

A Critical Review of Studies on Moral Judgment Development Using the Defining Issues Test in Arab Countries*

Ramadan A. Ahmed

Dept. of Psychology
College of Social Sciences
Kuwait University

Uwe P. Gielen

Dept. of Psychology
St. Francis College
New York, U.S.A.

Abstract

This article reviews published and unpublished studies employing Rest's Defining Issues Test (DIT) in various Arab countries. The DIT is based on Kohlberg's developmental theory of moral reasoning and aims to measure an individual's preference hierarchy for stage-typical moral arguments. On the basis of a detailed and cross-culturally oriented review of specific findings, it has been noted that a number of methodological problems make it difficult to arrive at definite conclusions about the usefulness of the DIT in Arab cultural settings. The results of the studies under review suggest that other Kohlbergian measures, such as Gibbs' Social Reflection Measure-Short Form (SRM-SF) may be more useful than the DIT for studies aiming to investigate the moral development of Arab respondents.

1. Kohlbergian Theory and the Defining Issues Test

The American psychologist Lawrence Kohlberg (1984) has developed a theory that conceptualizes the structures and processes underlying the development of moral reasoning. As a follower of Piaget, he assumes that development occurs through a series of cumulative stages that vary in complexity. Development is said to follow a sequence of stages and to occur in the same manner in all societies regardless of cultural differences. The speed and endpoint of development through the stages vary from person to person and from society to society, but the basic sequence of stages is assumed to be universal. Kohlberg emphasizes that his assertions apply only to the *structures* of moral reasoning whereas the specific *contents* of moral decision-making may vary both from person to person and from culture to culture.

Kohlberg's six stage model traces development from the pre-conventional stages 1 and 2, to the conventional stages 3 and 4, and finally to the principled stages 5 and 6. At the pre-conventional stages a person makes moral decisions from the point of view of an individual actor who tries pragmatically to avoid trouble, to satisfy his or her needs, and to follow practical rules of fairness, exchange, and social acceptability. At the conventional stages a person has internalized interpersonal expectations of societal virtues, religious prescriptions, and the duties, obligations, and rights prevailing in his or her culture or subculture. The person tries to be morally good by responding to the justified expectations of others and those of society. At the principled level of morality a person has developed inner, abstract principles of justice that are used to make decisions in difficult situations involving moral conflict. Such moral principles appeal to notions of human dignity, ultimate equality under the moral law, moral autonomy, and conceptions of an ideally just society. Moral decision-making is expected to be sharable or universalizable, representing an effort to reach consensus based on nonarbitrary social cooperation.

Kohlberg has developed a measurement instrument designed to assess the moral stages of reasoning, the Moral Judgment Interview (MJI). It is based on the interview method, relies on a very complex coding manual, and is quite cumbersome to use (Colby & Kohlberg, 1987).

Given the complexities of this measurement instrument, his student James Rest (1979) has developed a new technique to assess the development of moral reasoning stages, the Defining Issues Test (DIT). The DIT is an objective test containing six brief moral-political dilemmas which are similar to those employed in Kohlberg's MJI. After each dilemma, there

follow 12 moral arguments that could be used to solve the dilemmas. The arguments vary in structural complexity and represent Kohlberg's moral reasoning stages 2, 3, 4, 4^{1/2}, 5A, 5B, and 6. In addition, the test contains three so-called consistency checks in order to discover whether the respondent understands the meaning of the arguments and rates and ranks the arguments in a consistent and meaningful way. Included in the consistency check are some meaningless but lofty sounding arguments. If a person chooses several of these arguments he or she is eliminated from the subject pool because it is unclear whether the person fully understands the moral arguments contained in the test. The respondent is also asked to rate and to rank the same series of moral arguments. If a person rates and ranks the same arguments in a very different way he or she is also eliminated from the subject pool because it is unclear whether the person fully understands the test. Alternately, such a person may simply be careless in making moral choices. The "consistency checks" are of special importance in cross-cultural research because they help the researcher decide whether the DIT is an appropriate test in a given culture.

After the test-taker has rated and ranked all moral arguments contained in the test, his or her moral choices are integrated across all dilemmas and moral arguments by using an objective scoring procedure. The scoring procedure determines the percentage of the moral arguments chosen by the respondent for each of the moral stages. In this way, one can establish for each respondent a "moral judgment profile" representing all moral stages preferred by the respondent. However, the moral stages are unequally represented on the test. Conventional moral arguments (stages 3 and 4) are most frequently represented followed by postconventional arguments (stages 5A, 5B, 6) and preconventional arguments (stage 2).

Rest (1979) recommends the so-called P%-Score as the best overall indicator of a person's moral judgment maturity. The P%-Score refers to the combined percentage of moral arguments chosen by a person which represent stages 5A, 5B, and 6. In other words, the P%-Score indicates how much the person prefers principled moral arguments over arguments at the lower preconventional (stage 2 in the DIT) and conventional (stages 3 and 4) stages.

Another overall indicator which can be obtained from the test is the D-Score. It integrates a person's ratings of all presented moral arguments into an overall developmental score (Rest, 1986a).

Critics of Kohlberg and Rest such as Vine (1986) have argued that Kohlberg's and Rest's conceptions of principled morality are biased toward



a specifically Western, male, upper-class ideology of moral autonomy and individualism. The critics claim that Kohlberg's and Rest's tests of moral reasoning set up an ethnocentric standard of moral maturity that should not be applied to non-Western cultures. However, there exists a broad array of cross-cultural studies whose results are at variance with most claims of the critics. These studies have been reviewed by Gielen (1996a), Gielen and Markoulis (2001), Moon (1986), Walker (1988), and others. A special issue published by *World Psychology* provides an up-to-date overview of the strengths and weaknesses of Kohlbergian theory and research (Gielen, 1996b).

