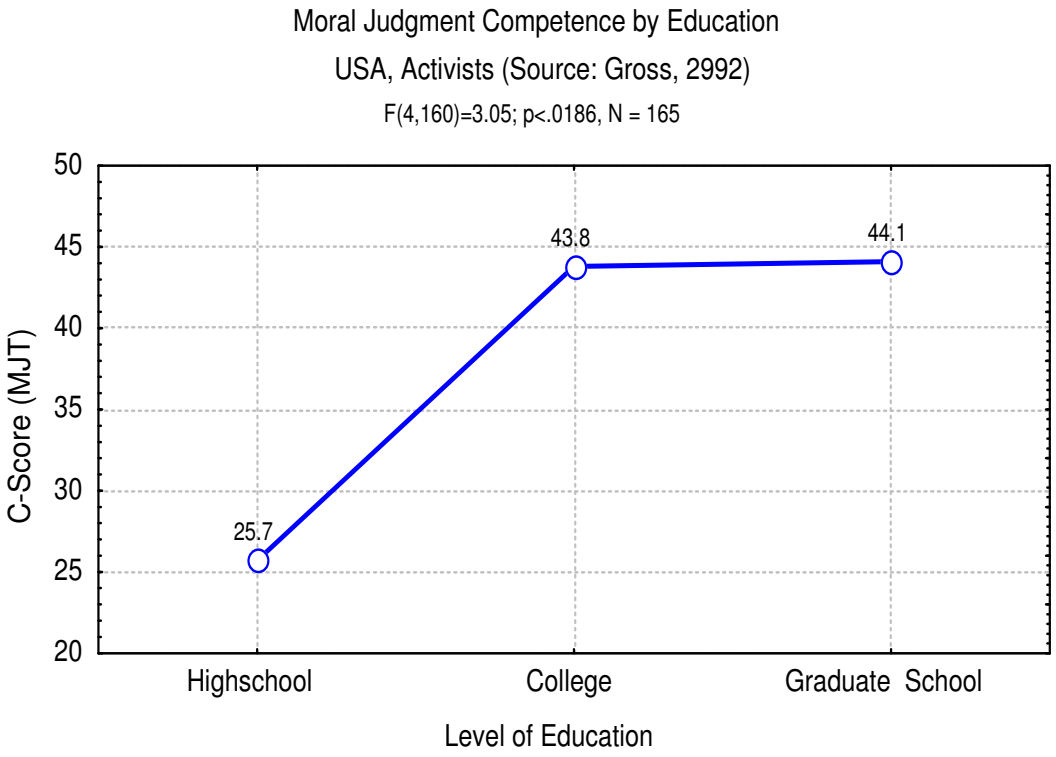


Figure 14

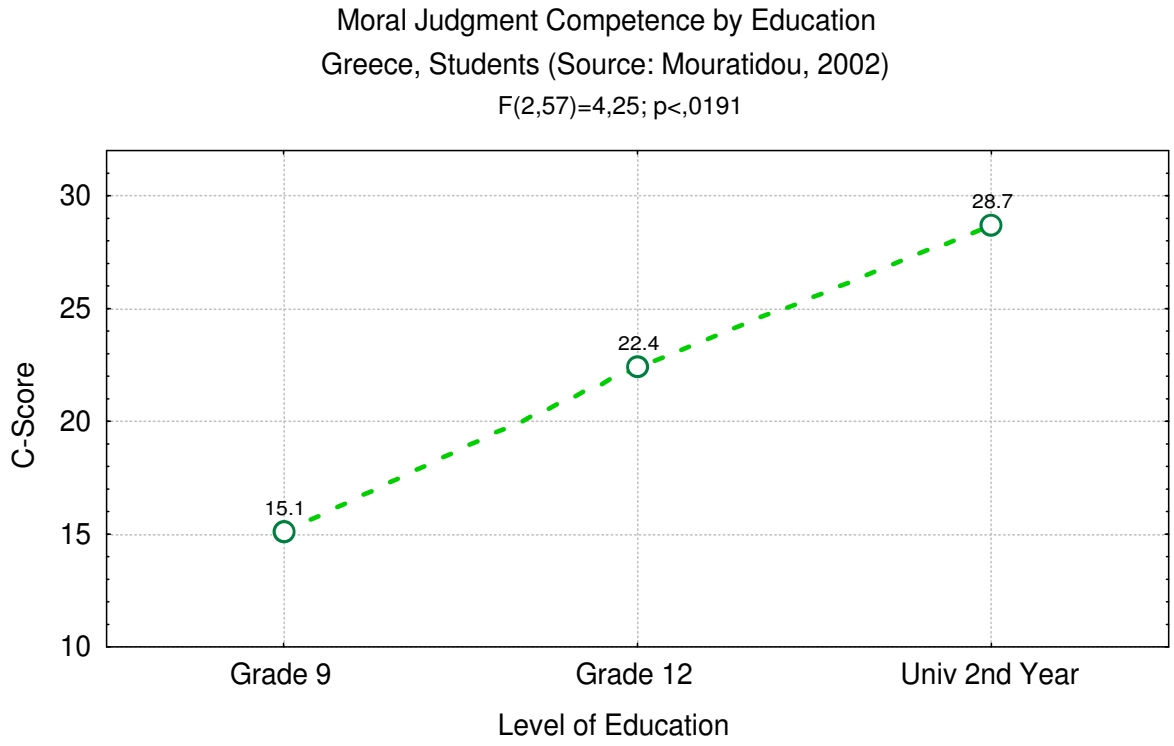


