

The Cross-Cultural Validity of the Moral Judgment Test (MJT) -  
Confirmation of 17 Cross-Cultural Adaptations <sup>1</sup>

Georg Lind <sup>2</sup>

University of Konstanz, Germany

Abstract

The Moral Judgment Test (MJT) has been developed 30 years ago to assess simultaneously moral attitudes and moral judgment competence in a cross-cultural study. Rigorous validation procedures were implemented to make sure that the MJT has the same (semantic and pragmatic) meaning for the participants in these countries and that, therefore, we can draw safe conclusions from comparisons of the data. The validity of the MJT was tested not only by checking its semantic equivalence by careful translations but also analyzing its semiotic equivalence using three empirical criteria derived from theory and research: a) Preference Hierarchy (Jim Rest), b) Affective-cognitive Parallelism (Jean Piaget), and c) Quasi-simplex Structure (Lawrence Kohlberg). In meanwhile, the MJT has been translated into 25 languages and used in even more countries and cultures. In this presentation, we will present findings from available transcultural validity studies of 17 cross-cultural adaptations of the MJT. The findings show that all new versions fulfill these three criteria and can thus be regarded as cross-culturally valid and be used for cross-cultural research. Moreover, these findings support in an unprecedented way the universal validity of these three assumptions of cognitive-developmental theory.

---

<sup>1</sup> Paper prepared for the MOSAIC conference in Konstanz, Germany, July 20<sup>th</sup> - 22<sup>nd</sup>, 2005. Parts of this paper have been already presented at other occasions.

<sup>2</sup> 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/>

The Moral Judgment Test (MJT)<sup>3</sup>

The *Moral Judgment Test* (MJT) has been developed 30 years ago to assess simultaneously moral attitudes and moral judgment competence in a cross-cultural study (Lind 1978; 1985; 2002a; Lind & Wakenhut, 1985). Rigorous validation procedures were implemented to make sure that the MJT has the same (semantic and pragmatic) meaning for the participants in these countries and that, therefore, we can draw safe conclusions from comparisons of the data (Lind, 1978). The validity of the MJT was tested not only by checking on its semantic equivalence by careful translations but also analyzing its semiotic equivalence using three empirical criteria derived from theory and research: a) Preference Hierarchy (Jim Rest), b) Affective-cognitive Parallelism (Jean Piaget), and c) Quasi-simplex Structure (Lawrence Kohlberg). In meanwhile, the MJT has been translated into 25 languages and used in even more countries and cultures. In this presentation, we will present findings from available transcultural validity studies of 17 cross-cultural adaptations of the MJT. The findings show that all new versions fulfill these three criteria and can thus be regarded as cross-culturally valid and be used for cross-cultural research. Moreover, these findings support in an unprecedented way the universal validity of these three assumptions of cognitive-developmental theory.

While there has been a long tradition of measuring attitudes in the moral domain, measuring competencies in this 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 (Kratwohl 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).<sup>4</sup> 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 feature made Kohlberg's definition suitable as a basis for creating culturally fair measure.

---

<sup>3</sup> 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/>.

<sup>4</sup> Though was not the first to use this concept. Already Charles Darwin spoke of moral competencies in the way we use this term today.

In order to measure a particular competence it is always necessary to define the kind of *task* for which this competence is needed.<sup>5</sup> Therefore, in order to measure *moral* judgment competence it was 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., Hawthorne & May, 1928; Milgram, 1974). However, for the ability to judge people, and rate arguments, in regard to moral criteria, no tasks had been specified. Lind and his colleagues considered several options (Lind, 1978; 1985a).

Inspired 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 (1973) 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 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 our 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.

In fact, the past 30 years of research with the moral judgment test provides much anecdotal and experimental evidence in support of these theories on the interplay of moral emotion and moral cognition (Lind, 1985, 1985b; 2002). The anecdotal evidence illustrates best, what the MJT is measuring:

- 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

---

<sup>5</sup> 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).

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)<sup>6</sup>. 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.
- 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.<sup>7</sup> At this point, the MJT’s index of moral judgment competence, the so-called C-score, will be zero.

---

<sup>6</sup> “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)

<sup>7</sup> 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.

- 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 seems to discover 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 should 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”). However, when conceiving the MJT, we rather believed that the C-score would vary according to the difficulty of the moral task (Lind, 1978). In fact, our studies up until some time ago had showed that participants get very similar C-scores on our two dilemma situation – the “workers dilemma” and the “mercy killing dilemma” with a slight advantage for the latter as seems to elicit more task solving motivation (Lind, 1985a, 1985b). Findings from recent systematic studies by Krebs et al. (1991) and Juujärvi (2003) confirm our theory, although they explain their findings in a different way. They argue that these differences are due to the difference between (unobservable) competence and (observable) performance. Our theory of difference in task difficulty of dilemmas explains these differences without resorting to “unobservable” entities. Like Habermas (1983), we believe that the concept of competence would be worthless if it was not real and observable.<sup>8</sup>
- More recently, in studies in Latin-America, we discovered a very new phenomenon that has direct impingement on the question of cross-cultural validity (Lind, 2000a; Bataglia et al., 2002; Bataglia et al., 2003). We discovered that one dilemma respondents get much lower scores than on the other. First it was assumed that because the respondents were mostly college students with a very affluent background, the workers dilemma depressed the C-score. It was

---

<sup>8</sup> "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.

hypothesized 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, many respondents in Latin American countries get very low C-scores on the mercy-killing dilemma. Today we have much evidence suggesting that the teaching of the Roman Catholic Church forbids true believers to deliberate about moral problems for which the church offers a divine solution or dogma. For a true believer, obeying God does not only mean to accept his will but also to refrain from any reflection about God's will. This phenomenon reminded us of the findings by Roland Wakenhut and his colleagues who found similar differences in their studies of German soldiers, who showed high C-scores in civilian dilemmas and lower C-score in military dilemmas that present very similar moral issues as the civilian dilemmas did. Wakenhut (1982) touted this phenomenon "moral *segmentation*". Moral segmentation shows that the C-score does not only reflect 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 phenomenon. It may require us to replace the C-score by a more refined index like the C-plus index which we had proposed some years ago (Lind, 1978). As authors, we decided against using the C-plus index because both scores were highly correlated, 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, anyone who wants to take such constraints into account can do so by using respective indices.

## 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).<sup>9</sup> It is assumed that this situation, in which counter-arguments have to be evaluated, elicit *self-protective* emotions as well as *moral* emotions, that is, the tendency to protect one’s own judgment, as well as the tendency to seek moral truth.

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 the competence to 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 competence, 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.<sup>10</sup> Technically, the C-score reflects the ratio of the response variance accounted for by the experimental factor “moral quality of the argument” by the total response variance.<sup>11</sup> The C-score ranges from zero,

---

<sup>9</sup> Kohlberg reduced his six stage-model at some point, but returns to it later again (Kohlberg et al., 1990).

<sup>10</sup> 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).

<sup>11</sup> 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).

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<sup>12</sup> 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 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.

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).

---

<sup>12</sup> 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.



