The Revised Critical Thinking Skills Scale for a Life-and-Death Course: Preliminary Scale Refinement

Huei-Lih Hwang1 • Huey-Shyan Lin2 • Hsiu-Hung Wang3*

ABSTRACT

Background: Death education involves acquiring knowledge, changing behavior, and developing proper views of life in both the affective and the value domains. Critical thinking that is honed through reflecting on life-and-death issues represents a way to reach these goals. Designing assessments able to measure college student content and critical thinking skills related to life-and-death issues is thus important. The Test of Critical Thinking Skills for Life-And-Death content (TCTS-LD) instrument requires the administration of additional tests to assess reliability and validity for future use in the assessment of perceptions on life and death.

Purpose: The purpose of this study was to refine the TCTS-LD.

Methods: A cross-sectional, descriptive design was used to recruit 715 college students in southern Taiwan. Three structured scales were administered in class to the participants. Data were collected in 2004 and 2006. Confirmatory factor analysis was applied to validate the structure of scales.

Results: Examination of the reliability of the three-factor and 15-item scale revealed a Kuder-Richardson coefficient of internal consistency of .54. The split-half reliability coefficients were .47 in the Spearman-Brown correlation and .40 in the intraclass correlation coefficient (ICC). The test-retest reliability coefficients (n = 22) were .58 in Pearson correlation and .56 in ICC. In addition to content validity verification by experts and face validity by students, the validity of this test was assessed using three methods, including (a) a comparable validity rating between this test and the TCTS-A (r = .34, p < .001); (b) a contrast-group technique with different responses to the instrument between those in education and nursing majors (r = .71, p < .01), with scores of 10.98 (SD = 2.42) and 9.82 (SD = 2.25), respectively; and (c) a confirmatory factor analysis confirming that TCTS-LD is related to the three dimensions of assumption, evaluation, and induction (χ2 = 81.800, p = .158, normed chi-square χ²/df = 1.169, comparative fit index [CFI] = .976, Tucker–Lewis index = .984, root mean square error of approximation [RMSEA] = 0.015). Three factors explained 31.19% of total variance for the revised TCTS-LD.

Conclusions: The revised TCTS-LD scale improved performance and effectiveness to a certain degree. However, reliability and construct validity must be further tested to permit its use as an evaluation tool.

Key Words: critical thinking skills, scale refinement, life and death.

Introduction

Life-and-death courses, developed from death education courses, have been designed mainly to help students face both life and death with respect and dignity. The four domains of death education include acquiring knowledge, changing behavior, providing affective soothing, and affirming basic human values (Corr, Nabe, & Corr, 2009). A value that influences behavior tends to create relevant patterns in the life and decision making of individuals. In Eddy and Duff’s (1986) article review also claimed that value clarification is not only a goal of death education but serves as guiding principles for future behavior. Furthermore, critical reflection on life-and-death issues can help individuals develop proper views of life, and critical thinking (CT) serves as a way to assist the attainment of value construction (Tsi, 2002). In a complex society, with the lack of universally accepted standards of value, it is necessary for death educators to devote themselves to improving the CT of their students to reach the goal of life-and-death course. Reasons for CT in life cited at the 27th International Conference on Critical Thinking (Scriven & Paul, 2007) included enhancing quality of thought and improving quality of life. CT is a transferable background knowledge skill that can help people develop meaningful decision making skills (Wen, 1997), especially on controversial issues such as euthanasia, suicide, life attitudes, meaning of hardship in life, funeral arrangements, and religious beliefs about death. Students who do not think critically may accept ideas without examining carefully what they mean. One of the researchers has found that many
students believe that life-and-death issues yield no “right” answers and that such are matters of personal belief, which need not be discussed or considered further. Those students expressed that individuals are qualified to make choices on the basis of their own preferences, which is not complete proper for dealing with life-and-death issues. Moreover, no life-and-death instruments have ever been given to students, who are left to deal with related issues on their own.

The need for CT on life-and-death issues is necessitated in light of the intertwined emotions, values, and cultural experiences involved. The goal of accepting the finality of life and facing life respectfully is not easy for students. This is particularly true for nursing students who must confront terminally ill patients and the struggle with their own personal difficulties with overwhelming emotions (Aradilla-Herrero & Tomás-Sábado, 2006). Nurses may have in their own mind many questions and doubts concerning life and death. Moreover, thinking skills and knowledge content are best learned together (National Mathematics Advisory Panel, 2008). In agreement with Silva’s (2009) statement, what students can do with knowledge is the essence of 21st century skills such as thinking, reasoning, teamwork skills, and proficiency in using technology. Although CT skills with domain-specific knowledge bases have been proposed to enhance CT abilities, little work has examined whether these CT skills works for life-and-death courses.

