Randomized controlled study of the impact of
the Konstanz method of dilemma discussion on moral judgement

Sanguan Lerkiatbundit, RPh, PhD
Assoc. Prof., Faculty of Pharmaceutical Sciences
Prince of Songkla University, Hadyai, Sonkla, Thailand

Parichat Utaipan, RPh, MSc
Lecturer, Siridhorn College of Public Health, Yala, Thailand

Chonlada Laohawiriyanon, PhD
Assist. Prof., Faculty of Liberal Arts
Prince of Songkla University, Hadyai, Sonkla, Thailand

Adisa Teo, PhD
Assist. Prof., Faculty of Liberal Arts
Prince of Songkla University, Hadyai, Sonkla, Thailand

Corresponding author:
Sanguan Lerkiatbundit, Assoc. Prof.
Department of Pharmacy Administration
Faculty of Pharmaceutical Sciences
Prince of Songkla University
Hadyai, Songkla, Thailand 90112
Tel and Fax: 66-074-428167
E-mail: lsanguan@makok.pharmacy.psu.ac.th
Randomized controlled study of the impact of the Konstanz method of dilemma discussion on moral judgement

Sanguan Lerkiatbundit, Parichat Utaipan, Chonlada Laohawiriyanon, Adisa Teo

Abstract

The main objective of the study was to determine the effects of the Konstanz method of moral dilemma discussion (KMDD) on moral competence in the allied health students. The study employed the Moral Judgement Test (MJT) as the instrument to monitor the change of moral judgement. The test was translated from English into Thai and validated in 247 students in grade nine, grade twelve, and first and second year pharmacy students. Overall, the scale satisfies four validity criteria: preference hierarchy, quasi-simplex structure of stage preference, affective-cognitive parallelism, and positive correlation between education and moral competence score (C-index). Test-retest reliability with one-month interval was 0.90. Thirty-eight pharmacy technician students and 45 dental nurse students at a community college participated in the study to investigate the impact of KMDD. The subjects were randomly assigned into control (n=41) or experimental groups (n=42). The experiment group participated in a 90-min KMDD once a week for six consecutive weeks. The students in the control group also met once a week for six weeks to discuss the topics not related to ethics. All subjects completed the MJT before and after the intervention and again 6 months later. At the start, the C-indexes between experiment and control groups were not different (0.21±0.13 and 0.25±0.16). The experiment group scored significantly higher the control did after the intervention (0.35±0.11 and 0.24±0.15), and six months later (0.34±0.11 and 0.24±0.14). The KMDD appears to be a practical and effective intervention for developing moral competence in the allied health students. The effect on moral competence persists at least six months after the intervention.
Randomized controlled study of the impact of
the Konstanz method of dilemma discussion on moral judgement

Moral judgment competence is “the capacity to make decisions and judgments which are moral (i.e., based on internal principles) and to act in accordance with such judgments”.

Kohlberg's theory has been dominant in guiding the investigation of impact of education on moral competence. In his theory, moral development proceeds through six stages within three levels. The first two stages are pre-conventional level. At this level, moral reasoning is based on one's interests and those of others one cares for. Stages three and four are conventional level. The basis of reasoning is the concern to maintain the social order. Moral judgments are guided by meeting the expectations of others or the authority and the obedience to rules. At stages five and six (post-conventional level), the reasoning is based on the universal principles for making choices among alternative courses of action that would be held by any rational moral individual. A central emphasis in the post-conventional level is on principles for choosing the most just options for individuals in society.

