Validation of Yoon’s Critical Thinking Disposition Instrument

Hyunsook Shin, PhD, APN, CPNP, 1, * Chang Gi Park, PhD, 2 Hyojin Kim, RN, MSN 1

1 College of Nursing Science, East-West Nursing Research Institute, Kyung Hee University, Seoul, South Korea
2 Department of Biobehavioral Health Science, College of Nursing, University of Illinois at Chicago, USA

Introduction

Critical thinking (CT) has been identified as a vital outcome for nursing education [1,2]. However, the lack of a valid instrument to measure nursing students’ CT abilities has resulted in limited assessment of students’ achievement in Korean academic programs, leading to ineffective academic mentoring [3].

Peter Facione [4] considered CT to have two dimensions, including “a frame of mind or a quest for thinking (disposition) and a set of operational cognitive skills.” Facione defined CT as “purposive, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as an explanation of the evidential and conceptual, methodological, criteriological, or contextual considerations upon which judgment is based” [4]. The American Philosophical Association’s consensus definition of CT led by Facione [4] includes the core cognitive skills and the disposition of CT. Therefore, current instruments measuring CT abilities generally assess either core cognitive skills or CT disposition. For example, the California Critical Thinking Skills Test (CCTST) [5] and Health Sciences Reasoning Test [6] measure core cognitive skills. On the other hand, the California Critical Thinking Disposition Inventory (CCTDI) [7] and Yoon’s Critical Thinking Disposition (YCTD) instrument [8] measure CT disposition. Previous studies on CT have found that most measurement instruments were not effective for use with nursing students because of either lack of instrument soundness or the problematic process of validation [9,10]. Considering that the CCTST was used as a nonspecific test for changes in CT, the CCTDI is recommended as a more reliable tool for measuring CT in nursing students in the assessment and planning of specific
Curriculum development activities and in counseling individual students [9,11]. Tools developed from other countries are limited in sensitively capturing the cognitive tendency associated with Korean students’ perception of every item [12]. With the increasing attention to CT in nursing education, the lack of reliable and valid evaluation tools targeting Korean nursing students’ CT abilities has been identified as one of the barriers to instructing and evaluating students in undergraduate programs [13].

Yoon [8] originally developed the YCTD based on the CCTDI for Korean nursing students. The subcategories of the YCTD are similar to those of the CCTDI, which was developed based on the American Philosophical Association definition of CT disposition. The seven subscales of the YCTD include objectivity, prudence, systematicity, intellectual eagerness/curiosity, intellectual fairness, healthy skepticism, and CT self-confidence [8]. Objectivity in CT is a tendency of eliminating personal biases, and prudence is the habit of seeing the complexity of issues. In addition, systematicity is the tendency of striving to approach problems in a systematic way and intellectual eagerness/curiosity is the tendency to want to know things. Intellectual fairness is the tendency of thinking with the viewpoints of others, while healthy skepticism is the habit of always seeking the best possible understanding of any given situation. Lastly, CT self-confidence is the tendency to trust reflective thinking to solve problems and to make decisions. The YCTD was developed in the form of self-assessment for Korean nursing students, and it has been identified as one reliable instrument that can be used to assess Korean nursing students’ CT abilities [14]. Based on an intensive review of studies of all the currently available CT instruments, Nair and Stamler [9] have reported an urgent need for examination of the instruments’ construct validity. In addition, Gregorich [15] addressed that self-report instruments should be evaluated to have the measurement invariance for meaningful comparisons across groups. Measurement instruments must be designed to yield replicable findings both cross-sectionally and longitudinally. A valid comparison of self-report instruments such as the YCTD requires that the constructs have a similar meaning across groups and time. Structural equivalence, which calls for an identical meaning of each item across groups or measurements, is difficult to evaluate [16]. Furthermore, Barbosa-Leiker et al. [17] suggested that verification of longitudinal invariance should be precedent before assessing whether observed change in certain measurement values with an intervention reflects true change or changes in evaluation or the construct structure over time.