Figure 13

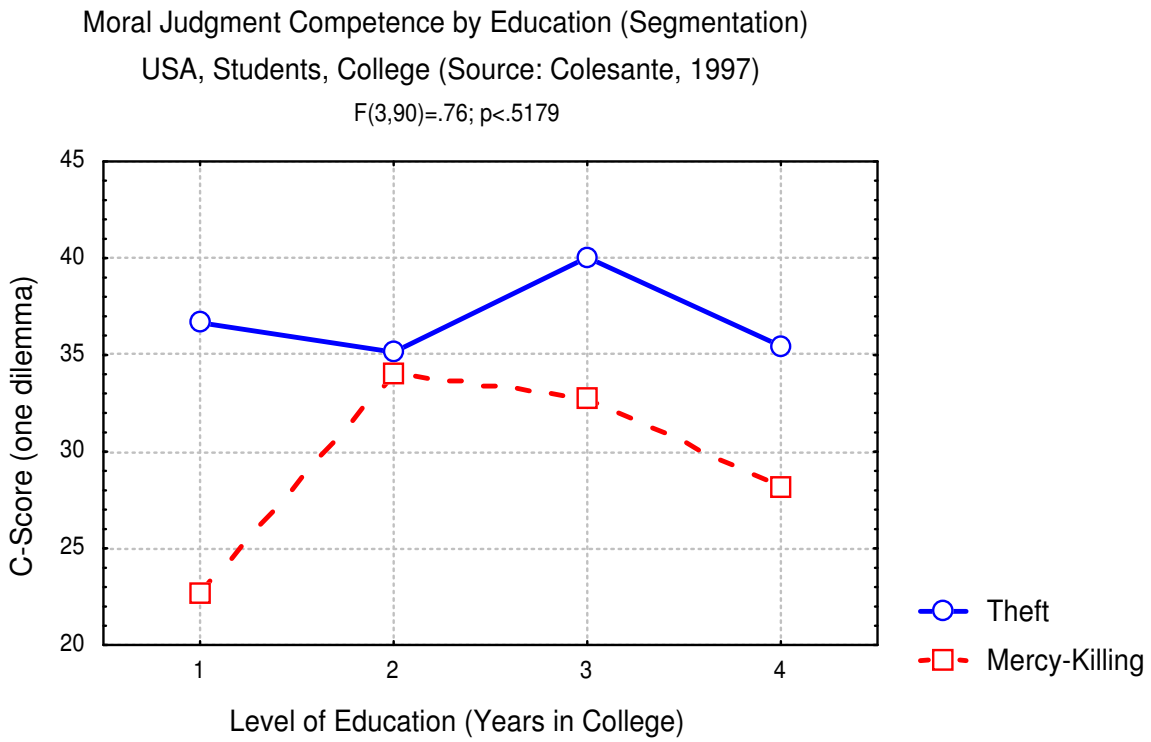


Figure 15

