The Moral Justification Scale: Reliability and Validity of a New Measure of Care and Justice Orientations

Linda S. Gump  
Psychological HealthCare

Richard C. Baker  
California School of Professional Psychology

Samuel Roll  
University of New Mexico

ABSTRACT. Research increasingly suggests that there are limitations to Kohlberg’s theory of moral development. Gilligan in particular has observed that Kohlberg’s theory considers abstract principled reasoning as the highest level of moral judgment, and penalizes those who focus on the interpersonal ramifications of a moral decision. Gilligan calls these justice and care orientations. The present paper describes the development of the Moral Justification Scale, an objective measure of the two orientations. The scale consists of six vignettes, of which two are justice oriented, two are care oriented, and two are mixed, incorporating both orientations. Construct validity was evaluated by expert judges and, overall, was high. Cronbach’s alpha was .75 for the Care subscale and .64 for the Justice subscale, indicating adequate internal consistency. Split-half reliabilities were as follows: Care, r = .72, p < .01; and Justice, r = .60, p < .05. Regarding test–retest reliability (approximately 2 weeks), r = .61, p < .05, for Care; r = .69, p < .05, for Justice. Neither subscale correlated significantly with the Marlowe-Crowne Social Desirability Scale. Thus, the Moral Justification Scale shows promise as an easily administered, objectively scored measure of Gilligan’s constructs of care and justice.


The Work of Kohlberg and Gilligan

Kohlberg (1981, 1985; Kohlberg & Kramer, 1969), using Piaget’s theories of cognitive and moral development as a starting point, developed a model of moral development with six stages. Kohlberg’s stages are grouped into three progressively higher levels: preconventional, conventional, and postconventional. People at the preconventional level (Stages 1 and 2), primarily children, conceive of rules and social expectations as external to the self. Moral decisions are made based on expectations of reward or punishment. At the conventional level (Stages 3 and 4), people subscribe to a morality of shared norms and values, centering on the needs of the individual and the rules and expectations of others. Interpersonal relationships and concern for others’ opinions are crucial (Stage 3). At Stage 4, obeying society’s laws becomes central. At the postconventional level (Stages 5 and 6), moral decision-making is based on principled reasoning. Stage 5 revolves around the utilitarian maxim, “the greatest good for the greatest number.” At Stage 6, people make decisions based on universal principles of justice, liberty, and equality, even if these violate laws or social norms.

Critics of Kohlberg’s model, most notably Gilligan (1982), have pointed out that his system is drawn from a Kantian philosophy that uses abstract principles: of justice as the basis of advanced moral reasoning. This penalizes those who focus on the interpersonal ramifications of a moral decision. Gilligan (1981) has argued that Kohlberg’s representation of women as fixated at Stage 3, which represents interpersonal morality, is flawed. Women’s reasoning, according to Gilligan (1982), is contextual and deeply tied to relationships, and Kohlberg has undervalued the equally valid Aristotelian moral concerns voiced by women (Vasudev, 1988). Thus, the emphasis on justice as the embodiment of morality appears to have underestimated the impact that interpersonal connectedness can have on moral decision-making.

Gilligan set out to test the validity of a care perspective, with the assumption that it is morally equivalent to the justice construct. Gilligan’s (1977) response to the apparent bias in Kohlberg’s theory toward the male (justice) perspective included an alternative stage sequence for the development of females’ moral reasoning. These stages are based on the degree of compassion and connection between self and others, manifested in the peace and harmony in relationships (Brabeck, 1983; Muuss, 1988).

Gilligan viewed women as progressing from initial selflessness (first level) to caring primarily for others (second level) and finally to an integration of concern for the needs of both self and others (third level). Her morality of responsibility emphasizes attachments, allows for both self-sacrifice and selfishness, and considers connections with others as primary, while Kohlberg’s morality of justice emphasizes autonomy, rules, and legalities, and considers the individual as primary. Gilligan and Attanucci (1988) have stressed that neither orientation is superior; rather, care and justice are complementary.

