

Empirical Findings on the Cross-Cultural Validity of the Moral Judgment Test (MJT)¹

Georg Lind²

DRAFT (April 24, 2003)

Abstract

The main goal of the present paper is to present data supporting the cross-cultural validity (*semantic* and *semiotic* equivalence) of translated versions of the Moral Judgment Test (MJT) developed by Lind (1978; 1985; 2002a; Lind & Wakenhut, 1985). Cross-cultural validity is defined and judged here on the basis of carefully designed observations in regard to four universalistic criteria, which have been derived from recent research and theorizing about the nature of moral judgment behavior: 1. Quasi-Simplex Structure 2. Hierarchical Preference Order 3. Affective-Cognitive Parallelism and 4. Correlation with education. The MJT has been translated into, and validated for, twenty different languages. To be certified as “culturally equivalent” each translated version had to meet these criteria. The findings will not only tell us something about the cross-cultural validity of this test but also about the universalisability of the underlying theory of moral judgment and behavior (Lind, 2002b).

In concluding, I will touch upon two new developments regarding (1) the use of the MJT in educational evaluation studies (i.e., the problem of test-weariness), and (2) the phenomenon of moral segmentation (that is, the subjects’ tendency to suppress their moral judgment competence in certain settings).

¹ Paper presented at the 2003 Annual Meeting of the American Educational Research Association, April 2003.

² Requests for his paper should be send to: Prof. Dr. Georg Lind, Department of Psychology, University of Konstanz 78457 Konstanz, Germany. Email: georg.lind@uni-konstanz.de <http://www.uni-konstanz.de/ag-moral/>

The Moral Judgment Test (MJT)

The main function of the MJT is to measure what Kohlberg (1964) has called *moral judgment competence*, that is "the capacity to make decisions and judgments which are moral (i.e., based on *internal* principles) and to act in accordance with such judgments" (p. 425; emphasis added). Note that this definition refrains from imposing specific moral values on each and every individual but requires only that each individual – regardless of culture – pursues its moral values in a consistent manner.³

In the *Moral Judgment Test* (MJT), created in the seventies, Lind (1978; 1985; 2002a; Lind & Wakenhut, 1985) we intended to measure to what degree a subjects reveals moral judgment competence when facing a moral task. The main moral task for the subject is to engage in a moral discourse by rating arguments speaking *in favor* and, especially, *against* her or his opinion on a fundamental moral dilemma. First, the subject is asked to give his or her vote on a decision made in a very difficult moral dilemma situation, then the subject is to indicate the degree to which she/he agrees or disagrees with several arguments presented in the test, both in favor and against the subject's opinion. The arguments reflect very different qualities or stages of moral orientation, each representing one of Kohlberg's six stages of moral orientation (Kohlberg, 1984).⁴ It is assumed that this situation elicits emotions, that is, that the subject becomes activated. To rate arguments in regard to one's own moral orientations or principles regardless of one's own opinion on the dilemma, has shown to be a very difficult task (Keasey, 1973). People have a strong tendency to rate others' arguments rather in accordance to their own opinion, agreeing to all arguments supporting their opinion and rejecting all arguments opposing it (Keasey, 1973). One explanation for this is peoples' strong tendency to avoid cognitive dissonance, see, e.g., Festinger (1957). Tests which do not incorporate a difficult moral task,

³ Note that the MJT has been constructed 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/> .

⁴ Kohlberg reduced his six stage-model at some point, but returns to it later again (Kohlberg et al., 1990).

cannot be regarded as tests of moral competence.⁵ Thus calculating a C-index for a test that contains no task, as Rest and his colleagues (Rest et al., 19..) did, is pointless.

The development of this moral competence typically proceeds like this: At the lowest phase, subjects see no qualitative differences in the arguments but rate them only on the basis of their agreement or disagreement with his or her opinion on a particular dilemma solution. First signs of an emerging moral judgment competence can be observed when subjects reject low stage (“bad”) arguments even though they support their opinion. They, so to say, dissociate themselves from “false friends.” Further on, they rate the supporting arguments more and more differentially but reject counter-arguments almost unanimously. In more advanced phases of the development of their moral judgment competence, subjects notice that some counter-arguments are actually quite acceptable even though they oppose their own opinion. At the most developed phase subjects rate counter-arguments as differentially as supporting arguments and do this in a morally consistent way. The C-score, ranging from 0 to 100 reflects this development numerically.

The theoretical claim that the C-score reflects a competence was subsequently tested empirically in a two experiments and longitudinal and intervention studies using highly rigorous criteria (Lind, 2002a; 2002b). It has been shown that the C-index cannot be simulated upward in experimental situations in which other test scores can (Emler et al., 1983), that is, that the MJT really seems to be a difficult task for the subjects (Lind, 2002a).

The MJT’s C-score has been designed to be *culturally fair*. In contrast to most, if not all other tests of moral development, the test’s index for moral judgment *competence*, the C-index, is not tied to (though based upon) a subject’s moral orientation but solely on his or her ability to apply his or her own moral orientation consistently to his or her judgment behavior.⁶ That is, the subject need not to subscribe to particular moral orientations but, as Kohlberg’s (1964) definition of this construct implies, needs only to apply his or her moral orientations consistently, to get a high C-score. Empirically, this

⁵ 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).

⁶ In some literature, the C-index is wrongly translated as “consistency”-index. The statistical formula which generates the C-index involves a certain kind of consistency measurement, namely consistency *in regard* to the subject’s preferred stage of moral orientation, however, it should be clear that moral judgment competence cannot be identified with mere “consistency” which could also mean moral “rigidity,” which is certainly not a sign of moral competence.

construct distinction hardly ever occurs as both aspects of moral judgment behavior correlate highly positive. Yet, only through making this clear distinction, we can observe this Affective-Cognitive Parallelism, and can notice when the parallelism is absent or breaks down. I will come back to this below. Through this feature, when using the MJT in cross-cultural research, we do not impose a certain way of defining the dilemmas in the MJT onto the subject. Thus it is more fair regarding individual and, of course, *cultural* differences to define the moral implications of a situation (Lind, 1995).

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 four 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. These – very rigorous – criteria are used in addition to more traditional strategies for securing cross-cultural validity (see, e.g., Gielen et al., 1996; Edwards, 1981).

Semantic and Semiotic Threats to Communicative and Cultural Validity

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 semiotic validity if we want to be sure that there is maximum communicative validity or, as Campbell (1963) has called it, "conceptual overlap."

By *semantic*⁷ 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-reading the items and by having six noted experts in the area of moral development research to critically examine each item.⁸

In the process of cross-cultural validation of the MJT, semantic validity was sought to be established through 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 *semiotic*⁹ 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 misscored. As Campbell (1963) argues this problem exists even in situations where the instructions and stimuli are very simple and hardly prone to semantic ambiguities, even 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. A single argument, taken by itself, Kohlberg (1958; 1981) argues, 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*.

⁷ 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).

⁸ I wish to thank all experts involved in this: Tino Bargel, Rainer Döbert, Thomas Krämer-Badoni, Gertrud Nunner-Winkler, Gerhard Portele, and Roland Wakenhut.

⁹ "From the Greek word semeion, meaning 'sign,' semiotics is the science of signs. It stresses that 'meaning,' even when it seems natural or inherent, is always the result of social conventions; also, it analyzes culture as a series of sign systems" (Culler, 1998).

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

The latter method is also used to check on the semiotic 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.

Invariant Universals as Criteria for Semiotic Equivalence and Cross-cultural Validity of the MJT

Usually, structural properties of a test or questionnaire are checked by a-theoretical criteria like “internal consistency” (or Cronbach’s alpha) or the numbers of dimensions produced by factor analysis. In contrast, the MJT’s construct validity is checked by structural analyses derived from theoretical knowledge.¹⁰

Like other methods for assessing moral development, it is based on universalistic moral philosophies (Barber, 1992; Küng, 1991; Annan, 1999; Habermas, 1981; Kohlberg, 1984; Kohlberg & Higgins, 1984), which claim to be universally valid and applicable everywhere and with every person, and thus not only allow but suggest comparative studies across various countries and cultures. For such a use, it is a critical question whether or not the method of observation and measurement is cross-culturally valid in the sense that the semiotic structure of adapted versions of this method

¹⁰ 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.

produces semiotically equivalent responses (Colby et al., 1987; Dana, 2000; Eckensberger, 1991, 1996; Gielen et al., 1986; Lorenço, 1996; Rest et al., 1999; Sullivan, 1977).

Universalistic claims are based on various philosophical sources including Kant's *Categorical Imperative* (1785), which maintains that behavior should be guided only by maxims which represent a general and universal moral law. The idea of universalism, in terms of moral principles that "prescribe human obligations" and rationality, supposes the possibility of an agreement on a moral point of view, without the need of achieving an agreement upon the "content of substantial principles of morality."

People from different cultures as well as people inside the same culture have different moral standards. However, universalistic moral philosophies assume that there are some common moral principles that are considered valid among various cultures and provide the basis for developing culturally fair measures of moral judgment competence. Universal citizenship, democratic principles and accountability can also be viewed independent of culture (Barber, 1992). As Walker (1995) points out, a mature moral decision making process demands notions of impartiality and universality. In line with Russel's definition of "intelligent thought" (1959), we understand the complete moral competence concept as the ability of arguing, enabling discussions, showing respect for different points of views and traditions, learning from others as pointed out by the "Dialogue of Civilizations and the Need for a World Ethic" (Annan, 1999) and as supported by the Universal Declaration for Human Rights (United Nations, 1948).

