

Cultural Fairness and the Measurement of Morality

Georg Lind

1995

Contact:

Prof. Georg Lind

University of Konstanz

FB Psychologie

78457 Konstanz

E-Mail: Georg.Lind@uni-konstanz.de

For further information and publications on this topic see

www.uni-konstanz.de/ag-moral/b-publik.htm

**Psychology of Morality &
Democracy and Education**

Paper presented at the AME-Meeting in New York, 1995

File: Lind-1995_Cultural Fairness MU-95.wpd / Revision: Jul 16, 2001

Cultural Fairness and the Measurement of Morality

Georg Lind

University of Illinois-Chicago

University of Constance, Germany

Abstract

Moral development is always measured by some standards and values. These standards typically originate in the value system of Western middle classes as most researchers and test makers have a Western middle class background. Therefore, we must ask whether the use of these tests in other classes and cultures is meaningful and whether a comparison across class and culture boundaries is fair. Obviously, this problem prompts another two questions, namely *a)* whether there is a universalistic definition of 'moral' and 'moral development' shared by all classes and cultures, and *b)* whether the test in question is a valid operationalization of this universalistic definition. We argue that Kohlberg's early definition of moral judgment competence comes close to the ideal of being universalistic. Yet his Moral Judgment Interview and most other tests are vulnerable to criticism because they measure people's moral judgment to a specific system of moral values. These tests measure a person's moral judgment competence by the standards set forth in Kohlberg's Stage model rather than by people's *own* moral ideals. The *Moral Judgment Test*, used in a wide array of cross-cultural studies, corrects this shortcoming through measuring persons' moral judgment competence by their own moral values.

1. *Introduction*

Cultural fairness continues to be an important issue in many areas of psychological research, spanning from research into mathematical and logical reasoning (Rogoff & Chavajay, 1995; Triandis, 1996) to research on moral development (Eckensberger, 1991; Snarey, 1995; Gielen & Markoulis, 1993). In moral psychological research, however, cultural fairness is not just one issue among many but is a core issue. “Culture” and “fairness” both are core moral categories. Moral and other values define the core meaning of culture. People can speak the same language but belong to different cultures, or speak different languages and belong to the same culture. People may live in the same country and belong to different cultures, or live in different countries or even continents and still belong to the same culture. So it is quite natural to ask whether moral research itself complies to the standards of fairness and mutual respect which we use when we measure these virtues in our subjects (Gielen, 1984; Simpson, 1974; Sullivan, 1977; Vine, 1984).

Culture is defined in many different ways: language, nationality, religion, geography, socioeconomic structure, behavioral habits, shared myths and so on. We may, on the one extreme side, chose a very wide definition of culture as “the totality of socially transmitted behavior patterns, arts, beliefs, institutions, and all other products of human work and thought” (The American Heritage Dictionary, 1992; see also Triandis, 1996). This all-encompassing definition tells us that the issue of cultural fairness may be at stake in any kind of comparison of people, not only of people who belong to big cultures but also to subcultures, social classes or even families. Yet this definition is too broad to be useful. More specific definitions relate either to a common *value* system (Child, 1954; Edwards, 1986; Eckensberger, 1991), or to a common system of skills, knowledge and *competencies* (Keesing, 1981).

We can distinguish three different possibilities as for the question of fair testing across cultures:

First, if all cultures share the same values, we had no problem of cultural fairness. However, this is hardly the case.

Second, if among different cultures *no* values would overlap, comparison regarding *any* standard would be *unfair*. All that we could do is to measure the degree to which people from other cultures happen to behave like us and to

solve the problems that we believe to be important. This case is also very unlikely.

Third, if different cultures would share *some* values but differ according to others, we could measure in a culturally fair way the degree to which people behave according to these shared standards but not according other values. This very likely case requires that we identify which values are shared, before we measure a particular trait or competence.

In this essay, I will assume that different cultures have at least some basic moral values in common, i.e., so-called universalistic moral values. To distinguish them from other, more relativistic values--like material values, conventional standards, personal values etc.--I will call them "moral principles." Moral principles can be distinguished from other values by applying Kant's *Categorical Imperative*: Act as if the principle on which your action is based were to become by your will a universal law of nature. Only a few values meet this criterion. Justice und respect for others are such "universalizable" principles. We may fairly assume that people in all cultures want to be treated justly and with respect. Otherwise, we had no reason to worry about "culturally fair" measurement. We may also assume that people in all cultures view it as legitimate that we expect them to treat *us* and our interests in a fair and respectful way.

These two principles entail the use of reason and discourse for solving conflicts of interest and opinion rather than violence and force. They entail, as Kohlberg (1964) called it, *moral judgment competence*, which he aptly defined as the "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).

The reader may object and point out that peaceful reasoning is less frequent than one should expect if justice and mutual respect would be universal moral principles. This objection is based on a wrong identification of moral values and behavior. If values were identical with behavior, we probably would have an ideal, peaceful world, and we would hardly need research and education in this domain. This gap between our moral principles and ideals on the one side and our behavior on the other, requires the study of moral competencies and their education.

In cross-cultural research then we are wrong when we *assume* that the values or standards of our measurement are shared by the foreign culture that we study. This error may not perturb us if our aim of measurement is merely to sort and select people who apply for a job or an education in our culture. By their application, we may argue, these people also adopt our value system. However, if the aim of our research is *understanding* the process of moral development in various cultures, this error might distort our findings.