Using measures designed to investigate more interpersonally oriented forms of moral reasoning, a number of studies (Gibbs, Arnold, & Burkhardt, 1984; Gilligan & Attanucci, 1988; Lyons, 1983; Pratt, Golding, Hunter, & Sampson, 1988; Rothbart, Hanley, & Albert, 1986) have found gender differences, with males primarily focusing on issues of justice and females primarily focusing on interpersonal issues. However, other studies have not found significant differences (Crow & Heatherington, 1989; Friedman, Robinson, & Friedman, 1987; Galotti, 1989; Pratt, Golding, Hunter, & Sampson, 1988; Walker, 1989).

There may also be a degree of ethnocentrism in Kohlberg’s theory. Miller and Bersoff (1992), for example, challenged Kohlberg’s claim of universality in moral reasoning, having found that Americans focus on justice considerations, while Hindus in India emphasize interpersonal considerations in rendering a moral decision. Snarey (1985) likewise concluded that Kohlberg’s model is specific to Western culture.

In short, there is reason to believe that Kohlberg’s method of studying moral judgment has certain limitations. Other systems—most of which are based on interviews—are cumbersome, and require considerable expertise in
administration and scoring. Thus, an easily administered, objectively scored measure of moral judgment, tapping both care and justice orientations, would have several advantages (e.g., it would have a greater chance of avoiding gender and cultural biases). The present paper describes such an instrument—the Moral Justification Scale—and provides data on reliability and validity.

Method

Participants

To recruit participants, sign-up sheets were posted at a large state university in Southern California. The study was described only as an investigation of how people make decisions, in order to provide general information while minimizing self-selection bias. Students received extra credit in psychology classes in return for their participation.

One hundred participants completed the research protocol. Since this was the first phase of research comparing Mexican Americans and Anglo Americans, certain inclusion criteria were employed. Using the Acculturation Rating Scale for Mexican Americans (Cuellar, Harris, & Jasso, 1980), Mexican Americans had to be classified as “very Mexican,” “Mexican-oriented bicultural,” or “syntonic bicultural,” not “Anglo-oriented bicultural” or “very Anglicized.”

The Anglo American group consisted of White, non-Hispanic individuals whose parents were not immigrants. All were European-origin “nonethnics,” who Spindler and Spindler (1990) have referred to as the “referred ethnicclass” (p. 33) of the United States, namely individuals of Anglo-Saxon or northern European Protestant descent, or who adhere to mainstream cultural practices.

The final sample consisted of 40 Mexican Americans and 40 Anglo Americans (20 females and 20 males in each group). Their ages ranged from 18 to 25 years ($M = 18.9, SD = 1.3$). A separate sample of 16 students (mean age = 29.1, $SD = 7.4$) was employed to investigate test–retest reliability.

Measures

Moral Justification Scale: The Moral Justification Scale (MJS; Gimp, 1994), which is the focus of the present study, consists of six dilemmas presented in the form of vignettes. Two involve justice-oriented situations, two involve care-oriented situations, and two are mixed, combining both orientations. For example, the dilemma dealing with the possible breakup of a couple was classified as having a care orientation, as it primarily involves relational issues (i.e., responsiveness toward another person) rather than issues of individual rights and reciprocity between individuals. This is not to say that a justice mode of moral reasoning cannot be used, only that the essence of the dilemma is highly interpersonal. The other care vignette involves a dating dilemma. The justice dilemmas involve cheating on an exam and denting a car. The mixed dilemmas dealt with the desire to shoplift to help a sibling and how to handle a friend’s drug problem.

The six vignettes were written to be of interest and important to college students. Furthermore, the names of all protagonists were common for both Anglo Americans and Mexican Americans (e.g., Ana, Michelle, and Tony). Each vignette can have either a male or female protagonist, allowing for counterbalancing. Participants respond to one care, one justice, and one mixed vignette with a male protagonist and one care, one justice, and one mixed vignette with a female protagonist.

After reading the vignette, participants are asked to take a moment to think about what the protagonist should do. They are presented with eight sentences that have been extracted from the vignette, four of which represent care concerns and four of which represent justice concerns. Each is followed by a 10-point scale, with anchors at 1 (not at all important) and 10 (very important); the participant is asked to indicate the importance of the item in making a judgment. An example of a care-oriented item is: “Tony was his closest friend, and Marcus didn’t want to hurt him by telling the teacher and getting him into trouble.” An example of a justice-oriented item is: “Julie glanced up at the sign on the wall which read: ‘Shoplifting is illegal.’” A complete justice-oriented vignette, Denying the Car Dilemma (female version), follows.