In the following, the authors review a series of studies on moral judgment and moral reasoning development carried out in Egypt, Saudi Arabia, Algeria, Kuwait, the Sudan, and among Palestinian Arabs using Rest's (1979) Defining Issues Test (DIT) as adapted to the milieu of the Arab world. The review also attempts to explain the discrepancies in the results obtained by psychologists when attempting to determine the patterns of moral reasoning development and socialization in various Arab countries. It is argued that methodological problems account for some of the discrepancies while other findings suggest that the DIT may not be a very good measure of moral judgment development in the Arab world. In addition, suggestions are made toward the further application of moral judgment measures in the Arab countries.

This article stands in the tradition of cross-cultural psychology which, following Ho and Wu (2001, p. 4), may be defined as the "scientific study of human behavior and mental processes, including both their variability and invariance, under diverse cultural conditions." One important aim of cross-cultural psychology is to investigate whether postulated psychological principles and processes—such as the development of moral reasoning—hold true in a broad variety of cultural settings, and whether such processes are influenced by cultural variables.

In the following, the authors review the results of research projects conducted in various Arab countries that are, at times, compared to the results of research projects conducted in Western and East Asian countries. Such cross-cultural comparisons must be done in a critical manner. Specifically, the DIT was developed in the United States and it is therefore important to ask whether the test provides comparable estimates of moral reasoning development among Arab and non-Arab respondents. For an international review of DIT studies, Gielen and Markoulis (2001) should be consulted.

2. A Review of DIT Studies Conducted in Arab Countries

In a pioneering, small-scale, cross-cultural study, Ismail (1976) found that cultural differences do exist in the moral development of American and Saudi Arabian Muslim respondents. As Bouhmama (1984) has pointed out, Ismail's findings may reflect some combination of religious as well as general cultural differences between the two groups of respondents.

In 1983, Webb and Steentsma administered the DIT to 259 male and female Christian and Muslim Palestinian university students. They found that the correlation between years of schooling and the P%-Score was very low, though statistically significant ($r = .16m$ $p < .05$). Age and P%-Score did not correlate significantly with each other. Practitioners of religion received lower scores than non-practitioners. The Palestinian students received an overall P%-Score of 29.35%. (Since the authors did not indicate an overall P%-Score, we computed the score from their data.) The P%-Score of 29.35% is much lower than the mean P%-Score typically reported for comparable North American samples which have been summarized by Rest (1986a, 1986b).

Jessup (1984) administered the DIT to 40 multinational students in 9th grade English classes at an American school in Kuwait. They were around 15 years of age. The purpose of the study was to investigate the moral development of multinational students in Kuwait in a culturally disparate educational setting. The study also aimed at determining both the effects of knowledge of the Kohlbergian theory of cognitive moral development gained through a short training period and of a request to "fake high" on the DIT's principled moral thinking items (P%-Score).

It was found that the respondents had a P% mean of 17.06 on the DIT pretest. This score falls within the range of the standard deviation for American peer groups although at the lower end. It was also found that the discussion of Kohlberg's moral stage theory did not significantly increase the participants' ability to select principled items. No significant advantages for the experimental group were found when examining the means of their post-test scores.

Finally, the pretest P%-Scores and D-Scores were examined for possible effects of sex, nationality, and language. No significant differences emerged in the statistical tests, although the relatively small number of respondents should be kept in mind in this context.

In general, the results of the study suggest that experience in an American educational setting may have accounted for the approximate comparability of multinational respondents to their American peers of

comparable age. The failure of the treatment to significantly enhance scores may reflect the duration of the treatment, and the inability to “fake high” supports the general theory of a sequence of cognitive stages in the development of moral judgment.

In 1985, El-Sheikh administered the DIT to 333 Egyptian male and female intermediate, secondary school, and college of education students ranging in age from 15 to 32.5 years. The results showed that participants preferred conventional moral judgments (stages 3 and 4 = 50%) over principled arguments (stages 5 and 6 = 20-30%) and preconventional arguments (stage 2 = 12%). Table 1 depicts the means and standard deviations for the various moral judgment scores of his respondents. They are reported separately for five age/educational groups and for males and females.

Table 1
Means and Standard Deviations of the Egyptian Sample
(N = 333) (El-Sheikh, 1985)

Moral Stages	Mean Age	N	Stage 2		Stage 3		Stage 4		P		P%* Score
			X	SD	X	SD	X	SD	X	SD	
Intermediate School 13-15 Years											
Males	15.00	49	4.18	2.19	15.10	5.01	21.10	5.42	12.10	3.86	20.23%
Females	15.00	40	4.50	3.06	16.90	5.16	17.30	6.55	13.90	5.66	23.08%
Secondary School 13-18 Years											
Males	17.00	57	4.93	3.76	16.00	4.69	17.00	4.93	15.50	5.70	25.83%
Females	17.50	64	5.10	3.20	17.00	4.80	19.00	4.38	11.10	3.77	18.50%
University Students 18-21 Years											
Males	22.00	23	3.37	2.24	13.60	4.82	21.40	5.42	14.70	4.23	24.53%
Females	21.50	19	3.70	1.74	14.40	4.37	24.00	4.09	13.10	3.18	21.90%
One Year Post-Graduate 21-22 Years											
Males	24.00	21	4.85	3.69	15.10	6.43	20.60	6.75	13.30	6.12	22.16%
Females	24.50	20	7.30	3.14	14.40	6.78	18.60	5.15	13.80	5.08	23.00%
Special Diploma 22-35 Years											
Males	32.50	23	3.83	3.35	10.50	5.20	21.30	6.53	18.00	5.56	29.93%
Females	27.50	18	6.87	4.34	15.80	6.36	15.80	5.97	13.00	0.43	21.67%

Note: P%-Scores were computed by the present authors based on El-Sheikh's (1985) study

An analysis of variance indicated that there were significant differences between males and females concerning stage 3 arguments in favor of the female subjects. A significant difference ($p < .01$) was found between males and females at stages 5 and 6 indicating that males preferred principled arguments more frequently than females.