Apart from this general sense of a core set of universally valid moral principles, we must be aware of the fact that, when translated into every-day decision making they can mean quite different things for people from different cultural backgrounds. Whereas in some cultures a specific situation might be viewed as invoking high moral principles, in other cultures they may not, or less so. For example, the choice between eating beef or pork is not seen as a moral problem by many of us, yet a Muslim will regard it as such. Note that in this case, Muslims and Non-Muslims clearly understand the words "beef" and "pork" in an identical way, so that there is not semantic problem involved in this communication. Yet, the moral (and hence semiotic) meanings of these two things are very different.

So when we observe how people deal with such situations in order to make inferences on their moral dispositions, we must make sure that the semiotic (not only the semantic) meaning of this situation is that which we (the researchers) think it is. Alike, if we compare observations in different cultures using the same experimental situation to elicit reactions, we must make sure that the semiotic

meaning structure is similar in different cultures. This cannot be checked on merely by looking at the semantic structure of the test (e.g., by back-translations). We must also look at the pattern of responses and correlation between responses.

Four Criteria for Semiotic or Structural Validity

Progress of measurement and progress of knowledge in a specific area of scientific inquiry usually depend on each other. While knowledge hardly progresses without advancements in measurement technology, the latter is itself very dependent on new gains in substantial knowledge about the objects that are to be measured. This is not only true for the natural sciences but also for the study of moral development and education. Kohlberg (1981) called this a “bootstrapping” process, in which the validation of moral development theory goes hand in hand with the validation of the instruments with which various aspects of moral development are measured.

In one phase of a science, the results of measurements are regarded as criteria for the truth or validity of a theory, while in another phase, well-established knowledge on some aspects of moral judgment and development can be taken as criteria for judging the validity of measurement instruments. Of course, during this bootstrapping process, great care must be taken that the findings does not result in tautologies, which immunizes theory and measurement against any criticism and, hence, against scientific progress.¹¹ When taken as criteria for testing the validity of an instrument, the theoretical knowledge used in this process must be developed, and must be consistently validated through research, *beforehand*. Nevertheless, positive findings in validation studies can also be regarded as supporting the theory, on which the measurement is based.

Our general idea then was to use universal constants of moral judgment behavior as criteria for testing the validity and the cross-cultural equivalence of the Moral Judgment Test. As these findings seem to be fairly well established we do not expect any substantial differences to show up in cross-cultural studies and hence can attribute any actual difference found to a lack of cross-cultural validity.

¹¹ Such a tautology or saving circularity would be to construct items for a test of moral development in a way which maximizes correlation between the test and, lets say, age. Then the test would be biased to confirm a test-age-relationship. Unfortunately, this is not an uncommon practice in psychometrics. With the MJT, we carefully avoided to use any criterion for validation which would be involved in the testing of substantial hypotheses in moral development research.

Four universal constants have emerged from empirical research over the past four decades:

1. Quasi-Simplex Structure of ordered stage correlations.
2. Hierarchical Preference Orders for the Kohlbergian stages.
3. Affective-Cognitive Parallelism.
4. Correlation of moral judgment competence with quantity and quality of education.

In the following, I will discuss these constants and describe the validation criteria which have resulted from them (for a more extensive discussion see Lind, 1978; 1985; 2002a; 2000b).

Criterion # 1: Quasi-Simplex Structure of ordered stage correlations

The validation criterion of “Quasi-Simplex Structure” refers to a special way that ordered variables may intercorrelate: 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 between them, . . . Such a pattern in quantitative data is called a 'simplex' by Guttman" (pp. 82-83).¹²

"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)

¹² Quasi-simplex is a simplex in which sizable error measurements are allowed (perfect simplex is reasonable only if measurement errors are negligible. I prefer the weaker prediction, though one could also argue otherwise. I wish to thank Debbie D. Reese for this clarification.

To my knowledge, [there is no statistical coefficient available to measure the rightness of this prediction with observation data](#). So we must, like a medical doctor who looks at an X-ray picture for signs of some disease, make a judgment based on visual inspection of the correlation matrices obtained in various studies.

Yet, we found two ways to assist this visual inspection with the help of statistical procedures. First, we used a method for reordering the correlations in order to maximize the Quasi-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 an indication of lack of test validity. However, this method is rather crude and insensitive. Now, 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. We predict, if the test is fully valid, 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 $4 \times 3 \times 2 \times 1 = 24$ ways in which the stages can be ordered. Hence, the probability that they are ordered in the predicted way by chance is $p = 1/24 = 0,042$.

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 | - |
| O | 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 usage of a stage in an interview 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 Quasi-Simplex-Structure from fictitious data (Figure 1) shows that they fit well though not perfectly. Studies using the MJT show a better fit to this criterion (for an example, see Figure 3).

Criterion # 2: Hierarchical Preference Order for the Kohlbergian stages

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.¹³ The subject is given arguments that resemble each of these stages, one argument pro and one speaking against the particular decision made in each of the two dilemma situations, 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.¹⁴

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 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.

¹³ See Kohlberg, Boyed & Levine, 1990.

¹⁴ I wish to thank Dr. Michael Huan, University of Missouri, for convincing me on this issue.

Criterion # 3: Affective-Cognitive Parallelism

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:

“Affective life, similar to intellectual life, is continuous adaptation, and both of these adaptations are not only parallel but interdependent, since sentiments express the interests and values of actions, intelligence constitutes the structure [. . .] Affectivity regulates the energetic aspect of action, of which intelligence provides the structure.” (Piaget, 1951, p. 220 - 221)

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

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.’ 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).

Criterion # 4: Correlation between moral judgment competence with quantity and quality of education.

For many decades, moral development tests have been validated by correlating its scores with the age of the subjects. A highly positive correlation with age was seen as the major indication of theoretical validity because it was believed that moral competencies develop in an invariant, always upward, sequence but never regress (Kohlberg, 1958; 1984; Rest, 1979).

However, age seems to be no longer a valid criterion for validating moral development tests because it is clear by now that moral competencies can regress, and that their development is fueled by quality and quantity of education rather than mere biological aging (Lind, 2002a; Rest, 1991). The illusion of age-correlation seems to have been caused by the fact that the first longitudinal and cross-sectional studies have been using only youth who attended school, college or university, so that the variable age and education were always confounded. Only when we studied youth who did not enter an academic career but entered the workforce at an early age, we could see that moral competence mostly erodes, which means that age by itself did *not* foster moral judgment competence. It seems that up until a certain stage of development, educational input is a necessary condition for moral growth. In most, if not all studies, the correlations between level of education and moral development have been higher than all other variables studied (Rest, 1979; 1991; Lind, 2002b).

However, this statement applies only to studies in which there was a substantial variation in the level of education and in studies done in countries where, presumably, education is of good quality. Therefore, we must be cautious to generalize it to all studies. Rather we must say that we expect the correlation between education and moral judgment competence to be high a) if the education is of good quality (i.e., that it is really designed to foster moral and democratic abilities rather than drill-skills), and b) if there is substantial variation in regard to educational experience in the studied sample.

Indeed, data from three longitudinal studies using three different methods of observation (Kohlberg’s Moral Judgment Interview, Rest’s Defining Issues Test, and Lind’s Moral Judgment Test) converge on supporting the hypothesis of a high correlation between moral development and level of

education. Because we now have studies showing clear regressions (Lind, 2002b), we can safely rule out that this correlation is spurious. Age is not a predictor of moral judgment competence unless it is associated with more and better education.

This finding helps us not only to set up a forth criterion for cross-cultural validity but also helps us to design a proper validation sample. In all four cases, we can expect the data to confirm to our criteria only if the sample contains a substantial variation of moral judgment competence. Without any variance of the scores we cannot get any correlation at all. So, to make sure that the validation sample is appropriate for testing the cross-cultural validity of the MJT, we mandate that the validation sample consists of three sub-samples representing three different levels of education, with the level about two years of education apart. In most cases, it seems, this guarantees that the variation of moral judgment competence in the whole sample is large enough for any correlation to show up if it is there (for more details on the mandated design of validation studies for the MJT, see <http://www.uni-konstanz.de/ag-moral>).

Findings from Cross-cultural Validation Studies

The MJT has been translated into many languages. So far, twenty language versions of the MJT have been thoroughly validated using the above described criteria and certified as cross-culturally valid: German, English, Dutch, Finn, Flemish, French, Hungarian, Hebrew, Italian, Lithuanian, Spanish, Portuguese, Russian, Singhalese, Tamil, Turkish. More translations and validations are in progress (e.g., in Arabic, Chinese and Japanese). A current list of translated and certified versions and of translations in progress can be found on this web-site: <http://www.uni-konstanz.de/ag-moral>).

The certification of each translated version has been based on specially designed validation studies as explained above. In each case, the findings and in many cases even the raw data were submitted for analysis and reviewed by the original author of the MJT. 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. In all instances, these hints helped the author(s) of the foreign language versions to identify the source of invalidity and improve the test. In some cases, technical flaws (like mistakes regarding data entry and processing or scoring)

were detected which accounted for a lack of validity. Only when the findings were showing a perfect or near-to-perfect fit, the new version was certified as cross-culturally valid.

The question we have asked is: Are these different language versions, which are used in very diverse cultures, cross-culturally valid? Is their semiotic meaning fully or only partially represented or not equivalent at all? In other words, can findings generated by these different versions of the MJT in different cultural settings be directly compared in any way or do we need to regard each version as idiosyncratic and only comparable with each culture?

