ASIA PACIFIC JOURNAL OF EDUCATION, 2017
https://doi.org/10.1080/02188791.2017.1386618

Developing and testing a scale of moral thinking and communication (MTC) functioning: a preliminary study and its implications for moral development and education

Chi-Ming Angela Lee[a] and Stephen J. Thoma[b]

[a]Department of Civic Education and Leadership, National Taiwan Normal University, Taipei, Taiwan; [b]The College of Education, The University of Alabama, Tuscaloosa, AL, USA

ABSTRACT
The purpose of this study was to develop and test a scale assessing students' moral thinking and communication (MTC) functioning as well as to explore the implications for moral development and education. The rationale of MTC functioning, including interaction of four independent competencies: moral awareness, moral judgement, moral discourse, and moral decision-making, is primarily adapted from L. Kohlberg's theory, Neo-Kohlbergian theory, J. Habermas's and K.-O. Apel's discourse ethics, and L. Hinman's ethical pluralism. There were two studies in the development of the measure. The resulting instrument contained four moral dilemmas differing by a focus on micro-moral and macro-moral levels. Findings from the studies indicated the MTC scale has acceptable psychometric properties and describe Taiwanese students as attending to the difficult problems of integrating local and international cultural norms. Additionally, the measure provides a detailed picture of students' moral thinking and communication functioning. These results are discussed in terms of implications for moral theory and assessment of moral education.

ARTICLE HISTORY
Received 27 March 2016
Accepted 31 August 2017

KEYWORDS
Moral thinking and communication functioning; Taiwanese students; scale development

1. Introduction: research purposes, theory rationale, and definitions

1.1. Research purposes

Modern moral education in more complicated and challenging societies should endeavour to cultivate students' competencies in moral thinking and communication (hereafter MTC). Without these efforts, a society may inadvertently promote blind obedience to political dogma or moral disengagement both in private and public spheres. Kohlberg (1975, p. 51) indicated "If moral development centers on a sense of individual justice, it becomes apparent that moral and civic education are much the same thing". In addition, moral philosophical considerations are "needed to conduct moral education and finally to enhance general human rights and welfare" (Han, 2014, p. 45). Since 1987 with the abolishment of martial law, Taiwanese political systems and socio-culture demands have gradually evolved from authoritarianism to democratization, pluralism, and liberalization. Concurrent with this cultural/political transition is the need to transform Taiwanese education from its traditional approach to a more modern and adaptable style that is more responsive to current educational trends that stress multiple and usable competencies. An important aspect of this transition is indicated by Taiwanese students' moral thinking and communication competencies, including interaction of four independent competencies:

CONTACT Chi-Ming Angela Lee t11023@ntnu.edu.tw

moral awareness, moral judgement, moral discourse, and moral decision-making (Lee, 2015). Using measures to assess the quality of students' moral thinking and communication competencies is an effective approach for evaluating the impact of moral education curriculum or programmes. Therefore, the main purpose of this research was to develop a sound scale for testing students' MTC functioning and as a means to improve moral teaching and learning, and then to explore the implications for moral development and education. In particular, the scale would be based on students' personal and social encounters in their real-life experiences as well as within a sociocultural context.

Consequently, we adapted insights from related theories and sought to develop a suitable research instrument to evaluate students' MTC competencies from age 12 to university level in Taiwan, an Asia-Pacific country with a blended culture of Eastern and Western elements, rather than simply translating other existing scales employed in other cultures and settings. Moreover, there is a need within the Taiwanese setting for an appropriate measure to evaluate the effectiveness of moral and character education in the Taiwanese educational system. Taiwan has historically emphasized moral education since the Nationalist government moved from China to Taiwan in 1949. More recently, in 1998, moral education was decentralized and there is interest in evaluating the various offerings and developing an understanding of best practices (Lee, 2009). Thus, a theoretically justifiable and empirically valid measure can provide information crucial to this task. To this end, we conducted two main studies: Study I was a pilot study of MTC functioning, containing four competencies: moral awareness, moral judgement, moral discourse, and moral decision-making. This study was designed to explore which strategy is more suitable for testing Taiwanese students' real-life experience. Study II builds from Study I by examining the reliability and validity of the MTC scales.

1.2. Theory rationale and definitions

This research describes the development of a new measure that is informed by competencies from four interrelated theoretical perspectives. From psychology, the measure draws from the legacy of L. Kohlberg's moral-developmental theory and the associated Moral Judgment Interview (MJI) (Kohlberg, 1986). Also from psychology, we attend to the Neo-Kohlbergian theory associated with research using the Defining Issues Test (DIT) and Four Component Model (FCM; moral sensitivity, moral judgement, moral motivation, moral character) (Rest, Narvaez, Bebeau, & Thoma, 1999a). From these two sources we identified "moral awareness" and "moral judgement" as two of the four competencies of MTC functioning assessed by the measure. Concretely speaking, moral awareness in the MTC functioning had broader meaning than moral sensitivity because it contained both sensitivity and empathy, that is, focusing on a competency of recognizing a moral issue/problem existed in a complex social situation (Rest et al., 1999a) as well as of arousing sympathy/empathy to care about self and others (Hoffman, 2000). The familiar construct of moral judgement forms the basis of the second competency. Briefly, moral judgement is a concept of judging right and wrong when faced with a moral dilemma. These concepts are described as developmentally ordered stages and attributed to these theories of I. Kant and J. Rawls based on principles of justice (Carter, 1986, p. 12; Kohlberg, 1984, p. 401).

In addition, the remaining two competencies were derived from philosophy, and, especially, the importance given to communication in the development of individual moral thinking. Specifically, the third competency "moral discourse" was influenced by J. Habermas's and K.-O. Apel's theory of discourse ethics. Habermas (1991) advocated that the purpose of discourse is oriented to reaching understanding in a reflective form within an intersubjective shared life world, which could transcend the "monological" procedure of examining maxims in accordance with I. Kant's categorical imperative and Kohlberg's post-conventional morality. Similarly, Apel (1999, pp. 46–48) argued that discourse ethics can be uncovered through transcendental-pragmatic reflection when we exist in a communicative community. The importance of moral discourse for problem solving and bridging knowledge gaps that the two scholars mentioned above has motivated us to investigate further this competency. Concretely, moral discourse described in the following studies involves evaluating and defending moral arguments, the process of dialogue and communication with each other on moral issues within a process of reasoned discussion and supported by a desire to search for identity of the self and one's place in the community (Habermas, 1991).

Moreover, since disagreement and difference are standard features of the moral landscape, the measure adopted the theory of ethical pluralism, proposed by L. Hinman. It stresses that the truth of a moral life is not unitary nor relative, but "many truths that are sometimes partial and sometimes conflicting" (Hinman, 2008, p. 49). Normative ethical theories, that is, utilitarianism, Kant's deontology, and virtue ethics (Mackinnon, 1995; Pojman & Fieser, 1990) may offer diversified moral standards and principles to use to make moral decisions; as a result, we involved "moral decision-making" as the last one competency of MTC functioning. Moral decision-making is defined as a competency of understanding multiple-perspectives (i.e., utility, duty, rights, and values), deliberative thinking on moral issues, and finally attaining the integrity of oneself or consensus-building with others regarding a tendency to moral action in private and public spheres (Hinman, 2008). In brief, the intention of this new measure was to continue the legacy of existing moral-cognitive instruments, and to broaden the scope of moral-cognitive competencies from personal self-thinking to communication with others, as well as the sphere of moral issues from everyday life to the structure/context.