The participants preferred preconventional moral arguments (stage 2) much less than the conventional and principled arguments. The percentage of the preconventional stages in any of the subgroups was not more than 12%. However, it should be kept in mind that the DIT contains more conventional than postconventional items, and more postconventional items than preconventional items.

Level of education appeared to contribute positively to the students' moral judgment level since according to the author's interpretation the better educated participants tended to receive higher moral judgment scores. In this context, El-Sheikh reports that the difference between intermediate and secondary school students on one hand and the university students on the other hand was significant at the .01 level. However, and in contrast to El-Sheikh's interpretation, the computation by the present authors of the mean P%-Scores for the different educational groups does not support the idea that clear developmental trends can be found in the results of El-Sheikh's study. For instance, the average P%-Score for secondary students is 21.95%, the average P%-Score for the university students equals 23.34%, and the average P%-Score for one-year postgraduate students is 22.57% (all scores computed by us on the basis of the data depicted in Table 1).

In 1985, Eissa used an adapted Arabic version of the DIT to study the impact of education on the development of moral judgment. He compared the responses of his small sample ($n = 15$) of senior males, majoring in Arabic, aged from 23-27 years, attending the College of Education, Tanta University, Egypt, and having a pre-college education at a teachers' institute (secondary level) with those of their peers ($n = 16$) having no college level training. Both groups were matched for the variables of sex, grade point average, and years of pre-college certification study. Eissa proposed that most adults reach a general plateau in the development of their moral judgment after their formal schooling and that previous research results in Western studies demonstrate a strong positive relationship between level of moral judgment and years of education (Rest, 1979, 1986a, 1986b).

Table 2
Distribution of Moral Stage Scores According to Eissa (1985)

Stage	College (N = 15)		Non-College (N = 16)	
	Mean	%	Mean	%
Stage 2	3.20	5.33%	5.87	9.78%
Stage 3	6.93	11.55%	13.94	23.23%
Stage 4	21.53	35.88%	24.75	41.25%
Stage 4 ^{1/2} (A)	-	-	-	-
Stage 5A	12.33	20.55%	6.31	10.52%
Stage 5B	2.53	4.22%	2.56	4.27%
Stage 6	5.26	8.77%	1.69	2.82%
P-Score (Principled)	20.13	33.55%	10.56	17.60%
Meaningless Items	2.46	-	1.75	-
P-Score (Overall Weight)	-	-	-	-

Note: Percentage were computed by the present authors based on Eissa's (1985) study.

A comparison between the two means of the P%-Score shows a significant difference in favor of the college group ($p < .001$). Their mean P%-Score was 20.13 (SD = 4.33) as contrasted with the mean P%-Score of the non-college group which was 10.56 (SD = 3.14). A further analysis of other stage scores indicated a significant change in the level of moral judgment according to the number of years spent in formal education at the College of Education. It should be added that the P%-Scores reported by Eissa are usually low when compared to scores typically obtained in the United States and elsewhere (Gielen & Markoulis, 2001).

In a later study, Eissa (1993) administered a locally adapted version of the DIT to 37 male and 40 female Egyptian secondary school students ranging in age between 15.4 and 17.3 years. The students also responded to Barron's Ego-Strength Scale, a scale that had shown good psychometric properties in previous Arab research. The correlation between the P%-Score and the Ego-Strength scale was $r = .59$ pointing to a strong relationship between the two variables. In addition, the researcher administered the DIT to 43 participants on two separate occasions. The test-retest reliability coefficient over a time period of two weeks reached a very convincing $r = .97$.

Ahmed, Gielen, and Avellani (1987) administered the DIT to 679 Sudanese arab male and female intermediate, secondary school, and

university students. Sixty-two percent of the protocols did not pass the standard inconsistency check recommended by Rest (1986a). These results suggest that the DIT was a very difficult and unusual test for most of the Sudanese respondents. P%-Scores and D-Scores showed low but statistically significant correlations with age for the remaining 147 male students, but for the remaining 102 female students the correlations were not significant. The results for the Sudanese students indicated that they strongly preferred conventional moral judgments (stages 3 + 4 = 57%) over principled arguments (stages 5A, 5B, and 6 = 26.6%) and preconventional arguments (stage 2 = 5.7%).

Compared to the results typically in American and East Asian studies, the Sudanese participants received much lower P%-Scores than usual (Gielen & Markoulis, 2001; Moon, 1986). For instance, the Sudanese college/university students in their junior and senior years received an average P%-Score of 24.5%, whereas college students from Hong Kong received an average P%-Score of 37.9% (Hau, 1983), from South Korea a P%-Score of 41.5% (Park & Johnson, 1984), and from Taiwan a P%-Score of 41.4% (Gielen, Miao, & Avellani, 1990).

Table 3
Distribution of DIT Moral Stage Scores in Sudan, Kuwait and the USA

Stage	Sudan		Kuwait		USA
	Mean Score	%	Mean Score	%	%
Stage 2	3.44	5.73%	3.12	5.20%	10.50%
Stage 3	12.15	20.25%	12.69	21.15%	20.70%
Stage 4	22.04	36.73%	21.70	36.17%	34.70%
Stage 4½ (A)	3.12	5.20%	3.33	5.55%	3.20%
Stage 5A	9.90	16.50%	9.48	15.80%	17.80%
Stage 5B	2.58	4.30%	2.80	4.66%	6.20%
Stage 6	3.50	5.83%	3.67	6.11%	2.50%
P-Score (Principled)	15.99	26.64%	15.92	26.53%	26.50%
Meaningless Items	3.26	5.43%	2.77	4.62%	4.50%
P-Score (Overall Weight)	18.17	-	22.75	-	18.80%
	N = 253 150 Males and 103 Females; Average Age: 21.3 years		N = 253 115 Males and 199 Females; Average Age: 20.65 years		N = 106 53 Males and 53 Females; Average Age: 16.2 years