Well, it seems the answer to this question is Yes. Already the sheer fact that it was possible to certify all translated versions as cross-culturally valid either on the first attempt or after some technical errors had been corrected and the semiotic meaning of problematic items have been adjusted, makes me believe that these test versions are cross-culturally valid – and that the criteria used are indeed universal constants. Note that each of the four criteria is rigorous in an unprecedented way. Taken together, these criteria are extremely unlikely to be met by chance alone. Because they address different aspects of the semiotic meaning structure of moral judgment behavior, they are very sensitive to any lack of construct validity, much more sensitive than most conventional psychometric criteria are. Moreover, they made it possible for an independent reviewer without any knowledge of the target language, to closely analyze the cross-cultural validity of translated versions of the MJT.

In all cultures in which the MJT has been used for research the four criteria have been very useful for detecting instances of *semiotic invalidity*, and also a number of merely technical errors in the validation studies. 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 (which I strongly recommend against!), 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.” The 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.

So, at least after the second round of validation study, all twenty versions agreed with the four validation criteria and could be certified as cross-culturally valid. The agreement with the first three criteria was perfect or nearly perfect in all cases. The agreement with the fourth criterion, correlation with education, was less satisfactory than we expected, but in no case was there a complete disagreement after all technical errors had been corrected and all semiotic problems had been solved.

Note that in no case was a test or a specific test item to be dismissed or revised because of not fulfilling the fourth criterion, as some critiques of the MJT suspect. This criterion, like the other three ones, rather serves as a “search light” for identifying obvious errors which impede semantic and semiotic validity. Moreover, the first three criteria of validity, as has been shown, do *not* include any substantial hypothesis which is to be studied by using the MJT, only the criterion of education-correlation of moral judgment competence does. Yet, in no case, the MJT was “optimized” to produce high correlations, and thus can be regarded as unbiased in relation to any hypothesis regarding the impact of education on moral development.

Our 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 correlations of a test score with certain criteria like stage invariance, age, education or political attitude. Such validation methods would make the test unsuitable for testing the very hypothesis anymore.¹⁵

The fact that four universal constants of moral judgment behavior hold true does not at all mean, that the level of moral judgment competence is the same in all countries. On the contrary, the C-index varies greatly between these cultures, and these variations can be regarded as valid signs of different levels of moral judgment competence, due to the fact that the meaning structure of the test has shown to be invariant across these cultures.

It should be noted however, that quality and quantity of education is indeed a very powerful determinant of moral development and that we must not disregard this in our interpretation of so-called cultural differences. Quantity and especially quality of education varies greatly between the cultures that have been studied (cf. Colesante & Biggs, 2003; Lind, 1986; Schillinger-Agati & Lind, 2003).

¹⁵ 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.

As the data shows (see Figures 8 through 14), the correlations vary strongly from country to country, [indicating that this 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 steady 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.

Maintaining validity in research and evaluation studies

Typically, we are not interested merely in the validity of a test but in the validity of test scores from a particular research or evaluation study. This broader scope of our interest necessitates also a broader view on the validity issue, including the purpose of the study, the conditions of taking the test, the problems involved in re-taking the test and the question of cultural norms.

The purpose of the study can very much affect the validity of the scores of the MJT. The MJT has been constructed to answer an important question of research or evaluation. This implies a) that the purpose of the study is explained to the subjects, e.g., that the focus is on testing a certain hypothesis or evaluating a certain educational method, but not her or him or some group of people, b) that the test is applied anonymously and no reference, whatsoever, can be made to the subject's test scores, and c) that the subject is told to respond to the questions of the test sincerely and carefully in order to obtain truthful data.

The conditions of the test-taking have to correspond to this purpose: Under disadvantageous conditions, the problem of faking test scores upward is minimal though they may cause selective drop-outs of the lesser capable subjects and thus inflate the mean test scores of the sample studied. The

problem of deflated test scores is more serious when the MJT is taken under the pressure of time or emotional stress. Therefore, the MJT should not be applied in connection with the evaluation of persons or groups of persons nor should it be given in a way that the subjects feel under the pressure of time. The subjects should be told that they can take their time (but should not feel obliged to spend too much time on each answer), and, when being part of a test battery, the MJT should not be given at the end, when many subjects feel under pressure to speed up their responses. When employing others persons to apply the test, make sure that they do not “explain” to the subjects that the test is to “measure your morality” or make other comments which makes the subjects suspect such a purpose.

When applying the MJT a second time (e.g., when using it in an evaluation study with a pretest-posttest-design), the problem is not, as we tended to believe, that the scores go up due to practice or social desirability¹⁶ but, as we learned, that the scores go down due to hidden messages that can undermine the validity of the test scores. If subjects come to believe a) that the re-testing is to check on their credibility and consistency of responding or b) that the researcher does mindless research, many or most subjects may not fill out the test as sincerely and carefully as is needed to produce valid test scores. Before we became aware of this problem, we observed that especially the subjects with high pretest score on the MJT tended to get very low scores on the posttest, thus masking the gains due to the educational intervention which we studied. Therefore, it is necessary to explain to the subjects the reasons for presenting the identical test a second time and to ask them to fill in the test as sincerely as the first time.

In some cultures, religious norms prevent subjects to deliberate about certain moral problems on which the religious authority has a firm judgment about. In these cultures, the subjects are not only told to agree to the way a moral dilemma is to be decided but also not to reason about this, not even affirmatively. For example, some churches or religious organizations firmly reject abortion and mercy killing, and also discourage reasoning about these issues as this may cause confusion or dissent in the believers. In several MJT studies in Latin America, where many people submit to the authority of the Roman Catholic church, subjects not only rejected mercy killing but also refused to reason about this

¹⁶ Contrary to what some textbook say, there is little evidence that a single test-taking can foster competencies so much to threaten the validity of a test. Bode and Page (1978) found that taking Kohlberg's *Moral Judgment Interview* twice resulted in somewhat lower rather than higher scores (the scores decreased 9 MMS points or 1,5 percent of the scale range). However, the validity of tests of moral attitudes may indeed suffer. In the same study, Bode and Page (1978) found that the Defining-Issues-Test scores went up 4,7 P-points, that is, 4,5 percent of the scale range within a short time range by merely applying the test twice without any intervention.

problem. So their C-scores on the mercy-killing dilemma of the MJT were very low, much lower than their C-scores on the so-called workers dilemma, which was not affected by church ruling (Bataglia et al., 2002; Lind, 2000; Schillinger-Agati, 2003). How to deal with this very recent finding, is still being discussed. On the one hand, these findings can be seen as indicating true lack of moral judgment competence regardless of what causes it. On the other hand, one may regard the restraining influence of certain religious norms as not indicating a lack of moral judgment competence but rather a legitimate way of adapting to certain cultures, that is, as *moral segmentation*. Obviously, this is not merely a questions of methodology and measurement but of highly philosophical issue. As long as this issues is not dissolved, I recommend to supplement the analysis of moral judgement behavior by a separate analysis of the C-scores for each dilemma, and by an additional dilemma which does not invoke religious norms. Presently, Patricia Bataglia and her colleagues are presently testing such an additional dilemma for the MJT.

Conclusion and Outlook

In an unprecedented effort, to secure cross-cultural validity, translated versions of the Moral Judgment Test have been submitted to rigorous validation studies using four empirical criteria derived from moral psychological research: 1. Quasi-Simplex Structure 2. Hierarchical Preference Order, 3. Affective-Cognitive Parallelism, and 4. Correlation with education. In all instances it was possible to produce a highly valid and equivalent version, either on the first attempt or after some revisions necessitated by an empirically driven, independent reviewing process based on these four theoretical criteria.

This finding makes us confident that differences in regard to the moral judgment competence found in various cultures in studies using the MJT cannot be discounted as measurement errors or lack of semiotic equivalence of the different test versions. As judged by the four rigorous criteria set up in this study, the validated and certified MJT-versions have shown to be highly equivalent in respect to their semantic and semiotic structure.

Moreover, it should be noted that hardly any non-scorable tests were 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 was well received in all countries/cultures. This contrasts favorably with DIT

studies in various cultures that have often produced many (up to 50 percent) non-scorable cases (see, e.g., Gielen et al., 1986).

The confirmation of cross-cultural validity of the MJT in so many instances supports also the claim that not only the test is valid but also the underlying dual-aspect theory of moral judgment behavior (Lind, 2002). There are universals on which we can rely, most importantly the consistently found Hierarchical Preference Order for the Kohlbergian stages of moral reasoning. There seems to be a universal preference for post-conventional moral principles – independent of culture.

In the course of validating the MJT in various cultures, an unexpected finding was made which cannot be discussed here at length also it certainly is of immediate relevance to the topic of cross-cultural validity, namely the finding that in some cultures, the moral judgment competence as revealed in the MJT seems to depend very much on the kind of dilemma presented. We call this phenomenon ‘moral segmentation’ as the subjects showed high moral judgment competence in one situation but not in the other (see 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, which allows us to detect and study it in detail (see Figure 14, which gives insights into the dynamics of students’ moral-cognitive development during college. Therefore, I strongly recommend for research and evaluation studies not only to look at the overall C-score but also analyze the C-scores for each dilemma in order to check on the possibility of ‘segmentation,’ especially when the score appear unusually low.