Semantic and Semiotic 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 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*<sup>13</sup> *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.<sup>14</sup>

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.

---

<sup>13</sup> 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).

<sup>14</sup> I wish to thank all experts involved in this: Tino Bargel, Rainer Döbert, Thomas Krämer-Badoni, Gertrud Nunner-Winkler, Gerhard Portele, and Roland Wakenhut.

By *semiotic*<sup>15</sup> 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 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.

---

<sup>15</sup> "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).

Three Invariant Universals as Criteria for Semiotic Equivalence  
and Cross-cultural Validity of the MJT<sup>16</sup>

Usually, structural properties of a test or questionnaire are checked by purely conventional criteria like “internal consistency” (or Cronbach’s alpha) or the numbers of dimensions produced by factor analysis. In contrast, the construct validity and cross-cultural equivalence of the MJT is checked using as criteria universally proven facts regarding the nature of moral judgment behavior.<sup>17</sup> We claim that these facts are so well established that any deviation of data will indicate a lack of validity rather than with the theory.

Three findings from many decades of moral judgment research seem to be so well proven that indeed we can consider them as facts. A fourth finding can be regarded as a fact when considering its dependency on certain conditions:

1. Kohlberg’s (1958) finding of a *Quasi-Simplex Structure* of ordered stage correlations.
2. Rests’s (1969) finding of *Hierarchical Preference Orders* for the Kohlbergian stages.
3. Piaget’s (1976; 1981) theory of Affective-Cognitive Parallelism, empirically corroborated by Lind (1978; 2002).

In the following, I will discuss these three validation criteria plus the criteria age and education, which we have been dropped from this list (see also Lind, 1978; 1985; 2002; 2004).

**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 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:

---

<sup>16</sup> 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.

<sup>17</sup> 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.

"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).<sup>18</sup>

"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 did not employ a statistical coefficient to measure the degree of fit of this prediction with observed data. So, like a medical doctor 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 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 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  $4! \text{ or } 4 \times 3 \times 2 \times 1 = 24$  ways in which the stages can be ordered. Hence, the probability that they are ordered in the predicted way by chance is  $p = 1/24 =$

---

<sup>18</sup> 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.

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	-
$\bar{x}$	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 Quasi-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).

### **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.<sup>19</sup> 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 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.<sup>20</sup>

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.

---

<sup>19</sup> See Kohlberg, Boyed & Levine, 1990.

<sup>20</sup> 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. 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 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)

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 1<sup>st</sup> 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).

**Other Criteria (dropped): Correlation of moral development with age and 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). 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. (See also the graphs in the appendix containing sample findings).

However, many studies have found undisputable signs of regression and, therefore, age cannot be regarded as a valid criterion for validating moral development tests any longer. 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 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 found clear evidence that moral competence mostly erodes, which means that age by itself did *not* foster moral judgment competence. It seems that up until a certain critical 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 (Lind, 2002; Rest, 1979; 1991).

Therefore, for some time it seemed 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, it still seems helpful in drawing a proper validation sample. The data will be an adequate basis for testing the validity only if the sample produces substantial variation of moral judgment competence. The size of variance of the C-score in the validation sample restricts the correlation that can be found. 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 guarantees that the variation of moral judgment competence in the sample is large enough for a meaningful validation analysis (for more details on the suggested 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).

Validation data are available from 17 of these adaptations of the MJT. In each case, the findings and often even the raw data were submitted for review to the 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. These hints 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. Only when the findings showed a perfect or near-to-perfect fit, new versions were certified as cross-culturally valid. A current list of translated and certified versions of the MJT can be found on this web-site: <http://www.uni-konstanz.de/ag-moral>).

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 three criteria were applied for confirming the test's *semiotic validity*, and to detect and correct 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 validation 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.

*Twenty translated versions agree with the three validation criteria that we have discussed above and, therefore, have been certified as cross-culturally valid. The agreement with the three criteria was perfect or nearly perfect in all cases. In no case was there a substantial disagreement after all technical errors had been corrected and all semiotic problems had been solved.*

Note that the items of the MJT were not empirically optimized to meet the validation criteria. In no case were the test items pre-selected on the basis of their correlation with age and education but have been selected solely on the basis of their agreement with Kohlberg's stages of moral reasoning. So the findings can be fully trusted as supporting the three propositions derived from the

theories of Piaget, Kohlberg and Rest.

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 make an instrument unsuitable for testing the very hypothesis anymore.<sup>21</sup>

The fact that three 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).

As the data shows (see Figures 8 through 14), the correlations vary strongly from country to country, indicating that this criterion is only of limited value as validation 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

---

<sup>21</sup> 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.

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 12<sup>th</sup> graders than in college students. On my inquiry I found out that the 12<sup>th</sup> 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.

### 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 three empirical criteria derived from moral psychological theories and research by Kohlberg, Rest, and Piaget, respectively:

1. Quasi-Simplex Structure 2. Hierarchical Preference Order, and 3. Affective-Cognitive Parallelism. 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 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 these 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 17 different countries also supports core assumptions of Kohlberg's cognitive-developmental theory (except the assumption of non-regression) and of Lind's supplementary Dual-Aspect Theory of moral judgment behavior (Lind, 2002): a) Regardless of cultural background (and regardless of age, gender, and education), people

prefer Kohlberg's stages of moral reasoning in the predicted order, b) these stages also constitute some developmental order as neighboring stages are more correlated than more distant stages, and c) the higher people's moral judgment competence, the more they prefer principle moral reasoning over lower stage reasoning (see also Snarey, 1985).

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 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.

One last finding 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 regression and their regression was much more pronounced than the increase by the other three quarters of the sample.

After some experimenting we discovered that this group of 'regressors' felt annoyed by 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 clearly state to the subjects at the post-test that they will have to fill out exactly the same test, and that they should do this as sincerely as they did the first time. We have no signs of test weariness anymore. On the contrary, the effect sizes of our moral education programs have went up tremendously (above  $r = 0.70$ ).

What does this all mean for education? Not all implications are that obvious, but two are. 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

- Apel, K.-O. (1990). *Diskurs und Verantwortung. Das Problem des Übergangs zur postkonventionellen Moral* (Original 1988). Frankfurt: Suhrkamp.
- 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.
- Bataglia, P., Schillinger-Agati, M., Lind, G. & , Quevedo, T.L. (2003). *Testing the segmentation hypothesis with an Extended Version of the MJT*. Poster presented at the meeting of the Association for Moral Education, Krakow, July 19, 2003.
- Blatt, M. & Kohlberg, L. (1985). The effect of classroom moral discussion upon children's level of moral judgment. *Journal of Moral Education*, 4, 129-161.
- 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 21<sup>st</sup> - 25<sup>th</sup>, 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.
- 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.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 4, 814-834.
- Hartshorne, H. & May, M.A. (1928). *Studies in the nature of character*. Vol. I: Studies in deceit, Book one and two. New York: Macmillan.
- Helkama, K., Uutela, A., Pohjanheimo, E., Salminen, S., Koponen, A., & Rantanen, Vantsi, L. (2003). Moral reasoning and values in medical school: a longitudinal study in finland. *Scandinavian Journal of Educational Research*, 47, 4, 399-411.
- 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, Finland: Helsinki University. Press.
- 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.
- 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.
- 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. (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. (2000a). *Off limits. A cross-cultural study on possible causes of segmentation of moral judgment competence*. Presentations at the AERA Conference. New Orleans.
- Lind, G. (2000b). 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. & 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.
- Milgram, S. (1974). Obedience to authority. An experimental view. New York: Tavistock.
- 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., 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.
- Schillinger-Agati, M. & Lind, G. (2003). *Moral judgment competence in Brazilian and German university students*. Paper presented at the AERA conference in Chicago, April 21<sup>st</sup> - 25<sup>th</sup>, 2003.
- 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.
- Wojcik, (1998). Semantics. In: *Grolier Multimedia Encyclopedia*. CD-Edition.