There is a significant and positive association between principled moral judgements and general moral behavior such as resistance to cheating, peer pressure, and unlawful or oppressive authority; "whistle blowing" on corruption; the keeping of contractual promises; non-aggression; and helping behaviors. It has also been found to be positively linked to community involvement and a sense of civic responsibility, ethical behavior in professions such as accounting, dentistry, and medicine. The positive relation between moral judgement and clinical performance have been reported repeatedly in physicians, nursing students, dentists, physical therapy students and pharmacists.
Thus, it appears that moral judgement is systematically associated with moral action. Growth in moral reasoning should enhance the possibility of moral action. However, moral reasoning alone may not be sufficient to determine moral behavior. Rest suggested that using principled reasoning to judge moral issues is only one of four components leading to moral action. The other three components are 1) moral sensitivity (be able to recognize that a moral dilemma exists, 2) moral motivation (place moral values above other values), and 3) moral character (have the strength of character to carry out the morally justifiable course of action).12

Dilemma discussion appeared to be particularly effective in fostering moral judgement. The average effect size of this method in college students was 0.51. Previous research identified three characteristics of effective moral discussion: conflict, stage disparity and transactive discussion. The chosen dilemma for discussion should generate disagreement among the participants. Consensus on solution to dilemma reduces the likelihood that students would challenge or respond to one another's reasoning and thus reduced the impact of the discussion on morality competence. Stage disparity among participants should be on the order of one-half stage. This stage disparity is about what one finds among students in the classroom. The third condition is transactive discussion which is characterized by listeners' efforts to integrate the speaker's statements into one's own framework before generating a response. It involves the attempt to extend the logic of the speaker's argument, refute the assumptions of the speaker's argument, or provide a point of commonality or resolution between the two conflicting positions.

Recently, Lind proposed the Konstanz method of dilemma discussion (KMDD). The main feature of the method (described in method section) is designed to intensify transactive discussion among participants. The German study in the students with major psychology and education showed the KMDD was more effective than the traditional
discussion method\textsuperscript{16}. Furthermore, the KMDD seems to produce an earlier observable impact, compared to the traditional method. Three to ten 120-minute sessions of the KMDD over several weeks were sufficient to produce the significant impact on moral judgement\textsuperscript{16}. The study in medical students reported that a minimum of 20-29 hours of traditional method discussion was needed to achieve a significant gain in moral competence\textsuperscript{17}. However, the findings on the KMDD stem from a study with small groups of subjects, mostly university students\textsuperscript{16}. There is a need to further investigate the method. The KMDD has never been tested in any health professionals or subjects in any other countries other than German. We conduct the study to determine the immediate and long-term effects of the KMDD in the allied health students.

**Methods**

The article consists of two sub-studies. Study one addresses the validity and reliability of the measure of moral competence employed in the study. Study two describes the impact of KMDD on the development of moral judgement.

**Study one: Validation of the Moral Judgement Test (MJT)**

**Instrument and C-index**

Kohlberg's theoretical work has led to the development of a number of instruments to estimate an individual's level of moral judgment such as the Moral Judgment Interview\textsuperscript{18}, the Defining Issues Test\textsuperscript{19} and the Moral Judgement Test (MJT)\textsuperscript{20}. This study used the Moral Judgement Test (MJT) for several reasons. First, a large body of literature supports the validity of the test\textsuperscript{20}. Second, its built-in counter arguments enable the probe of subjects’ ability to apply moral principles consistently. As a result, the test is able to separate cognitive aspects of moral judgment from affective aspects (To which degree does a subject
prefer each of the six Kohlbergian stages of reasoning?)\(^{20}\). Third, MJT is shorter, usable with a wider age range and sensitive to education-induced change. The subjects cannot fake the scores upward in experimental situations\(^{20}\). Fourth, MJT has been used in subjects from various cultures including Asian (e.g., China, Philippines, and Sir Lanka), which is similar to the subjects in this study.

There are two dilemmas in the MJT, each followed by 12 items. One of the dilemma is in Appendix 1. The items in the MJT are grouped into six pro and six con arguments representing Kohlberg's six stages. The participants are asked to rate their acceptability on a 9-point Likert scale. The stage-consistency or C-index, ranging from 0-1 indicates the level of moral judgement competence. C-index is the portion of variance of responses to 24 items of the MJT that could be uniquely attributed to moral stage. The calculation of C-index is based on an ANOVA (2 dilemma*2 kinds of statement (pro/con)*6 moral stages) of each individual participant's pattern of responses\(^{20}\).