One last finding regarding the validity of the MJT deserves mentioning here, which comes from the use of the MJT in evaluation studies. Like other competence tests, the MJT is highly sensitive to certain test taking conditions that restrict the subjects, like time pressure, test-taking fatigue, or mistrust. For example, we found that the mean C-score of a sizable proportion of subjects sharply dropped from pretest to post-test. Interestingly, this drop was mostly confined to subjects with initially rather high scores. Was this what statisticians call ‘regression to the mean?’ Obviously not, because only about one fourth of the sample showed this drop and the drop was much more pronounced than the increase by the other three quarters of the sample.

After some experimenting we discovered that this group of ‘droppers’ was very critical of the second test-taking. Some of them said they felt double-checked and mistrusted because they had to fill out the same test twice. So we now announce to the subjects at the port-test that they will go to fill out exactly the same test, and ask them to do this as sincere as they did the first time. We have no

signs of test weariness anymore. The effect sizes of our moral education programs have went up tremendously as the overall mean scores are not deflated anymore by some subjects whose scores fell off on the posttest.

What do our findings mean for education? I think, two implications are obvious. First, the MJT can indeed be used to compare the effects of educational methods and of aspects of 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” values, but rather a need to teach children 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).

References

- Annan, K. (1999). *Dialogue of civilizations and the need of a world ethic*. Lecture at the Center for Islamic Studies, Oxford, UK, June 28th, 1999.
- Barber, B.R. (1992). *An aristocracy of everyone. The politics of education and the future of America*. New York: Ballantine Books.
- 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.
- Blatt, M. & Kohlberg, L. (1985). The effect of classroom moral discussion upon children's level of moral judgment. *Journal of Moral Education*, 4, 129-161.
- Bode, J. & Page, R. (1978). Comparison of measures of moral judgment. *Psychological Reports*, 43, 307-312.
- 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.
- 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.
- Eckensberger, L.H. (1991). Moralische Urteile als handlungsleitende soziale Regelsysteme im Spiegel der kulturvergleichenden Forschung. In: A. Thomas, Ed., *Einführung in die kulturvergleichende Psychologie*, pp. 259-295. Göttingen: Hogrefe.
- Eckensberger, L.H. & Zimba, R.F. (1996). *The development of moral judgment*. In P. Dasen & T.S. Saraswathi, eds., *Handbook of cross-cultural psychology*, Vol. 3. Boston, MA: Allyn & Bacon.
- 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., 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.
- 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. (1981). *Theorie des kommunikativen Handelns*. [Theory of communicative action.] Frankfurt: Suhrkamp.
- Keasey, C.B. (1973). Experimentally induced changes in moral opinions and reasoning. *Journal of Personality and Social Psychology*, 26, 30-38.
- 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. (1976). Moral stages and moralization: The cognitive- developmental approach. T. Lickona, Ed., *Moral development and behavior: Theory, research and social issues*, pp. 31-53. New York: Holt, Rinehart & Winston.
- Kohlberg, L. (1981). *The meaning and measurement of moral development*. The Heinz Werner Lecture Series, vol. 13. Worcester, MA: Clark University Press.
- Kohlberg, L. (1984). *Essays on moral development, Vol. II, The psychology of moral development*. San Francisco, CA: Harper & Row.
- Kohlberg, L., Boyd, D., & Levine, C. (1990). The return of Stage 6: Its principle and moral point of view. In: T.E. Wren, (Ed.), *The moral domain: Essays in the ongoing discussion between philosophy and the social sciences*, pp. 151-181. Cambridge, MA: MIT Press.
- Küng, H. (1991). *Global Responsibility. In Search of a New World Ethic*. New York: Crossroads (Original: Projekt Weltethos. München: Piper, 1990).
- 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. (1985). *Inhalt und Struktur des moralischen Urteilens*. [Content and structure of moral judgment.] University of Konstanz: Unpublished doctoral dissertation.
- Lind, G. (1985a). The theory of moral-cognitive judgment: A socio-psychological assessment. In: G. Lind, H.A. Hartmann & R. Wakenhut, Eds., *Moral development and the social environment. Studies in the philosophy and psychology of moral judgment and education*, pp. 21-53. Chicago: 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). *Off limits – A cross-cultural study on possible causes of segmentation of moral judgment competence*. Paper presented at the for the annual scientific meeting of the American Educational Research Association, April 24-28, 2000, New Orleans, LA.
- Lind, G. (2002a). 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. (2002b). *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung* (Can morality be taught? Research findings from modern moral psychology). Berlin: Logos-Verlag.
- 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.
- Lourenco, O. (1996). Reflections on narrative approaches to moral development. *Human Development*, 39, 83-99.
- Nagl, W. & Walter, H.G., Eds. (1986). *Konstanzer Statistisches Analyse System KOSTAS. Bericht 37*. Zentrum I Bildungsforschung. University of Konstanz, Germany.
- Piaget, J. (1965). *The moral judgment of the child* (Original 1932). New York: The Free Press.
- 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.
- Piaget, J. & Inhelder, B. (1969). *The psychology of the child*. Basic Books (Original 1966).
- 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. (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., Turiel, E. & Kohlberg, L. (1969). Level of moral development as a determinant of preference and comprehension of moral judgment made by others. *Journal of Personality*, 37, 225-252.
- Russel, B. (1959). Philosophy and politics. E. Egnor & L.E. Dennon, eds., *The basic writings of Bertrand Russell*, pp. 454-467. New York: Simon & Schuster, Inc.
- Schillinger-Agati, M. & Lind, G. (2003). *Moral judgment competence in Brazilian and German university students*. Paper presented at the AERA conference in Chicago, April 21st - 25th, 2003.
- Snarey, J. (1985). The cross-cultural universality of social-moral development: A critical review of Kohlbergian research, *Psychological Bulletin*, 97, 202-232).
- Sullivan, E.V. (1977). A study of Kohlberg's structural theory of moral development: A critique of liberal social science ideology. *Human Development*, 20, 352-376.
- Trow, M. (1976). Higher education and moral development. *AAUP Bulletin* 4, 227-238.
- United Nations (1948). *Universal Declaration of Human Rights*. New York.
- 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.
- Wojcik, (1998). Semantics. In: *Grolier Multimedia Encyclopedia*. CD-Edition.