Note: Data for the USA are taken from Gielen et al. (1986) Data for the Sudan are taken from Ahmed et al (1987). Data for Kuwait are taken from Gielen et al. (1992)

Table 4
Distribution of DIT Moral Stage Scores for Three Countries
(Kuwait, Sudan, and Taiwan)

Stage	Kuwait	Sudan	Taiwan
Stage 2	5.20%	5.73%	6.88%
Stage 3	21.15%	20.25%	22.67%
Stage 4	36.17%	36.73%	25.07%
Stage 4½ (A)	5.55%	5.20%	3.56%
Stage 5A	15.80%	16.50%	24.66%
Stage 5B	4.66%	4.30%	7.15%
Stage 6	6.11%	5.83%	6.36%
P-Score (Principled)	26.57%	26.63%	38.17%
Meaningless Items	4.62%	5.43%	3.64%
P-Score (Overall Weight)	–	18.17%	22.08%
	N = 314 115 Males and 199 Females; Average Age: 20 years, 7 months	N = 253 150 Males and 103 Females; Average Age: 21 years, 3 months	N = 506 187 Males and 319 Females; Average Age: 17 years, 8 months

Note: Data for Sudan are taken from Ahmed, Gielen, and Avellani (1987). Data for Taiwan are taken from Gielen, Miao, and Avellani (1987). Data from Kuwait are taken from Gielen, Ahmed, and Avellani (1992). All Ss passed the standard DIT consistency checks. With the exception of D-Scores, moral judgment scores are expressed in percentages.

The authors also investigated the relationship between parental behavior (as perceived by the students) and moral reasoning skills. For male students, the father's acceptance was positively related to moral judgment development, while the father's aggression and indifference were negatively related to moral judgment. For female students, the mother's warmth and acceptance were negatively related to moral judgment development, while the mother's aggression, indifference, and rejection were positively related to moral judgment development. The results for the female Sudanese were surprising and difficult to interpret.

In a second study, Gielen, Ahmed, and Avellani (1992) administered the DIT to 685 Kuwaiti male and female intermediate, secondary, and university students. 54.1% of the protocols were rated as inconsistent. Consequently, the data analysis included only protocols from the 314 participants who successfully passed the consistency check.

The results for the Kuwaiti students (see Tables 3 and 4) indicate that they preferred conventional moral arguments (stages 3 and 4 = 57.32%) over principled arguments (stages 5A, 5B, and 6 = 26.57%) and preconventional arguments (stage 2 = 5.20%). It should be noted in this context that the DIT contained more conventional than principled or preconventional arguments. The obtained results, therefore, partially reflect the overall construction of the DIT.

A comparison between the samples from Kuwait and Sudan demonstrates a highly similar overall distribution of moral stage scores (Tables 3 and 4). The average P%-Scores for the two countries are almost identical (Sudan = 26.64%; Kuwait = 26.57%). These P%-Scores are, however, only slightly lower than the average P%-Scores are, however, only slightly lower than the average P%-Score of 29.35% that Webb and Steentsma (1983) reported for a sample of 259 male and female Palestinian arabs. In 1985, along the same lines, El-Sheikh reported average P%-Scores ranging from 20.23% to 29.93% for 333 intermediate, secondary school, and college and graduate level male and female students from Cairo, Egypt. In 1985, Eissa found in his Egyptian samples an average P%-Score of 10.56% (SD = 4.1) for the non-college group and an average P%-Score of 20.13% (SD = 4.33) for the college group.

It was expected that the Kuwaiti sample would show a somewhat higher P%-Score than their Sudanese counterparts due to the influences of modernization. But, contrary to the authors' expectations, the results appeared to reflect similar rather than different sociocultural influences for the two samples.

Compared to the Taiwanese study (Table 4), the Kuwaiti and Sudanese students received much lower P%-Scores, but higher stage 4 scores; this is consistent with the idea that the Kuwaiti and Sudanese students tended to adopt a conservative ideology.

In a subsequent study, the Algerian psychologist Bouhmama (1989) administered the DIT to 90 Muslim students, ranging in age from 18 to 43 years, who resided in England. The students came from many different countries including Arab countries such as Bahrain, Egypt, Iraq, Jordan, Kuwait, Libya, Morocco, Saudi Arabia, Syria, the United Arab Emirates as well as non-Arab countries such as Bangladesh, Iran, Nigeria, etc. Ninety out of Bouhmama's 104 students (or 86.54%) passed to the consistency check. It should be added that the students were allowed to take the test

home, a procedure that provided them with an unlimited amount of time to complete the questionnaire and presumably also with the opportunity to ask others about the meaning of difficult items. The author divided the students according to four educational levels: senior high school level, bachelor's degree level, master's degree level, and Ph.D. level. However, an ANOVA indicated no significant differences between the moral judgment scores of the four educational groups. This finding is important since it suggests that the moral reasoning of Muslim students residing in England did not become similar in its structure to that of their English peers.

Eissa's Arabic adaptation of the DIT was also employed in a study by Ben Abdel-Aziz (1989) who administered the test to 260 male and female Saudi university students. While students majoring in different academic subjects did not differ from each other in their moral judgment levels, there was a significant effect due to level of education: Seniors received higher scores on stages 5A and 5B while freshmen were more likely to endorse moral arguments representing stages 2 and 3. There were no gender differences except that male students favored stage 5A arguments more frequently than female students.

Bouhmama (1989) administered the DIT to 100 Algerian male and female psychology students aged between 19 and 25 years old (mean age was 21 years). The overall distribution of stage scores obtained in Bouhmama's study is depicted in Table 5.