2. Fairness, the Two Kinds of Measurement and Validity

The distinction between two basic aims or purposes of measurement relates directly to the two types of measurement that Coombs, Dawes and Tversky (1970) have made: *a*) measurement as a *technique* for obtaining information that lets us as accurately as possible predict the criterion behavior or disposition that we are interested in, and *b*) measurement as a *criterion*. Measurement as a technique is often chosen when the valid test of behavioral disposition that we want to measure is difficult to apply or very costly. Such measurement techniques do not need to possess content or theoretical validity but only high *predictive* validity. That is, they must correlate very highly with the actual criterion that they only *indicate*. A classical example is the *Minnesota Multiple Phases Inventory* (MMPI) which was constructed merely from studies that screened many test items for high empirical correlations with the mental illness records of large samples (Meehl, 1988). The authors of the MMPI claim that this test can predict mental illness better than most clinical psychologists and psychiatrists do, though many items have no psychological meaning or resemblance of the disorders they seem to indicate. Measurement as a technique does not even require a well-developed theory about the nature of the criterion (for example, about the nature of mental illness). It merely requires agreed-upon categories (for example, for classifying persons as schizophrenic, phobic, depressive, or normal). A diagnostic system that is merely measurement-as-a-technique may work very effectively. They may work because techniques such as the MMPI define both the predictor *and* the criterion, and have thus some kind of “self-referenced-validity.” Yet we have no way of checking its theoretical validity if we have no theory about these phenomena. Statistical

analyses that have no grounding in theoretical expectations provide no basis for *understanding* the nature of those mental disorders, and does not tell us anything about the *theoretical validity* of a measure. Moreover, there is no *fair* and *objective* way of comparing these measures cross-culturally because the empirical correlations, on which these measures are based, reflect many, relevant and irrelevant, facts like the number of mental hospitals, the health insurance system, and the admission policies of the hospitals in a country.

If we want to advance our understanding of human behavior and to test our theories about it, we need *criterion measurement*. Criterion measurement cannot, of course, be validated in the usual way, that is, it cannot be correlated with some criterion because there is no valid criterion. For example, before Kohlberg (1958) designed his measure of level of moral judgment, we had no criterion available for validating tests by means of correlation. Instead he derived hypotheses about the nature of moral judgment (quasi-simplex structure of stage inter-correlations, invariant stage progression) and used these predictions as a test of his instrument's validity. When empirical studies corroborated these two predictions, Kohlberg concluded that both his theory and his measure were valid. Because in this validation, theoretical considerations are decisive, we call this *theoretical validity*, as distinct from *empirical validity*, which means merely the correlation with established criterion behavior. Such simultaneous validation of theory and measure, which Kohlberg called *bootstrapping*, however, has its drawbacks:

The first drawback is that, if prediction fails, we are left with no unambiguous cues whether the theory *or* the measure is invalid. "Saving circularity" makes the theory and the measure immune against empirical critique and falsification. For example, if the prediction of invariant sequence fails, that is, if cases of regression occur, we may conclude either that the theory is false *or* that the measure is invalid. When Kohlberg and Kramer (1969) found major cases of regression, they concluded that had to change the theory. Kohlberg augmented his six Stages of Moral Development by a Stage 4½, which could account for cases that appeared to regress from Stage 4 to Stage 2. Yet later they concluded that the finding of regression invalidated the measure rather than the theory (e.g., Kohlberg & Higgins, 1984). They then interpreted cases of regression as signs of "measurement error." So he and his colleagues worked for more than ten years to revise their scoring procedure that

would prevent cases of regression to occur. A high correlation with age became the ultimate criterion for the measure's validity.¹

A test is theoretically valid if, and only if, it really measures what its author intends to measure. Therefore, we can solve the problem of theoretical validity only from our knowledge about the phenomena measured, and not merely from statistical analysis. The development of valid measures and theories is mutually dependent on each other. We couple them in a "bootstrapping process," in which the design of measures draws heavily on the knowledge that we have accumulated about some human behavior, and we test new theories through using those measurement tools. Good measurement tools, it seems, thus contain and reflect all reliable knowledge that is available at some point of time (see Popper, 1968, p. 62).

Both types of measurement imply completely different validation strategies in psychological research and, in particular, in cross-cultural studies. Measurement-as-a-technique implies merely correlational studies. Measurement-as-a-criterion implies many different kinds of checks; it requires *a*) that we prove that the theoretical assumptions on which it is built are empirically true, and *b*) that the content and the design of the measurement are unambiguously linked to that theory. For example, a test of behaviors *structures* must always elicit more than one response because only then we can assess a *relationship* between behaviors. For example, if a man espouses a principled reason for a particular decision we can only infer that he *knows* such reasons. Yet only if he behaves consistently, that is, if the same principled reason lets him reconsider his decision, we can infer that he can act according to principled reasons.