Julie, a 16-year-old, is having difficulty making a decision. A few days ago, Julie’s younger sister, Susanna, had a minor accident with their parents’ car. Susanna only had her learner’s permit and wasn’t allowed to drive without her parents. She had taken the car out anyway and had driven around the neighborhood while her parents were away for the day. Upon her return, Susanna had accidentally dented the car a little by running into a telephone pole while attempting to park. Julie wasn’t sure what she should do about this situation, but was seriously considering telling her parents that it was she who had caused the dent instead of Susanna in order to avoid a great deal of family fighting.

Having been in a lot of trouble recently, Susanna had strained her parents’ relationship, as they frequently fought over what to do about her being bad. Julie didn’t want to cause further strain on her parents’ marriage by telling them about Susanna’s latest blunder. On the other hand, she didn’t want to have to lie about what happened either, as she felt that lying was wrong. Julie was also concerned about covering up for Susanna in this way. She always tried to set a good example for Susanna and didn’t want Susanna to think that she could just break the rules whenever she pleased. Besides, Julie didn’t want to damage her positive relationship with her parents—she’d worked hard on being close with them and didn’t want to risk hurting them by losing their trust. She also worried about her relationship with Susanna. Julie and Susanna had stood up for each other equally in times of trouble in the past, especially when it came to getting in trouble with their parents.

Julie began to wonder what might happen if she told her parents the truth. She had always believed that her parents were too strict in certain areas, and thought that this would probably be one of those areas. More than likely, she thought, their punishment of Susanna would be quite severe, and she didn’t want her little sister to have to suffer. If they thought that she had dented the car instead of Susanna, however, they would probably just laugh about it, as she had her driver’s license but was just learning how to drive. Besides, Susanna kept it secret last year when Julie skipped school so that she could go to the beach with friends, and she felt she owed Susanna for this. But Julie also realized that the loss of self-respect she would experience might not be worth going along with the story she and Susanna had made up.

Please take a moment to think about what Julie should do, then turn the page and answer the questions.

If you were to make a decision, think about how important each of these statements would be. On a scale from 1 (not at all important) to 10 (very important), please circle the number ($[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]$) that indicates how important you think each of the following ideas was in making your decision.
Article 20  The Moral Justification Scale: Reliability and Validity of a New Measure of Care and Justice Orientations

1. Julie didn’t want to cause further strain on her parents’ marriage by telling them about Susanna’s latest blunder.
2. Julie didn’t want to have to lie about what happened, as she felt that lying was wrong.
3. Julie didn’t want Susanna to think that she could just break the rules whenever she pleased.
4. Julie didn’t want to damage her positive relationship with her parents—she’d worked hard on being close with them and didn’t want to risk hurting them by losing their trust.
5. Julie and Susanna had stood up for each other equally in times of trouble in the past, especially when it came to getting in trouble with their parents.
6. More than likely, Julie thought, her parents’ punishment of Susanna would be quite severe, and she didn’t want her little sister to have to suffer.
7. Susanna kept it secret last year when Julie skipped school so that she could go to the beach with friends, and Julie felt she owed Susanna for this.
8. Julie realized that the loss of self-respect she would experience might not be worth going along with the story she and Susanna had made up.

The Care subscale score is based on responses to the four care items for the six vignettes. Scores across the 24 items are averaged, and thus range from 1 to 10. The Justice subscale is scored similarly.

Miller Social Intimacy Scale. The Miller Social Intimacy Scale (MSIS; Miller & Lefcourt, 1982) is a 17-item self-report inventory designed to assess the level of social intimacy experienced in marriage or friendships. Respondents are asked to rate, on a 10-point scale, their relationship with a spouse or closest friend. Responses are summed to produce a Total score, as well as scores for two subscales, Frequency and Intensity.