Considering that the structural validity and group invariance of the YCTD have not yet been established, this study aimed (a) to validate a proposed seven-factor model of the YCTD, a CT disposition measurement instrument currently used in Korean nursing research and education, and (b) to examine the multigroup measurement invariance of the YCTD across different groups and time periods using cross-sectional and longitudinal data in order to compare the structure of responses to CT instrument across these different student groups.

Methods

Study design

The study was a validation study of the YCTD using cross-sectional and longitudinal data from a multisite, pretest, posttest study on the effect of nursing education.

Participants

According to a power analysis for confirmatory factor analysis (CFA), a sample size of 345 participants was adequate for the study [18]. A total of 350 baccalaureate nursing students were recruited at three universities in Seoul, Korea. Students enrolled in a pediatric nursing practicum between February and December in 2012 and 2013 were included as study participants. Specifically, data from three universities with 248 participants (100 at school A, 75 at school B, and 73 at school C) were used for cross-sectional analysis, multi-group measurement invariance test between groups, and data from one university with 168 senior participants (95 in 2012 and 73 in 2013) were used for longitudinal analysis by using pre-practicum and postpracticum, premeasurement and post-measurement invariance test within a group. Most students (95.0%) were female.

Data collection

At each university, student participants were introduced to the study and were asked to use the YCTD to evaluate their CT before and after a clinical pediatric nursing practicum which included an integrated simulation curriculum. The simulation curriculum in this practicum was developed to enhance participants’ CT [19,20]. All 350 nursing students undergoing the practicum were asked to participate in the study, among them 345 students completed both the YCTD pretest and posttest.

Instruments

The YCTD [8,12] was used to measure the participants’ levels of CT in the pretest and posttest. The instrument consists of 27 items and uses a 5-point Likert scale ranging from 1 (strong disagree) to 5 (strong agreement). The instrument’s seven subscales include objectivity, prudence, systematicity, intellectual eagerness/curiosity, intellectual fairness, healthy skepticism, and CT self-confidence. Yoon’s original study [8] reported the instrument’s construct validity and reliability for Korean nursing students; the explained variance for the factor analysis was 52.0%, and the instrument reliability using Cronbach’s coefficient was .84. The YCTD was found to have strong reliability in several previous studies [14,21–23], and the Cronbach’s coefficient score in the present study was .84.

Statistical analysis

CFA was used to validate the seven-factor model proposed by Yoon’s original study. We employed the structural equation model in STATA version 13.0 to estimate the first-order CFA [18]. Considering that the YCTD was developed based on the CCTDI, a theory-driven instrument, CFA was used to test the YCTD’s construct validity. The overall model fits were estimated by the following statistics: the chi-square statistic and associated probability (p), the root mean square error of approximation (RMSEA) index, the standardized root mean squared residual (SRMR), the comparative fit index (CFI), the coefficient of determination, Akaike’s information criterion (AIC), and the Bayesian information criterion (BIC). We have adapted Bentler and Bonett’s [24] criterion of CFI > .90 as indicative of acceptable model fit, and Browne and Cudeck’s [25] criterion of RMSEA < .05 as “close fits”, between .05 and .08 as “reasonably close fit”, and > .10 as “an unacceptable model”. RMSEA indicates the average standardized absolute value of the difference between the observed covariance matrix elements and the covariance matrix elements implied by the given model, with smaller values reflecting better fit. Use of the AIC and BIC is recommended when models are compared and when the model with the smaller AIC and BIC values is preferred. We used a second-order CFA to identify any latent general factor named “critical thinking”
and the value of Cronbach \( \alpha \) to test the internal consistency reliability of the YCTD.

The second step was to evaluate whether the measurement structure of the YCTD was invariant with cross-sectional data between groups and longitudinal data within a group. Cross-sectional measurement invariance between groups was performed using multigroup CFA to compare the performance between groups. Measurement invariance between groups must be confirmed before comparing group performances. To assess correct intervention effects without perception change or response shift within a group, premeasurement and postmeasurement invariance within a group was also performed using the multigroup CFA. Longitudinal measurement invariance must be confirmed to correctly evaluate the intervention effect within a group. The stability of the structure of the instrument was assessed using four measurement invariance methods. First, configurational invariance was assessed by determining whether the same congeneric measurement model, which indicates the same factors and patterns of loadings across groups, provides a reasonable goodness-of-fit to data from each group. Second, factor loading invariance (weak invariance) that concerns the degree to which a one-unit change in the underlying factor is associated with a comparable change in measurement units for the same given item in each group. Third, scalar invariance (strong invariance) was assessed to identify any difference in item intercepts when holding the latent variable mean constant at zero indicating that the particular differential items yield different mean responses for individuals from different groups (but who have the same value on the underlying factor). Fourth, residual invariance (strict invariance) was examined because the invariance of unique errors indicates that the levels of measurement error in item responses are equivalent across groups.