¹ "We wish," says Kohlberg already in 1958, "to provide evidence for ... age differences in various formal attributes of moral thinking" (p. 17). In his latest statement (together with Ann Higgins, 1984) he reiterates this: "Our efforts to define a stage structure clearly differentiating structure from content, and moral judgment from ego development, were motivated by failure of my earlier (Kohlberg, 1958) stage definitions and scoring method to meet the upward invariant sequence hypothesis of stage theory, a defect particularly apparent in charting development after high school." (p. 426). From this and other statements it is quite clear that, for Kohlberg, "data on longitudinal sequence tell us less about truth of theory than about the construct validity of a test based on the theory. [...] Before you try to explain data of change and development with a cognitive-developmental theory, make sure your data can be observed with a measure you have made to fit the sequence rule" (Kohlberg 1984, p. 424).

With the distinction between the two types of measurement, two kinds of moral perspectives are associated: Measurement-as-a-technique leans toward an *ethnocentric* perspective. We use it mostly for selecting or placing people in a certain track of education or career, to optimize the functioning of an organization. Measurement-as-a-criterion implies a *universalistic* point of view. It is used when we want to learn something about the nature of human behavior, about the structure of attitudes in a particular culture, or about an individual personality. This prerequisites norms or standards for measurement on which everyone can agree.

3. Compensatory versus Universalistic Approaches

One can object that the above distinction is too simple and that measurement-as-a-technique can be modified to account for cultural differences just like universalistic approaches or maybe even better.

In its purest form, measurement as a technique disregards any cultural differences, because a selector will see such differences merely as drawbacks, for example, an employer or a school system, and wants to weed them out by the measurement. *Compensatory* views urge us to make the test *fair* for test-takers from different cultural, ethnic or class backgrounds. Compensatory arguments may have different grounds. Some argue *deontologically* that everybody, or at least everyone in our society, should be given a fair chance to make up for specific disadvantages. Therefore, in this view, disadvantages should be compensated for by adding the overall difference between the worse-doing and the best-doing ethnic groups to the individual scores.

Although the compensatory approach to cross-cultural measurement has a great charm, some of these approaches reflects an *ethnocentric* ethics. Rather than fostering their abilities to the point at which they compete successfully with the members of other cultures we fake their test scores to make them appear as if they could compete with the members of other cultures. As a consequence, people may themselves overestimate their abilities and might be overwhelmed by the expectations of others.

When doing measurement in foreign cultures we often take great pains in properly translating our tests to the different cultural experiences but we hardly bother to accommodate the standards of measurement. That is, we do not adapt our criteria for selecting the items or questions, and for scoring the subject's answers. We are often not even aware of the fact that we employ particular rather than universal standards and that these standards may be unfair to the cultures we study.

If we are aware of the particularity of our standards and the differences in other cultures, we usually have one of three options: (1) either we take a *relativistic* perspective and try to accommodate fully to the cultures that we research, or (2) we take a compensatory perspective and try to compensate these differences through some bonus system, or (3) we take a universalistic perspective and use only those standards for measurement which can be thought to be universally valid on ethical and empirical grounds. Each of these perspectives has its merits and its pitfalls. Let us look at them in turn.

If taken seriously, the radical *relativistic* point of view, aptly advocated in many anthropological publications, implies that there is no way of accommodating our cognitive schemes to other cultures, and therefore, there is actually no possibility of doing cross-cultural research. All measurement standards, it is argued, are relative to the culture in which they are conceived. For example, people argue that a white American upper-middle class male psychologist can only conceive measurement standards that are white, upper-middle class, male and American. Whatever such a person comes up with for measuring the moral development of nonwhite, non-American, female, or lower class subjects, his measurement instrument is bound to be culturally biased and *unfair*, and must lead us to wrong conclusions. Although only a few, if any, psychologists endorse such a radical relativistic view in full, some seem to sympathize with it (see, e.g., Shweder, 1982; Snarey, 1985; 1995; Simpson, 1974; Sullivan, 1977).

While relativists rightly point out the pitfalls of ethnocentric and unfair cross-cultural research, I think, they overstate the case. We should not compare people from different cultures in regard to *any* standard which comes to our mind. Many of these standards are specific for our culture and would be an unfair basis for assessing their moral development. However, this fact does not

preclude the possibility that *some* value standards are *universally* valid and thus may provide a fair basis of comparison.

Like the relativistic perspective, the *compensatory* perspective is founded in the belief that all standards for measuring moral development are relative to a particular culture but that some, or most, particularities can be compensated through using the particular moral standards of each culture for measurement, or by a bonus system. To my knowledge, the bonus method of compensating cultural differences is not used in moral development research, so we do not need to explore its merits and shortcomings. However, some suggest the use of culturally specific moral standards. Snarey (1995) urges us to adapt our measurement instruments to different cultural *voices*. Specifically, he maintains that the *communitarian* morality, to be found in many folk societies and in the working-class cultures of our own society, are fully equivalent to the principled morality of the Stages 5 and 6 of Kohlberg's model of moral development. According to Snarey, *communitarian* morality reflects a *Gemeinschaft*- rather than a *Gesellschaft*-type of morality, and it is, as he maintains, only *different* from the middle-class, liberal ideology which is reflected in Kohlberg's stage model but not less developed. Snarey's proposal is based on the sociological theory of *Ferdinand Tönnies*, a German sociologist who published his work at the beginning of this century. Tönnies' distinction between *Gemeinschaft* and *Gesellschaft* has gain some fame among social scientists and influenced our thinking more than many of us are aware of. Among many, the recent *communitarian* movement, initiated by *Armitai Etzioni* (see, e.g., 1993), another German born sociologist, has been very much influenced by this distinction. Etzioni even believes that a communitarian society needs yet to be achieved.