**Translation, pretest and validation sample**

The translation of the MJT-English version into Thai language followed the standard translation-back translation method\(^{21}\). The process was iterative until Thai version seemed to preserve the original meanings. The resulting version was tested in three pharmacy students using the think aloud method\(^{22}\). Each student read the questions and described what he/she was thinking in the process of filling the questionnaires (e.g. the understanding of the questions, how he/she made a decision in judging each reasoning or item etc.). The questionnaire was further corrected according to the information obtained from this method. Unfortunately, the pretest of the resulting questionnaires in 624 students (grade 7-12, first-fifth year pharmacy students and graduate students) failed to meet the validity criteria as described below. The interview with some subjects revealed that dilemmas in MJT were
difficult moral tasks and attention to take the test seriously was crucial for the validity of the results.

As a result, we conducted further think aloud test until no new misunderstandings were found. The number of subjects participated in think aloud test was 10. Finally, the instrument was validated in 4 convenient groups of subjects: 88 students in grade 9, 39 in grade 12, 24 first year pharmacy students and 96 fourth year pharmacy students. The researchers explained to the subjects the objectives of the study, asked for the voluntary participation and then emphasized the importance of paying attention to the test taking. The subjects completed the test in classes or took home the test and later returned the questionnaires within 1 week. There was no time pressure for filling the MJT.

**Validation criteria, reliability and data analysis**

Four validity criteria derived from the cognitive-developmental and the dual-aspect theory of moral behavior\(^1,20\).

1. Preference hierarchy: The preferences for the six stages of moral reasoning (calculated by summing four items on the same stage) should be ordered as predicted by the theory. The reason on the higher stages should be preferred over the reason on the lower stages.

2. Quasi-Simplex structure of stage preference: the correlation between the preferences of neighboring stages (like four and five) should be higher than those between more distant stages like four and one. Principle component factor analysis of preference data with two-factor solution and varimax rotation should reveal the the factor loadings of stages one to six being orderly aligned on a simplex arc.

3. Affective-cognitive parallelism: A person with a high level of moral judgment is more likely to accept the higher stage reasoning and reject the lower stage reasoning. C-
index (indicator of moral judgement) should show a negative correlation with preference scores for stages 1 and 2 and a positive correlation with stages 5 and 6 preference.

4. C should correlate positively with education or number of years in schools.

Test-retest method was employed to assess reliability of the MJT-Thai version. Ten subjects completed the test twice one month apart. The correlation of C-indexes indicates the reliability of the test.

Study two: the KMDD impact on moral competence

Subjects and randomization

The subjects were 38 pharmacy technician students and 45 dental nurse students in their first year at a community college in the Southern Thailand. The participation was on a voluntary basis. The students agreed to participate after being told the objectives and details regarding the study. The subjects from two programs were pooled and then randomly assigned into control (n=41) or experimental groups (n=42) based on the computer-generated random number.

Intervention in the experimental group

The intervention procedure or KMDD\textsuperscript{23} as proposed by Lind had been pre-tested in the fourth year pharmacy students in the university nearby and in the second year students in pharmacy technician program at the study site. The KMDD used in the study is described below:

1. The researchers distribute the handout describing one dilemma to the class. The students read the dilemma and make the decision individually if the action taken by the protagonist is right or wrong. The researcher asks the subjects to note down the reasons they use in judging the action.
2. The researchers briefly ask the class what the conflicting situation is in the dilemma to make sure that the dilemma is clear and every student fully understands it.

3. Students vote on the dilemma solution. Then, the class is split into two large groups (pro and con groups) based on their opinion on whether the protagonist is right or wrong.

4. Each large group (pro and con) forms the small groups of 5 or 6 students. Members of the small groups take turn to inform the other members the reasons they use. Then, the small group discussion starts to find the consensus on the appropriate reasons for judging the action in the dilemma.