Although many standardized CT instruments have been developed and used, some researchers have suggested that they are not sufficiently domain specific (Stone, Davidson, Evans & Hansen, 2001), show little sensitivity to student progress (Pike, 2001), and are largely unrelated to important aspects of nursing education (Tanner, 2005). Researchers have further claimed that CT is not transferable across domains, but domain-specific knowledge can enhance CT ability (Botti & Reeve, 2003; Duncan, 2007; McPeck, 1990). This concern has thus prompted assessment experts to suggest that locally developed CT measures may better serve program goals than commercially developed instruments (Simpson & Courtney, 2002). Although such arguments offer valuable insights into CT scale content, developing a content-specific scale could be needed to fully address the issues that are raised. Instruments such as the Cornell Critical Thinking Test, the Watson–Glaser Critical Thinking Appraisal (WGCTA), and the Test of Critical Thinking Skills for Adults (TCTS-A) are based on CT skills only (not for life-and-death–related knowledge). Therefore, this present study has been designed to measure how students put life-and-death knowledge into practice. Objectives and content must also be carefully considered in such an evaluation.

CT, described as a cognitive skill, consists of recognition of assumptions, deduction, induction, interpretations, and argument evaluation (Watson & Glaser, 1994). Specific instruments (e.g., Cornell Critical Thinking Test, WGCTA, and TCTS-A) have been developed and widely used for measurement of CT in each of the five dimensions. However, identification of relevance plays a pivotal role in clarifying relationship between the premise and the statements that engage in CT skill (Kramer, 1993) or conceptual acquisition of acquiring knowledge (Anderson, et al., 2001). Researchers in this area have based their approaches on the five dimensions and added considerations of relevance to design the 22-item TCTS-LD (Hwang, Yu-Chi, Lin & Teng, 2007). The original instrument, however, did not use a robust test such as a confirmatory factor analysis (CFA) to assess construct validity. The reliability (Kuder–Richardson [KR-20] coefficient of .51) suggests that further testing is needed. Therefore, the purposes of this study were to revalidate the TCTS-LD by expanding and revising its contents and to examine how the revised TCTS-LD, incorporating domain-specific knowledge bases, served as an assessment tool for a life-and-death course.

**Methods**

The TCTS-LD refining process used in this study was divided into two phases and took 2 years to complete. First, we developed test items and made revisions on the basis of the results of an expert panel discussion. In the second phase, we evaluated the reliability and validity of the revised instrument. A cross-sectional, descriptive design was used to recruit 715 college students in southern Taiwan. Data were collected between May 2004 and May 2006. Although most people recommended a minimum of 300 participants for a factor analysis to achieve factor structure stability regardless of the participant-to-item ratio (Tabachnick & Fidell, 2001), in this study we adopted the largest sample size (greater than a 20:1 participant-to-item ratio) to produce more accurate factor solution (Costello & Osborne, 2005).

**Phase 1: Refining the Scale Items**

This phase consisted of two steps: developing additional items and assessing internal consistency. In Step 1, individual discussions with three academic experts in philosophy and two in critical education were undertaken. The process of developing multiple-choice test items for CT adopted the model described by Morrison, Smith, and Britt (1996). Ten items distributed evenly across six dimensions were designed for respondents with a reading level of ninth grade or higher (the same as TCTS-LD), consistent with the WGCTA Form S (Watson & Glaser, 1994). To shorten the form and to make the number of answers consistent, we selected to have three multiple-choice responses for each test item.

Step 2 examined content and face validity. Content validity was confirmed on the basis of the link between the revised version of the TCTS-LD and the opinions of the expert panel. After the initial item pool was completed, the researcher invited four experts from different fields of CT teaching (including two philosophy teachers, one CT teacher, and one Chinese teacher) to examine item relevance, CT skills, and clarity of wording as well as the
appropriate of the scale questions and answers. The experts evaluated the content validity of these items using a 4-point Likert scale that ranged from 1 (not relevant, clear, or appropriate) to 4 (relevant, clear, or appropriate). Three of 10 items were given ratings of 2 by two experts. Through individual discussion with these two experts, changes were made to enhance item consistency in line with the CT principles. In addition, face validity was established by asking four students to complete the test and comment upon its ease of use.

**Phase 2: Field Testing the Scale**

This phase aimed to evaluate the reliability and validity of the scale developed during the previous phase. Two field-test steps were employed. First, a pilot test was used to refine test questions and test procedure. A total of 170 nursing students in a junior college nursing program were recruited as participants. After establishing content and face validity, the TCTS-LD was administered to these students to assess instrument reliability. The students were also asked to provide feedback that detailed their personal opinions regarding taking the revised TCTS-LD. On the basis of student feedback, we rewrote two statements to improve clarity of meaning.