At a closer look, *communitarian* society and morality can hardly be regarded as being as developed as, or even superior to, modern democratic, law-based *Gesellschaft*, or society. *Tönnies*, while being an outspoken social democrat and opponent to the Nazis, his romantic theory of social life helped the enemies of the weak democracy in Germany of that time, finally to destroy it. *Etzioni*'s agreeable theory of a better society seems to be as romantic, and in its consequences as anti-democratic, as Tönnies's plea for *Gemeinschaft* (see Kahne, 1995). As can be inferred on the basis of the few narratives Snarey (1995) provides for in support for his compensatory proposal, it seems doubtful

whether it is philosophically tenable and empirically sound. For demonstrating the moral equivalence of the communitarian voice to postconventional moral reasoning, he quote a working-class subject in Kohlberg's longitudinal study. In responding to the *mercy killing dilemma*, this man said that he wishes "that we could circumvent the court system or the legality of shutting off the machine" (p. 124). Unquestionably, this is, as Snarey asserts, a *classic communitarian statement*. In the eyes of many people this is also a very sympathetic statement. Many of us feel sometimes like this man. But it is certainly not a *postconventional* moral judgment in the sense of Kohlberg's theory. Even scoring it as a Stage 4 response, I believe, would cast doubt on the consistency of Kohlberg's Stage because it expresses a clearly negative attitude toward the democratic legal system.

Moreover, *compensatory* approaches may not really solve the problem but may just shift them in that they merely substitute one ethnocentric bias with another one. How can one white, male middle-class American *know* that his measurement standards are less biased and more fair than the standards of another white, middle-class American?

4. Universalistic Ethics as Standards for Measurement

The last perspective, the *universalistic* perspective acknowledges that people of different cultures, and even people *within* one culture, indeed differ in respect to their moral standards and even with regard to their cognitive structures. But, from this perspective, it is also maintained that there are ways of reaching universally valid, and hence culturally fair standards of measuring moral development. Universalists maintain that there are at least some common moral principles that are hold valid across all cultures, and hence provide a common ground for constructing culturally fair measures of moral development.

Universalists clearly distinguish between *moral* values, which are thought to be universal, from *non-moral* values, which are thought to be relative to a particular society. A test constructed to measure those non-moral values can not be called a fair test of moral development. For example, a person's happiness, tradition, language, inclination to "do favors," solidarity with same-class

people, and many other values which cannot be regarded as *moral* values, are relative to a particular culture, and hence, cannot be taken as a basis for measuring moral development across cultures.

From a universalistic point of view the construction of a culturally fair test of moral development must be based on *universally valid* standards, that is, on standards which are fair for *everybody* regardless of the particular non-moral values hold by a person with different culture, class-membership, and so on.

How can we know that a value is a truly *moral* value, that is, a value that is universally valid and can thus be made the basis for constructing a measure of moral development? There are two strands of justification for arriving at an answer to this important question. One strand employs rational judgment, the other empirical evidence. None can replace the other, both, as it turns out, are needed as they supplement each other in bringing about a culturally fair test. Rational judgment cannot totally replaced by empirical evidence, nor can it be made the sole basis of cross-cultural measurement.

5. The *Rational* Justification of Cross-Cultural Measurement of Moral Development

The rational strand of justifying the universalistic claim is derived from the definition of moral values as those principles of behavior which can be universalized. The probing question is that of the *Golden Rule* (Don't do to others what you do not want them do to you) and Kant's *Categorical Imperative*, which says: Let your behavior guide only by those *maxims* which you can wish to become the basis of a generally binding law. Accordingly, one can probe into the universalizability of one's standards for measuring moral development by asking: *Could one wish that the standards I employ in scoring the behavior of subjects from other cultures, be made a universally valid criterion for assessing everybody's moral development, including my own - behavior?*

This Golden Rule of culturally fair measurement implies concrete courses of action: *Take only those standards which can be thought to apply always, regardless of the conditions you live in. Take only those which you also could want to be measured by yourself.*

The six “stages” or modes of moral reasoning suggested by Kohlberg (1981) may serve as a starting point for such a rational justification. Kohlberg founds this stage-model on long philosophical traditions as well as on rational argument. He points to many eminent philosophers like Kant and Habermas, sociologists like Comte and Hobhouse, and psychologists like John Dewey and Jean Piaget, who made similar claims. He argues that the Stages 1 to 6 reflect a sequential ordering modes of moral reasoning from low to high. Each stage, he Kohlberg reasons that each stage can be thought to emerge from a previous stage as providing more adequate, and more just solutions for solving moral problems. He argues that “principled morality” (Stages 5 and 6) provides an ideal endpoint of moral development.

Hardly any other measure of moral development has been that well justified on rational-philosophical grounds. Think, for example, of the *behavioristic* standards which suggest that the degree of a person’s morality can be simply inferred from the fact whether he or she keeps or breaks socially defined rules like “Do not steal.” Of course, this is important rule but can one really infer from such kind of behavior alone the moral development of a person or even of a culture?