Cronbach alpha coefficients of .91 and .86 were reported by Miller and Lefcourt (1982). Test–retest reliability over a 1-month interval (r = .84, p < .001) and 2-month interval (r = .96, p < .001) reflected stability in scores over time (Miller & Lefcourt, 1982). Convergent validity was demonstrated using the UCLA Loneliness Scale (Russell, Peplau, & Fergusson, 1978), r = .65, p < .001, and the Interpersonal Relationship Scale (Guernsey, 1977), r = .71, p < .001.

Marlowe-Crowne Social Desirability Scale. The Marlowe-Crowne Social Desirability Scale (SDS; Crowne & Marlowe, 1960) was developed to control for participants who seek to present themselves in an exaggeratedly favorable or unfavorable light. It consists of 33 items, with true-or-false response categories. Crowne and Marlowe reported internal consistency of .88; the test–retest correlation after one month was .89. Construct validity was established through significant positive correlations with the L and K scales of the MMPI (Crowne & Marlowe, 1960).

Acculturation Rating Scale for Mexican Americans. The Acculturation Rating Scale for Mexican Americans (ARSM; Cuellar, Harris & Jasso, 1980) is a self-report questionnaire designed to assess level of acculturation in both nonclinical and clinical populations. Each of the 20 items on the scale is assigned a value ranging from 1 (extremely Mexican oriented) to 5 (extremely Anglo oriented). The dimensions assessed are language familiarity and usage, ethnic pride and identity, ethnic interaction, cultural heritage, and generational proximity. Five groups of Mexican Americans are identified according to level of acculturation: very Mexican, Mexican-oriented bicultural, syntonic bicultural, Anglo-oriented bicultural, and very Angloized.

A coefficient alpha of .88 was obtained for a nonclinical sample. Concurrent validity was demonstrated using the Behavioral Acculturation Scale (Szapocznik, Scopetta, Kurtines, & Aranalde, 1978), rho = .76, p < .001, and the Biculturalism Inventory (Ramirez, Cox, & Castaneda, 1977), rho = .81, p < .001.

Procedure

In a psychology laboratory on the college campus, participants signed a consent form and completed the MJS, MSIS, SDS, and ARSMA in groups of up to 14. Most took about one hour. Upon finishing, participants were debriefed.

Results

Validity of the Moral Justification Scale

To assess construct validity, the 48 MJS items (six vignettes, each with eight items) were rated by eight judges as to whether they represented the care or justice constructs. Judges were clinical psychologists who had taught graduate courses in ethics or developmental psychology, or doctoral candidates in clinical psychology who had taken courses in these areas. The judges were first provided with a brief summary of Gilligan’s concepts of care and justice.

For 35 of the items, there was unanimous agreement by the eight judges that the care and justice constructs were accurately represented. There was agreement by seven of the judges for 10 items. The three items for which there was agreement by fewer than seven judges were not included in subsequent calculations of MJS scores.

Concurrent validity was assessed via comparison of Care subscale scores and Miller Social Intimacy Scale scores. The MSIS measures interpersonal intimacy and the Care subscale taps interpersonal considerations in moral judgment. While the underlying constructs overlap, they do so only moderately. Therefore, only a modest correlation was expected and, in fact, found (r = .22, p < .05).

Reliability of the Moral Justification Scale

Internal consistency (Cronbach, 1951), split-half reliability, and test–retest reliability were computed for a separate sample of 16 students. Internal consistency and split-half reliability were based on the first administration only, whereas test–retest reliability was calculated using both administrations.

Internal consistency. Cronbach’s alpha was .75 for the Care subscale and .64 for the Justice subscale, indicating adequate internal consistency. Because of counterbalancing, it was possible to calculate Cronbach’s alpha by sex of the protagonist in each vignette. Alpha levels were expected to be lower than those for the full subscales and, in fact, ranged from .26 to .71.

It was also possible to calculate Cronbach’s alpha based on dilemma content (care, justice, or mixed). Alphas ranged from .31 to .57, reflecting marginally acceptable reliability. Examining dilemma content (care, justice, or mixed) by sex of protagonist (male or female) by subscale (Care versus Justice) produced alphas ranging from .84 to .47, with a mean of .12.