2. Study I: a pilot study of MTC functioning

2.1. Research questions of Study I

The purpose of this pilot study was to construct both the scoring rubric and the optimal structure of the proposed scale of moral thinking and communication. To this end, our methods were sensitive to Taiwanese culture as well as students' personal and social encounters in real-life experiences. The initial question informing this work focused on one moral dilemma describing a person who must decide whether or not to report to the authorities that his/her best friend cheated in an examination (called the cheating story). The focus on a cheating dilemma was informed by students who were asked to generate situations that were most troubling to them. We suspect the interest in cheating is due in part to the strict and intensive examinations which occur often in Taiwanese school culture. Using this dilemma, the focus of Study I was to explore the similarities and differences in MTC functioning as evaluated by a set of diversified assessment strategies, including A. perspectives on a particular moral dilemma (called students' Dilemma score), B. self-reporting MTC functioning (called students' Self-reported score), C. interviews (called students' Interview score) and D. ratings by the participant's teachers (called teachers' Rating score) (see "Method of Study I" as below).

In order to judge the merits of one measurement type over another we employed a "known groups" approach (Rest, Narvaez, Bebeau, & Thoma, 1999b, p. 74). In short, this approach argues that one aspect of the validity of a measure is its ability to distinguish groups which ought to be different on the proposed construct. Applied to the current study we note that the moral competencies under study are claimed to be both developmental and influenced by educational experiences. Based on these assumptions we would expect to find measures of these moral competencies ought to differentiate university students from high school students given that these groups both differ in age and in educational experiences. Finding differences between these known groups supports the measure (and similarly shores up the support for the theory). Therefore, the specific research questions were: (1) Is there any difference in MTC functioning between students' Dilemma score, Self-reported score, Interview score, or teachers' Rating score? (2) Is there any difference between university students, senior high students, or junior high students' responses to the individual on Dilemma scores, Self-reported scores and Interview scores? (3) Is there any difference between estimates of moral awareness, moral judgement, moral discourse, and moral decision-making for university students, senior high students, or junior high students?

2.2. Method of Study I

2.2.1. Constructing a rubric

The main data gathering methods we employed in Study I were six symposia, questionnaires, and student interviews from August 2012 to February 2014. Guiding these data collection steps was a rubric for moral thinking and communication functioning that was based on related theories and existing assessments. This rubric was assessed and modified through workshops with 10 scholars (whose majors are psychology, curriculum and instruction or statistics) and six elementary and high school teachers. These informants were asked to critique the initial structure of the rubric based on their expertise and experience according to the purpose and goals of this project, theoretical framework and definitions on the four competencies provided by this project. The resulting modified rubric contained four competencies of MTC functioning with 12 criteria and a four-level quality index representing various degrees of proficiency from the lowest to the highest for MTC functioning (see Table 1).

2.2.2. Developing the instruments

Following the development of the rubric we developed a semi-structured student questionnaire and teachers' questionnaire through an iterative process of meetings with faculty and research assistants to ensure these instruments corresponded to the four competencies outlined in the rubric. The first drafts of the measures were further modified after a trial testing for 12 students from two high schools and two universities. Following these steps the resulting student questionnaire included three sections. In part I, students were asked to provide demographic information. In part II, students were asked to consider the cheating story by addressing 12 open-ended questions (e.g., Q1. "What kind of problem do you think that the person of the story is encountering?"). The scoring of these 12 questions was on a "1 to 4" scale and then summed up to create an overall score labelled the Dilemma score. In part III students were asked to complete 12 questions in order to assess students' self-reported moral thinking and communication functioning (e.g., Q1. Do you agree that you are usually aware of the moral issues of current news? (1) totally disagree; (2) somewhat disagree; (3) somewhat agree; (4) totally agree) and were labelled the Self-reported score. In addition, we developed a teacher questionnaire to evaluate their students, which was similar to part III of the students' questionnaire. The questionnaire contained a request to rate student performance on the same 1–4 scale used by the students (e.g., Q1. Do you agree that your students attend to moral issues and current news? (1) totally disagree; (2) somewhat disagree; (3) somewhat agree; (4) totally agree). The score was labelled the teacher Rating score. Each measurement system was scored using the common rubric in order to estimate students' proficiency on the four competencies of MTC functioning. This assessment strategy allowed us to compare performance across competencies that are assumed to tap different processes and unique aspects of moral functioning. Specifically, the use of a common rubric provides a means to assess which and to what degree the three assessment strategies provide different proficiency estimates.

2.2.3. Visiting schools

We visited 16 schools from five junior high schools, six senior high schools, and five universities in the northern, central, eastern, and southern parts of Taiwan. In total we assessed 90 students, including 44 male and 46 female students, and 30 junior high students (mean of age = 14, SD = 1), 30 senior high students (mean = 17, SD = 1), and 30 university students (mean = 20.5, SD = 1.5) were invited to fill in the student questionnaires. After students finished the questionnaire, research staff interviewed students in small groups using the same items presented in part III of the questionnaire to better comprehend the ways in which participants understood the questions. These interviews were transcribed and then consequently rated according to the rubric. The resulting interview score ranged from scores one to four to code and summed the scores of qualitative data according to the rubric of moral thinking and communication functioning (Level A scored 4, Level B scored 3, Level C scored 2, and Level D scored 1). In addition, we requested 16 teachers from each school to fill out the teacher questionnaire as well.

Table 1. The rubric of MTC functioning.

		Quality of four levels			
		1-Level D	2-Level C	3-Level B	4-Level A
Four competencies	12 criteria (c)	No/Limited proficiency	Some proficiency	More proficiency	High proficiency
moral awareness	Be sensitive to (c1), empathetic (c2) and to care about (c3)moral phenomenon and contextual situation	Fails to DO...	DOES some…	Clearly DOES…	Successfully DOES…
moral judgement	To understand with the depth and breadth (c4), to analyse advantages and disadvantages (c5), and to have reasoning for moral events and issues (c6)	Fails to DO...	DOES some…	Clearly DOES…	Successfully DOES…
moral discourse	To defend for moral arguments (c7), to have a moral dialogue with others (c8), and form a commitment on shared moral community (c9)	Fails to DO...	DOES some…	Clearly DOES…	Successfully DOES…
moral decision-making	To possess multiple-perspectives on moral belief (c10), to deliberatively choose for moral alternatives (c11), and to know how to reach democratic consensuses of moral life (c12)	Fails to DO...	DOES some…	Clearly DOES…	Successfully DOES…

2.2.4. Data processing

Data from the open-ended questions and interviews were transcribed for coding based on the rubric. Prior to the coding, eight teachers from different educational levels and two research assistants were trained in the coding process during a day-long meeting. The training provided coders with the theory behind the measure and practice with the rubric. Once trained, the raters coded the student questionnaire and group interviews. Internal consistency reliability (Cronbach's alpha) was estimated for part II and part III of student questionnaires and group interviews. The resulting values were 0.80~0.94 and well within acceptable standards.[1]

2.3. Findings of Study I

Differences in proficiency estimates across the three data gathering methods. The three student summary scores, including the "Dilemma score", "Self-reported score", and "Interview score" were used in the analysis. Each score contained four sub-scores based on four competencies of MTC functioning: moral awareness (ma) (items 1–3), moral judgement (mj) (items 4–6), moral discourse (md) (items 7–9) and moral decision-making (mm) (items 10–12). As described in Table 2 our findings indicated that students' self-reported scores were rated highest followed by the dilemma score and the students' interview score. Using one-way repeated measures (competencies as the within-subjects factor) with *post hoc* comparisons we found statistically significant differences between the different scores and competencies with the exception of the comparisons between the dilemma scores and the interview scores for the moral discourse and moral decision-making assessments. Taken together, the most striking finding was the higher scores participants gave to themselves (see Table 2). Given this tendency for students to elevate their own competencies, we concluded that the students' self-report strategy was less suitable for the MTC scale. By contrast the dilemma and interview strategy provided data that more closely reflected our expectations.