Second, a convenience sample of 725 students was recruited for psychometric testing. To conduct known-group validity testing, participants were recruited from 10 classes (n = 625) of nursing college students, which included three life-and-death study classes (n = 79) and three classes (n = 100) of general university students. A total of 715 students completed the revised TCTS-LD and Critical Thinking Disposition Scale (CTDS); 449 students (nursing = 58%, LD class = 20%, education = 22%) took the revised TCTS-LD, CTDS, and TCTS-A. For participants who completed all three scales, education majors (51.1% aged 20 years, ranging from 20 to 23 years) were on average 1 year older than nursing majors (48.3% aged 19 years, ranging from 19 to 46 years). Twenty-two students were recruited on the basis of their willingness to take a retest 2 weeks later. Although the sample represented varied grade levels and two university majors, most were from nursing colleges (86.3%). The age of the sample averaged 21.6 years, with a standard deviation of 2.76. Data collection followed a prescribed protocol for test administration.

**Instruments**

In addition to the revised TCTS-LD, three more scales were used. A demographic questionnaire collected data on gender, age, major, school performance during the previous semester, and highest education level of the student’s parent(s). Demographic data were used to examine subject construct validity. The CTDS, on the basis of two tests (Ennis, 1985; Watson & Glaser, 1994), was used to assess concurrent validity and to measure personal intent to apply CT. The 21-item scale consisted of four dimensions, including reflective thinking, curiosity, open-mindedness, and analytic ability. Questions were rated on a 5-point Likert scale ranging from 1 (never) to 5 (always). Cronbach’s alpha coefficient of the CTDS was .88 (n = 100; Yeh, Chen, Hsieh, & Yeh, 2001). In this present study, the Cronbach’s alpha coefficient was .87. The TCTS-A with 25 items was used to examine criterion-related validity (Yeh et al., 2001). The scale with four response options has previously been used to measure general CT ability of college students. Participants were asked to answer each statement with one number that best expressed their postviewing judgment. The KR-20 coefficient for the TCTS-A was .67 (n = 326), and concurrent validity was r = .2 (p < .05) with Chinese literature. In our study, the KR-20 coefficient was .89 (n = 448).

**Procedures**

Data were obtained in classroom settings. Teachers distributed tests and read instructions before the students began to respond. Students in southern Taiwan colleges and universities who consented to participate received the revised TCTS-LD first, followed by standardized oral instructions and sheets. Students then finished the CTDS and filled in demographic information on provided answer sheets. The TCTS-A was administered in accordance with participants’ willingness and available time. Data sheets were returned to teachers during class time. Respondents took from 30 to 60 minutes to complete the questionnaires. Subjects for the test–retest voluntarily took the TCTS-LD again in the researcher’s office 2 weeks after completing the first round of data collection. They are free to write down comments or to talk to researchers about their thoughts on answering the TCTS-LD.

**Ethical Considerations**

After review and approval by our school’s Research Development Committee (No. 95-08), a convenience sample was adopted. Before beginning the study, researchers informed potential participants about the study and obtained their written consent. Confidentiality was ensured by assigning a code number to each participant lieu of their name. Students were assured test score confidentiality and were assured that their participation would not impact upon their school record. Participants were further informed that they could leave the study at any time.

**Data Analysis**

Item analysis was performed to provide information about item difficulty and a discrimination index. CFA with categorical factor indicators was used to test the extent to which data supported the factor structure for test items. The Statistical Package for the Social Sciences for Windows (Version 12.0; SPSS Inc., Chicago, IL) and the Mplus
Version 5.2 software (Muthén & Muthén, Los Angeles, CA) was used for the analysis.

**Results**

**Item Description**

We conducted a test to evaluate course content and to assess CT. According to the life-and-death course content and CT definition, 10 items were added and four items were reworded, resulting in a final 32-item version of the TCTS-LD in six dimensions (Appendix). All test items were based on situations encountered in daily life and actual course settings. The items consist of two types of content: neutral (i.e., custom and science issues) and controversial (e.g., religious and social issues). Each item was allowed three response options, one of which was identified as “correct.” One point was assigned for each correct response given. For all well-structured items, correct answers were identified by experts with expertise in the content area. The total potential score range was from 0 to 32, with higher scores reflecting stronger CT skills with regard to life-and-death content.