Table 5
Distribution of Moral Stage Scores According to Bouhmama (1989)

Stage	Males (N = 50)		Females (N = 50)		All (M + F) (N = 100)	
	Mean	%	Mean	%	Mean	%
Stage 2	8.53	14.22%	6.32	10.53%	5.32	8.87%
Stage 3	33.82	56.37%	29.54	49.23%	35.02	58.37%
Stage 4	31.00	51.67%	27.18	45.30%	40.88	68.13%
Stage 4 ^{1/2} (A)	8.00	13.33%	3.46	5.77%	10.32	17.20%
Stage 5A	9.00	15.00%	17.12	28.53%	13.24	22.07%
Stage 5B	2.84	4.73%	8.33	13.88%	3.08	5.13%
Stage 6	7.33	12.22%	11.12	18.53%	8.34	13.90%
P-Score	15.27	25.45%	24.83	41.38%	25.52	42.20%

Note: Percentage were computed by the present authors based on Bouhmama's (1989) study.

It is difficult to interpret the results contained in this table because Bouhmama does not report his DIT data according to standard procedures. Adding all mean stage scores in Table 5 gives a result of 116%, a percentage that is clearly impossible. (The total of stage scores does not add up to 100% as it should.)

A similar problem arises when one compares Bouhmama's reported P%-Scores against the summed-up scores for stages 5A, 5B, and 6 separately for males and for females. For males, the reported average P%-Score is 15.27%, but the summed-up scores for stages 5A + 5B + 6 are equal to 19.17%. For females, the reported P%-Score is 24.83% whereas the summed-up scores for stages 5A + 5B + 6 are equal to 36.57% (compare to numbers reported by Bouhmama, 1989, in Table 5).

Another question arises when one compares his reported P%-Scores ($P\% = 25.52$) to the summed-up scores of stages 5A, 5B, and 6 (combined equals 24.66%). In reality, these two scores should be equal. Therefore, Bouhmama's scores cannot be compared to scores obtained in other DIT studies.

Kamel (1991) investigated relationships between stage of moral judgment, religious attitudes, and locus of control among 450 male and female Muslim university students ranging in age from 19.2 to 21.6 years. The participants in the study were 4th level students majoring in Arabic, English, French, mathematics, and biology. Research instruments included the DIT, a locally devised scale for assessing religious attitudes, an adaptation of Rotter's Internal vs. External Locus of Control Scale, and a locally devised index to assess the students' socioeconomic status. The results of the study point to important relationships between religious attitudes, locus of control, interaction between these variables, and level of moral judgment.

Habib (1991) sought to identify various characteristics of moral development among 100 male and 100 female second grade students enrolled in the College of Education, Tanta University, Egypt. In addition to S. A. El-Sheikh's adaptation of the DIT and a set of general background questions, the author administered Arabic adaptations of various self-concept, social competence, and shyness scales. The internal validity of the DIT was assessed on the basis of intercorrelation matrices between the various stage scores separately for males, females, and the overall sample. The resulting patterns of intercorrelations provided modest statistical support for the structural coherence of the DIT. The male and female students did not differ from each other in their stage scores

although there were some gender differences in the overall factorial structure of the DIT results. Students who were either enrolled in scientific subjects or in literature did not differ from each other in their performance on the DIT, but there were significant correlations between P%-Scores and two tests of social shyness.

Amir Khan (1992) conducted a study to examine moral stages and values, correlations between moral judgment levels and social and religious values, and other issues. The participants in the study included 92 Saudi and 109 non-Saudi male and female students receiving a secondary school education. The students responded to the DIT, to a locally constructed questionnaire assessing six values, and to questions regarding the students' socioeconomic background. The distribution of stage scores in the overall sample resembled that found in other Arabic studies, i.e., stage 4 scores were very high, P-scores fairly low, and stage 2 scores very low. There were no significant stage score differences between the Saudi and non-Saudi samples, nor did gender or SES have a significant impact on the students' level of moral judgment. To both groups of students religious values were of great importance followed by social and theoretical values. However, endorsement of religious and social values was unrelated to the students' moral judgment scores.

Ibrahim (1992) was interested in the relationship between moral judgment development (P-Score index) and dogmatism as measured by a locally devised scale. The participants in his study included 120 male and 120 female students ranging in age from 15.6-17.9 years. They came from urban and rural areas in Egypt and were enrolled in general, religious, and technical-industrial secondary schools. The research results pointed to a significant and substantial negative correlation ($r = -.69$) between the students' P-Score and their level of dogmatism. The male students achieved a higher mean P-Score than the female students and, in addition, they scored lower on the dogmatism scale. Some significant differences in moral judgment level were also found: Those males (but not females) enrolled in general secondary schools received higher P-Scores than their counterparts attending the religious education schools.

In a well-controlled study, Shafie's (1994) compared the moral judgments of 216 Egyptian adolescents enrolled in general secondary schools with those of a comparable group of 216 adolescents enrolled in the well-known Al-Azhar secondary school. The students were evenly divided between males and females, generally 15-16 years of age, and responded to a background questionnaire, the DIT (as based on M. R.

Eissa's Arabic adaptation), and Cattell's Scale for Intelligence. Students scoring below a cut-off score of 90 points were removed from the sample to make the results more meaningful. Of the original sample of 576 respondents, 144 or 25% did not pass the DIT consistency checks, and they were subsequently removed from the data analysis.

The results of Shafie's study pointed to important moral judgment differences among the students. Those students coming from higher socioeconomic backgrounds received considerably higher P-Scores than those coming from middle-class backgrounds. Students from lower-class backgrounds tended to receive the lowest scores. More surprising, students from rural backgrounds scored higher on the P index than students coming from urban areas. The results of the study also included a significant gender difference favoring female over male students. Finally, a correlational analysis indicated a substantial correlation ($r = +.45$) between level of moral judgment and level of intelligence, a finding which replicates results previously obtained in Western studies (Rest, 1979, 1986a).