Criterion #1: Quasi-Simplex Structure

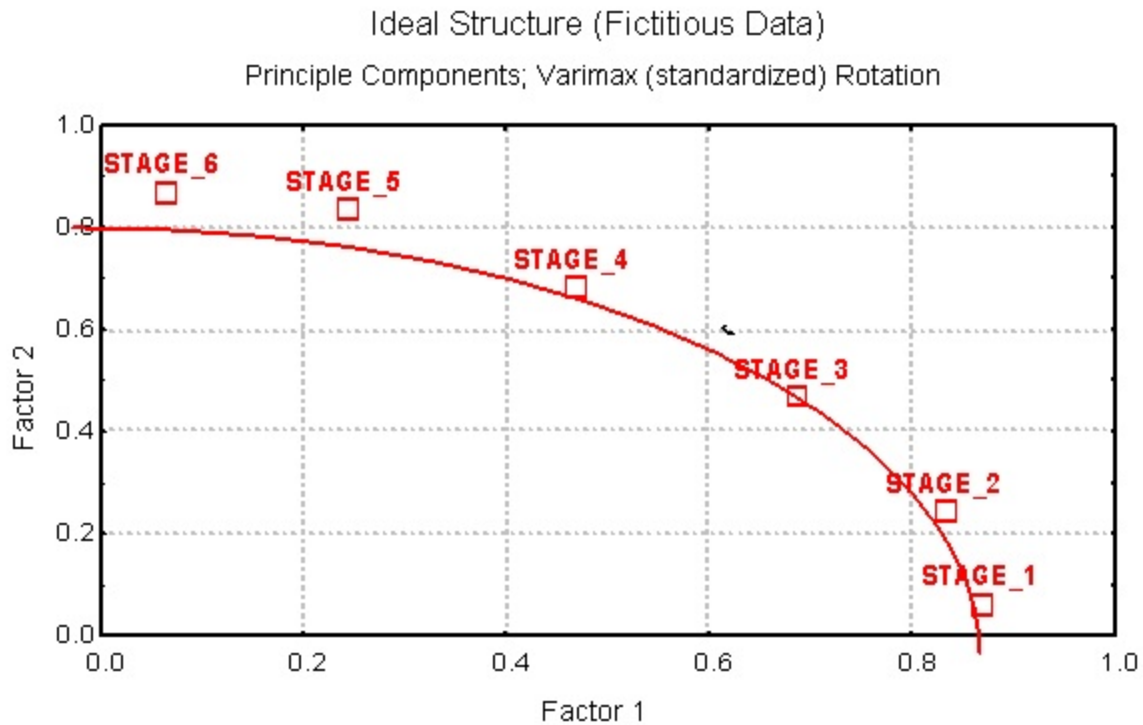


Figure 1

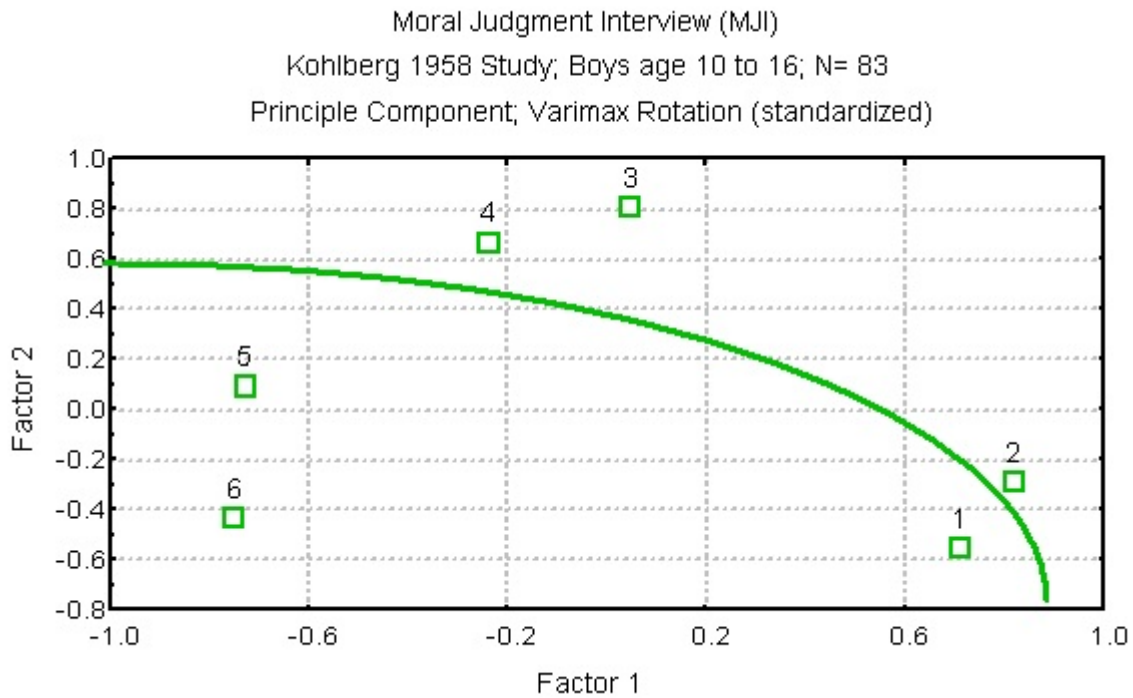
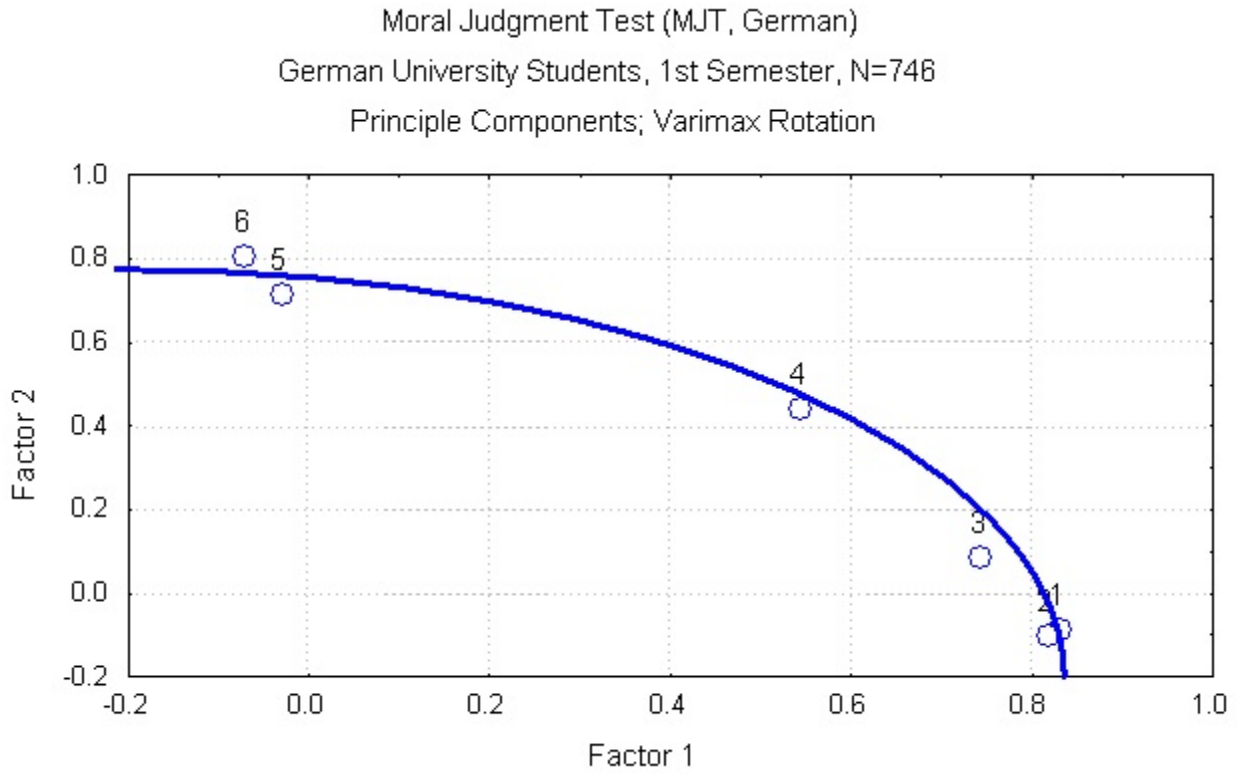