**TABLE 1. Item Analysis for the TCTS-LDS**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Item</th>
<th>Difficulty Index</th>
<th>Discriminate Index</th>
<th>Item-Factor Correlation $r$</th>
<th>KR-20 If Item Deleted ($n = 715$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n = 170$</td>
<td>$n = 715$</td>
<td>$n = 170$</td>
<td>$n = 715$</td>
</tr>
<tr>
<td>Assumption</td>
<td>1</td>
<td>.77</td>
<td>.76</td>
<td>.64</td>
<td>.59</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.62</td>
<td>.52</td>
<td>.74</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.81</td>
<td>.79</td>
<td>.58</td>
<td>.60</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.57</td>
<td>.53</td>
<td>.94</td>
<td>.85</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.61</td>
<td>.65</td>
<td>.46</td>
<td>.66</td>
</tr>
<tr>
<td>Evaluation</td>
<td>6</td>
<td>.78</td>
<td>.79</td>
<td>.72</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>.63</td>
<td>.66</td>
<td>.77</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>.67</td>
<td>.67</td>
<td>.77</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>.15</td>
<td>–</td>
<td>.13</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>.88</td>
<td>.92</td>
<td>.32</td>
<td>.29</td>
</tr>
<tr>
<td>Relevance</td>
<td>11</td>
<td>.95</td>
<td>–</td>
<td>.09</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>.42</td>
<td>.45</td>
<td>.70</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>.73</td>
<td>.62</td>
<td>.42</td>
<td>.56</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>.71</td>
<td>.62</td>
<td>.54</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>.92</td>
<td>–</td>
<td>.10</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>.37</td>
<td>.36</td>
<td>.61</td>
<td>.59</td>
</tr>
<tr>
<td>Induction</td>
<td>17</td>
<td>.72</td>
<td>.70</td>
<td>.70</td>
<td>.70</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>.32</td>
<td>.30</td>
<td>.92</td>
<td>.91</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>.87</td>
<td>.85</td>
<td>.32</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>.73</td>
<td>.70</td>
<td>.65</td>
<td>.63</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>.84</td>
<td>.77</td>
<td>.45</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>.74</td>
<td>.72</td>
<td>.52</td>
<td>.51</td>
</tr>
<tr>
<td>Deduction</td>
<td>23</td>
<td>.28</td>
<td>.28</td>
<td>.55</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>.88</td>
<td>.84</td>
<td>.36</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>.45</td>
<td>.38</td>
<td>.64</td>
<td>.81</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>.55</td>
<td>.51</td>
<td>.96</td>
<td>.89</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>.28</td>
<td>.25</td>
<td>.59</td>
<td>.75</td>
</tr>
<tr>
<td>Explanation</td>
<td>28</td>
<td>.40</td>
<td>.47</td>
<td>.32</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>.55</td>
<td>.63</td>
<td>.55</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>.28</td>
<td>.26</td>
<td>.60</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>.30</td>
<td>.39</td>
<td>.47</td>
<td>.70</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>.48</td>
<td>.49</td>
<td>.48</td>
<td>.72</td>
</tr>
</tbody>
</table>

* $p < .05$. ** $p < .01$. 
Item Analysis and Reliability Estimates

Using item analyses, three test items of the revised TCTS-LD that met the two criteria of item discrimination index less than .3 and item-factor correlation less than .3 were deleted after the pretest (n = 170). In the actual test (n = 715), the KR-20 coefficients of Assumption, Evaluation, Relevance, Induction, Deduction, and Explanation subscales were .44, .36, -.001, .40, -.006, and .13, respectively. In addition, three subscales in which the KR-20 coefficients were less than .2 and were still not higher than .3 if some items in a subscale were deleted. This item analysis resulted in three factors with a total of 15 items. Item response distributions reflecting proportion of items correct (difficulty index) were .30 to .92 (Table 1), with an average difficulty index higher than .58, indicating that the test was not difficult for participants to complete.

Validity Investigation

To confirm the hypothesized three-factor structure of the revised TCTS-LD as a valid measure of CT skill, we applied CFA with categorical factor indicators. As expected with this large sample and model, the chi-square–associated p value was greater than the .05 significance level ($\chi^2 = 81.800$, $p = .158$). Other indices also reached acceptable levels (normed chi-square $\chi^2/df = 1.169$, CFI = .976, Tucker–Lewis index = .984, RMSEA = .015). CFA results supported the revised TCTS-LD, composed of 15 items (Figure 1) that reflected the three dimensions of assumption, evaluation, and induction. These three dimensions explained 31.19% of total TCTS-LD variance.

In addition to verification of content validity by experts and face validity by students, four types of validity evaluations, introduced below, were performed to determine the validity of the revised TCTS-LD. Criterion-related concurrent validity was substantiated when the revised TCTS-LD was correlated with TCTS-A ($r = .27$, $p < .001$) and CTDS ($r = .07$, $p < .05$). As for contrasted validity, a contrast-group technique found significantly different responses to the instrument between education and nursing majors ($t = 2.71$, $p < .01$), with scores of 10.98 (SD = 2.42) and 9.82 (SD = 2.25), respectively. The correlation $r$ for variables less than .3 represents that they are different concepts. In this test, correlations between factors ranging between .08 and .28 indicate that factors are independent. In other words, the revised test exhibited divergent validity. On the other hand, correlations between individual factors and the overall TCTS-LD ranging from .51 to .74 indicate that the revised test exhibited convergent validity. A correlation matrix between all factors and the overall TCTS-LD is shown in Table 2.