5. The researchers ask one group to present to the class the reasons they use for dilemma solving. Then, the presenter chooses another student from the other side (i.e., pro student chooses a student in the con group, and vice versa) to become the next presenter. The chosen student has to express his/her opinion on the following questions: How acceptable are the reasons just given by the last presenter? What is the principle one use to judge them? and What should be the best reasoning for solving the dilemma?. Then, the presenter again chooses another student from the other side, and the discussion is repeated with the same procedure. This part of discussion takes 20-25 minutes. During the discussion, the researchers take the role of listeners or moderators to facilitate the discussion and avoid expressing their opinion. The researchers also record all major pro and con arguments on the transparency and show them to the class.

6. All subjects return to their small groups and resume the discussion on the following points: How acceptable are the reasons given by the other side as listed on the transparency? How do the groups judge them? Do some of them make the group rethink their own opinion? What should be the appropriate reasons for solving the dilemma?

7. Both sides appoint a speaker, who summarizes the outcomes of the discussion and comments on the other side’s arguments.
8. The class takes a final vote on if the protagonist is right or wrong.

9. The researchers thank the students for participation and encourage them to join next discussion.

   The discussion was about 90 min and took place once a week for six consecutive weeks. The dilemmas for the six discussions were modified from the literature\textsuperscript{23,24} and are briefly described here: 1) Should a doctor take the organ from the dead body without the permission to help the severely ill patients in case of emergency? 2) Should a student inform the teacher that one’s best friend involves in drug dealing? 3) How do one choose one of the two polices to defuse the dangerous bomb? 4) Should a father suffocate his own crying baby to save the lives of his wife and two daughters from the terrorists? 5) Should one tell the police the address of his/her best friend who shoplifts? and 6) Should a judge give a permission to torture the prisoner for information to save the lives of many people? These dilemmas were chosen based on the results from pre-testing the dilemma in the pharmacy students. The chosen cases are semi-real dilemmas causing a strong conflict between moral principles within and among the students. Furthermore, they do not affect the self-identity and hence, do not prohibit the free expression of the idea.

   The students in the control group met in a small group of 5 or 6 once a week for six successive weeks. The researchers asked them to discuss on the topics not directly related to ethics such as their opinion on the violence of welcome party for freshmen, quality assurance policy of the College, their academic problems, and their ideal characteristics of graduates. The discussion in both group was regarded as an extra-curriculum activity, not a part of any courses.

Data collection and data analysis

   All subjects completed the MJT in the class for three times: one-week before the intervention, at the end of the intervention, and six months later. There was no time pressure
for completing the MJT, and the importance of taking the test seriously was emphasized. C-scores across three points were compared using a split plot repeated measure ANOVA with an adjusted F-test because for the violation of sphericity assumption\(^{25}\). The comparisons between groups at each time were Bonferroni adjusted test.

## Results and Discussion

### Study one: Validity study

Table 1 displays the results from validity study. Theoretically, the participant should prefer the higher stage of reasoning to the lower stage of reasoning. The pattern of preference in table 1a (column 2) was consistent to the theory with some discrepancy. The preference for stage 6 reasoning was lower than those of stages four and five. Table 1b shows the correlation matrix of stage preferences. The pattern of matrix was consistent to a quasi-simplex structure. Overall, the correlation between the neighboring stages was higher than those between more distant stages were. The loadings from the factor analysis of preference data orderly located from stage one to six on a simplex curve (Figure 1) as predicted with a minor diversion in stage five and six. The last column of table 1a shows affective-cognitive parallelism. The preferences for the stages five and six were positively correlated with the C-index, while those for the stages one and two were negatively correlated with C-index. The size of correlations of C-index and preference for stages 3 and 4 was in between. Table 1c shows C-index increased with the education levels. The correlation between years of schooling and C-index was 0.35 (P<0.001). For reliability, the correlation between C-indexes measured at one-month apart was 0.90.