Age/education trends and differences between competencies by assessment strategies. The second series of analyses focused on students' scores by age/education levels. Results from these analyses indicate that senior high school students' moral awareness and moral judgement sub-scores of self-report data received the highest scores among different educational levels (junior high and university). There were no statistically significant differences between the moral discourse and moral decision-making sub-scores by the different educational levels using the self-report data (see Table 3). Students' dilemma scores were assessed for differences across the four competencies by assessment type. With respect to the moral awareness competency of dilemma data, university students' sub-scores were highest,

Table 2. MTC scores and sub-scores of students from different assessment strategies.

	Competency				
Assessment strategies	Moral awareness Mean (SD)	Moral judgment Mean (SD)	Moral discourse Mean (SD)	Moral decision-making Mean (SD)	MTC scores Total Mean (SD)
self-report (N = 89)	3.07 (.48)	2.85 (.49)	3.01 (.62)	3.05 (.54)	3.00 (.41)
dilemma (N = 89)	2.59 (.59)	2.42 (.49)	2.07 (.50)	2.29 (.44)	2.34 (.33)
interview (N = 83)	2.19 (.44)	2.03 (.52)	2.16 (.46)	2.21 (.42)	2.16 (.35)
Self-report vs. dilemma p-value	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
Effect Size Cohen's d	d = .90	d = .88	d = 1.67	d = 1.54	d = .66
Self-report vs. interview p-value	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
Effect Size Cohen's d	d = 1.91	d = 1.62	d = 1.56	d = 1.74	d = .74
Dilemma vs. interview p-value	$p < .001$	$p < .001$	n. s.	n. s.	$p < .001$
Effect Size Cohen's d	d = .77	d = .77	d = −.19	d = −.19	d = .26

Table 3. MTC scores of students from different assessment strategies in different educational levels.

Different assessment strategies and their four competencies	University (1) N = 30 Mean(SD)	Senior high (2) N = 30 Mean(SD)	Junior high (3) N = 29(self-report, dilemma) N = 23(interview) Mean(SD)	Total N = 89 Mean(SD)	F	(p-value)	Scheffe Post Hoc Test
Self-awareness (self-a)	2.89(0.52)	3.22(0.42)	3.11(0.44)	3.07(0.48)	4.049	(.021)	p* < 0.05 (1) < (2)
Self-judgement (self-j)	2.80(0.43)	3.04(0.52)	2.70(0.48)	2.85(0.49)	4.041	(.021)	p* < 0.05 (2) > (3)
Self-discourse (self-d)	2.88(0.60)	3.22(0.61)	2.94(0.60)	3.02(0.62)	2.754	(.069)	n.s.
Self-decision-making (self-m)	2.90(0.53)	3.13(0.58)	3.13(0.50)	3.05(0.54)	1.815	(.169)	n.s.
Dilemma-awareness (dilemma-a)	2.91(0.49)	2.41(0.62)	2.44(0.51)	2.59(0.59)	7.985	(.001)	p** < 0.01 (1) > (2),(1) > (3)
Dilemma-judgement (dilemma-j)	2.38(0.57)	2.52(0.43)	2.37(0.44)	2.42(0.49)	.938	(.395)	n.s.
Dilemma-discourse (dilemma-d)	1.83(0.42)	2.12(0.58)	2.28(0.38)	2.07(0.50)	6.861	(.002)	p* < 0.05 (1) < (2), (1) < (3)
Dilemma-decision-making (dilemma-m)	2.13(0.37)	2.47(0.45)	2.26(0.46)	2.29(0.44)	4.637	(.012)	p* < 0.05 (1) < (2)
Interview-awareness (interview-a)	2.34(0.44)	2.04(0.49)	2.19(0.28)	2.19(0.44)	3.738	(.028)	p* < 0.05 (1) > (2)
Interview-judgement (interview-j)	2.29(0.51)	1.90(0.55)	1.90(0.39)	2.03(0.52)	6.383	(.003)	p** < 0.01 (1) > (2),(1) > (3)
Interview-discourse (interview-d)	2.16(0.43)	2.22(0.56)	2.10(0.37)	2.16(0.46)	.499	(.609)	n.s.
Interview-decision-making (interview-m)	2.34(0.42)	2.20(0.48)	2.08(0.30)	2.21(0.42)	3.084	(.051)	n.s.

Note: All F tests were assessed using (2,80) degree of freedom.

while their moral discourse sub-scores were lower than both senior high students and junior high students. Additionally, senior high students' moral decision-making sub-scores of dilemma data were higher than the other educational levels (see Table 3). There were no statistically significant differences on the moral judgement sub-score dilemma data. As expected, university students' interview-based moral awareness and moral judgement sub-scores were the highest scores among different educational levels. However, there were no statistically significant differences on moral discourse and moral decision-making interview scores among different educational levels (see Table 3).

Differences between assessment strategies and age/education levels. The third analysis using a one-way repeated measures ANOVA with competencies for each assessment strategy as the within subject factors and age/education as the between-subjects factor was used to test interactions between factors. *Self-report scores*: Moral competencies assessed by a self-report methodology indicated lower sub-score for the moral judgement competency (see Table 3). There was no statistically significant interaction between the self-reported scores and educational levels ($F_{(6,258)}$ = 1.572, p > 0.05). *Dilemma scores:* Findings associated with the dilemma methodology indicated that the differences between competencies varied by age/educational grouping ($F_{(6,258)}$ = 9.962, p<0.001). Specifically, university students' scores were highest on the moral awareness competency, but lower than high school students' scores on scores measuring moral discourse and moral decision-making competencies (see Table 3). *Interview scores*: Differences in competency estimates were also observed for this method. University students' scores were statistically different from both junior high and senior high school students. Interactions between competency and age/education ($F_{(6,240)}$ = 3.419, p<0.01) indicated that moral discourse scores were lowest for university students, while the moral judgement scores were the lowest for high school students (see Table 3). Interestingly, university students scored higher on moral awareness than the other two groups on the interview assessment.

Table 4. MTC sub-scores of teachers' rating students.

Sub-scores	Mean	SD
Moral awareness	2.94	0.44
Moral judgement	2.81	0.47
Moral discourse	2.81	0.47
Moral decision-making	2.60	0.57

Teacher ratings of MTC functioning. The fourth analysis focused on the 16 teachers' rating students' MTC functioning. Following the repeated measures ANOVA approach used in the previous analyses, teachers differed in their ratings across the students' competencies. Specifically, teachers found students' performance was higher in moral awareness followed by moral judgement, moral discourse, and moral decision-making respectively. This finding is consistent with a descending pattern. Interestingly, the average teachers' rating was lower than students' self-reported scores but higher than students' dilemma scores and interview scores (see Table 4).