Reliability

Three forms of reliability testing, measuring internal consistency, split-half, and test–retest reliabilities, were adopted.

The three subscales with 15 items were tested with the KR-20 for internal consistency. The KR-20 coefficients ranged from .36 to .44 and for the total scale .54, which reached an acceptable level (Guieford, 1965). The split-half reliability for the total scale was examined using the Spearman–Brown coefficient (.47) and the ICC (.40), which indicated medium (Cohen, 1988) and average (Cicchetti, 1994) levels, respectively. Test–retest reliability was established by retesting 22 students at 2 weeks after initial testing. Results
Table 2: Correlations Between the Three Subscales of Revised TCTS-LDS Items (N = 715)

<table>
<thead>
<tr>
<th>Correlation r</th>
<th>CT1</th>
<th>CT2</th>
<th>CT3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT2</td>
<td>.084*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT3</td>
<td>.282***</td>
<td>.121**</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.724***</td>
<td>.513***</td>
<td>.744***</td>
</tr>
</tbody>
</table>

Note. CT1 = assumption detection; CT2 = argument evaluation; CT3 = inductions.
*p < .05. **p < .01. ***p < .001.

showed a significant Pearson correlation coefficient of .58 (p < .001) and an ICC of .56 (p < .001), which indicated high (Cohen, 1988) and average (Cicchetti, 1994) levels, respectively. In sum, reliability of the revised 15-item TCTS-LD was deemed acceptable.

Discussion

The 22-item TCTS-LD was initially developed to assess domain-specific knowledge, with CT skills associated with six dimensions. Using a reliability test to select three dimensions with acceptable KR-20 coefficients and a CFA to confirm the three dimensions in this study, we shortened the TCTS-LD to 15 items. CFI, NNFI, and RMSEA all reached acceptable levels for model-fit criteria and offered the best fit to data. The present work provided additional evidence for TCTS-LD construct validity. However, three dimensions as relevance, interpretation, and deduction were deleted. There might be that the relevance identification between the premise and the statement in the relevance dimension is also related to the skill of evaluating arguments in evaluation. Item description lengths and location in the deduction and interpretation dimensions might tend to overload participants. The sequence of test items (both two dimensions arranging at the end of the test) may cause fatigue and result in poor concentration or motivation to read test items carefully.

In addition, a comparison of the original and revised versions of TCTS-LD found a similar guess rate on both tests (a mean of 0.27 for the original version and 0.26 for the refined version). Students were not familiar with tests requiring CT, and interpretation of statements with distracters was conceptually difficult (e.g., familiar terms and statements used in the same context, despite examples being provided in each of the six sections). We found that participants failed to consider terms that were familiar to them, such as the three items involving Buddhism’s reincarnation, folk religious customs, and popular terms. They might answer these questions without carefully reading and thinking about their responses. Some participants told the researchers that they deduced the meanings of some test items using socially defined and enforced practices. The interpretation section was more than likely to occur within specific cultural and political contexts and was linked to the world in which subjects lived. Therefore, participants answered questions with biased or prejudiced responses, which corroborated claims made by experts at the 27th Annual Conference on Critical Thinking, who suggested that much of our thinking is biased or downright prejudiced (Scriven & Paul, 2007). Some participants told the researcher that answers were too similar to make a choice; in some cases, correct answers were not included among choices, or personal views tended to interfere with judgment when giving answers.

Yarbrough (1992) indicated reliability and validity as not solely inherent in the test but were related to characteristics such as items, students, and testing situation. On the students’ side, daily fluctuations in student mood and diverse environmental classroom conditions, such as various semesters, course sections, or venues for test taking, may have influenced participant concentration, motivation, and response accuracy. As TCTS-LD scores ranged from 4 to 20, with a mean of 13.6, a score negatively skewed (skew = −0.486), we may deduce that some participants did not take the test seriously, resulting in 12 extremes (scores lower than 6). Pike and Philipp (1989) advised that student characteristics such as motivation be taken into careful consideration when evaluating test scores, particularly when there is no penalty for poor test performance.

Although recognizing the lack of established standards in this area, we were able to show a substantial correlation with TCTS-A, a developed test of general CT skills for adults in Taiwan. Regarding the evidence of the above concurrent validity, this test supported a positive relationship with CT disposition, corroborating findings by many researchers (McCarthy, Schuster, Zehr, & McDougal, 1999; Profetto-McGrath, 2003), although other studies could still fail to identify a positive relationship because of certain study limitations (Albert, Albert, & Radsma, 2002). The KR-20 coefficient on the revised 15-item TCTS-LD is slightly higher than the original 22-item TCTS-LD. However, the coefficient of .54 is still lower than .70. We consider this as a preliminary study and would suggest further research in this area as follows to examine reliability and the TCTS-LD factor structure.