Overall, the MJT-Thai version is reliable, and satisfies four validity criteria with a few exceptions, especially those regarding stage 6 items. They concern the most abstract form of reasoning. This characteristic may prevent the subjects from having clear
understanding of stage 6 items and negatively affect their rating of preferences. The future attempt to refine the instrument (Thai version) should focus on the stage 6 items. We decided to use the MJT-Thai in the intervention study because it is reliable and meets nearly all of the stringent validity criteria.

**Study two: the KMDD impact on moral competence**

The subjects in the experiment and control groups were similar in their gender, major, and age (Table 2). The majority of subjects were female with age between 17-24 years. The participants were very cooperative. The number of subjects participated in six dilemma discussions ranged from 39-42. Those in control were 39-41. The dilemmas seemed to perform as expected. The class split well into the pro and con groups. The subjects in pro group ranged from 14-29 over the six dilemmas. Those in the con group ranged from 13-28. Class splitting is an essential element for effective dilemma discussion. Consensus on the outcome reduced the likelihood that students would challenge or otherwise respond to one another's reasoning and thus reduced the impact of the discussion on students' existing notions of morality.

Table 3 displays the C-indexes over three times of data collection. The ANOVA revealed the significant interaction between times and group with F(1.09,88.35)=14.65, P<0.001. As a result, we compared C-scores between groups at each time of data collection. At the beginning of the study, the C-indexes between experiment and control groups were about the same (P=0.179). However, the experiment group scored higher significantly after the intervention, indicating the significant impact of KMDD on the development of moral competence. The C-index maintained its level at the sixth month after the intervention. The C-index in the control remained the same through out the study period. The finding increases the internal validity of the study because it rules out the possibility of the contamination of the intervention, history and maturation effect.
The results indicate that the significant impact of the KMDD could be seen with six 90-minute discussions or about 9-10 hours of exposure to discussion. The KMDD seems to be an interesting alternative to the traditional method of discussion, which a minimum of 20-29 hours of discussion was needed to achieve a significant gain in moral competence17. The effect size of KMDD, as calculated by the ratio of the C-index improvement (from time 0 and 6 months after intervention) to the standard deviation at time 0 was 1.0. It is higher than those of traditional method (0.51) as reported from the synthesis of research in college students3. However, the comparison should be interpreted with cautions because the previous research and the present one are different in various aspects such as using different instrument, subjects and exposure time. A more valid comparison would be the head to head comparison of two methods in the same study.

In this study, the intervention increased the C-index for 14 points. A large improvement may result from the ceiling effect because subjects in the experimental group showed a rather low level of C-index at the beginning of the study (Scores less than 0.19 are considered low, 0.20-0.29 are medium and those more than 0.30 are high20). However, the result is similar to those reported by Lind in the study using the KMDD in 42 German samples with an increase of C-index by 12-20 points within one semester16.

In conclusion, The KMDD is an effective and practical method for teaching moral competence. The effect of six 90 minutes discussion sustains for at least six months. It should be applicable to a wide range of allied health professionals. The method is not labor-intensive, requiring one moderator for a group of 40 students because there is no need for close supervision of the discussion in each group. The give and take among peers is central to the method. There is no data on the maximum class size for using the KMDD. The class size of allied health students could be as large as 100. Further study on the effects of class size and the impacts of the method is needed.
Table 1: The results on validity testing of the MJT-Thai (N=247)

a) Mean stage preference and correlation between C-index and preferences

<table>
<thead>
<tr>
<th>Stage</th>
<th>Mean±SD</th>
<th>r between C-index and stage preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage1</td>
<td>-2.63±5.39</td>
<td>-0.41</td>
</tr>
<tr>
<td>Stage2</td>
<td>-2.87±5.46</td>
<td>-0.41</td>
</tr>
<tr>
<td>Stage3</td>
<td>1.60±4.75</td>
<td>-0.09</td>
</tr>
<tr>
<td>Stage4</td>
<td>2.59±4.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Stage5</td>
<td>4.03±4.33</td>
<td>0.38</td>
</tr>
<tr>
<td>Stage6</td>
<td>2.03±3.80</td>
<td>0.17</td>
</tr>
</tbody>
</table>