2.4. Discussions of Study I

(1) Although the three assessment strategies of Study I were based on the same MTC rubric (see Table 1), we find differences between the estimated competencies. Specifically, we find self-reported scores were consistently higher than dilemma scores which were in turn, higher than interview scores. There were no statistically significant differences between the dilemma and interview scores regarding moral discourse and moral decision-making competencies. In addition, teachers' rating scores of students were lower than students' self-reported scores but higher than students' dilemma scores and interview scores. Taken together, it appears that the self-reporting assessment strategy overestimates students' competencies. The other two assessment strategies utilizing a dilemma story and a group interview approach obtained similar results for moral thinking and communication functioning. It is also interesting to note that the dilemma approach worked as well as the more labour intensive interview approach when both focus on the same dilemmas and focus on similar questions. This finding is reminiscent of the observed relationships between the DIT-1 and Kohlberg's interview measure (Thoma, 2006, p. 77).

(2) Comparing the different educational levels on the three assessment strategies, we found that senior high school students' moral awareness and moral judgement sub-scores were the highest scores among different educational levels on the self-reported scores. By contrast, university students' scores were the lowest across groups. One wonders whether these results indicate that high school students are not as introspective and overinflate their own skills as compared with university students. On the dilemma scores, a different pattern was observed. Specifically, university students' moral awareness sub-scores were the highest across all groups but moral discourse sub-scores were the lowest among different educational levels. A similar pattern was noted on the interview scores. Here university students' moral awareness and moral judgement sub-scores were highest among different educational levels. Taken together, on the assessments that focus on production, students' moral awareness and moral judgement competencies increased with advanced educational levels.

(3) Based on the rationale of the MTC functioning model, we wondered whether there were certain patterns that link the four competencies: moral awareness, moral judgement, moral discourse, and moral decision-making. According to our empirical findings on the total scores and sub-scores of students' responses to the dilemma story, the general pattern indicated a descending order from moral awareness, moral judgement, and moral discourse to moral decision-making. However, we found that there were variations from this general pattern for different educational levels and score types. Responding to dilemmas, university students scored higher on moral

awareness competency, but lower than high school students' scores on moral discourse and moral decision-making competencies. Additionally, the patterns of interview scores showed that university students' scores were statistically different from both junior high and senior high school students. The moral discourse sub-scores were the lowest for university students, while the moral judgement sub-scores were the lowest ones for high school students. Taken together, the interview results suggest that university students are more hesitant to evaluate and defend their moral arguments, while high school students have less competency in justifying their moral principles. For example, university students usually said that they need to respect all opinions and they do not want to evaluate others' viewpoints or defend their own moral beliefs. In contrast, high school students are quick to say what is right or wrong but found it harder to defend their positions.

3. Study II: three assessments to test the properties of the MTC scale

3.1. Research questions of Study II

Based on the previous findings of Study I, we adopted the dilemma assessment strategy as the MTC scale's focus for Study II. This decision was based on two factors: first the dilemma approach provided competency scores that best fit a clear teaching and learning pattern descending from moral awareness, moral judgement, and moral discourse to moral decision-making. Second, dilemma scores were most likely to demonstrate age/educational trends. Following this decision, the goal for Study II was to assess the psychometric properties of the resulting measurement of the MTC functioning scale. Prior to this assessment three additional dilemmas were developed along with fixed items representing the four competencies. Once developed and piloted, the major questions associated with Study II were: (1) Do the findings support the validity of the MTC scale and show a descending order from moral awareness, moral judgement, and moral discourse to moral decision-making? (2) Is the MTC scale sensitive to age-educational group differences? (3) Is the MTC scale sensitive to educational interventions? In addition, Study II also addressed the reliability of the MTC scale. Specifically, two reliability estimates were computed: (1) the test-re-test reliability of the MTC scale, and (2) internal reliability of the MTC scale.

3.2. Instruments of Study II

The main method employed by the research group in Study II utilized three data gathering waves during a period from September 2013 to July 2014. The MTC functioning scale was designed to include four versions. Each version combined two moral dilemma stories chosen from two main subcategories: micro-morality and macro-morality. Micro-morality refers to the particular face-to-face relations that people have in everyday life; by contrast, macro-morality describes the formal structure of society, as defined by institutions, rules, and roles (Rest et al., 1999b, p. 291). The resulting stories consist of the original cheating story described in Study I representing a micro-morality issue, a story focused on a decision whether or not to terminate a pregnancy (labelled the abortion story, micro-morality), a story describing the choices between environmental vs. economic development considerations (labelled the environment story, macro-morality), and a story describing the arguments for waging a pre-emptive war to prevent genocide (labelled the war story, macro-morality) (see Appendix 1 for the four stories).

Each story is followed by nine questions: the first question requests participants to choose whether they support or oppose the protagonist's moral action. Following this action choice, participants were asked eight questions representing the four competencies of MTC functioning. Each competency is assessed using two approaches: rating and ranking of items representing the four competencies. For the rating task 12 items were presented to participants for rating using a 5-Likert scale (i.e., ranging from 1–5 with 1 being strongly disagree and 5 being strongly agree). Following the rating task, participants were asked to rank the four most important items.

For this study one micro-morality and one macro-morality story was paired to form four versions of the MTC: Version-I – included the cheating story and environment stories; Version-II – included the abortion and war story; Version-III – included the cheating and war story; and Version-IV – included the abortion and environment story. The four stories were vetted by four scholars and four teachers who were asked to assess the content of the stories through discussion and reaching consensus. Preliminary validity was addressed using a formal process of testing the face validity of the items as well as content validity through the use of a two-way specification tables (see Appendix 2 sample table of moral awareness competency of cheating story).

3.3. Participants of Study II

Sample 1: Participants included 91 (63 valid questionnaires, male = 27, female = 36) Taiwanese students from two junior high schools (aged 13–15), two senior high schools (aged 16–18) and two universities (aged 19–22). This sample was sub-divided into two groups each receiving a different combination of MTC versions at two testing dates separated by a three-week interval. Specifically, Group 1 (N = 33 valid) was administered the Version-I of the MTC scale and re-tested with the Version-III, whereas Group 2 (N = 30 valid) received the Version-II and then re-tested with the Version-IV.

Sample 2 ： Participants included two sophomore classes (104 students and 67 valid questionnaires, male = 17, female = 50) of Taiwanese undergraduate students (mean of age = 20, SD = 1) at the university level. Both samples received a pre-test and post-test of two versions of the MTC at the beginning and end of an 18-week semester. All students were enrolled in one of two Ethics courses, including MTC model applied on social issues and taught by one of the researchers during the semester (from September 2013 to January 2014). Class-I (N = 24 valid) participants were assessed using the Version-I and Version-IV of the MTC scale for the pre-and post-tests respectively. Class-II (N = 43 valid) were tested on the same schedule but were assessed on the Version-II and Version-III of the MTC at the pre- and post-test.

Sample 3 ： Participants included 1,363 (936 valid questionnaires, male = 403, female = 529, unfilled = 4) students from 12 junior high schools, 10 high schools, and 20 universities (including undergraduate and graduate levels) in the northern, central, eastern, and southern parts of Taiwan and tested from March to July 2014. This sample was sub-divided into four groups according to their educational levels and each group was tested by the differing MTC four versions. Junior high students (259 valid questionnaire) took Version-I, including the cheating and environment story. Senior high students (282 valid questionnaire) took Version-II, including the abortion and war story. Undergraduate students (237 valid questionnaire) took Version-III, including cheating and war story. Graduate students (158 valid questionnaire) took Version-IV, including abortion and environment story.