Criterion #1: Quasi-Simplex Structure

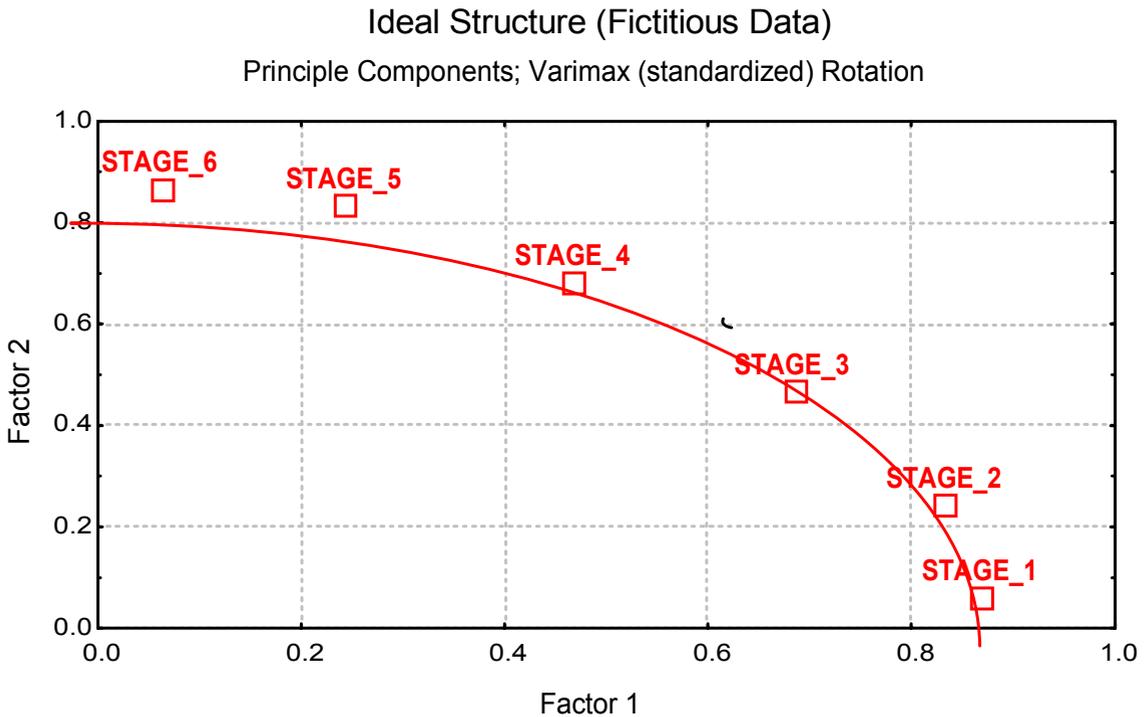


Figure 1

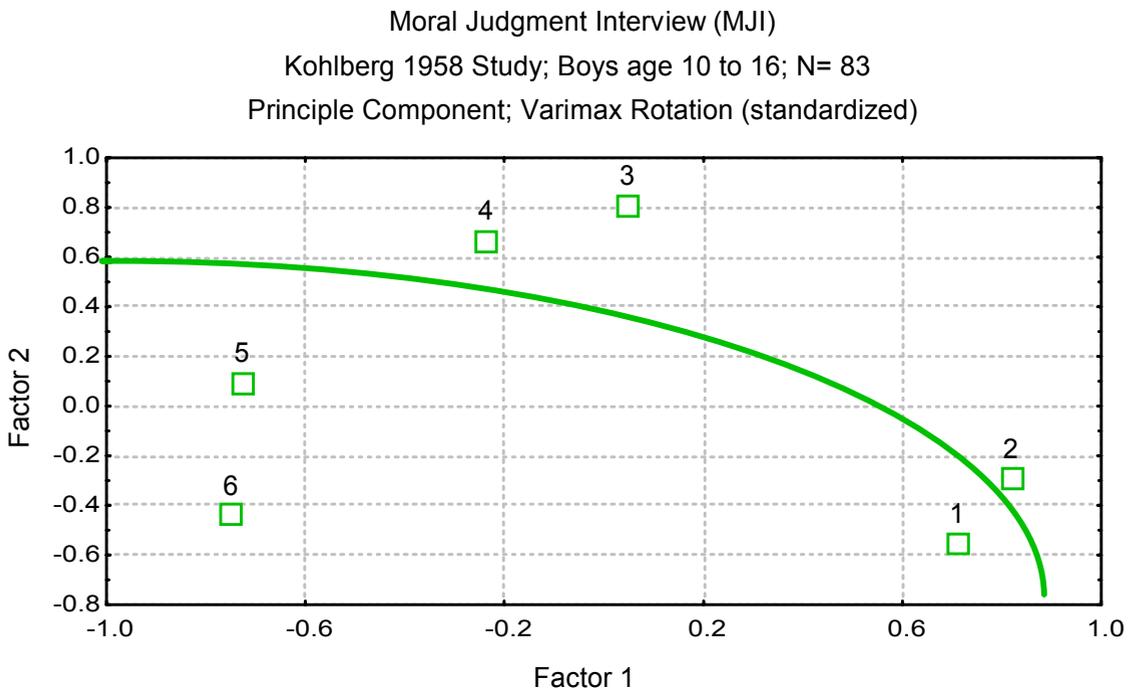
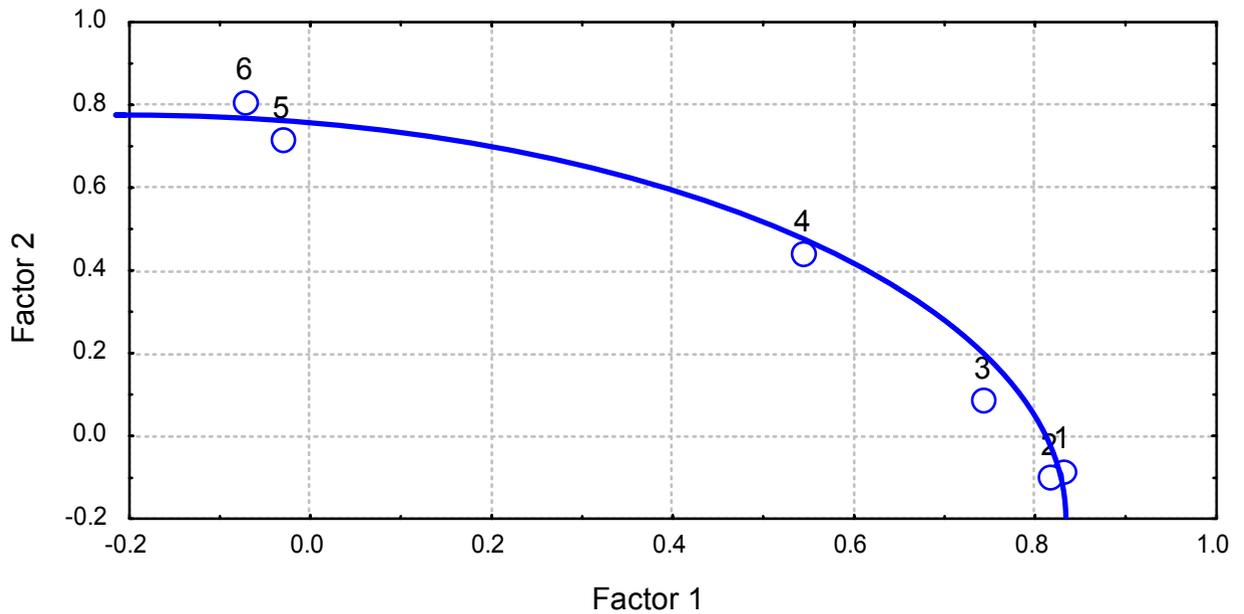


Figure 2

Criterion #1: Quasi-Simplex Structure (cont'd)

Moral Judgment Test (MJT, German)
German University Students, 1st Semester, N=746
Principle Components; Varimax Rotation