The answer can only be yes, if we can assume that stealing reflects nothing but the morality of the thief. This assumption stands on weak grounds. One would have to presume that thieves steal because they think that stealing is right to do. But no thief can want to make theft an universally accepted rule; no thief wants to be robbed him- or herself. What makes people become thieves seems to have mostly non-moral reasons that reside partly within the person, and partly outside. An example for a non-moral reason for stealing residing inside a person is his or her inability to understand the importance of rules that protect property. There are many very poor people who do not transgress the law, because they understand that the law, as badly as it may often be enforced, protects also their property and their life. But many, lacking parents and education, may not be able to understand that, and therefore see no reason why they should keep the law.

An example for a non-moral reason residing outside a person is extreme poverty which may make people think that the moral obligation to preserve their own life and the life of their family members, outweighs the obligation to respect the property of other people. Shortly after World War II, when food was

very scarce in Germany, and people were starving, many people who had never transgressed the law before and who never did it afterwards, stole food at least sometimes. Another, contemporary example is given by Nucci (1995) who reports about a recent study of Brazilian adolescents who were involved in many forms of unlawful behavior. It is noteworthy that they found hardly youth who thought that his behavior was morally right, but nearly all felt that it was immoral. These findings are not surprising given the fact, as Turiel (1983) and Nunner-Winkler and Sodian (1988) found in their research, that most children know, understand and appreciate moral principles already at a very young age, much earlier than the studies by Piaget and Kohlberg have suggested.

Another example is the cheating behavior of the kids who took part in Hartshorne and May's (1928-30) experiments cannot be taken as a fair and unambiguous basis for the measurement of these children's moral character. Ironically, the experimenters themselves lied to the children in order to do their experiments. to be able to observe their unmoral behavior. The researchers felt they had good moral reasons for doing so; but the children might also have given them good reasons for their rule-breaking behavior, if only they had ask them.

In sum, the rate of rule-keeping and rule-breaking behaviors in a society cannot be regard as a fair index of the moral development of that culture because is reflects not only the morality of the actor but also his or her ability to understand and to apply moral rules, and the conditions of that person's life.

6. Applying the Golden Rule of Measurement

What then can be taken is a valid and culturally fair index of moral development, if neither *moral behavior* in the sense of rule-keeping behavior not *moral attitudes* and *values* qualify for this?

In line with the proposal Kohlberg (1964), I suggest to use moral judgment *competence* as an index for cross-cultural comparison of moral development. Moral judgment competence is defined by Kohlberg (1964) 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).

This definition has two big advantages over the other definitions. First, moral competence is defined here in regard to the individual's *own* moral standards and thus fairer as in those definitions which define moral development purely in terms of outside moral standards (for an apt critique of externalistic approaches of moral measurement, see already Pittel & Mendelsohn, 1966). Any measurement instrument based on this definition thus takes the particular moral standards of a culture into account by virtue of the test's design. Kohlberg's own *Moral Judgment Interview* (Colby et al., 1987) attends to the moral standards of the individual to some degree, but this is not done very explicitly, and not as much as his definition of moral judgment competence seems to imply. As many critiques noted, a subject *has* to prefer Stage 5 and 6 moral values in his or her reasoning for getting a high score on Kohlberg's developmental scale.

We have, therefore, designed a new test, the *Moral Judgment Test* (MJT) in which the subject can prefer any of the six stages for basing his or her reasoning on, but has merely to demonstrate through his or her competence to apply his or her moral principles *consistently* (Lind, 1978; 1993; Lind & Wakenhut, 1985).

With the MJT it is not necessary to defend the universal validity of the stage ordering because the scoring of this test is neutral regarding any particular stage ordering. However, like any competence test, the MJT cannot be without normative reasoning. Why should we want that people have a high moral judgment competence? Why can we assume that this normative decision is universally valid and thus fair when applied to different cultures?

The answer to these questions is based on the believe that unless people have developed a high competence in applying their moral principles and solving the moral dilemmas which they inevitably will encounter in their life, they will not be able to solve such conflicts and problems in a peaceful, nonviolent way, but must use brutal force and wars in order to succeed (see Habermas, 1983; Lind, 1993). Further support for this claim is found in the empirical findings of the past twenty years of research, which will be summarize below.

7. Universal Empirical Findings as Standards of Measurement

The standards resulting from the above considerations may be called *philosophically fair*. In many instances, philosophical fairness is probably all we can ask for, and is the only kind of fairness we can achieve. But you may still have doubts as to whether thought experiments alone can produce standards for measuring moral development which are really and under all circumstances fair. Each of us has only a limited ability, and a limited inclination, to put ourselves fully into the shoes of someone else, or even into that of another culture. Thus, it would be of great value if the outcome of such thought-experiments, could be checked empirically if such a check is available.

In his review article of cross-cultural research using the Kohlberg-interview method, Snarey (1985; 1995) concludes that Kohlberg's claim has only partially been supported by empirical research. In most longitudinal studies subjects move through the developmental stages in the predicted order but their development seems to stop at different Stages. Frequently this variation in the terminal point of moral development (when measured with the MJI) is seen as being linked to differences in the cultural or class background of the subjects (Snarey, 1995; Edwards, 1986). This finding, therefore, is mostly interpreted as showing that Kohlberg's standards for measuring moral development are *biased* toward Western cultures, liberal middle-class ideologies or alike, being thus unfair to the actual level of moral competence people with other cultural and ideological backgrounds are capable of (Simpson, 1974; Sullivan, 1977; Snarey, 1995; Vine, 1983).