Figure 2

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



Source: Lind, 2002

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

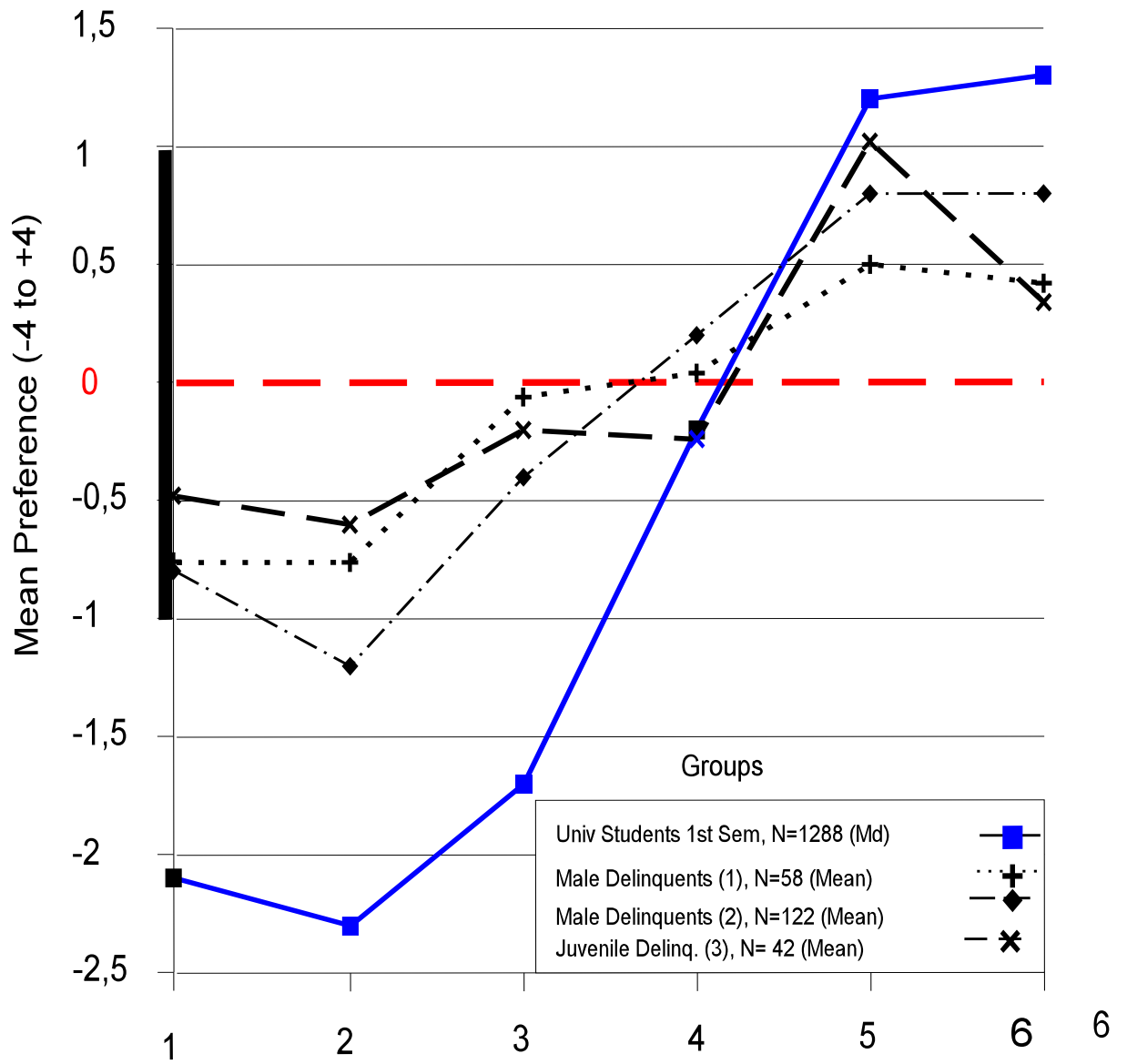


Figure 4