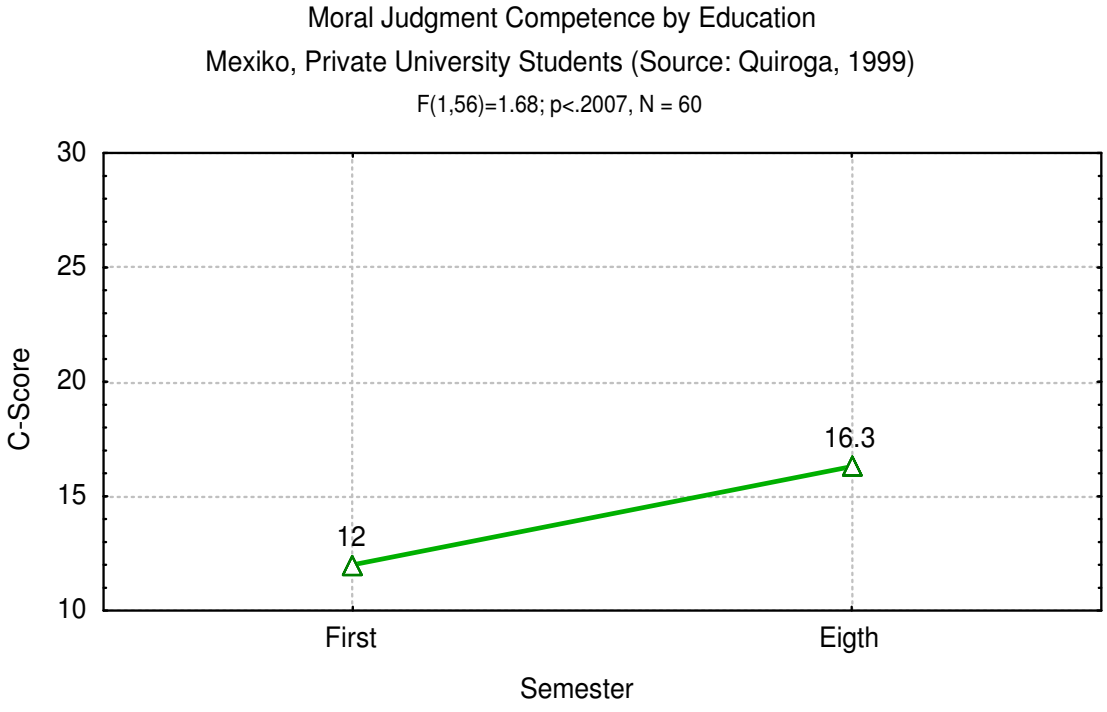


Figure 16

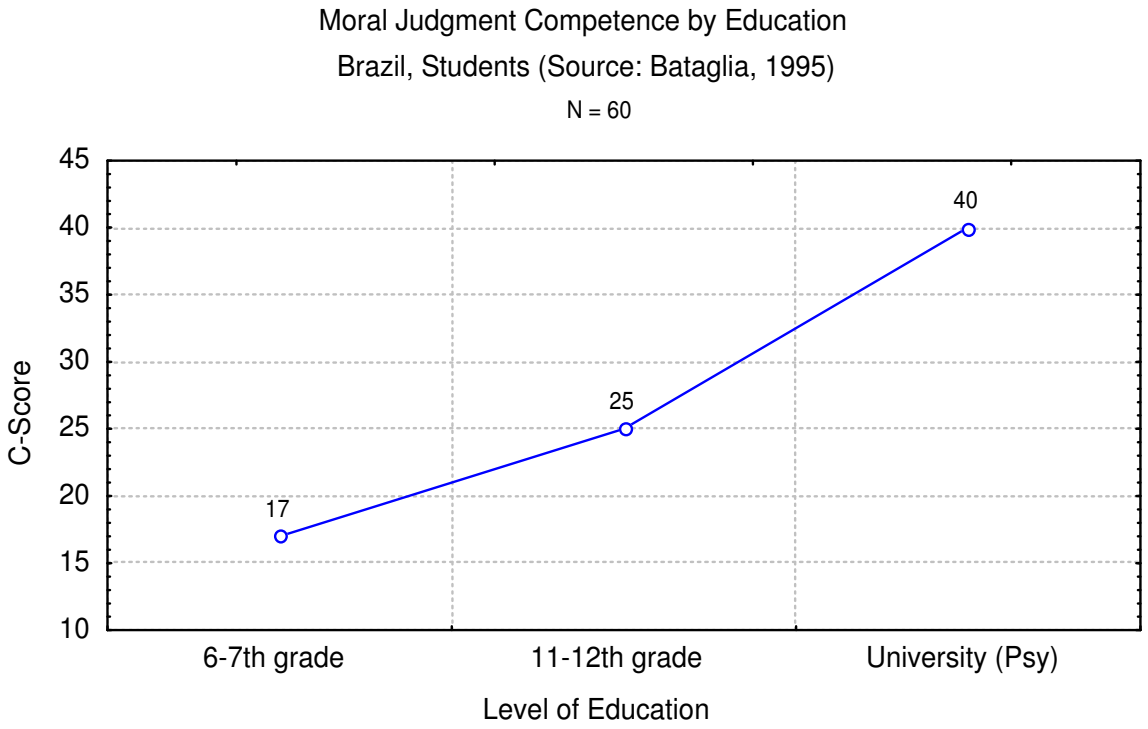