First, researchers may integrate the vignettes and the statements in relevance with the evaluation dimension because the meaning of hardship and suicide issues in the relevance dimension represents death education goals. Second, researchers may reduce the number of instruments and item description lengths in the induction and interpretation dimensions as well as insert them in a location other than at the end of the test. Third, participants need to be reminded more effectively not to choose answers by intuition but rather in accordance with the statements included in each section. Fourth, measures can be taken to reduce the degree of successful guessing, such as giving enough time for completing tests, refining items with higher difficulty indexes, and designing scale items with more numbers of response alternatives, such as including five possible answers ranging from definitely false, most likely...
false, insufficient information, most likely true, to definitely true. Finally, CT skills will not be acquired without incorporating broader value issues. To educate students to think critically and to prevent them from being biased and deceived by statements, promoting recognition of fallacies related to life-and-death issues may be a productive activity.

Limitations
The limitations of this study were related to the design of the test items. First, the two test item formats could potentially confuse students. One focused on argument judgment, for example, the deduction and the evaluation dimensions, whereas the other focused on underlying meaning, for example, assumption recognition, relevance, induction, and interpretation. Second, the convenience sampling design mainly recruited participants from one college. Generalization might be limited in this population.

Conclusions
Defining CT components in relation to life-and-death issues is an ongoing challenge in the discipline of education as well as infusing CT skill to life-and-death education. This refined 15-item scale achieved progress in terms of construct validity. However, because of certain limitations, present findings should be interpreted with caution, and further testing is needed to confirm these results. This assessment will not solve all the problems of life-and-death issues in clinical practice and daily life of nursing students. For those dedicated to improving day-to-day learning and long-term student learning outcomes, designing assessments that measure both content and CT skills on life-and-death issues for college students is paramount. Further work on CT tools will bring life-and-death education one step closer to document the effects of thinking skills on students’ lives.

References


APPENDIX.
The 32-Item Version of TCTS-LD

I. Recognition of assumptions: which assumption is indicated about the following 5 statements?

1. I do not know how long I can live, or the goals I can reach. Therefore, I must value my time.
   A: Life is changeable.  B: Life is hard.  C: Life is like a show (unreal).

2. If life was not fun anymore, I would rather die. The meaning of life
   A: is looking for fun.  B: discovering the commitment of existence  C: looking for death

3. I agree with euthanasia because both patients and family can get out of the abyss of misery.
   A: People must die someday.  B: Death is the beginning of another life.  C: Death allows extrication from a predicament.

4. We have to bear an insult to survive.
   A: This insult must be tolerated.  B: There is no other choice but to bear the insult.  C: This humiliation is necessary for life.

5. According to a Hakka custom, a coffin is left in the house until the funeral ceremony. However, this family moves the coffin
   out of the new house out of fear of inviting a bad luck for the bereaved.
   A: Respecting the deceived is the best.  B: Death is a taboo.  C: The custom cannot be overturned.

II. Evaluation: Which argument (1, 2) or both are strong on the questions 6 through 10?

Example: Should we arrange a unified test schedule for the students?

Argument (1): Should. According to statistic findings last year, later testers have higher scores than earlier. Students also
reported that later test takers seek test items from those who took the test earlier.

Argument (2): Should not. By taking all the tests in 1 week, it will decrease the learning effects by overly pressuring students.

Choice A: Argument (1) is strong.  Choice B: Argument (2) is strong.  Choice C: Both arguments (1) and (2) are strong.

Analysis: The best answer is A. Argument (1) is strong because the inference is based on objective data findings and unified
 test schedule is fair for all test takers. Argument (2) is weak because there are multiple potential reasons for
loosing learning effects.

6. Do people should have the right to kill themselves?

Argument (1): No, killing oneself is against nature. Life comes from nature and no one has the right to take it away.

Argument (2): Yes, suicide is an option for people overwhelmed with sorrow. People make decisions on life or death
 depending on their mood.

7. Can people go to heaven after death?

Argument (1): Yes, The spirit can go to heaven after death if you have lived in accordance with the teachings of the Christian
 Bible.

Argument (2): No, people die ever day. If those spirits stay in heaven, then it will be too crowded to stay.

8. Should people fear death?

Argument (1): Yes. Only people who fear death will strive for fame and fortune.

Argument (2): No, if people fear death and refuse to talk about death, people cannot master life nor plan a “living will.”

9. Which one is strong on the question about human can live forever?

Argument (1): Maybe, in an objective view. People can prolong life by scientific advances as people have instincts upon
 which they can rely.