Source: Lind, 2002

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

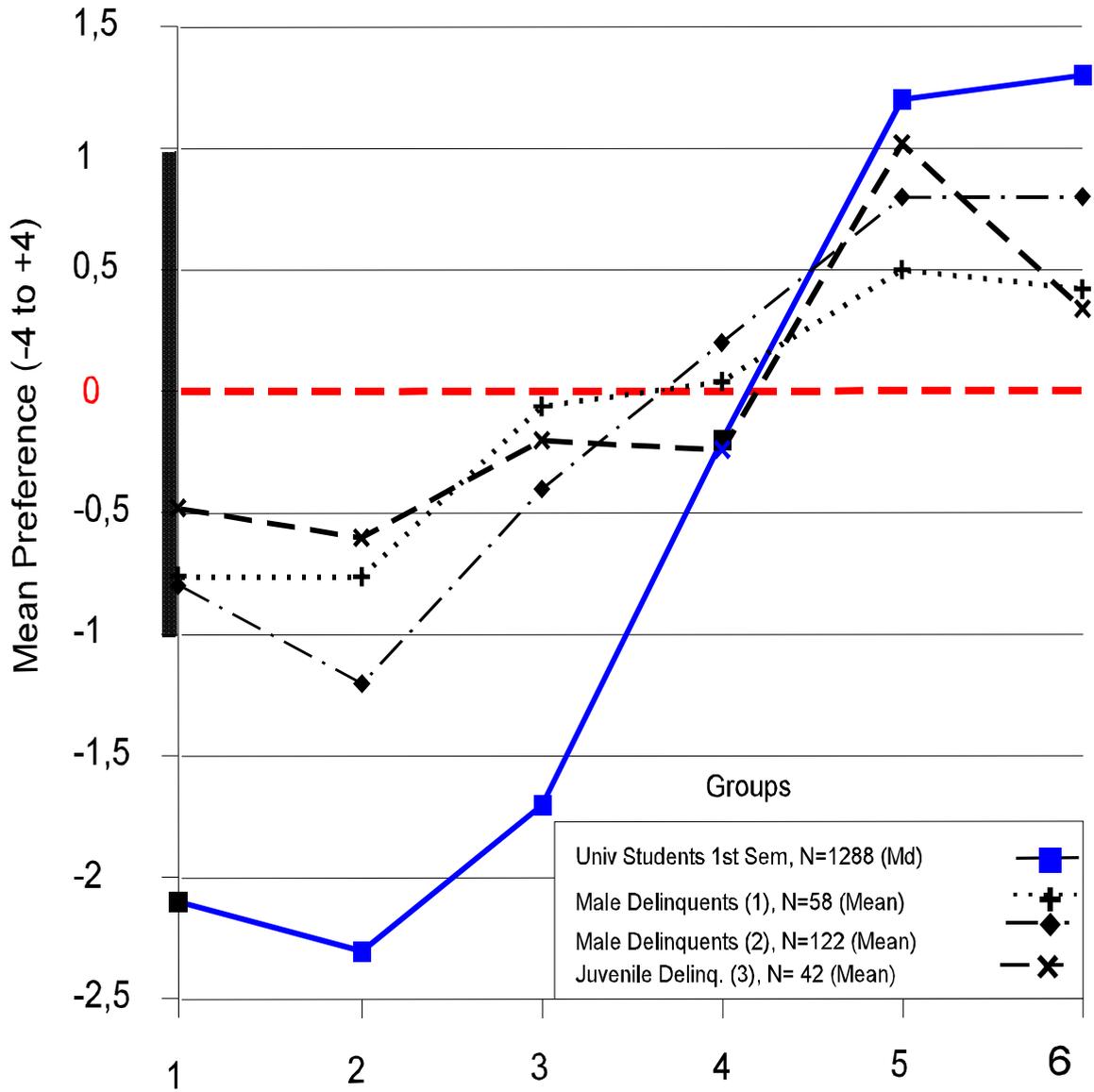


Figure 4

Criterion #3: Affective-Cognitive Parallelism

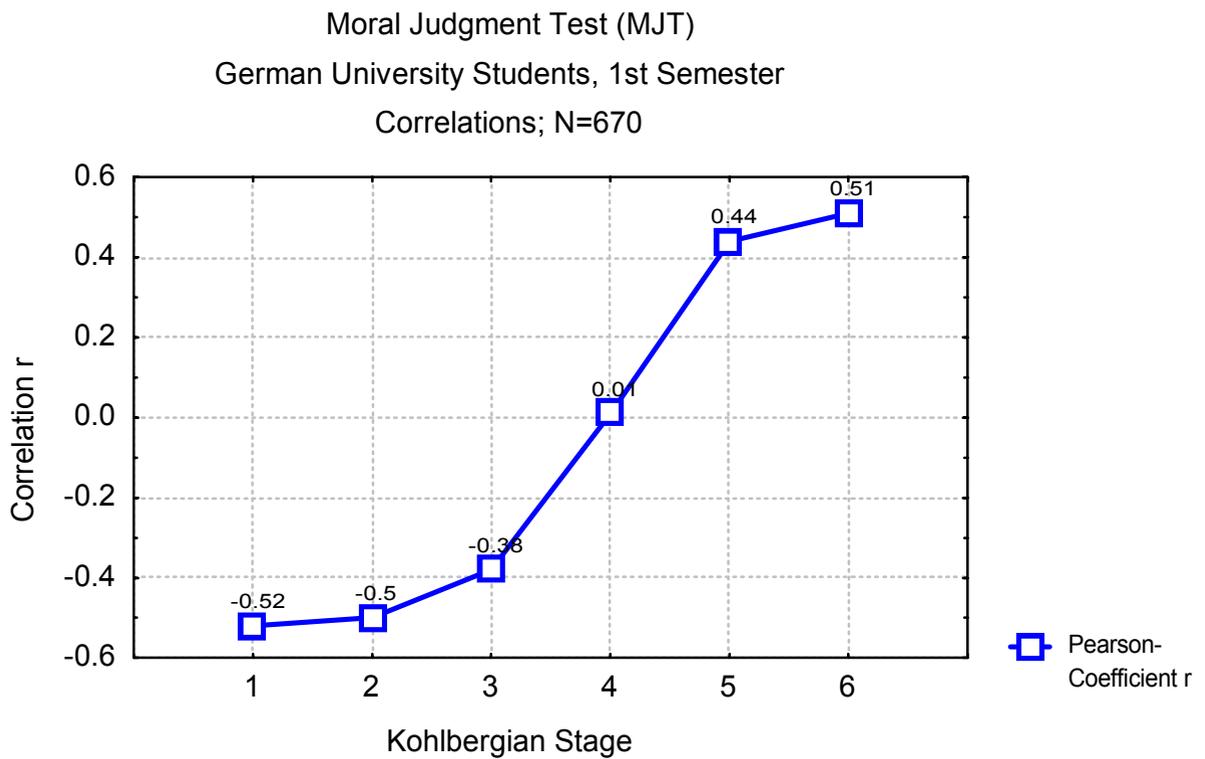


Figure 5

Criterion #4: Substantial Positive Correlation between Moral Judgment Competence and Quantity/Quality of Education – Evidence from Three Longitudinal Studies

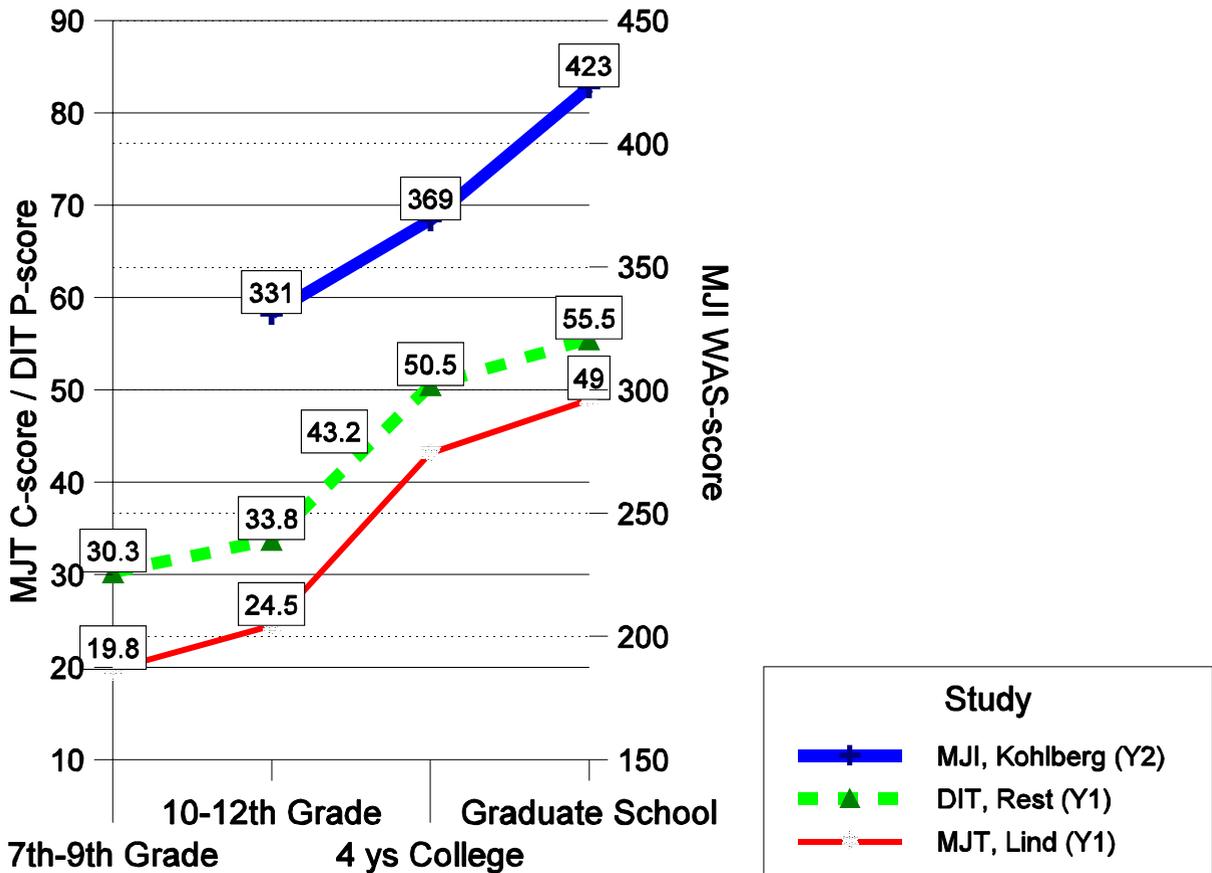


Figure 6

Criterion #4 (cont'd): Educational Correlation...