b) Correlation among stage preferences

<table>
<thead>
<tr>
<th></th>
<th>Stage1</th>
<th>Stage2</th>
<th>Stage3</th>
<th>Stage4</th>
<th>Stage5</th>
<th>Stage6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage2</td>
<td>0.57</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage3</td>
<td>0.48</td>
<td>0.41</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage4</td>
<td>0.26</td>
<td>0.23</td>
<td>0.37</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage5</td>
<td>0.01</td>
<td>0.10</td>
<td>0.27</td>
<td>0.33</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Stage6</td>
<td>0.15</td>
<td>0.20</td>
<td>0.29</td>
<td>0.36</td>
<td>0.37</td>
<td>1</td>
</tr>
</tbody>
</table>

c) C-index by education

<table>
<thead>
<tr>
<th>Education</th>
<th>N</th>
<th>Schooling years</th>
<th>Age</th>
<th>C-index±SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 9</td>
<td>88</td>
<td>9</td>
<td>14-16</td>
<td>0.16±.12</td>
</tr>
<tr>
<td>Grade 12</td>
<td>39</td>
<td>12</td>
<td>17-19</td>
<td>0.21±.14</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; year student</td>
<td>24</td>
<td>13</td>
<td>18-20</td>
<td>0.22±.13</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; year student</td>
<td>96</td>
<td>16</td>
<td>20-23</td>
<td>0.28±.15</td>
</tr>
</tbody>
</table>
Figure 1: Factor loadings along two components (factors) derived from the factor analysis of preference data.
Table 2: Characteristics of subjects in the study of the impact of the KMDD

<table>
<thead>
<tr>
<th></th>
<th>Experiment</th>
<th>Control</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>42</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>male</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>33</td>
<td>37</td>
</tr>
<tr>
<td>Dental nurse program</td>
<td>22</td>
<td>23</td>
<td>0.90&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Pharmacy technician</td>
<td>20</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>18.81±0.97</td>
<td>18.80±1.19</td>
<td>0.98&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>(18-23)</td>
<td>(17-24)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>: P-value of chi-square test with the continuity correction.

<sup>b</sup>: P-value of independent t-test.
Table 3: C-indexes of the subjects

<table>
<thead>
<tr>
<th></th>
<th>Start</th>
<th>After intervention</th>
<th>6 months after the intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment (N=42)</td>
<td>0.21±0.13</td>
<td>0.35±0.11</td>
<td>0.34±0.11</td>
</tr>
<tr>
<td>Control (N=41)</td>
<td>0.25±0.16</td>
<td>0.24±0.15</td>
<td>0.24±0.14</td>
</tr>
</tbody>
</table>

P-value\(^a\) 0.179 <0.001 0.001

\(^a\) Bonferroni adjusted test
References


15. Lind G. Moral dilemma discussion - The Konstanz method. Paper presented at the meeting of the Association for Moral Education (AME); 2002 Nov 7-10; Chicago.


21. Streiner DL, Norman GR. Health measurement scales: A practical guide to their


23. Lind G. The Konstanz method of dilemma discussion (KMDD) [online]. 16 September
2004 [cited 5 November 2004]. Available from URL: http://www.uni-konstanz.de/ag-
moral/moral/dildisk-e.htm.

    Harper&Row; 1981.

Appendix 1: The sample of one dilemma and its items from the MJT

Acknowledgements

The authors thank Professor George Lind, for granting the permission to translate and use the MJT in various phases of the study and for his kind help and suggestion with the validation and the KMDD. The authors also thank the students who were data collectors and the participants in this study.