Argument (2): Not possible. Death is a natural rule for living creatures. Humans are a living creature, and so cannot
 overcame death.

10. Which one is strong on the question about the need for people to have a goal for life?

Argument (1): A goal for life is needed, as human beings are the wisest of all creatures. Therefore, they need goals to live
 with direction and value.

Argument (2): There is no need to think of goals. Human beings live by instincts, such as eating when hungry and sleeping when tired.

(continues)
APPENDIX.

The 32-Item Version of TCTS-LD, continued

III. Relevance: What the relation (support, against, irrelevant) between the premise and statements from 11 to 16?

Premise of vignette: “People cannot live without difficulty. One can accept instead of reject hardship if one finds meaning within the hardship itself.”

11. Awareness and mind change can transform hardship.
12. Request a Taoist priest to change the fate of those trapped in a “sea of woes.”
13. People who commit suicide will go to hell and never be reborn.
14. It is a wrong choice to kill oneself to avoid suffering.
15. Experiencing hardship can facilitate people to realize meaning of life.
16. Life without difficulty would be an “imperfect” life.

IV. Induction: Which conclusion best fits the following statements?

17. According to statistics, suicide is the third leading cause of death those aged 15 to 24 years in Taiwan.
   A: suicide rate is high for teenagers in Taiwan.
   B: Teenagers like to kill themselves to show sovereignty over their body.
   C: Taiwan’s teenagers have lower tolerance to setbacks.

18. One research indicated not only that dying patients enjoy the peace obtained by forgiving others, but also staff that are able to forgive their victimizer can help patients relieve their worries.
   A: Those who cannot forgive themselves cannot forgive others.
   B: Those staff that hold an old grudge against someone will not bring peace to dying patients.
   C: Dying patient who cannot forgive God or others need medical staff to straighten them out.

19. Students are aware of their own mortality and think more about such issues by writing living wills and planning their funeral.
   A: Writing a living will and funeral planning are common teaching activities.
   B: Awareness of the inevitability of death and thinking about life’s meaning is a focus for education.
   C: Writing living wills and funeral planning encourage students to confront death and further explore the meaning of life.

20. Ancient Egyptians made mummies to preserve the body because they believed spirits return to the world. In the modern world, some people also seek to extend life or earn immortality.
   A: People pursue immortality at all times and in all countries.
   B: Denying death is a goal for modern people.
   C: People are afraid of being forgotten and try to leave as much evidence of their existence as they can.

21. Many characters in legend are not proved by science, such as devils in animal forms, angels, Yama, or Satan. However, people are afraid of encountering disaster for not believing these myths in Taiwan.
   A: religious belief does not need scientific proof;  B: people believe that wishes will come true if sincerely believed;
   C: people would rather believe it as true than nothing at all.

22. People cannot reject that God does not exist. Science cannot prove that God does exist. People still believe that God does exist and live by the creed.
   A: God’s existence is a mystery;  B: People love God;  C: People who have faith in God do not need scientific proof.

V. Deduction: choose the most logically followed statement (1, 2) or both are for premise of 23 to 27.

Example: Human beings are animals. Corpses of animals decay.
Statement (1): the corpse of the priest will decay and stink.
Statement (2): Those decayed creatures are animals.
Choice A: Statement (1) is strong.  Choice B: Statement (2) is strong.  Choice C: Both statements (1) and (2) are strong.
Analysis: The best answer is A. Statement (1) is strong because priest is human which is also an animal. His body will decay after death. Statement (2) is weak because not only animals but also plants decay.

23. I believe Buddhism’s six ways of samsara. Only those who practice Buddhism can reach Nay-Pan where spirits can escape from the samsara of life and death.
   Statement (1): Doing more good things can keep evil creatures at bay and keep one away from living hell.
   Statement (2): Practicing charitable and pious deeds can let you go to Western Paradise after death to escape the life-and-death cycle.

24. I believe that people have nothing after death. Therefore,
   Statement (1): I care much more about happy living now then at other times.
   Statement (2): I want to know whether my deceived relatives live alright or not.
**APPENDIX.**

*The 32-Item Version of TCTS-LD, continued*

<table>
<thead>
<tr>
<th>25. Most of people who die of traffic accidents are motorcyclists. Not wearing a helmet is a cause of death for some drivers.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement (1): some drivers who do not wear helmets die in traffic accidents.</td>
</tr>
<tr>
<td>Statement (2): people who wear helmets die less than those who do not.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>26. If one applies reincarnation to explain the trials of life, then he/she has clearer thoughts than others.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement (1): Those who do not apply reincarnation to explain their suffering have thoughts that are less clear than others.</td>
</tr>
<tr>
<td>Statement (2): Those unclear thinkers do not apply reincarnation to explain the trials of life.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>27. Death acceptance is a necessary condition for a good death.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement (1): If people forbid talk of death, they cannot die in peace.</td>
</tr>
<tr>
<td>Statement (2): Accepting death by love and wisdom allows people to have a good death.</td>
</tr>
</tbody>
</table>