2.1 Relationship of P%-Score and Gender

While no systematic gender differences regarding stages 2 through 6 emerged among the Sudanese, Kuwaiti, Saudi, and Palestinian Muslim students, some gender differences were found in El-Sheikh's Egyptian study: women endorsed conventional stage 3 arguments more frequently than men ($p < .05$); however, the P%-Score differences favored men ($p < .01$). Similarly, in Ibrahim's (1992) research project, male secondary school students from Egypt received higher P-Scores than their female counterparts. In contrast, in Shafie's (1994) study of Egyptian secondary school students, the female students received higher P-Scores than the male students, while in Habib's (1991) study of Egyptian college students, no significant gender differences emerged with respect to P%-Scores.

In the case of Palestinian Christians, women received a higher mean P%-Score than the men (P%-Scores = 30.3% and 25.4%, respectively).

In general, gender differences have been unsystematic and fairly small in the various Arab studies, sometimes slightly favoring male Muslim students and sometimes female Muslim students in regards to the P%-Score. In Western and East Asian (Taiwan, Hong Kong, and South Korea) studies, female students have tended to receive somewhat higher P%-Scores than male students (Thoma, 1986). Consequently, those Arab studies favoring males diverge in their results from the gender differences typically found in comparable Western and East Asian research.



2.2 Relationship of P%-Scores and D-Scores to age and Level of Education in Arab Studies

P%- and D-Scores increase slightly with age for male Kuwaiti and Sudanese respondents, but not for females. Webb and Steentsma (1983) reported an insignificant correlation between age and P%-Score among Palestinian Arabs. The same results were found in the Egyptian samples (Eissa, 1985; El-Sheikh, 1985). In addition, there was a modest increase of state 4-Scores with increasing age among male students (Tables 3 and 4).

The comparison between El-Sheikh's (1985) study in Egypt and the two studies conducted respectively in Sudan and in Kuwait (Ahmed et al., 1987; Gielen et al., 1992) demonstrates a highly similar distribution of moral stage scores in the three studies. The three studies are also similar in terms of the relationship between education/age and the P%-Score. The three studies show very weak or insignificant relationships between education/age and the P%-Score. In other words, none of these three studies shows a clear developmental trend for principled moral reasoning. The obtained correlations between age/education and the P%-Score or D-Score are simply too weak. It is important to add in this context that El-Sheikh (1985) employed a different adaptation of the DIT when compared to that employed by Ahmed et al. (1987) and by Gielen et al. (1992). Thus, an absence of clear developmental trends was found in three Arab countries using two different adaptations of the DIT.

Only for Sudanese males (Ahmed et al., 1987) and for male and female students from Saudi Arabia (Ben Abdel-Aziz, 1989) did the level of education correlate significantly with the P%-Scores. Webb and Steentsma (1983) reported no correlation between P%-Scores and scores on matriculation examination among the Palestinian Arabs. In the Egyptian samples (El-Sheikh, 1985) significant correlations were reported between level of education and morals cores at stages 3 and 4 only. Shafie's (1994) finding of a clear correlation between levels of intelligence and levels of moral judgment is, however, quite compatible with results obtained in the USA.

The high to very high inconsistency rates found in three studies conducted in Sudan (Ahmed et al., 1987 – 62%), Kuwait (Gielen et al., 1992 – 54.1%), and Egypt (Shafie, 1994 – 25%) indicate that the DIT was a very difficult and unusual task for the respondents and that the DIT was poorly understood by many of them. Several of the other studies including Arab participants did not use the standard consistency check, a procedural omission that throws some doubt on the reliability of their findings.

It appears that the DIT contains some moral issues that are more or less unfamiliar to many Arab respondents, reflecting cross-cultural differences in religion, politics, cultural traditions, and socialization patterns. In this context, the following two moral dilemmas contained in the DIT may be especially inappropriate for an Arab adaptation of the test: 1) The Doctor's Dilemma: Euthanasia is not known in Arab societies and it is against Islam;

2) The Newspaper Dilemma: The issue of freedom of speech in high school newspapers is an unfamiliar topic in Arab school systems. (This dilemma was not used by either Ahmed, Gielen, and Avellani, 1987 or Gielen, Ahmed, and Avallani, 1992 in their studies.)

It is noted that only men are represented as the main characters in the DIT dilemmas, and this factor may have affected the moral evaluations by both male and female participants (Eissa, 1992, 1997a, 1997b). However, many DIT studies have found slight gender differences on the P%-Scores favoring female over male respondents (Thoma, 1986) suggesting that the gender of the DIT characters is not of crucial importance. It is also noted that a person's performance on the DIT, like that person's performance on other objective tests, may be strongly influenced by his or her intellectual level, verbal comprehension, and social class background (Shafie, 1994).

3. Conclusions

The present survey of studies using the DIT with Arab respondents has underlined that a number of methodological problems makes it difficult to arrive at clear-cut conclusions about the usefulness of the DIT in Arab cultural settings. Most importantly, the high rates especially of Kuwaiti and Sudanese students that failed the three standard inconsistency checks suggests to us that the DIT may not be a reliable and valid assessment instrument of the moral judgment skills of many Arab students at various educational and age levels. Various other studies reviewed above did not employ the inconsistency checks and/or contained other methodological shortcomings. In addition, as reported by Gielen et al. (1992), informal discussions with some Kuwaiti respondents showed that the students found many of the DIT's moral arguments to be "strange" and not easily understandable.