This critique needs careful evaluation before any conclusions can be drawn. In some respects it seems to have serious shortcomings itself. In other respects it does not seem radical enough:

1. This critique does not deal adequately with Kohlberg's claims. He never claimed that everybody will reach the same stage of moral development but that everybody is passing through the stages in the same order.

2. In hardly any of the critical analyses, the role of education is examined. The rare studies which include level of education, indicate that not differences in culture or class-membership but level of education accounts for the differences in the speed and termination of moral development.

3. Most critics focus solely on the Moral Stage index and ignore all other aspects of moral development which have been researched and which may be universally valid, such as the hierarchy of stage preferences.

4. In hardly any of the critical appraisals, the more radical critique by Pittel and Mendelsohn (1966) is acknowledged. They convincingly argued that the question of fair measurement is not confined to foreign cultures but extends to any other person. How can we fairly assess another person's morality at all? Pittel and Mendelsohn's (1966) suggest a way of a measuring moral development which would let us, at least partly, take a person's own moral principles into account. They recommend tests in which "items should not deal with differences between right and wrong but should [. . .] be concerned with 'conflicts between right and right or those between wrong and wrong'" (p. 26). They concluded, "to date no one seems to have constructed a test of moral attitudes based exactly on these prescriptions" (p. 26).

This was in 1966. Today, we have several test at hand which take up their proposal. For example, the C index of the MJT is designed to compare the person's responses in regard to "differences between right and right" as well as to "differences between wrong and wrong," rather than just in regard to right or wrong. The MJT's C index is computed *independently* of a person's particular moral value system, so it is not bound to culturally specific ways of solving moral dilemmas. A subject can get a high moral competence score regardless to the level of moral reasoning he or she prefers. The C index reflects a person's ability to apply consistently his or her *own* moral values to a decision making process. More precisely, the C index reflects the subject's competence to evaluate of arguments no matter his or her own opinion about that solution. In other words, the C index does not reflect a person's opinion on moral issues, nor does it reflect his or her preferences for certain stages of moral reasoning. It measures the degree to which a person basis his or her judgment on *moral* considerations rather than on the consideration whether a given argument agrees or disagrees with her or his own opinion.

8. Empirical Research

Taking together, the findings of research with cognitive-developmental measures of moral development show that moral development has some aspects which are universally valid (see Lind, 1993; in press):

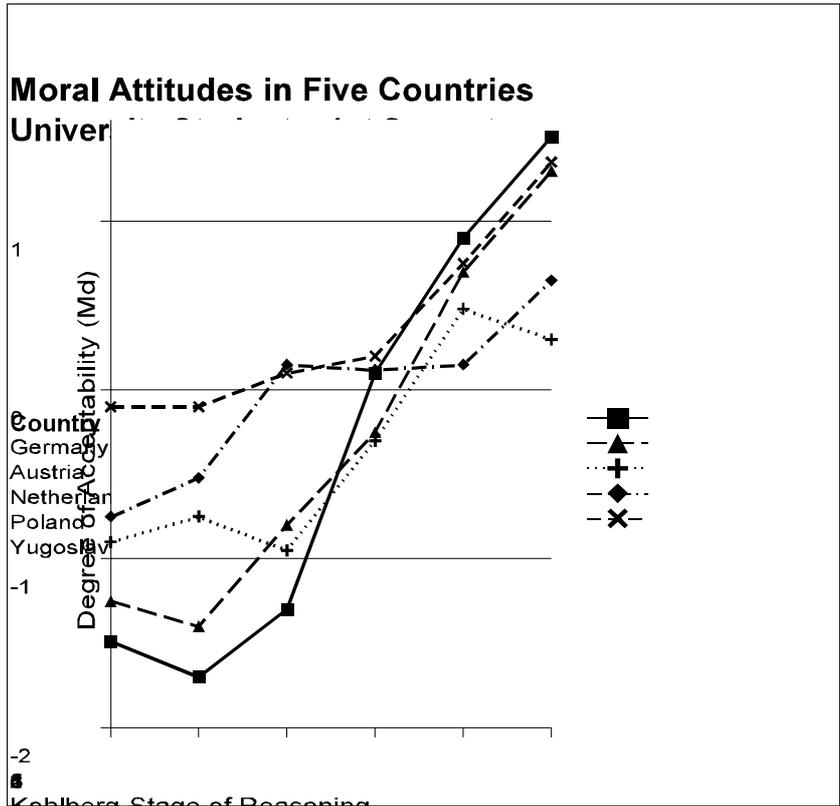


Abb. 2 Source: Lind (1993), p. 145

- Across all cultures studied, the preferences of the six Kohlbergian levels or stages of moral reasoning are indeed ordered as theoretically predicted (with stage 6 being preferred most, stage 5 second most etc.), though some small inversions of stage preferences (especially between stages 1 and 2, as well as between stages 5 and 6) occur (see graph).