Criterion #3: Affective-Cognitive Parallelism

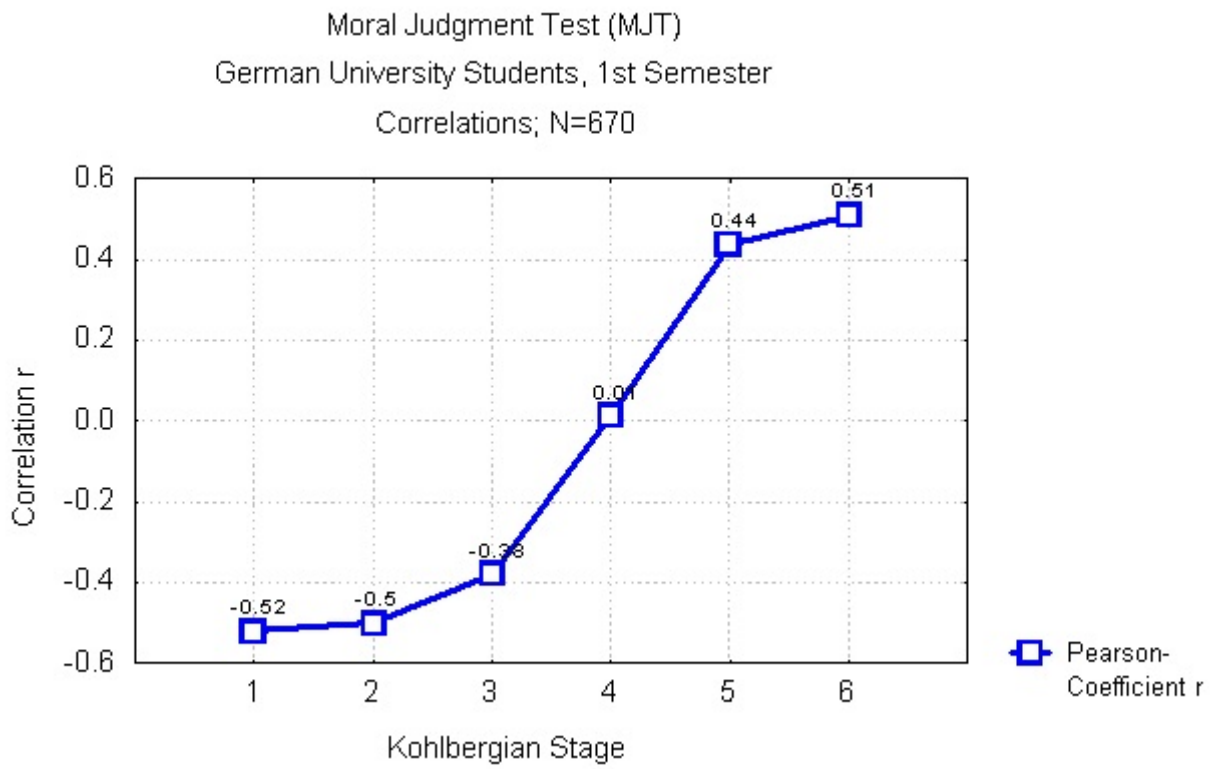


Figure 5

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

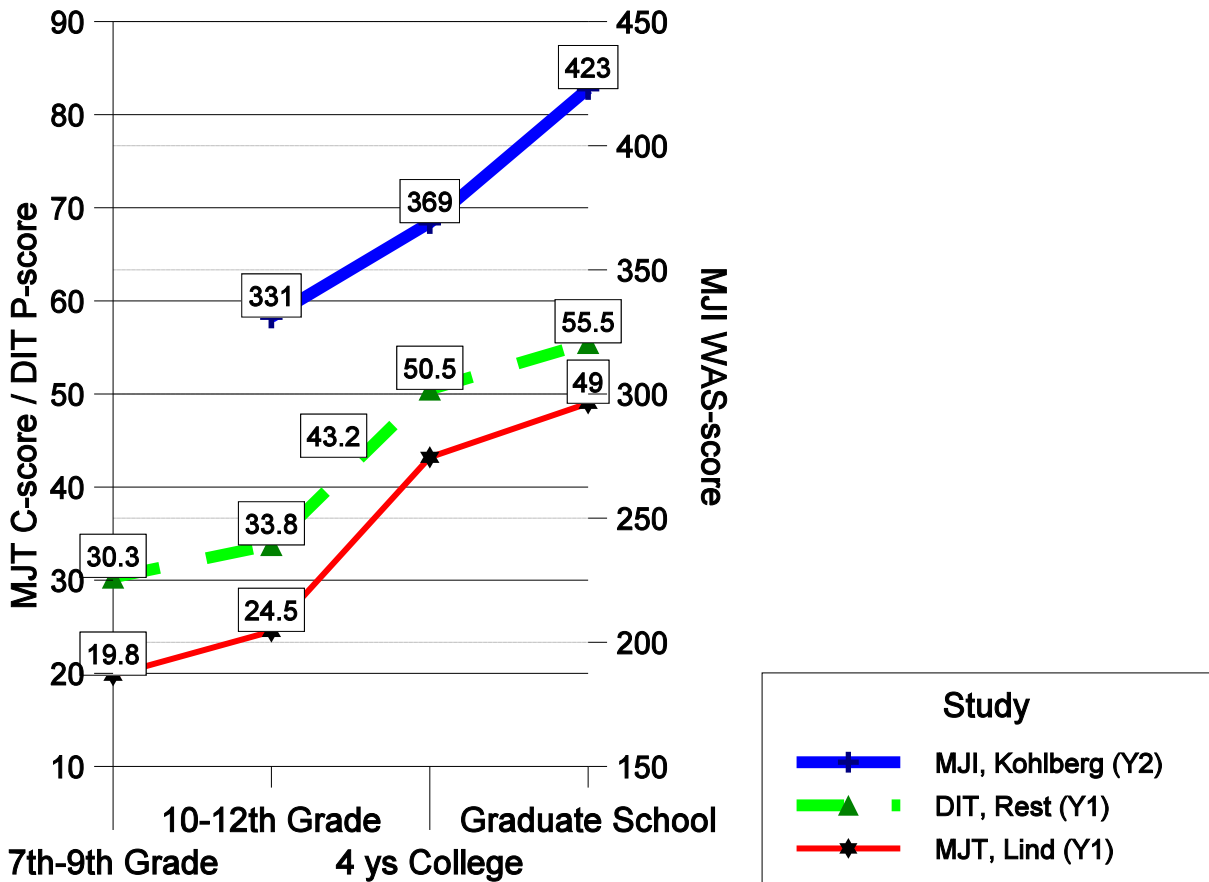


Figure 6