**VI. Interpretation:** Choose the statement that is more reasonable for the following vignettes.

<table>
<thead>
<tr>
<th>28. Those who believe life without a next life are selfish and careless of others.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: People without belief in a next life tend toward evil conduct.</td>
</tr>
<tr>
<td>B: There is an argument the between next life and no more life after life.</td>
</tr>
<tr>
<td>C: Views toward next life are associated with personal conduct.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>29. If hospice care provided holistic care for the dying, then dying patients would not seek euthanasia. Assuming most of the dying choose euthanasia in one country. Which one can be referred to the above meaning?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Hospice care does not satisfy the needs of the dying in this country.</td>
</tr>
<tr>
<td>B: Euthanasia is legal in this country.</td>
</tr>
<tr>
<td>C: Most of the dying patients support euthanasia in this country.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>30. In Taiwan, people offer sacrifices to ancestors on festive occasions to ensure the family’s well-being. Which of the following refers to Amy, who offers pork, fruit and drink to her ancestors today?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Amy might have encountered some trouble and need to ask a favor of her ancestors.</td>
</tr>
<tr>
<td>B: Compared with other nations, people in Taiwan remember their ancestors with more gratitude.</td>
</tr>
<tr>
<td>C: People often eat food that is offered to ancestors in Taiwan.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>31. Vignette “Arranging a writing a living will activity in class can decrease fear of death and death taboos as well as promote meaning of life insights among students.”</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Students who do not write a living will not change their views on life and death;</td>
</tr>
<tr>
<td>B: Students who finish their living will talk more about death issues with family;</td>
</tr>
<tr>
<td>C: There is a need to have a living will writing activity to change student views of life and death.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>32. Vignette “To facilitate living with dignity, health insurance reimburses hospice care to relieve pain and soothe the mind of terminal patients”</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: People live without dignity because the government does not take good care of terminal patients;</td>
</tr>
<tr>
<td>B: People choose hospice care because it can bring them peace in spirit and relief from pain;</td>
</tr>
<tr>
<td>C: By reimbursing for hospice care, the government encourages people to choose hospice care to live with dignity.</td>
</tr>
</tbody>
</table>
生死批判思考量表—量表初步修訂

黄慧莉  林惠賢  王秀紅

背景  
生死教育乃透過協助學習者面對死亡，探索生命意義，進而建立健康人生觀。這種不斷地反思死亡、生命是需要學習者以批判思考能力來協助達成。因此，在評估生死教育課程時，需要有批判思考技巧融入設計的生死議題來追蹤學生的學習。過去發展22題的「生死批判思考量表」（TCTS-LD）未呈現效度資料，且信度係數為.51，有必要修訂量表並測試之。

目的  
修訂TCTS-LD以評量生死學有關課程的批判思考技能。

方法  
採用橫斷性描述研究設計，以兩年間募集南臺灣715位大專學生同意為研究對象。參與學生於課堂中需填完三種量表（批判思考意向 [CTDS]，成人版批判思考測驗 [TCTS-A]，修正版TCTS-LD）。

結果  
TCTS-LD 15個題目之內在一致性信度KR-20係數為.54，折半信度之斯皮爾曼-布朗係數為.47，組內相關係數為.40，再測信度（n = 22）之皮爾森相關係數為.58（p = .004）、組內相關係數為.56（p = .003）。題目採用以下三種效度檢核方法：(1)與成人版的批判思考測驗呈顯著相關（TCTS-A; r = .27, p < .001），(2)國立大學教育系與私立五專四年級護理科兩群學生的平均分數呈顯著差異（t = 2.71, p < .01），(3)驗證性因素分析得出：假設、評鑑、歸納等3個因素（χ² = 81.800, p = .158; Normed chi-square χ²/df = 1.169; CFI = 0.976; TLI = 0.984; RMSEA = 0.015），總共可解釋31.19%的生死批判思考變異量。

結論  
此生死批判量表修訂雖有某種程度進展，為提升生死教育效能，增加護理學生處理生死議題的困境，未來仍須以此15題量表持續進行信度、效度上的測試，以利大專生生死教育相關課程之評量，以及教學策略之研發。

關鍵詞：批判思考能力、量表修訂、生死學。

1國立台南護理專科學校護理科助理教授暨高雄醫學大學護理系博士班學生  2輔英科技大學護理學院護理管理系副教授  3高雄醫學大學護理學院護理學系教授

受文日期：99年2月3日  修改日期：99年6月10日  接受刊載：99年8月19日

通訊作者地址：王秀紅       80708高雄市三民區十全一路100號