Taken together, the results of the studies under review suggest to the authors that other Kohlbergian measures may prove to be more useful than the DIT for studies attempting to investigate the moral development of Arab respondents. For instance, Gibbs, Basinger, and Fuller (1992) have recently introduced a relatively brief production measure of moral judgment, the

social Reflection Measure-Short Form (SRM-SF). The measure contains 11 short-answer items that address seven sociomoral values: contract, truth, affiliation, life, property, law, and legal justice. It is obvious that all of these values have some meaning in Arab culture. Because the SRM-SF does not employ any moral dilemmas nor any prepackaged moral arguments, it may be a more easily adaptable and more meaningful test for Arab respondents than the DIT. One practical disadvantage of the SRM-SF, however, is that the respondents' answers must be coded with the help of a fairly intricate coding guide not easily understood by researchers whose native language is other than English. In addition, the SRM-SF can only be coded for the pre-conventional and conventional stages of moral reasoning, but not for the post-conventional stages. (Gibbs rejects the latter for theoretical reasons.) Unlike the DIT, then, the SRM-SF cannot be used to investigate whether post-conventional forms of moral argumentation are recognized and preferred in a given population. In this context, the authors are now developing a research project to ascertain whether the SRM-SF can be usefully employed to assess stages of pre-conventional and conventional moral judgment among Egyptian students of varying ages and educational attainment.

Another approach to moral judgment development in Arab society has been taken by Eissa (1992, 1997a, 1997b). The author developed a Judgment Preponderancy Scale (JPS) based on four moral decision stories. Each of the stories is followed by 11 arguments designed to help respondents in determining the basics of their moral decision making. The JPS is similar to the DIT in its general structure and, like the DIT, it may not be a very powerful instrument for identifying the developmental aspects of moral decision making in Arab society. Nevertheless, any attempt such as that of Eissa to identify the indigenous aspects of moral decision making is of considerable significance for the progress of Arab psychology.

The authors would like to add that future research studies on moral development should be longitudinal whenever possible, and they should be designed to investigate the effects of important cultural variables on the development of moral reasoning, as Nisan and Kohlberg (1982) have done with Turkish males. The relationship between socialization practices, personality traits, cognitive structures, value systems, moral reasoning, and religious beliefs and practices need to be studied further (Kamel, 1991; Shafie, 1994; Webb & Steentsma, 1983). However, the few studies that have been conducted, especially those by Ahmed (1994), Ahmed et al. (1987), Gielen et al. (1992), Ahmed and Gielen (1991), and Kamel (1991) give us some insight into the differential impact of the perceptions of

parental behavior and religious beliefs and practices on moral development. In addition, the relationship between moral judgment scores and theoretically relevant moral behaviors should be an important topic for Arab investigations (Kamal, 1986).

Kohlbergian theories and research methods attempt to identify the universal features of moral reasoning and moral judgment processes, and how these might develop throughout a person's lifespan. These features are of crucial importance if we are to understand those more or less rational aspects of morality which are shared (or could be shared in the future) by humanity. At the same time, Kohlbergian theories and research methods are less suitable for discovering the more culture-specific aspects of morality. As psychologists we are at present far removed from understanding either the more universal or the more culture-specific aspects of moral development in the various Arab countries. A sound psychological understanding of moral development is surely difficult to achieve, yet it is precisely in this area that psychologists could make a real contribution to the spiritual and ethical welfare of Arab societies.

WORKS CITED

- * Some studies reviewed in the present study were discussed briefly in a previous and unpublished general paper entitled *A Review of Arab Research Studies on Moral Judgment* which was presented by the same authors at the Regional Convention of the International Association for Cross-Cultural Psychology, Debrecen, Hungary, July 1991. Moreover, the present paper covers Arab research studies on moral judgment using the Defining Issues test which have been conducted since 1990.
- Ahmed, R.A., Gielen, U.P., & Avellani, J. (1987). "Perceptions of Parental Behavior and the Development of Moral Reasoning in Sudanese Students. *Growth and progress in cross-cultural psychology* Kagitçibasi (Ed.) Lisse (The Netherlands): Swets and Zeitinger. (pp. 196-206).
- Ahmed, R.A., & Gielen, U.P. "A Review of Arab Studies on Moral Judgment Development." Paper presented at the Regional Convention of the International Association for Cross-Cultural Psychology, Debrecen, Hungary. 1991, July.
- Amir Khan, M.H. Moral Judgment and Values: A Comparison Study Between Saudis and Non-Saudis in the city of Jeddah, the Western Region of Saudi Arabia." *Journal of the University of Umm el-Qura*, 4.6 1992. 252-295 (in Arabic).
- Ben Abdel-Aziz, B.H.S. "Level of Moral Judgment in Male and Female Students at the University of Umm el-Qura, Mecca." Unpublished master's thesis, University of Umm el-Qura, Saudi Arabia (in Arabic) 1989.
- Bouhmama, D. "Assessment of Kohlberg's Stages of Moral Development in Two Cultures." *Journal of Moral Education*, 13.2, 1984 124-132.
- "The relation of formal education to moral judgment development." *The Journal of Psychology*, 122. 2, 1988 155-158.