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

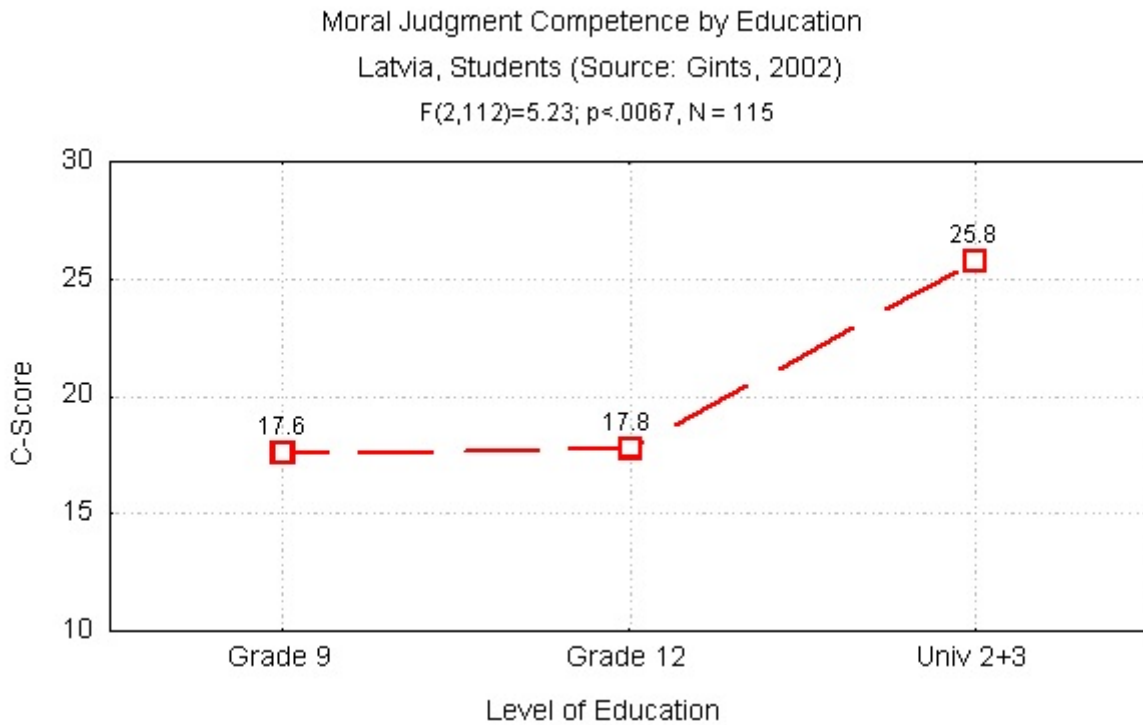


Figure 8

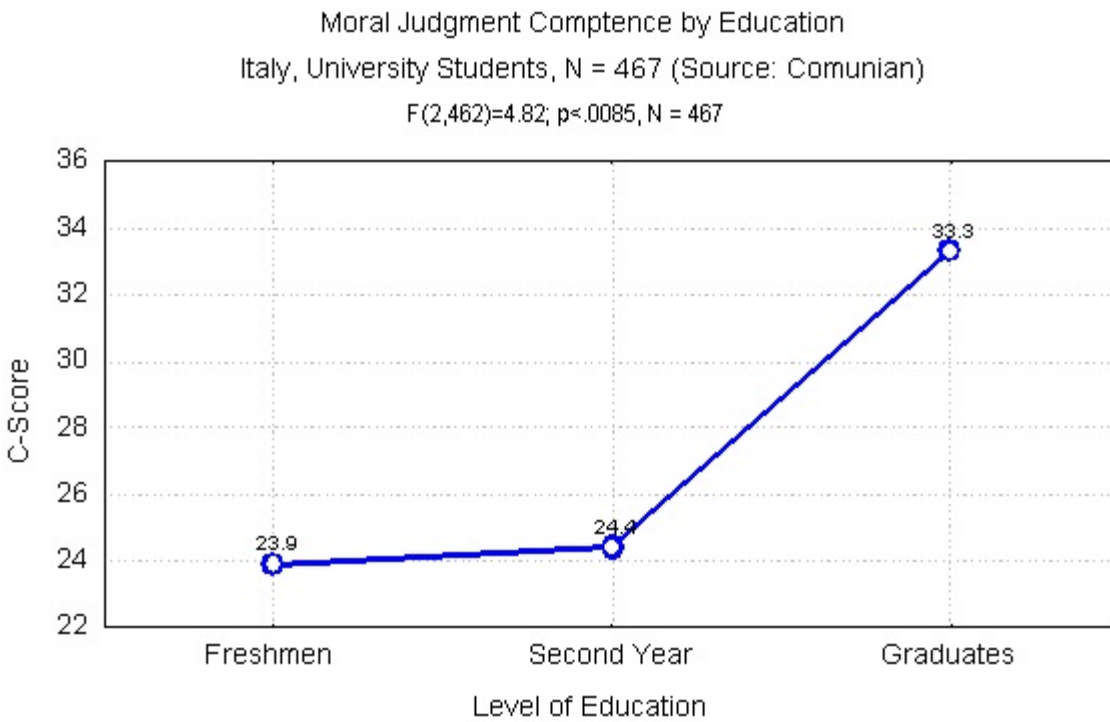


Figure 7

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