Figure 17

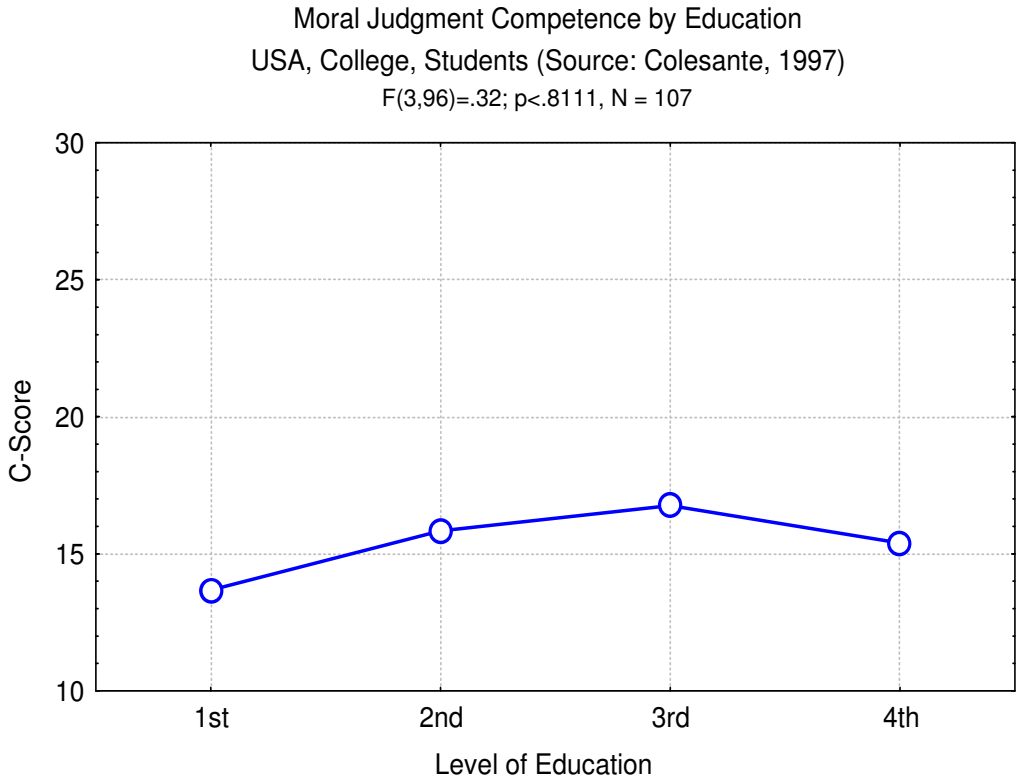


Figure 18