3.4. Computation and scoring indexes of Study II

Responses to the questionnaires were summarized by focusing on participants' responses to the "ranking" questions across the different stories. The resulting score was labelled the PL1 index. The PL1 index is informed by the P score of Defining Issues Test (Rest et al., 1999a). Following the DIT model, this score focusing on the ranking data attends to the proportion of items selected by the participant that represents "more" (Level B) and "high" (Level A) proficiency of MTC functioning. The 12 items of rating and ranking for each competency of MTC scale correspond to a pattern from low to high proficiency from D, C, B to A the same as the rubric (see Table 1) of the cheating story in Study I (see Table 1). The ranking items listed by the participants would be scored "4" for the first most important item, "3" for the second, "2" for the third, and "1" for the fourth. The PL1 index is summed up by the individual scores of A, B, C, D and computed by [(A + B)/(A + B + C + D)]*100% (see Appendix 2). In addition, we had a very rigid selection for valid questionnaires to ensure their quality, that is, any missing data invalidated the questionnaires.

Table 5. The Cronbach's alpha of internal reliability of four stories in Sample 3.

Story	Internal reliability (each story consisted of five items including ma, mj, md, mm and total)
Cheating	Cronbach's alpha = 0.81 (N = 496, Group1 and Group3)
Abortion	Cronbach's alpha = 0.71 (N = 440, Group2 and Group4)
Environment	Cronbach's alpha = 0.77 (N = 417, Group1 and Group4)
War	Cronbach's alpha = 0.81 (N = 519, Group2 and Group3)

Note: Total valid questionnaires 938.
Group1: 259 junior high school students; Group2: 282 senior high school students; Group3: 237 undergraduate students; Group4: 158 graduate students.

3.5. Findings and discussions of Study II

Reliability of MTC scale – test/re-test and internal consistency. Test-retest reliability was computed for participants in Sample 1. These correlations were found to be moderate to high (0.61~0.79) on the cheating story of the MTC scale ($r = 0.72$, $p<0.01$, $N = 33$) and on the abortion story ($r = 0.67$, $p<0.01$, $N = 30$). Similarly, internal consistency reliability estimates (Cronbach's alpha) assessed within Sample 3 indicated acceptable values on all the four stories (0.71 and higher see Table 5).

3.5.1. Validity of MTC scale – pattern, type, and version

Results from Sample 1 addressed the pattern of responses across the four MTC competencies. Taken together we found that moral awareness typically was scored the highest, followed by moral discourse, moral decision-making, and moral judgement scores (see Table 6). These results are consistent with the findings from Study I which indicated the presence of a descending order with moral awareness typically scored the highest, followed by moral discourse, moral decision-making, and moral judgement. Therefore, these results provide evidence of face and content validity for the MTC scale. In addition, we wondered whether scores differed by dilemma type. To assess this question we combined the samples and formed two groups based on whether participants completed the micro-moral and macro-moral stories. Using a paired sample T-test we found there was no statistically significant difference between the scores of micro-moral or macro-moral stories (see Table 7).

3.5.2. Validity of MTC scale – sensitivity to educational interventions

Using Sample 2, students' participation in an Ethics course is served as an educational intervention for this research. According to statistical results, all the post-test scores were higher than the pre-test ones with the majority of these differences showing statistically significant difference on the PL1 index. These findings indicated that the MTC measure was sensitive to an educational intervention. Specifically, there were statistically significant differences between pre- and post-test of the environment story ($t_{(23)}$= -12.95, $p<0.001$, $N = 24$) and war story ($t_{(42)}$=-7.22, $p<0.001$, $N = 43$). These were the two stories used in the intervention study (see Table 8).

3.5.3. Validity of MTC scale – differentiation of age/educational groups

The findings of Study I indicated that the differences between competencies varied by age/educational grouping. In addition, students' moral awareness and moral judgement competencies increased with advanced educational levels. However, in Sample 1 of Study II we found that there were no statistically significant differences among the four competencies of MTC scale by educational levels as defined by university, senior high, and junior high school. This unanticipated finding may have been due to a small sample and limited educational range. To overcome these limitations we assessed age/educational trends using Sample 3, which had larger samples than used in Sample 1 and included students from junior high school through graduate school to assess age/educational trends. The finding indicated that there were 13 items in 15 comparisons with statistically significant differences for educational levels. Among them, 10 comparisons revealed that the higher educational level students scored higher than the others, except the undergraduate students' scores. Additionally, these comparisons, showing the

Table 6. MTC scores and sub-scores in four versions in Sample 1.

	Version I group 1- test	Cheating story		Environment story		Version III group 1-retest	War story		Cheating story	
	N	Mean	SD	Mean	SD	N	Mean	SD	Mean	SD
moral awareness	33	72.42	29.48	80	15.41	33	78.48	19.7	77.88	24.59
moral judgement	33	66.06	25.12	71.52	26.12	33	74.85	19.22	65.76	22.5
moral discourse	33	69.09	26.14	78.18	15.7	33	80.61	21.64	72.73	24.01
moral deci-sion-making	33	67.88	28.91	70.61	21.93	33	75.15	20.78	68.48	26.82
total	33	68.91	20.39	75.08	9.69	33	77.27	14.12	71.89	16.27
	Version II group 2-test	Abortion story		War story		Version IV group 2-retest	Environment story		Abortion story	
	N	Mean	SD	Mean	SD	N	Mean	SD	Mean	SD
moral awareness	30	81	20.74	81	15.17	30	83	16.43	85.67	18.13
moral judgement	30	67.67	24.17	74.33	25.42	30	79.33	19.64	73.97	23.29
moral discourse	30	70.33	24.42	77.33	15.96	30	83.67	13.26	95.33	10.74
moral deci-sion-making	30	69.67	28.71	70	21.34	30	85	15.7	83.33	17.49
total	30	82.81	13.78	80.92	13.75	30	82.75	9.2	84.54	11.28

Table 7. Paired samples statistics between micro-macro stories in Sample 1.

		Mean	N	SD	t-value	sig	Cohen's d (absolute value)
Pair1	microMA	79.05	126	24.01			
	macroMA	80.56	126	16.70	−.664	n.s.p=0.508.	0.07
Pair2	microMJ	68.25	126	23.73			
	macroMJ	74.92	126	22.72	−3.143	**$p = 0.002$	0.29
Pair3	microMD	76.59	126	24.50			
	macroMD	79.92	126	16.99	−1.273	n.s $p = 0.205$	0.16
Pair4	microMM	72.14	126	26.46			
	macroMM	75.08	126	20.77	−1.199	n.s $p = 0.233$	0.13
Pair5	microTotal	76.72	126	17.14			
	macroTotal	78.87	126	12.15	−1.461	n.s $p = 0.147$	0.14

*ma-moral awareness, mj-moral judgment, md-moral discourse, mm-moral decision-making.

Table 8. The validity of sensitivity to educational interventions focused on pre- and post-tests in Sample 2.