- "Moral judgment of a sample of psychology students at the University of Oran (Algeria)." *The Educational Journal* (Kuwait), 6.21, 1989 107-133 (in Arabic).
- Colby, A., and Lawrence Kohlberg. *The Measurement of Moral Judgment* (Vols. 1-2). New York: Cambridge University Press.
- Eissa, M.R. "The relationship between higher education and the level of moral judgment using a selected sample of students in the College of Education, Tanta University, Egypt." *Journal of the Social Sciences* (Kuwait), 13.2, 1985. 117-131 (in Arabic).
- "The Judgment Preponderancy Scale (JPS)." *Education: A Journal for Educational Research* (Faculty of Education, Al-Azhar University, Egypt). 26 (1992). 285-330 (in Arabic).
- "The Relationship Between Level of Moral Judgment and Ego-Strength." *Educational Journal* (Kuwait), 8 (28) 1993. 61-115 (in Arabic).
- "The Judgment Preponderancy Scale (JPS): An Objective Scale for Measuring Level of Moral Judgment. Part 1" (*Instructions*) 1997a Kuwait: Dar el-Qalem (in Arabic).
- "The Judgment Preponderancy Scale (JPS): An Objective Scale for Measuring Level of Moral Judgment. Part 2" (*Questionnaire*) 1997b Kuwait: Dar el-Qadem (in Arabic).
- El-Sheikh, S.A. "A Study of Moral Thinking in Egyptian Adolescents and Adults." E.A. Abou Hatab, Ed. *Yearbook of Psychology* (Vol. 4) 1985. Cairo: The Anglo Egyptian Bookshop pp. 123-169 (in Arabic).
- Gibbs, J.C., Basinger, K.S., and Fuller, R. *Moral maturity: Measuring the Development of Sociomoral Reflection*. Hillsdale, NJ: Erlbaum 1992.
- Gielen, U.P. "Moral Reasoning in Cross-Cultural Perspective: A Review of Kohlbergian Research." *World Psychology*, 2 (3-4) 1996a 313-333.
- (ed). *World Psychology: Focus on Lawrence Kohlberg*, 2.3-4, i-viii, 265-496 1996b.
- Ahmed, R.A., & Avellani, J. "The Development of Moral Reasoning and Perceptions of Parental Behavior in Students From Kuwait." *Moral Education Forum*, 17. 3. 1992. 20-37.
- Cruickshank, H., Johnston, A., Swanzy, B., and Avellani, J. 1986. "The Development of Moral Reasoning in Belize, Trinidad-Tobago and the USA." *Behavior Science Research*, 20. 1986. 178-207.
- & Markoulis, D. "Preference for Principled Moral Reasoning: A Developmental and Cross-Cultural Perspective." *Cross-Cultural Topics in Psychology*. L.L. Adler & U.P. Gielen 2nd ed. Westport, CT: Praeger. 2001.
- Miao, E.S.C.Y., & Avellani, J. "Perceived Parental Behavior and the Development of Moral Reasoning in Students From Taiwan." *Proceedings of CCU-ICP International Conference: Moral Values and Moral Reasoning in Chinese Societies* 1990. pp. 464-506. Taipei: Chinese Culture University.
- Habib, M.A. "The Relative Importance of the Characteristics of Moral Development: A social-Cognitive Approach for Solving Male and Female Adolescents' Problems." *Psychological and Educational Measurement and Evaluation* (Giza Strip, Palestine), 1.3. 1991. 45-92 (in Arabic).
- Hau, K.T. *A Cross-Cultural Study of a Moral Judgment Test (The DIT)*. Unpublished master's thesis, Chinese University, Hong Kong. 1983.
- Ho, D.Y.F., & Wu, M. Introduction to Cross-Cultural Psychology. In L.L. Adler & U.P. Gielen (Eds.), *Cross-Cultural Topics in Psychology* (2nd ed., pp. 3-13). Westport, CT: Praeger. 2001.

- Ibrahim, I.A.A. "The Relationship Between Level of Moral Judgment and Level of Dogmatism in Samples of Secondary School Students. Thesis, Tanta University, Egypt. 1992. (in Arabic).
- Ismail, M.A. "A Cross-Cultural Study of Moral Judgment: The Relationship Between American and Saudi Arabian University Students on the Defining Issues Test." *UMI* 77-5105. 1976.
- Jessup, S.E.B. "The Effects of Knowledge of Moral Development stage Theory and a Request to Fake on the Defining Issues Test scores of Adolescents in a Multi-National American School in Kuwait." *DAI* 45.1984.158A-159.
- Kamal, A.A. "Evaluative Notes on Kohlberg's Theory of Stages of Moral Development." *Annals of the College of Education, University of Qatar* (Qatar), 5.5 1986.414-431. (in Arabic).
- Kamel, M.M. "Some Variables Related to the Level of Moral Judgment in a Sample of University Students. *Proceedings, 7th Annual Convention of the Egyptian Association for Psychological Studies* (pp. 286-304). Cairo: The Anglo-Egyptian Bookshop. 1991.
- Kohlberg, L. "Essays on Moral Development." Vol. 2: *The Psychology of Moral Development*. New York: Harper & Row. 1984.
- Moon, Y.L. "A Review of Cross-Cultural Studies on Moral Judgment Development Using the Defining Issues Test." *Behavior Science Research*, 20.1986.147-177.
- Nisan, M., and Kohlberg, L. "University and Cross-Cultural Variation in Moral Development: A Longitudinal and Cross-Cultural Study in Turkey." *Child Development*. 53.1982. 865-876.
- Park, J.Y. and Johnson, R.C. "Moral Development in Rural and Urban Korea. *Journal of Cross-Cultural Psychology*, 15.1984.35-46.
- Rest, J. *Development in Judging Moral Issues*. Minneapolis, MN: University of Minnesota Press. 1979.
- *Manual for the Defining Issues Test: An Objective Test of Moral Judgment Development*. (3rd ed.). Minneapolis, MN: Minnesota Research Project. 1986a.
- *Moral development: Advances in Research and Theory*. New York: Praeger. 1986b.
- Shafie, A.M. [Comparing] the Moral Judgment of Adolescents With General Education and With Al-Azhar Education. Unpublished Thesis, Institute for Higher Studies on Childhood, Ain Shams University, Egypt. 1994. (in Arabic).
- Thoma, S.J. "Estimating Gender Differences in the Comprehension and Preference of Moral Issues." *Developmental Review*. 6.1986.165-180.
- Vine, I. "Moral Maturity in Sociocultural Perspective: Are Kohlberg's Stages Universal?" *Lawrence Kohlberg: (Consensus and Controversy*. eds. S. Modgil and C. Modgil. London: The Falmer Press 1986. 431-450.
- Walker, L.J. The Development of Moral Reasoning. *Annals of Child Development*, 5.1988.33-78.
- Webb, R.J., and Steentsma. I.R. "A Study of Moral Reasoning by Palestinian Arabs." Paper presented at the 91st Annual Convention of the American Psychological Association, Anaheim, California, U.S.A. 1983, August.

* * *