- In all studies, independent of culture, the stage preferences are inter-correlated as to form a “quasi-simplex.” Neighboring stages (for example, stages 5 and 6) correlate higher than more distant stages (for example, stages 4 and 6).
- Affective-cognitive parallelism: The stage preferences correlates in a predicted manner with the MJT measure of moral judgment competence, i.e., the preference for the highest stages correlates highly positively with the competence score, the preferences for the lowest stages correlates highly negatively with that score, and the other preferences measures show correlations in between these extremes.
- Correlation with level of education. The median or mean moral judgment competence score varies in between 10 points and 50 points, and this variation depends mostly on the degree of education the subjects have. The correlations of MJT scores with many other variables studied, like subjects' age, gender, socio-economic level, and culture, have shown to be mostly spurious, that is, they can be almost completely accounted for by level of education.

All these findings could be replicated in numerous independent studies (for a summary, see Lind, 1993) and agree with most studies using other measurement instruments like the MJI and the DIT. Because they are so highly replicable, these findings are now used as empirical criteria for the 'cross-cultural validity' of the *Moral Judgment Test*. Although we regard theoretical criteria as the ultimate standards for validating the MJT, we recommend to check them before using a translated MJT version in research. In fact, they have proven to be very useful in detecting inadequate translations of test items when creating foreign language versions. Moreover, these findings support the claim that we can measure moral development cross-culturally in a valid and fair way.

9. Discussion

The issue of cultural fairness is an important topic of moral development research. Psychological and educational measurement always involves the application of behavioral standards. Therefore, the question is this: How can we measure moral development without imposing some particular views on other people? Which standards of measurement are legitimate, that is, can be

justified by a universal measurement principle? I have discussed four different perspectives on this issue, the ethnocentric, relativistic, compensatory and the universalistic view. Though each perspective has its merits and good reasons, I argued that only some kind of universalistic perspective can provide a fair basis for cross-cultural moral research. Kohlberg's six stages of moral reasoning seems to provide such a universalistic standard not only for philosophical reasons (Kohlberg, 1981) but, more importantly, also for empirical reasoning. Philosophical reasons can always be challenged on philosophical grounds. Empirical findings, if well replicated, appear to be a more stable grounding.

Cross-cultural research of the past twenty years supports this claim. Most notably, and most astonishingly, almost all cultures show a similar ranking of the moral ideals described in Kohlberg six-stage-model. Nearly everybody agrees on which level fundamental moral dilemmas are to be discussed. This fact is important for measurement because it provides universalistic standards for scoring. And it is important for education because it provides a common basis for teachers and students to engage in a moral education process.

However, People differ greatly regarding their ability to act upon their *own* moral values or principles. These differences might be largely due to differences in the quality and quantity of education they have experienced. They are hardly due to culture. In fact, most studies, as imperfect they are, seem to show that so-called cultural differences are mostly due to educational differences (Lind, 1993; in press).

So, we need to discuss the question of cultural fairness in the light of social inequality and educational opportunities. Empirical research, we have seen, shows that most, if not all, people want to act upon their moral ideals. Therefore, it seems *fair* to measure their ability to live up to their moral ideals. (This does not free us from the problem of validity, though.) We must, however, reflect see those educational differences on the background of growing educational disparities within and between cultures (Triandis, 1996). These disparities seem to be responsible not only for a great amount of the differences in mathematical and language proficiencies (Darling-Hammond, et al., 1996) but also for most variation in moral and civic competencies. These new research findings should compel us to think about policies through which we can close these disparities within the next decade.

References

- Child, I. (1954). Socialization. In G. Lindzey, Ed., *Handbook of Social Psychology. Vol. II*, pp. 655-692. Cambridge MA: Addison-Wesley.
- Colby, A., Kohlberg, L., Speicher, B., Hewer, A., Candee, O., Gibbs, J. & Power, C. (1987). *The measurement of moral judgment: Vol. 1 & 2*. New York: Cambridge University Press.
- Coombs, C.H., Dawes, R.M., & Tversky, A. (1970). *Mathematical Psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Darling-Hammond, L. & Aness, J. (1996). Democracy and access to education. In R. Soder, Ed., *Democracy, education, and the schools*, pp. 151-181. San Francisco, CA: Jossey-Bass.
- 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
- 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.
- Etzioni, A. (1993). *The spirit of community: Rights, responsibilities, and the communitarian agenda*.
- Gibbs, J.C. (1979). Kohlberg's moral stage theory. A Piagetian revision. *Human Development*, 22, 89-112.
- Gielen, U. (1984). Cross-cultural research on moral reasoning and Stage 5: Comments on Ian Vine's paper (pp. 77-84). In G. Lind, ed., *Beiträge zum Zweiten Konstanzer Werkstattgespräch über Moral und Umwelt*. Universität Konstanz: Bericht.
- Gielen, U.P. & Markoulis, D. (1993). Preferences for Principled Moral Reasoning: A Developmental and cross-cultural perspective (pp. 73-87). L.L. Adler & U.P. Gielen eds., *Cross-cultural topics in psychology*. Westport, CT: Praeger.
- Habermas, J. (1983). *Moralbewußtsein und kommunikatives Handeln*. Frankfurt: Suhrkamp.
- Hartshorne, H. & May, M.A. (1928-30). *Studies in the nature of character. Vol. I: Studies in deceit. Vol. II: Studies in service and self-control. Vol. III: Studies in organization of character*. New York: Macmillan.
- Hogan, R. (1970). A dimension of moral judgment. *Journal of Consulting and Clinical Psychology* 33, 307-316.
- Kahne, J. (1995). Lacking democracy: Etzioni's 'spirit of community.' Unpublished manuscript, University of Illinois-Chicago, College of Education.