	M	mean	N	SD	T	df	sig	p-value	Cohen's d (absolute value)
Pair1	environment-ma	79.58	24	19.44	−2.02	23	0.06	n.s.	0.56
	environment (post)-ma	88.75	24	12.62					
Pair2	environment-mj	78.33	24	17.36	−1.18	23	0.25	n.s.	0.40
	environment (post)-mj	85	24	15.6					
Pair3	environment-md	71.67	24	21.4	−5.06	23	0	***	1.45
	environment (post)-md	95.42	24	8.84					
Pair4	environment-mm	68.75	24	26.26	−2.48	23	0.02	*	0.69
	environment (post)-mm	84.17	24	17.67					
Pair5	environment-total	71.56	24	8.72	−7.89	23	0	***	2.00
	environment (post)-total	88.33	24	8					
Pair6	war-ma	45.58	43	17.63	−12.95	42	0	***	2.64
	war (post)-ma	90	43	16.04					
Pair7	war-mj	74.19	43	17.89	−1.4	42	0.17	n.s.	0.30
	war (post)-mj	79.3	43	16.53					
Pair8	war-md	73.02	43	23.86	−3.33	42	0	**	0.53
	war (post)-md	84.65	43	19.68					
Pair9	war-mm	78.84	43	17.49	−2.4	42	0.02	*	0.40
	war (post)-mm	85.12	43	13.87					
Pair10	war-total	67.91	43	10.41	−7.22	42	0	***	1.40
	war (post)-total	85.12	43	13.87					

p.s.1.*p < 0.5; **p < 0.01; ***p < 0.001; 2. ma-moral awareness, mj-moral judgement, md-moral discourse, mm-moral decision-making.

validity of sensitivity to educational levels, more clearly correspond to the micro-moral stories (cheating story or abortion story) than the macro-moral stories (environment story or war story) (see Table 9).

In sum, we constructed the MTC functioning scale, which combines two moral dilemma stories chosen from micro-morality (cheating story or abortion story) and macro-morality (environment story or war story), and further conducted three samples to examine the validity and reliability of the MTC scale besides the face and content validity we tested during the development of the four stories. Firstly, when comparing the scores of the four competencies of MTC functioning, we found that moral awareness was typically scored the highest, followed by moral discourse, moral decision-making, and moral judgement scores in Sample 1, while Sample 3 revealed another pattern of connecting MTC competencies, descending from moral awareness, moral judgement, and moral discourse to moral decision-making. Secondly, results showed that there were statistically significant differences among the four competencies of MTC scale between educational levels, including graduate, undergraduate, senior high, and junior high students; however, we needed to be concerned with the moral competencies of undergraduate students because of their lower scores than expected. Moreover, the MTC scale was shown to be sensitive to educational interventions in a university population. Finally, the re-test reliability of the MTC scale showed middle to high correlation (0.61~0.79). Similarly, the internal reliability of the

Table 9. The validity of sensitivity to educational levels in Sample 3 of Study II.

	Junior high	Senior high	Undergraduate	Graduate		
	(1) N = 259	(2) N = 282	(3) N = 237	(4) N = 158		
Story and competencies	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	F(p-value)	Scheffe Post Hoc Test
Micro-story moral awareness	73.98(24.59)	79.72(20.16)	74.34(21.95)	83.10(14.32)	8.88(0.00)	$p < 0.001$ (1) < (2),(1) < (4) (2) > (3),(3) < (4)
Micro-story moral judgement	71.24(22.81)	74.89(22.98)	74.17(20.21)	75.19(23.26)	1.59(0.19)	n.s.
Micro-story moral discourse	66.10(30.05)	66.70(16.77)	72.76(23.51)	75.51(17.76)	8.43(0.00)	$p < 0.001$ (1) < (3),(1) < (4) (2) < (3),(2) < (4)
Micro-story moral decision-making	61.97(29.62)	75.11(22.58)	68.09(24.53)	66.96(23.48)	12.33(0.00)	$p < 0.001$ (1) < (2),(2) > (3)
Micro-story moral competencies(all)	68.32(18.85)	74.10(13.23)	71.62(15.75)	75.19(11.13)	9.08(0.00)	$p < 0.001$ (1) < (2),(1) < (4)
Macro-story moral awareness	78.15(19.93)	80.14(20.18)	72.95(25.58)	77.41(21.60)	4.77(0.00)	$p < 0.001$ (2) > (3)
Macro-story moral judgement	74.36(23.02)	75.32(17.00)	68.06(21.81)	65.51(26.02)	10.46(0.00)	$p < 0.001$ (1) > (3),(1) < (4) (2) > (3),(2) > (4)
Macro-story moral discourse	71.51(23.54)	71.70(26.14)	64.81(24.43)	68.04(16.87)	4.66(0.00)	$p < 0.001$ (1) > (3),(2) > (3)
Macro-story moral decision-making	64.48(24.08)	71.91(23.12)	65.84(25.98)	72.53(22.85)	6.71(0.00)	$p < 0.001$ (1) < (2),(1) < (4), (2) > (3)
Macro-story moral competencies(all)	72.12(15.18)	74.77(14.64)	66.22(18.26)	70.87(13.89)	13.18(0.00)	$p < 0.001$ (1) > (3),(2) > (3), (3) < (4)
Two stories moral awareness	152.12(34.48)	159.86(32.92)	144.57(40.21)	160.51(27.25)	10.74(0.00)	$p < 0.001$ (2) > (3),(3) < (4)
Two stories moral judgement	145.60(34.91)	150.21(31.12)	140.65(35.55)	140.70(38.29)	4.22(0.01)	$p < 0.05$ (2) > (3)
Two stories moral discourse	137.61(42.57)	138.40(31.79)	133.19(42.02)	143.54(25.42)	2.53(0.06)	n.s.
Two stories moral decision-making	126.45(44.25)	147.02(35.86)	130.52(46.10)	139.49(34.88)	13.33(0.00)	$p < 0.001$ (1) < (2),(1) < (4), (2) > (3)
Two stories moral competencies(all)	140.44(28.71)	148.87(23.99)	137.24(29.66)	146.06(19.92)	10.01(0.00)	$p < 0.001$ (1) < (2),(2) > (3), (3) < (4)

Note: All F tests were assessed using (3, 932) degree of freedom.

four stories' scores were equal to or above 0.71. Given evidence to support the MTC's reliability and validity, it seems reasonable to conclude that the measure has acceptable psychometric properties.

4. General discussions and implications for moral development and education

The purpose of this paper was to describe and test a multi-faceted measure of MTC functioning based on L. Kohlberg's theory, Neo-Kohlbergian theory, J. Habermas's and K.-O. Apel's discourse ethics, and L. Hinman's ethical pluralism as a means to improve moral teaching and learning, based on the Taiwanese context. A number of discussions and implications are listed as below.

First, according to the research, findings of two studies of the MTC scale, which was based on Taiwanese students' real-life experience, revealed preliminary support for the reliability and validity of an item-based moral thinking and communication functioning scale. Given that the MTC scale was established in the Taiwanese context, an open question is whether the results described above are limited to this context or have a more general utility. We would note that other measures of moral functioning have demonstrated an ability to provide useful data in contexts other than the ones in which they were constructed (Lind, 2008; Rest et al., 1999b; Snarey & Walker, 2004; Thoma, Derryberry, & Crowson, 2013). Thus we think it beneficial to explore the similarity and difference between Taiwan and other countries and in other contexts.