**Ethical considerations**

Institutional Review Board approved the research prior to the study. Instructions and a written consent form were provided voluntarily. Personal participant information was kept confidential by using individual identification numbers. The participants were informed on the data protection and usage before participating in this study. In addition, the original author of the YCTD permitted us to use the tool for this study.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>Cronbach ( \alpha ) with item deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td>1. I think I can get through any complicated problem.</td>
<td>3.73</td>
<td>0.03</td>
<td>.834</td>
</tr>
<tr>
<td></td>
<td>15. I willingly solve a complicated problem.</td>
<td>3.52</td>
<td>0.03</td>
<td>.833</td>
</tr>
<tr>
<td></td>
<td>20. When I have a question, I try to get the answer.</td>
<td>3.77</td>
<td>0.03</td>
<td>.835</td>
</tr>
<tr>
<td></td>
<td>21. I'm trying to understand how the unknown thing works.</td>
<td>3.84</td>
<td>0.02</td>
<td>.833</td>
</tr>
<tr>
<td></td>
<td>22. When I confront a problem, I try hard to find an answer until solving it.</td>
<td>3.62</td>
<td>0.03</td>
<td>.831</td>
</tr>
<tr>
<td>Fairness</td>
<td>8. I explain reasons if I don't agree with others.</td>
<td>3.90</td>
<td>0.02</td>
<td>.836</td>
</tr>
<tr>
<td></td>
<td>24. I willingly accept the proved truth though having different opinion.</td>
<td>4.16</td>
<td>0.02</td>
<td>.841</td>
</tr>
<tr>
<td></td>
<td>26. I evaluate fairly either my opinion or others opinions.</td>
<td>3.95</td>
<td>0.03</td>
<td>.837</td>
</tr>
<tr>
<td>Objectivity</td>
<td>1. I have a reasonable proof.</td>
<td>3.90</td>
<td>0.02</td>
<td>.836</td>
</tr>
<tr>
<td></td>
<td>6. I think any opinion needs to have a reliable reason to insist.</td>
<td>4.23</td>
<td>0.02</td>
<td>.837</td>
</tr>
<tr>
<td></td>
<td>8. I explain reasons if I don't agree with others.</td>
<td>4.00</td>
<td>0.02</td>
<td>.836</td>
</tr>
<tr>
<td>Prudence</td>
<td>2. When I am questioned, I think twice before I give my answer.</td>
<td>3.79</td>
<td>0.03</td>
<td>.835</td>
</tr>
<tr>
<td></td>
<td>4. I tend to make a decision hastily without considering a matter fully.</td>
<td>2.83</td>
<td>0.04</td>
<td>.843</td>
</tr>
<tr>
<td></td>
<td>14. I tend to act rashly and carelessly when I face a difficulty.</td>
<td>2.83</td>
<td>0.04</td>
<td>.847</td>
</tr>
<tr>
<td></td>
<td>18. I don't rush to judgment.</td>
<td>3.62</td>
<td>0.03</td>
<td>.837</td>
</tr>
<tr>
<td>Skepticism</td>
<td>7. I prefer to think differently from others and routines.</td>
<td>3.56</td>
<td>0.03</td>
<td>.839</td>
</tr>
<tr>
<td></td>
<td>10. Although something is already set firmly, I have questions on it.</td>
<td>3.65</td>
<td>0.03</td>
<td>.840</td>
</tr>
<tr>
<td></td>
<td>12. I continually evaluate whether my thought is right or not.</td>
<td>3.68</td>
<td>0.03</td>
<td>.840</td>
</tr>
<tr>
<td>Systematicity</td>
<td>16. When I see the world, I see it with a questioning mind.</td>
<td>3.51</td>
<td>0.03</td>
<td>.838</td>
</tr>
<tr>
<td></td>
<td>18. I don't rush to judgment.</td>
<td>3.60</td>
<td>0.03</td>
<td>.833</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>.842</td>
</tr>
</tbody>
</table>