Split-half reliability. Split-half reliability of the Care and Justice subscales was also examined (exclusion of three items, as previously noted, prevented equal representation of dilemmas in each half). For the Care subscale, r = .72, p < .01 (correlation between halves was .91 using
Kuder-Richardson Formula 20). For the Justice subscale, $r = .60$, $p < .05$ (.75 using Kuder-Richardson Formula 20).

Test retest reliability. Approximately 2 weeks after the first administration, the MJS was readministered to the same 16 participants. Test-retest correlations for the Care subscale ($r = .61$, $p < .05$) and the Justice subscale ($r = .69$, $p < .05$) indicated adequate reliability.

Social Desirability
To evaluate whether the 80 participants responded to the vignettes honestly, they were administered the Marlowe-Crowne Social Desirability Scale. The correlations between the SDS and the Care subscale ($r = -.19$) and the Justice subscale ($r = -.17$) were not significant. Thus, the MJS did not appear to be influenced by social desirability pressures.

Conclusion
The Moral Justification Scale may be a useful alternative to measures based on Kohiberg’s system. For example, in clinical work with delinquent children, for whom the nature and style of moral reasoning are important issues, the MJS can be of special help. Those with the capacity to consider the needs and feelings of others (care-oriented) would be distinguished from those with more of a justice orientation. Separate interventions could then be designed and implemented.

Tracking the progress of different groups would itself be of interest, in order to determine whether outcomes differ. Further, with the current focus on values-related education, the MJS could be used to measure how various educational programs produce changes in care or justice orientations, or both.

In accord with the work of Gilligan (1982), Guisinger and Blatt (1994), and other theorists, cross-cultural studies should be undertaken with the MJS. In particular, it would be interesting to determine whether cultures under stress show declines in levels of moral reasoning.

References


Address correspondence to: Linda S. Gump, Ph.D., Clinical Psychologist, Psychological HealthCare, PLLC, 110 West Utica, Oswego, NY 13126.

Exercise for Article 20

Factual Questions
1. Have all previous researchers found gender differences when using more interpersonally oriented forms of moral reasoning?
2. The final sample consisted of how many students?
3. The Moral Justification Scale consists of how many vignettes?
4. The Marlowe-Crowne Social Desirability Scale was developed to control for participants who seek to do what?
5. The judges were in unanimous agreement on how many of the 48 items?
6. What was the value of the correlation coefficient for the relationship between Care subscale scores and Miller Social Intimacy Scale scores?
7. Was the relationship between the social desirability scores and the Care subscale scores significant?
Questions for Discussion

8. In order to recruit participants, do you think it was appropriate to describe the study as only "an investigation of how people make decisions"? Explain. (See lines 88–90.)

9. How important is the complete vignette in helping you understand this study? Would the report be just as strong if the vignette was omitted? Explain. (See lines 157–228.)

10. What do you think the researchers mean when they state that "... participants were debriefed"? (See lines 283–284.)

11. How helpful is it to know the test–retest reliability of the MJS? Do you think it has adequate reliability? (See lines 333–337.)

12. Has this study convinced you that the MJS is reasonably valid? Explain.

Quality Ratings

Directions: Indicate your level of agreement with each of the following statements by circling a number from 5 for strongly agree (SA) to 1 for strongly disagree (SD). If you believe an item is not applicable to this research article, leave it blank. Be prepared to explain your ratings.

A. The introduction establishes the importance of the study.
   SA 5 4 3 2 1 SD

B. The literature review establishes the context for the study.
   SA 5 4 3 2 1 SD

C. The research purpose, question, or hypothesis is clearly stated.
   SA 5 4 3 2 1 SD

D. The method of sampling is sound.
   SA 5 4 3 2 1 SD

E. Relevant demographics (for example, age, gender, and ethnicity) are described.
   SA 5 4 3 2 1 SD

F. Measurement procedures are adequate.
   SA 5 4 3 2 1 SD

G. All procedures have been described in sufficient detail to permit a replication of the study.
   SA 5 4 3 2 1 SD

H. The participants have been adequately protected from potential harm.
   SA 5 4 3 2 1 SD

I. The results are clearly described.
   SA 5 4 3 2 1 SD

J. The discussion/conclusion is appropriate.
   SA 5 4 3 2 1 SD

K. Despite any flaws, the report is worthy of publication.
   SA 5 4 3 2 1 SD