Second, the measurement development in Study I and the samples of the MTC scale in Study II indicated two patterns: one pattern showed that the moral awareness competency was associated with the highest average scores, followed by moral discourse, moral decision-making, and finally moral judgement in Sample 1 of Study II; the other pattern revealed a descending order from moral awareness, moral judgement, moral discourse, and finally moral decision-making scores in Sample 3 of Study II. This finding raises the question of a potential neglected focus on four competencies of moral thinking and communication within Taiwanese moral education. The moral judgement, moral discourse, and moral decision-making scores were lower than the expected scores possibly because of the particularity of Taiwanese culture which emphasizes strict national examinations for entrance to higher educational levels but at the same time neglects to teach moral competencies. One might argue that the low scores on the moral judgement scale may be a function of a flaw in the assessment. However, we would note that the measure appears sensitive to moral competencies as evidenced by the Ethics course intervention presented in Sample 2 of Study II. Specifically, the findings indicated a positive effect of exposure to training and guidance in formulating moral judgements and other competencies. These findings support the moral competencies scale and increase the likelihood that the obtained findings suggesting lower scores in the general population are an accurate reflection of these competencies. The intervention study also supports the MTC functioning model and measurement system as an evaluative tool for moral and character education.

Third, regarding the validity of sensitivity to educational levels in Sample 3 of Study II, some of the stories indicated clear trends (cheating and abortion) while other stories did not clearly indicate age/educational trends (environment and war). Interestingly, the presence and absence of clear age/educational trends were associated with macro- or micro-moral stories. We are also interested in why the undergraduate students' scores were not uniformly higher than the senior high school students' scores. Although speculative, one possible explanation for the findings follows from the finding of relativism associated with the college-aged population. For instance, King and Kitchener (1994) noted the high degree of relativism in the university student's thinking about the nature of knowledge. Similarly, Kohlberg noted and was concerned that "the phenomena of relativism had also been critical to our locating principled moral thinking as a moral stage occurring in youth and adulthood" (Kohlberg, 1984, p. 446). By contrast, high school students appear more accepting of positions advocated by society as evidenced by social norms and structures (King & Kitchener, 1994). It may be that senior high school students were more comfortable responding with socially acceptable responses whereas the undergraduate students approached the items with a more critical stance and were more equivocal. It is important to note that undergraduate students had strengths in moral awareness and decision-making particularly in the interview and dilemma assessments where one must interpret and problem solve. These strengths are more consistent with other measures of cognitive moral developmental measures. It is also important to note that the years 2012 to 2014 when these data were collected were associated with intense new social movements occurring in Taiwan, for example, land justice, environmental justice, labour rights, minority rights, and gender equality, among others. While high school students had many chances to systematically and comprehensively learn related knowledge and judgement regarding current news on moral issues from mandatory curricula in school (e.g., Social Studies, Civics and Society), most undergraduate students superficially learned these issues from social media and did not have a formal way to develop their understanding of these issues. This explanation exactly corresponded to the students' interview of the pilot study in Study I of this research and speaks to the concern that in the absence of targeted discussion, simple socialization is not sufficient to develop moral understanding (Bebeau, 2002).

Finally, the development of the MTC scale described in these studies provides a measure that is unique in the field. Unlike the more traditional measures such as the DIT and Moral Competence Test (MCT, former MJT), the MTC scale broadens the focus of moral functioning to include discourse and awareness as well as distinguishing between moral judgement and moral decision-making competencies. As such, it may be a more useful measure in the assessment of high school and university students' moral functioning. Furthermore, the sensitivity of the MTC scale to moral education (Ethics) interventions found in Study II suggests the measure as an evaluation tool for educational interventions at these age and education levels. These strengths support the use of the measure in other cultures and settings. The MTC scale revealed success in capturing moral competencies even though there are some limitations. We note that the MTC requires significant effort on the part of participants as well as time to complete the various sections. The higher than expected rate of invalid questionnaires raises the question that there is something systematic in who completes and who abandons the measure prematurely. Exploring more closely how participants – particularly younger students – complete the measure is important for evaluating and improving the quality of a scale. To expand the support for the MTC measure we suggest additional studies that assess different strategies for developing scores. For instance, it may be profitable to explore alternative indexes derived from the rating data. An advantage of using rating data over the current focus on the ranking data is that different statistical approaches can be used to test the scales such as exploratory or confirmatory factor analyses. Our focus on the ranking data was supported by findings that measures in the moral domain have shown advantages for ranking data presumably because the participant must consider the set of items as a whole (Thoma, 2006). In addition to different scoring procedures we also support additional validity studies to further explore the psychometric properties of the measure. A particularly profitable approach would be tests of the MTC's convergent validity with other established measures in the field such as the Defining Issues Test. We would expect the MTC to relate most strongly to the judgement scale and less so the others and in so doing provide assessments of both convergent and divergent validities.

Note

1. The formula of Internal consistency reliability is:Cronbach's alpha = m(Meanr)/[1 + (Meanr)(m-1)] (where Meanr = the mean of all inter-coder correlations for a given variable and m = the number of coders) (Neuendorf, 2002, p. 166).

Acknowledgements

We wish to thank APJE anonymous referees for their insightful comments and suggestions. Finally, we would like to express my appreciation to all the participants, people consulted and research assistants in this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This article was a research project funded by the Taiwan Ministry of Science and Technology and assisted by the Academic Paper Editing Clinic, National Taiwan Normal University.

Notes on contributors

Chi-Ming Angela Lee, PhD, is Professor of Department of Civic Education and Leadership, College of Education, National Taiwan Normal University, Taipei, Taiwan. She has served as an editorial board member of the Journal of Moral Education since 2005. Lee focuses on empirical research regarding moral development, moral atmosphere and professional curricula in civic and moral education, both creation and improvement of their implementation in schools. She has published seven books in Chinese (i.e., "Character/Moral Education and School Culture-Building", 2011) and numerous articles in Chinese and in English on these topics.

Stephen J. Thoma, PhD, is Professor of educational psychology and University Distinguished Research Professor at The University of Alabama, Alabama, USA. He is also a fellow of the American Psychological Association and the American Education Research Association. Thoma is internationally known for his research in moral psychology and psychometrics, particularly known for creating and validating various measurement strategies in the social/moral domain. Thoma has published numerous papers related to the "Defining Issues Test" as well as moral judgement development.

References

Apel, K.-O. (1999). *The response of discourse ethics—to the moral challenge of the human situation as such and especially today*. Leuven, Belgium: Peeters.

Bebeau, M. J. (2002). The defining issues test and the four component model: Contributions to professional education. *Journal of Moral Education, 31*(3), 271–295.

Carter, R. (1986). Does Kohlberg avoid relativism? In S. Mogdil, & C. Mogdil(Eds.), *Lawrence Kohlberg: Consensus and controversy* (pp. 9–20). Philadelphia, PA: Falmer.

Habermas, J. (1991). *Justification and application: Remarks on discourse ethics*. (C. P. Cronin, Trans.). Cambridge, MA: MIT.

Han, H. (2014). Analysing theoretical frameworks of moral education through Lakatos's philosophy of science. *Journal of Moral Education, 43*(1), 32–53. doi:10.1080/03057240.2014.893422

Hinman, L. (2008). *Ethics: A pluralistic approach to moral theory* (4th ed.). Belmont, CA: Thomson.

Hoffman, Martin L. (2000). *Empathy and moral development*. Cambridge, MA: Cambridge University Press.

King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass.

Kohlberg, L. (1975). Moral education for a society in moral transition. *Educational Leadership, 33*(1), 46–54.

Kohlberg, L. (1984). *Essays on moral development: The psychology of moral development*. San Francisco, CA: Harper Row.

Kohlberg, L. (1986). A current state on some theoretical issues. In S. Modgil & C. Modgil (Eds.), *Lawrence Kohlberg: Consensus and controversy* (pp. 485–546). Philadelphia, PA: Falmer.