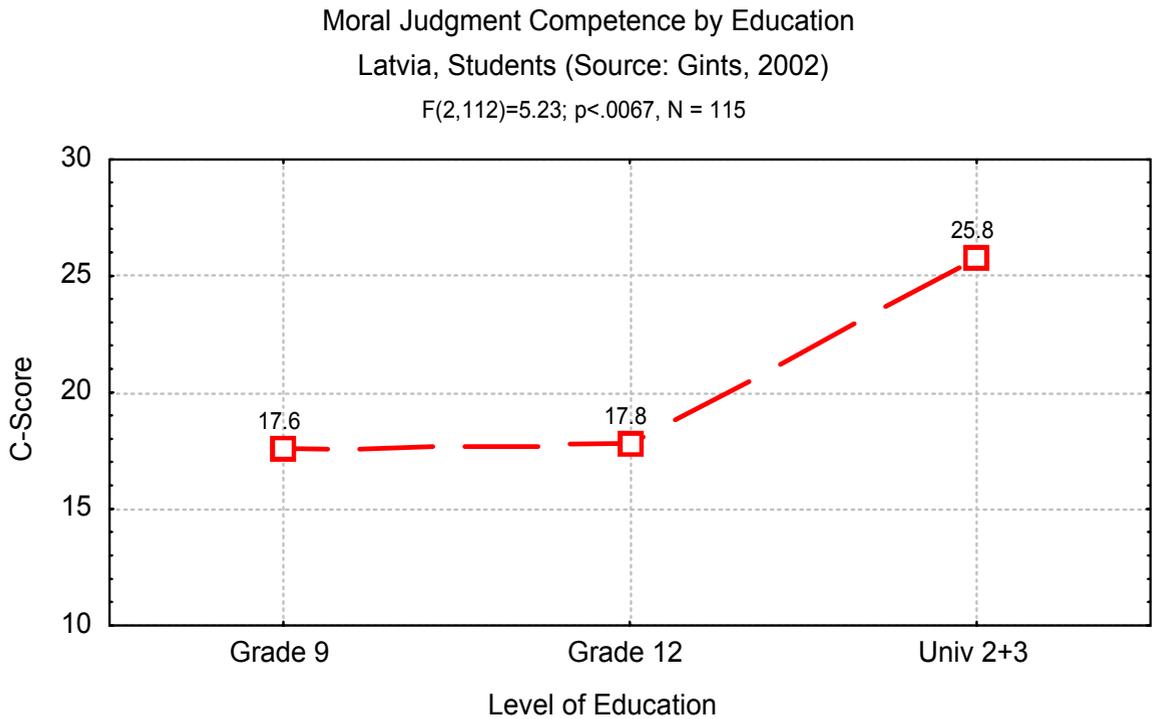


Figure 8

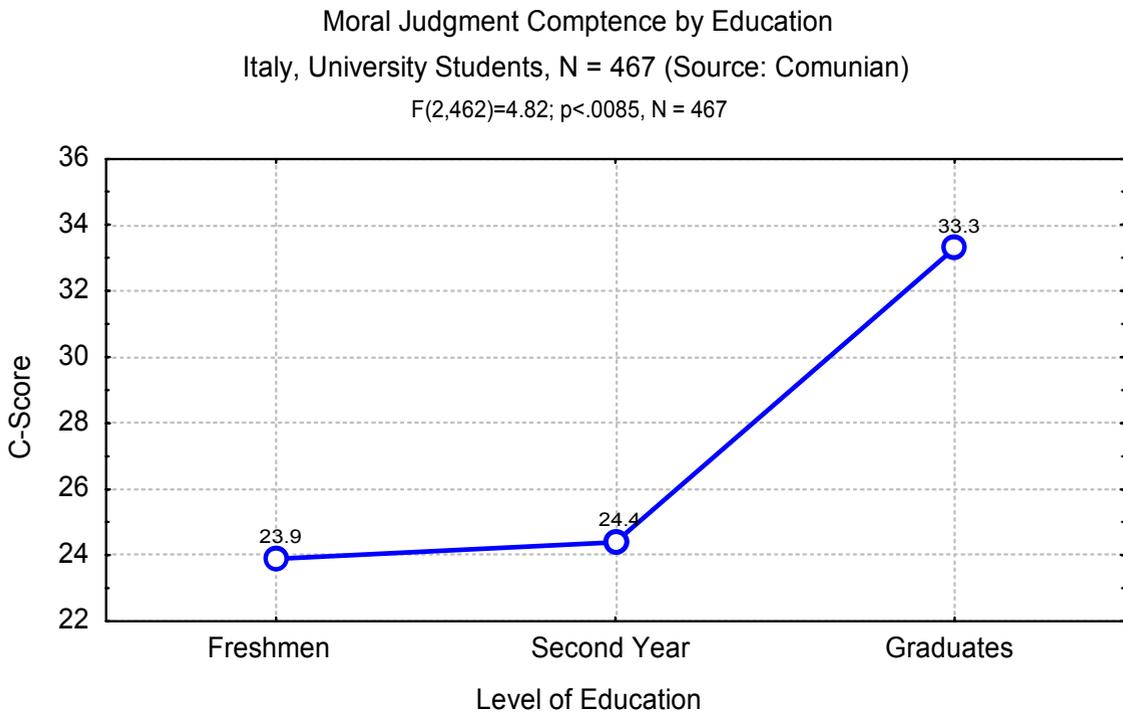


Figure 7

Criterion #4 (cont'd): Educational Correlation...

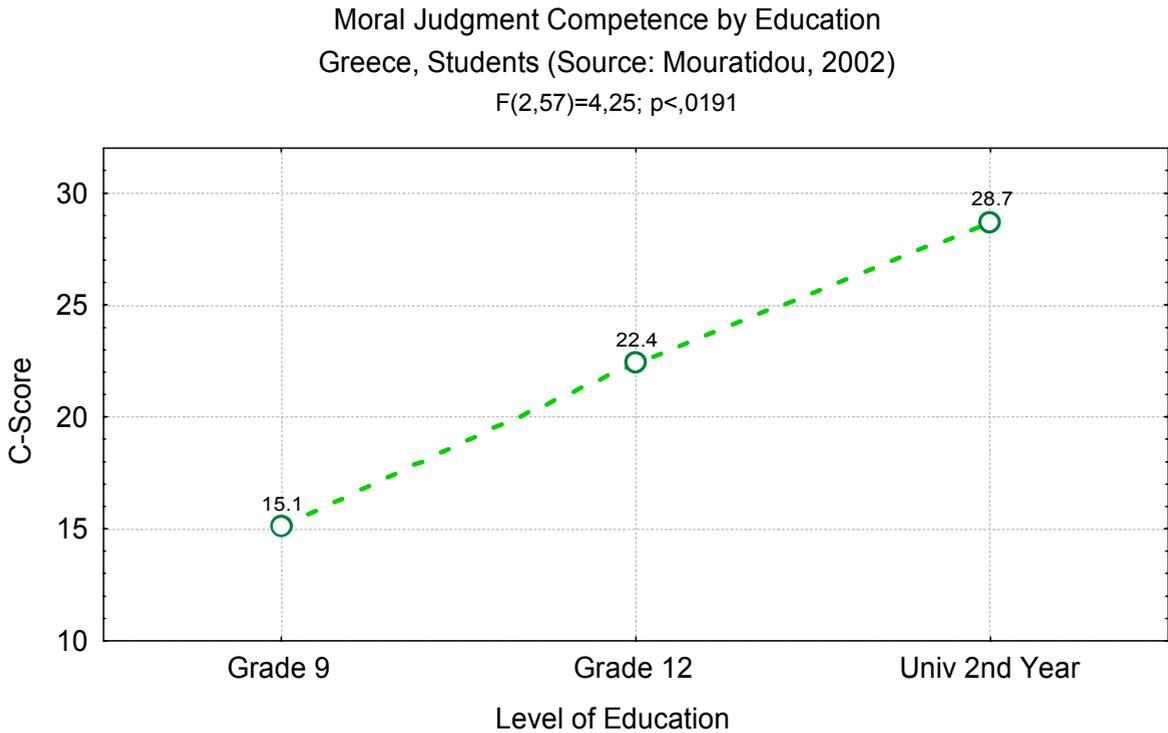


Figure 9

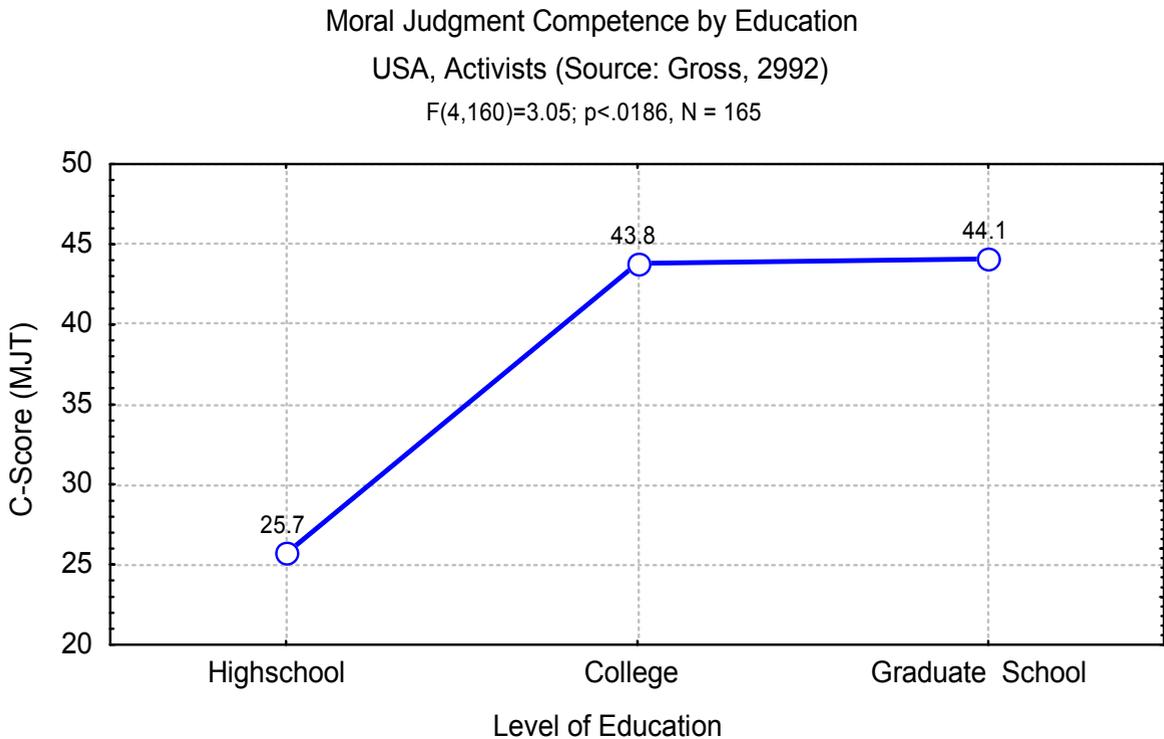


Figure 10

Criterion #4 (cont'd): Educational Correlation...