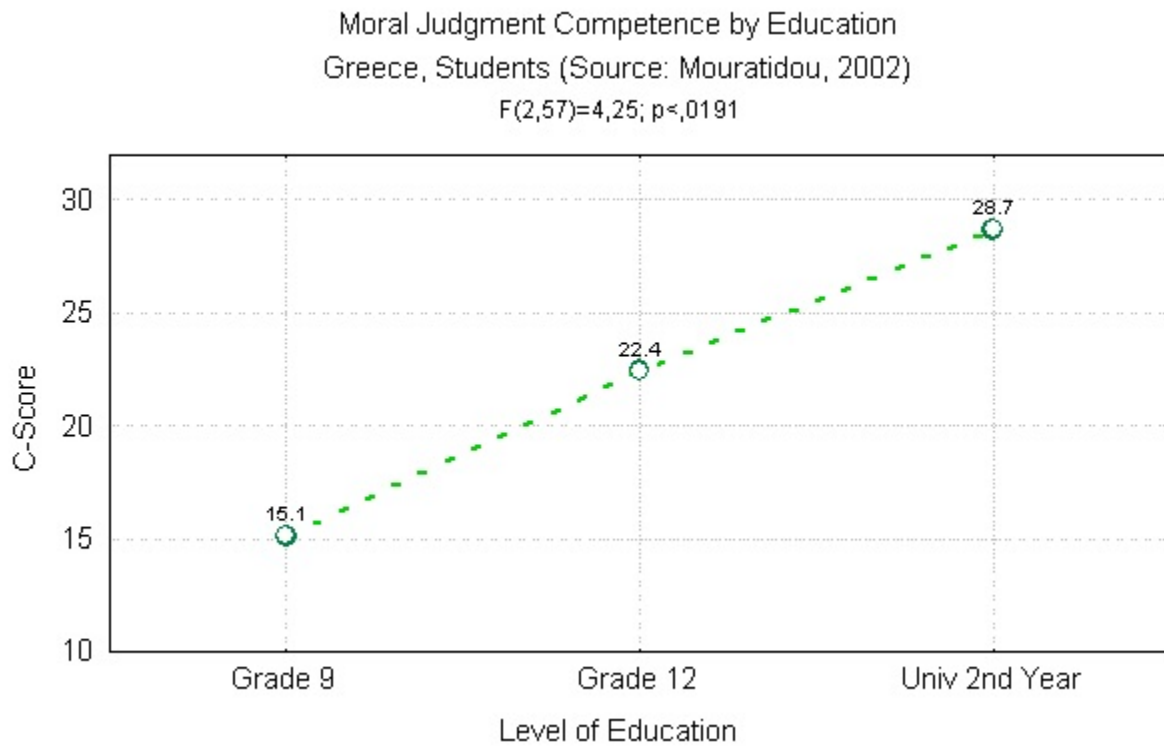


Figure 9

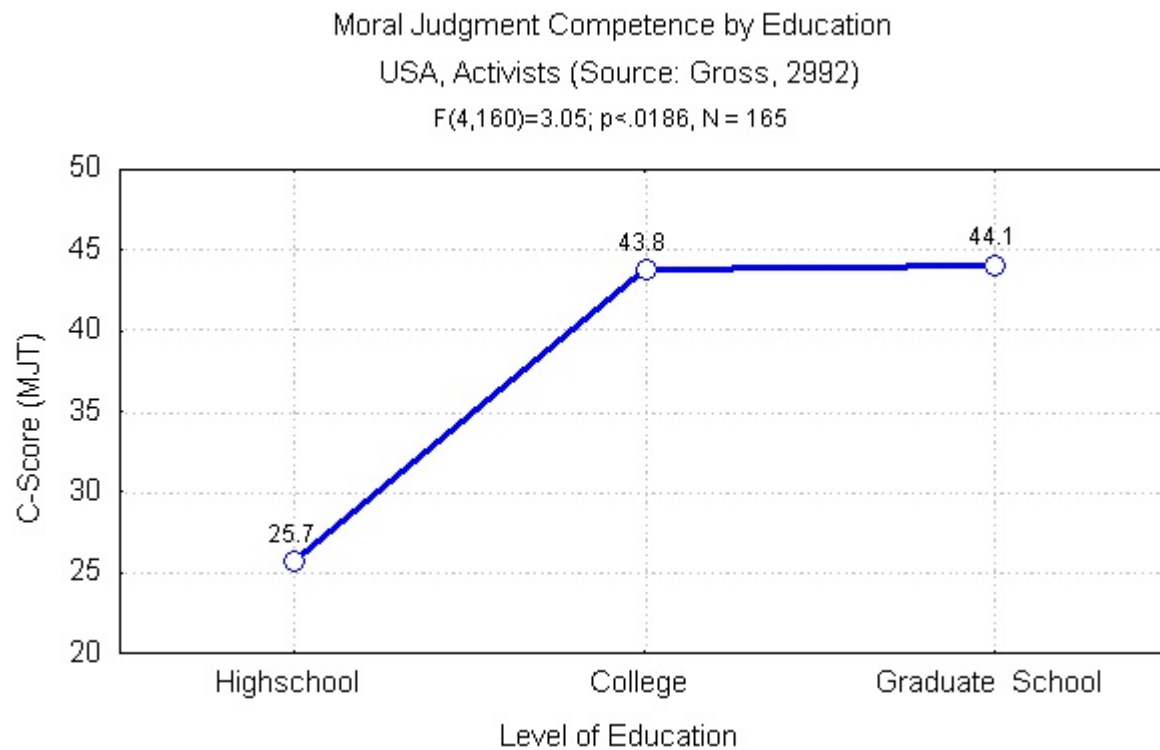


Figure 10



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

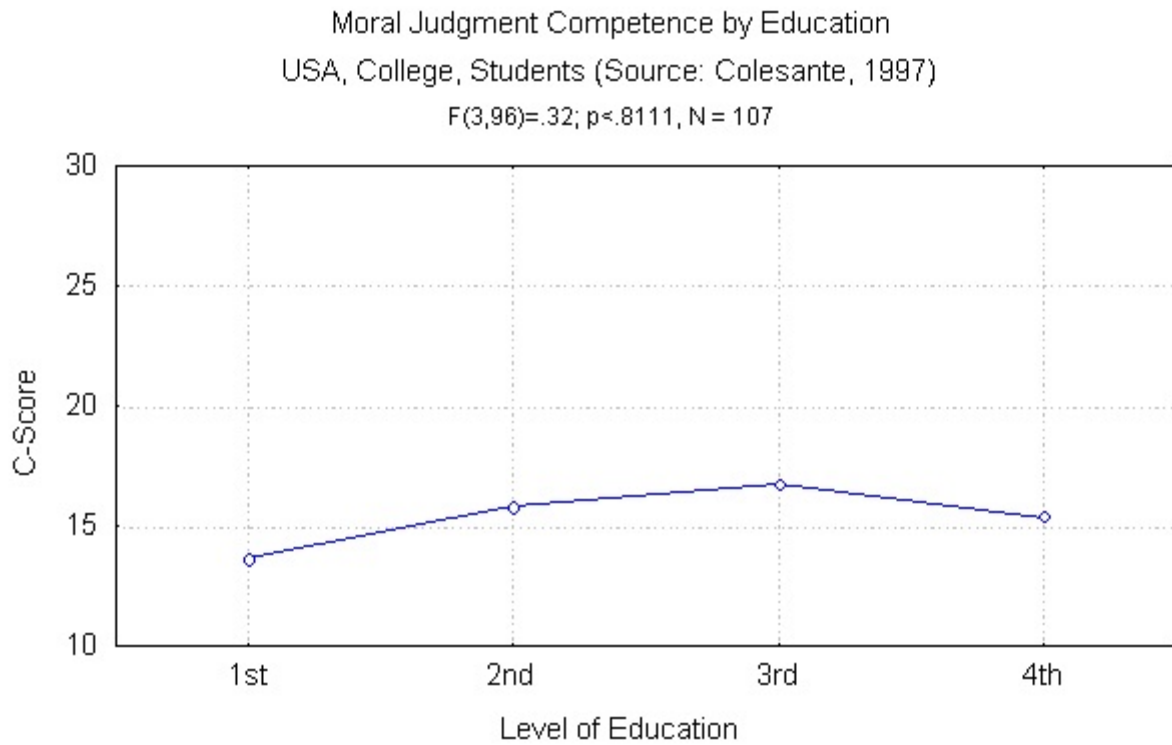


Figure 11