Lee, C.-M. (2009). The planning, implementation and evaluation of a character-based school culture project in Taiwan. *Journal of Moral Education, 38*(2), 165–184. doi:10.1080/03057240902792686

Lee, C.-M. (2015). Moral thinking and communication (MTC) competencies as the core of character and moral education. *Secondary Education, 66*(4), 6–17. doi:10.6249/SE.2015.66.4.01 (in Chinese)

Lind, G. (2008). The meaning and measurement of moral judgment competence: A dual-aspect model. In D. Fasko & W. Willis (Eds.), *Contemporary philosophical and psychological perspectives on moral development and education* (pp. 185–220). Creskill: Hampton Press.

Mackinnon, B. (1995). *Ethics: Theory and contemporary issues* (2nd ed.). Belmont, CA: Wadsworth.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Pojman, L., & Fieser, J. (1990). *Ethics: Discovering right and wrong* (7th ed.). Boston, MA: Wadsworth.

Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (1999a). A neo-Kohlbergian approach: The DIT and schema theory. *Educational Psychology Review, 11*(4), 291–324.

Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (1999b). *Postconventional moral thinking: A neo-Kohlbergian approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Snarey, J., & Walker, V. S. (2004). Conclusion: Primary values and developing virtues of African American ethics. In V. S. Walker & J. R. Snarey (Eds.), *Race-ing moral formation: African American perspectives on care and justice* (pp. 130–146). NY: Teachers College Press.

Thoma, S. (2006). Research using the defining issues test. In M. Killen & J. Smetana (Eds.), *Handbook of moral psychology* (pp. 67–91). Mawah, NJ: Lawrence Erlbaum Associates.

Thoma, S. J., Derryberry, W. Pitt, & Crowson, H. M. (2013). Describing and testing an intermediate concept measure of moral thinking. *European Journal of Developmental Psychology, 10*(2), 239–252. doi:10.1080/17405629.2013.766129

Appendix 1

Story 1. Reporting Cheating Story

Bruce and Jack are good friends at school. Both of them were straight A students in high school. However, Bruce missed many classes due to illness this semester. Jack discovered that Bruce had cheated on several exams. A few classmates noticed as well, but their teacher did not.

I. Under this situation, do you think Jack should tell the teacher the truth?

□1. He should tell the teacher. □ 2. I can't make a decision. □ 3. He shouldn't tell the teacher.

II. Could you imagine Jack's problems or feelings on these matters? Please indicate your responses to the following statements

	Disagree……Agree				
Items	1	2	3	4	5
1. It is not a big deal.	□	□	□	□	□
2. This is a dilemma between promoting justice and pressure from classmates.	□	□	□	□	□
3. If you do not report it, it would be against your own interests.	□	□	□	□	□
4. This is an issue of complying by school rules or betraying friends.	□	□	□	□	□
5. I would feel a struggle between unfairness and friendship.	□	□	□	□	□
6. I would feel inconvincible but don't want to damage the friendship.	□	□	□	□	□
7. No particular feelings, remain calm.	□	□	□	□	□
8. I would feel frustrated, so just face the problem myself.	□	□	□	□	□
9. Society has its own elimination mechanism, so let Bruce worry about himself.	□	□	□	□	□
10. Beware of the overall impact on society.	□	□	□	□	□
11. Everyone is central to the community, and he/she has to try to make an influence.	□	□	□	□	□
12. I don't care because it happens all the time.	□	□	□	□	□

III. Please select 4 items from Part II which most closely match your perspective on Jack's problems and feelings and rank them in order.

First _____ Second_____ Third______ Fourth______

IV. & V. What do you think about cheating on exams and reporting friend's cheating? Please indicate your responses to the following statements. (The following 12 items rated and ranked omitted.)

VI. & VII. What kind of reasons do you think are most convincing to the public when you discuss with others whether to report a friend cheating or not? Please indicate your responses to the following statements. (The following 12 items rated and ranked omitted.)

VIII. & IX. What are your standpoints and decisions on this issue, which will lead to your intended action? Please indicate your responses to the following statements. (The following 12 items rated and ranked omitted.)

Story 2. An Abortion Story

Sunny aims to be a musician as her parents have made great efforts on her cultivation since she was a child. Due to her talents and efforts, she has gotten numerous awards in major music competitions. When she was a 12th grade student, Sunny met Brian, who is the same age and from a less well-off family, at a party. After dating for a few months, they had protected sex. However, both of their parents agree that the two should take a break from their relationship from the new semester begins. At the same time, Sunny received a scholarship notification from a famous conservatory overseas and is requested to enroll within a few days. Meanwhile, Sunny went to the hospital because of physical discomfort and found that she was three months pregnant. As a result, she considered getting an abortion. Do you think Sunny should get an abortion in this situation?

Story 3. An Environmental Story

William is the village head of Beautiful Village, which possesses a unique landscape and rich ecological resources. The village has attracted many tourists from abroad. However, there is only one connecting road in the village, so the traffic is inconvenient. The government plans to build a new connecting road to reduce traffic jams, thus improving tourism. Some villagers worry the new infrastructure will lead to environmental damage and reduce their quality of their life. Therefore, they initiate a protest action and invite William to join them.

Story 4. A War Story

West clan and East clan are two hostile ethnic groups in the Cosmos nation. The two clans have long fought over religious belief and resource. Since West clan became the ruling party, they launched the genocide against the East clan. The East clan asked the Earth nation for help, but all their interventions in this dispute and the imposition of economic sanctions were ineffective. So the Earth Nation decides to start a war in order to halt the genocide.

Appendix 2

Appendix 2 an example of moral competencies corresponding to the rubric: moral awareness competency (sections II & III) of cheating story

II. Could you imagine Jack's problems or feelings on these matters? Please indicate your responses to the following statements
*Rating [Labeled "1"-Level D No/Limited proficiency; labeled "2"-Level C Some proficiency; Labeled "3"-Level B More proficiency; labeled "4"-Level A High proficiency]

	Four levels based on the rubric			
	D	C	B	A
12 Items of section II	1	2	3	4
1. It is not a big deal	v			
2. This is a dilemma between promoting justice and pressure from classmates				v
3. If you do not report it, it would be against your own interests	v			
4. This is an issue of complying by school rules or betraying friends			v	
5. I would feel a struggle between unfairness and friendship				v
6. I would feel inconvincible but don't want to damage the friendship		v		
7. No particular feelings, remain calm			v	
8. I would feel frustrated, so just face the problem myself		v		
9. Society has its own elimination mechanism, so let Bruce worry about himself		v		
10. Beware of the overall impact on society			v	
11. Everyone is central to the community, and he/she has to try to make an influence				v
12. I don't care because it happens all the time	v			

*Ranking:

III. Please select 4 items from Part II which most closely match your perspective on Jack's problems and feelings and rank them in order.

First______ Second_____ Third______ Fourth______

Computation steps [e.g., First item 5, Second item 2, Third item 11, Fourth item4;
Step1: ensure the levels of rubric for each ranking items, [e.g., item 5(A), item 2(A), item 11(A), item4(B)];
Step2: list the scores of four levels A, B, C, D and scored the first order "4", the second order " 3", the third order "2", and the fourth order "1" [e.g., A = 4 + 3 + 2, B = 1, C = 0, D = 0];
Step3: We set up the benchmark on B, so the score of PL1-ma-cheating story is [(A + B)/(A + B + C + D)]*100, [e.g., PL1 = [(9 + 1)/(9 + 1 + 0 + 0)]*100 = 1*100 = 100]