Note. YCTD — Yoon’s Critical Thinking Disposition.
Results

Reliability

Each item of the YCTD in this study presented in Table 1 was evaluated for their reliability using the coefficients of Cronbach’s. The reliability with the Cronbach’s score was .84. The scores of Cronbach’s with item deleted ranged from .83 to .85 among items.

CFA

We employed CFA to test a seven-factor model of the YCTD originally suggested by Yoon in 2008 [12]. Initially, to identify the possibility of having different factor structures across groups, the configurual model was tested, in which factor loadings could differ among groups. The unstandardized factor loadings for the seven-factor model found in Table 2.

The overall fit of the seven-factor model (Table 3) was acceptable based on the fit indices (CFI = .81, RMSEA = .06, SRMR = .07) though its chi-square value ($\chi^2 = 1,079.70, p < .001$) was significant. The significance of the chi-square value may not be applicable in some cases. For models with about 75–200 cases, the chi-square test is a reasonable measure of fit. But for models with more cases (400 or more), the chi-square test is almost always statistically significant [24,26].

To confirm this observation, the fit of a single-factor model (the second-order CFA with one latent composite factor on top of the seven factors) was tested. The overall fit of the model (Table 3) was not adequate with data as evidenced by its significant chi-square value ($\chi^2 = 2,058.74, p < .001$) and the other values for fit indices (CFI = .57, RMSEA = .09, SRMR = .08). Table 3 shows model fit statistics for the seven-factor model and multigroup CFAs between groups and within a group.

The seven-factor model presented acceptable fit statistics to data: CFI = .81, RMSEA = .06, and SRMR = .07. In addition, the values of AIC and BIC in the seven-factor model were smaller than that in the second-order factor model, which indicates that the first-order seven-factor model is the best model. Then, the seven-factor model was tested in different groups. Figure 1 shows a diagram for the seven-factor CFA of the YCTD.

Measurement invariance of the seven-factor model

The seven-factor model for the YCTD was assessed for the measurement invariance by starting from the baseline configurual invariance followed by factor loadings, intercepts, measurement error variances, and correlated measurement errors. Table 4 shows the measurement invariance findings of the seven-factor model. In steps 1, the configurual invariance, which allows all the other invariances, including the factor loadings, intercepts, measurement error, and correlated measurement errors. Table 2 shows the within-group and between-group unstandardized CFA results for the seven-factor model for the participants in this study. The fit of this model was acceptable between groups ($\chi^2 = 1,750.84(909), p < .001; \text{CFI} = .76; \text{RMSEA} = .08; \text{SRMR} = .09$) and within a group ($\chi^2 = 1,124.81(606), p < .001; \text{CFI} = .76; \text{RMSEA} = .07; \text{SRMR} = .10$), which establish the configurual invariance of the proposed model.

In the next step, the factor loadings were constrained to be equal across cross-sectional groups and longitudinal groups. The constrained model fit between groups was not significantly different from that of the configural model ($\chi^2(df) = 52.18(40), p = .094$) and the fit of the model was acceptable ($\chi^2 = 1,803.02(949), p < .001; \text{CFI} = .76; \text{RMSEA} = .08; \text{SRMR} = .10$). In addition, the difference between the constrained model within a group and unconstrained model was not significantly different ($\chi^2(df) = 27.92(20), p = .111$) and the fit of the model within a group was also acceptable ($\chi^2 = 1,152.73(626), p < .001; \text{CFI} = .76; \text{RMSEA} = .07; \text{SRMR} = .10$). The factor loadings were invariant across cross-sectional groups and longitudinal measurements within a group. The scalar invariance (intercept) as the third step was not acceptable within a group because the model fit $\chi^2$ value was significantly different from that of the previous model ($\chi^2(df) = 74.02(27), p < .001$), whereas it was acceptable between groups ($\chi^2 = 1,885.70(1,003), p < .001; \text{CFI} = .75; \text{RMSEA} = .07; \text{SRMR} = .10$; $\chi^2(df) = 82.68(54), p = .073$). In the fourth step, residual invariance was examined between groups and the model fit chi-square value was significantly different between model 4 and model 3 ($\chi^2(df) = 125.30(54), p < .001$). In testing results of final invariances across the multigroups, the model fit chi-square values between groups ($\chi^2 = 77.43(56), p = .031$) and within a group ($\chi^2(df) = 34.42(28), p = .187$) are not significantly different across the groups.