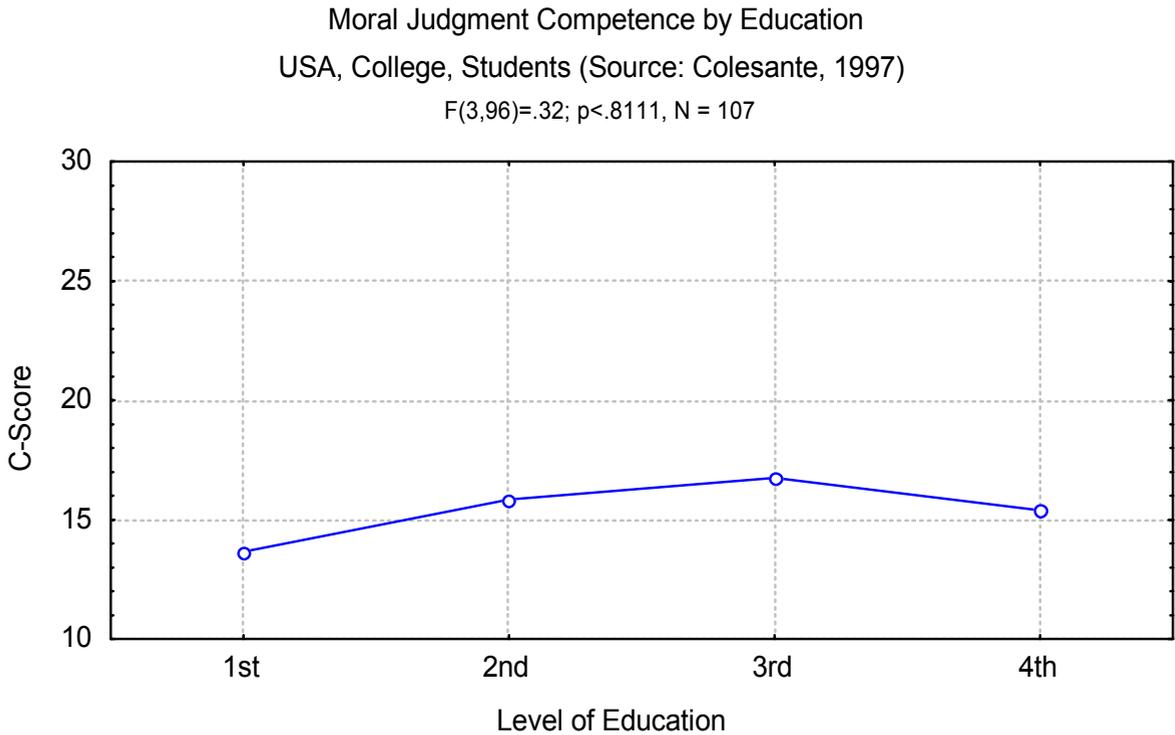


Figure 11

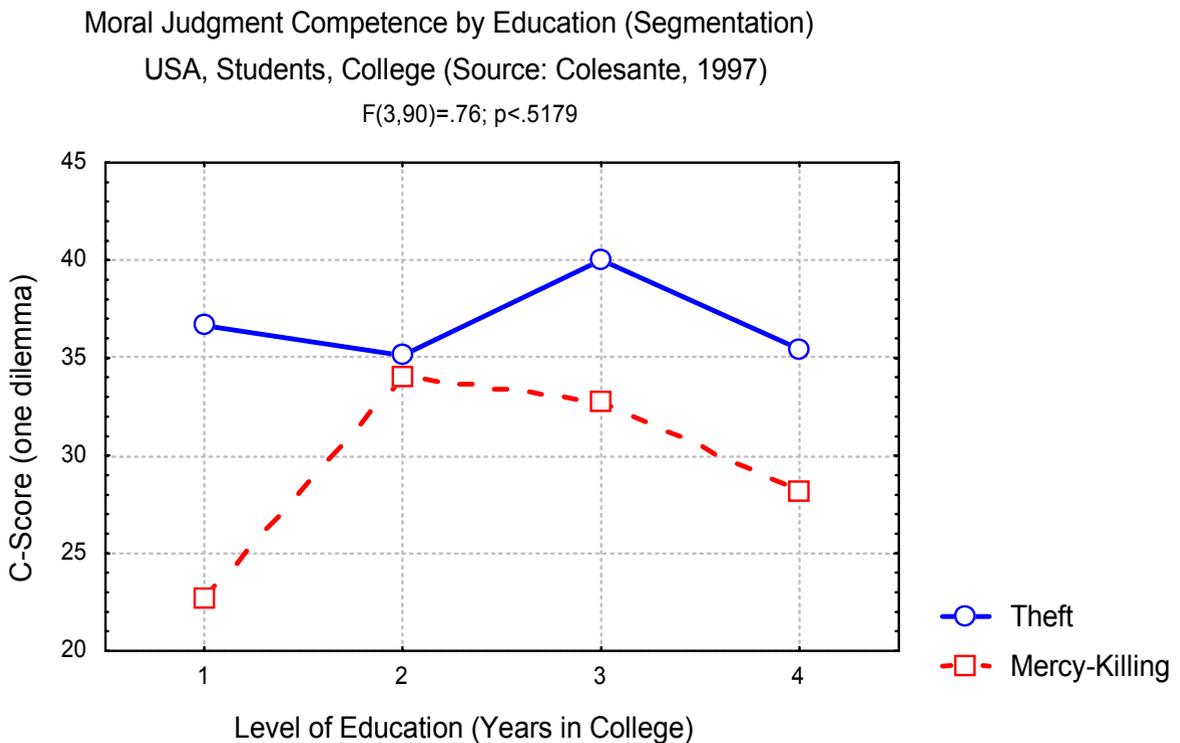


Figure 12

Criterion #4 (cont'd): Educational Correlation...

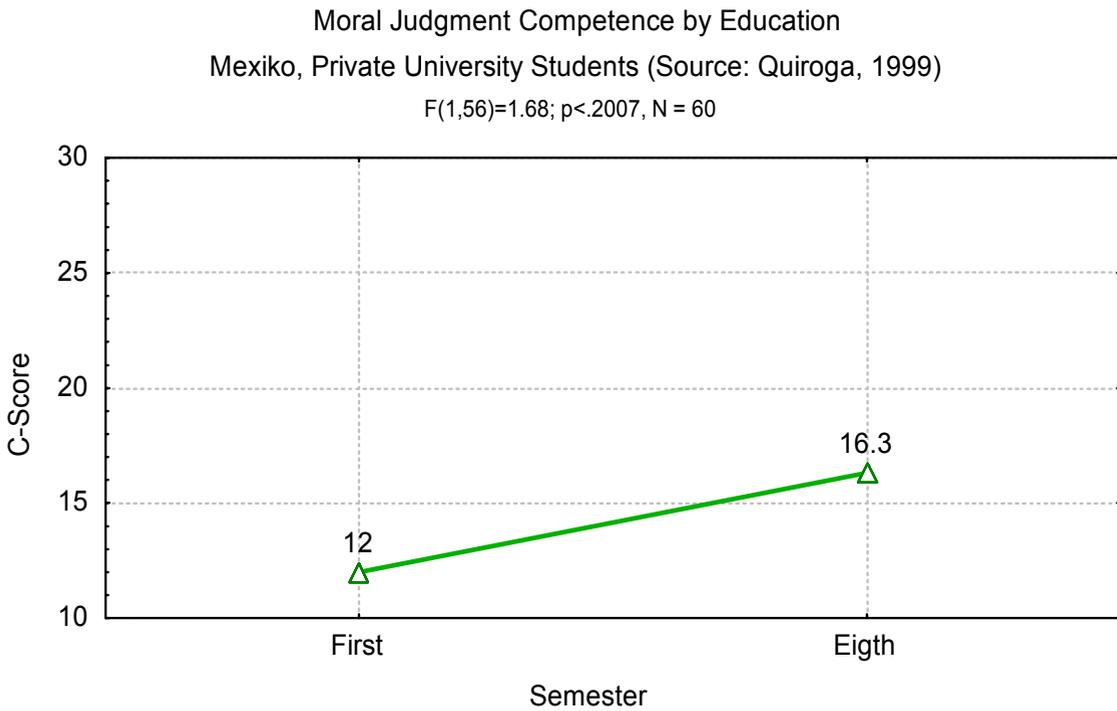


Figure 13

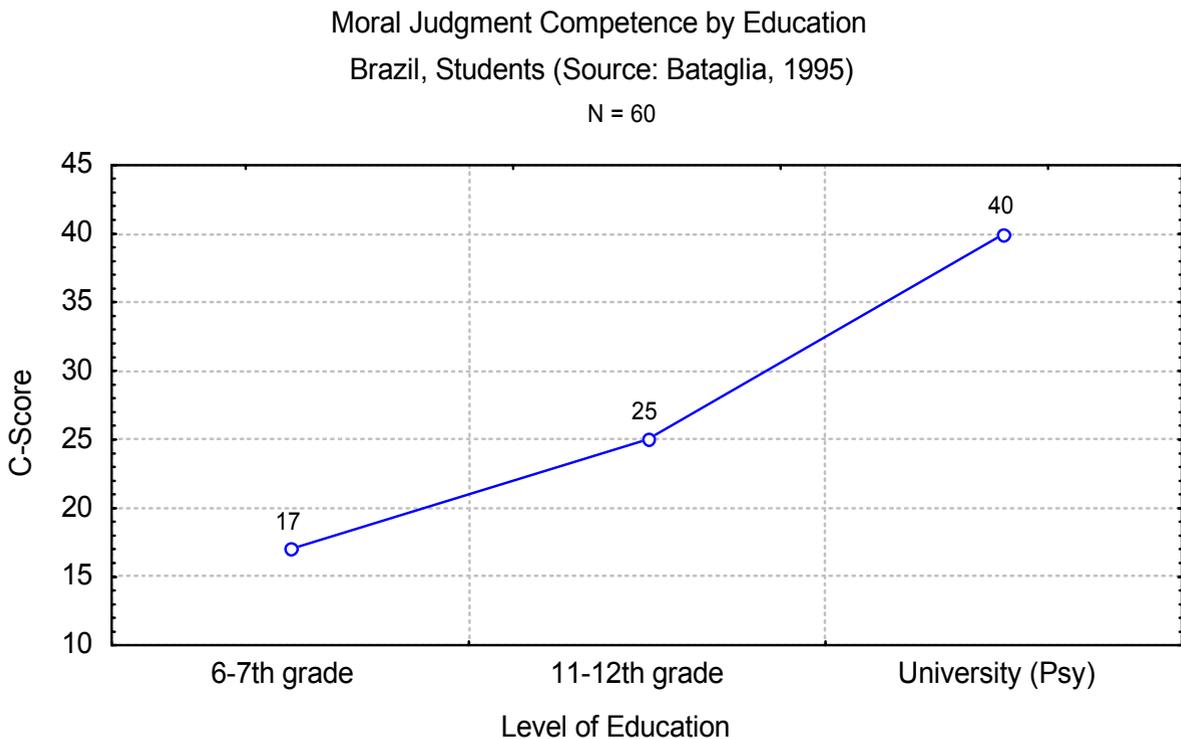


Figure 14