- Keesing, R.M. (1981). Theories of culture. In R.W. Casson, Ed., *Language, culture and cognition*, pp. 42-66. New York: Macmillan.
- Kohlberg, L. (1964). Development of moral character and moral ideology. In M.L. Hoffman & L.W. Hoffman, Eds., *Review of Child Development Research, Vol. 1*, pp. 381-431. New York: Russel Sage Foundation.
- Kohlberg, L. (1981). Essays on moral development. Volume I: The philosophy of moral development. San Francisco: Harper & Row.
- Kohlberg, L. & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral. *Human Development, 12*, 93-120.
- Kohlberg, L. & Higgins, A. (1984). Continuities and discontinuities in childhood and adult development revisited - again. L. Kohlberg, *The psychology of moral development* (pp. 426-497). San Francisco, CA: Harper & Row.
- Lind, G. (1985). The theory of moral-cognitive judgment: A socio-psychological assessment. In G. Lind, H.A. Hartmann & R. Wakenhut, eds., *Moral development and the social environment. Studies in the philosophy and psychology of moral judgment and education*. Chicago: Precedent Publishing Inc., S. 21-53.
- 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. (1993). *Moral und Bildung. Zur Kritik von Kohlbergs Theorie der moralisch-kognitiven Entwicklung*. Heidelberg: Asanger.
- Lind, G. & Wakenhut, R. (1985). Testing for moral judgment competence. (pp. 79-105). In: G. Lind, H.A. Hartmann, & R. Wakenhut, eds., *Moral development and the social environment*. Chicago: Precedent Publ. Inc.
- Lind, G., *The psychology of moral judgment competence*. Hillsdale NJ: Erlbaum (in press).
- Maller, J.B. (1944). Personality tests. In: J. McV. Hunt, ed., *Personality and the behavior disorders* (pp. 170-213). New York: Ronald Press.
- Meehl, P. E. (1988). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50*, 370-375.
- Niemczyński, A., Czyżowska, D., Pourkos, M. & Mirski, A. (1988). The Cracow study with Kohlberg's moral judgment interview: Data pertaining to the assumption of cross-cultural validity. *Polish Psychological Bulletin, 19*, 43-53.
- Nucci, L. (1995). Moral development and character formation. In: H.J. Walberg & G.D. Haertel, eds., *Educational psychology: Effective practice and politics*. Berkeley, CA: McCutchan (in press).
- Nunner-Winkler, G. & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development, 59*, 1323-33.
- Pittel, S.M. & Mendelsohn, G.A. (1966). Measurement of moral values: a review and critique. *Psychological Bulletin, 66*, 22-35.

- Rogoff, B. & Chavajay, P. (1995). What's become of research on the cultural bias of cognitive development?. *American Psychologist*, 50, 859-877.
- Simpson, E.L. (1974). Moral development research: A case study of scientific cultural bias. *Human Development* 17, 81-106.
- Shweder, R. (1982). Beyond self-constructed knowledge: The study of culture and morality. *Merrill-Palmer Quarterly*, 28, 41-69.
- Snarey, J. (1985). The cross-cultural universality of social-moral development: A critical review of Kohlbergian research. *Psychological Bulletin*, 97(2), 202-232.
- Snarey, J. (1995). In a communitarian voice: The sociological expansion of Kohlbergian theory, research, and practice. In: W.M. Kurtines & J.L. Gewirtz, eds., *Moral development: An introduction*. (pp. 109-133). Boston, MA: Allyn & Bacon.
- 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.
- The American Heritage Electronic Dictionary (1992), Third Edition. Houghton Mifflin Company.
- Triandis, H.C. (1996). The psychological measurement of cultural syndromes. *American Psychologist*, 51, 407-414.
- Turiel, E. (1983). *The development of social knowledge. Morality and convention*. Cambridge, MA: Cambridge University Press.
- Vine, I. (1984). Moral maturity in socio-cultural perspective: Are Kohlberg's stages universal? In: S. Modgil & C. Modgil, Eds., *Lawrence Kohlberg: Consensus and controversy*. Brighton GB: Falmer Press.

Notes

1. Pittel and Mendelsohn (1966) criticized the research of their time as being confined to "externalistic." By that they meant that the standards for judging someone's morality were taken solely from sources outside the person being measured and that the person's own moral standards were never taken into account. Their critique applies to many otherwise very unlike research paradigms. Within the behavioristic research paradigm, a subject's behavior is judged as right or wrong only in regard to the external social standards that the researcher holds (see, e.g., Hartshorne & May, 1929). Similarly, within the attitude paradigm, the statements are selected, and the subject's responses are scored, only in regard to standards external to him or her (see, e.g., Hogan, 1970). Even within the clinical interview paradigm, advocated by Jean Piaget and Lawrence Kohlberg, the standards for scoring answers as high or low on a moral development scale are chose solely from an externalistic point of view but are not checked for the degree to which they may coincide with the subject's own standards. Pittel and Mendelsohn did not rule out any externalistic considerations or elements in the process of scale construction but they were concerned with the fact that most, if not all, measurement of a person's morality was done from an externalistic point of view, not caring for the moral standards which the subjects would think to be legitimate. They believed that this negligence contributed much to the slow progress of moral research.