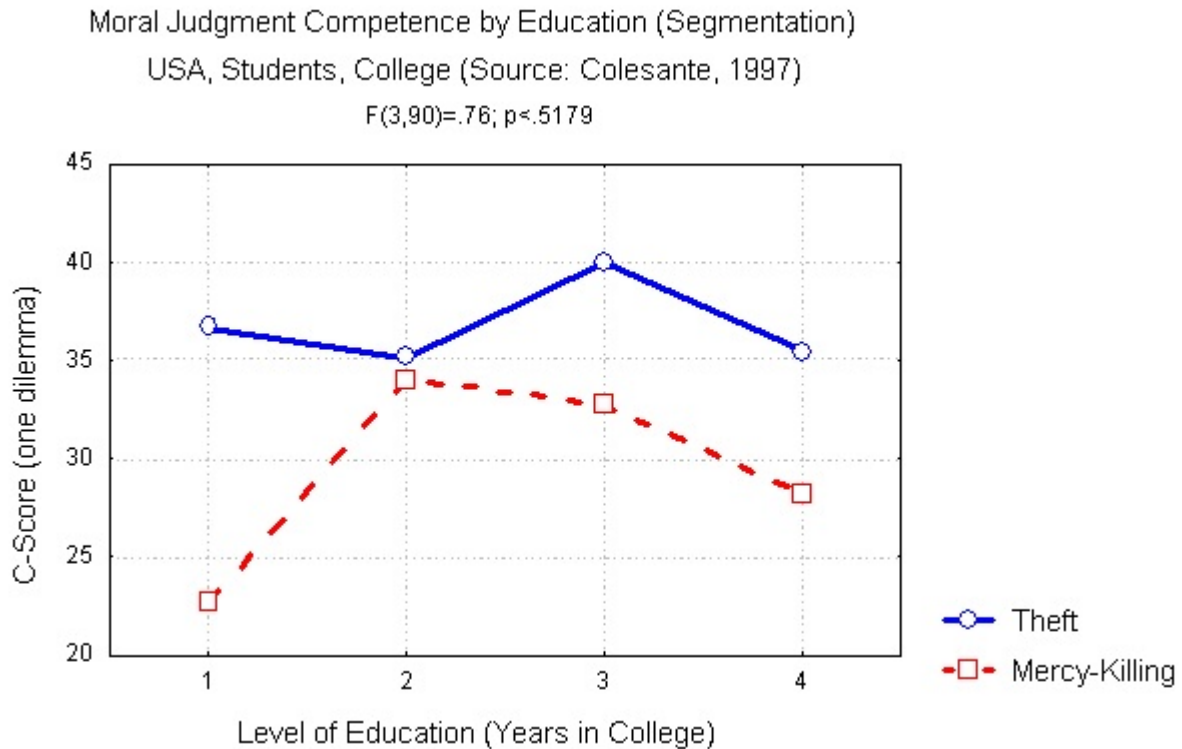


Figure 12

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

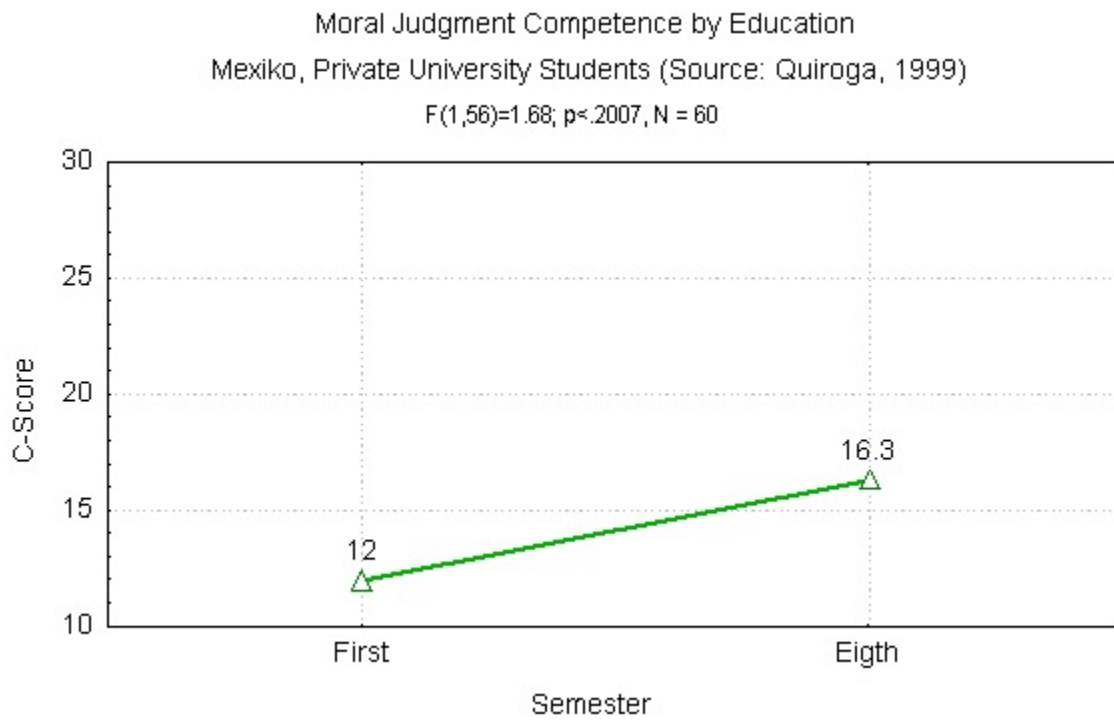


Figure 13

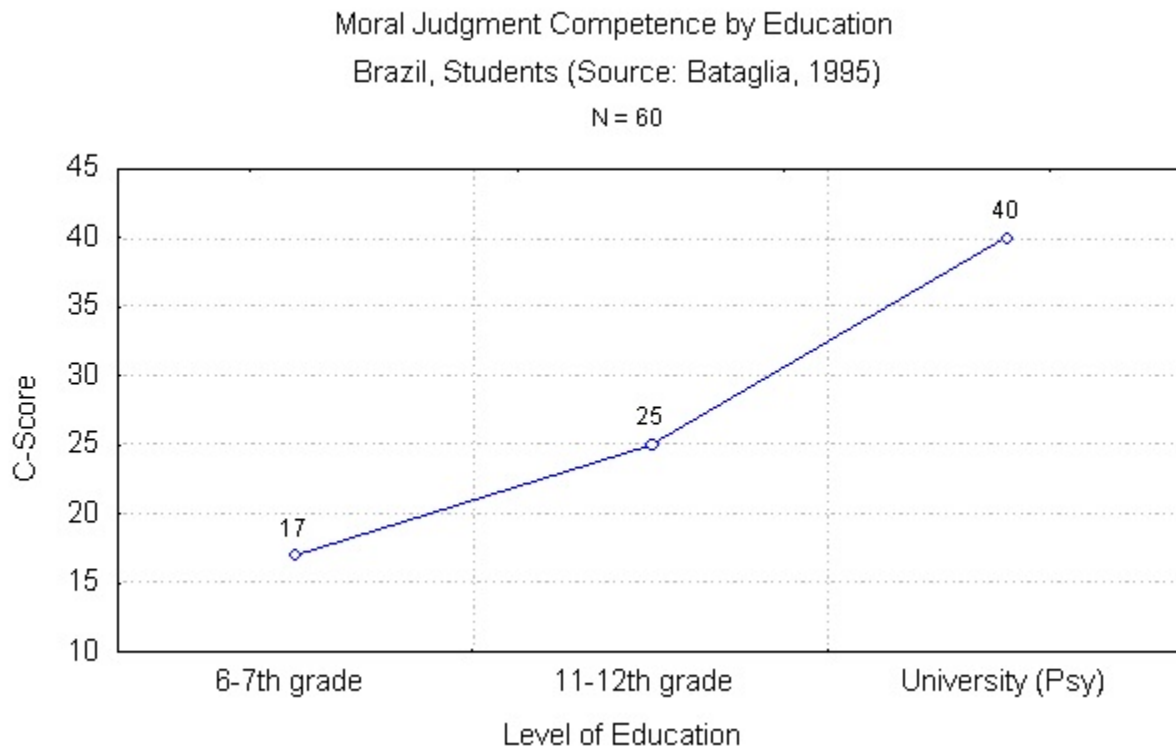


Figure 14