In this study, CFA demonstrated that the proposed seven-factor model fit the data longitudinally and cross-sectionally, supporting configurual invariance. Multigroup CFA results concerning sequential factor loading, intercept, and residual invariance demonstrated adequate model fit, but the longitudinal invariance was not fully presented, whereas the cross-sectional invariance was the level of strong invariance.

Discussion

This study supports the cross-sectional and longitudinal validity of the YCTD as well as the construct validity of the seven-factor model of the YCTD in Korean nursing students. This model

---

**Table 3. Model Fit Statistics for CFAs.**

<table>
<thead>
<tr>
<th>Model fit index</th>
<th>Overall</th>
<th>Between groups (n = 248)</th>
<th>Within a group (n = 168)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Seven-factor</td>
<td>One-factor</td>
<td>School A (n = 100)</td>
</tr>
<tr>
<td>Chi-square</td>
<td>1,079.70</td>
<td>2,058.74</td>
<td>542.85</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.063</td>
<td>.091</td>
<td>.074</td>
</tr>
<tr>
<td></td>
<td>(.059–.067)</td>
<td>(.087–.094)</td>
<td>(.064–.084)</td>
</tr>
<tr>
<td>AIC</td>
<td>34,446.29</td>
<td>35,383.33</td>
<td>3,844.92</td>
</tr>
<tr>
<td>BIC</td>
<td>34,903.09</td>
<td>35,746.09</td>
<td>4,078.55</td>
</tr>
<tr>
<td>CFI</td>
<td>.81</td>
<td>.57</td>
<td>.80</td>
</tr>
<tr>
<td>SRMR</td>
<td>.07</td>
<td>.08</td>
<td>.10</td>
</tr>
</tbody>
</table>

Note: AIC = Akaike’s information criterion; BIC = Bayesian information criterion; CFI = coefficient of determination; CFA = confirmatory factor analysis; DFI = comparative fit index; RMSEA = root mean square error of approximation index; SRMR = standardized root mean squared residual.
structure demonstrated strong invariance between groups (i.e., configural, factor loading, and intercept combined) but weak invariance within a group (i.e., configural and factor loading combined). Failed scalar (intercept) invariance within a group could be interpreted with discussions in previous studies [27,28]. They suggested that intercept differences may reflect actual mean group differences, not measurement bias. In the present study, students’ CT abilities were measured prepracticum and postpracticum and their scores of CT gains using the YCTD were significantly increased. This actual mean changes might affect the intercept within a group, which was consistent with Millsap’s interpretation [27] in their study.

Our results indicate that the YCTD is a valid tool for measuring nursing students’ CT in Korea since the estimates using the seven-factor model in this study were within the acceptable range. The YCTD was developed based on the CCTDI with the seven subcategories of CT. When we performed the additional testing of exploratory factor analysis with the to identify a good model, we
found that one factor model of the YCTD supported the possibility of one strong latent factor on top of the seven different factors as a second order factor. We found that the the first order seven-factor model is the best model for the YCTD as well as corresponding to the theoretical framework of the original YCTD.

Regarding the measurement invariances across multigroups, the findings showed longitudinal weak invariance and cross-sectional strong invariance because the test of equality of the intercepts within a group and residual between groups failed. Even though the model fit chi-square values between groups and within a group are not significantly different in the testing results of final invariances across the multigroups, change in the CT values using the YCTD over time should be interpreted with caution because measurement properties of the indicators may partly contribute to the change preintervention and postintervention [17]. Further study is recommended using the Yoon and Mill sap’s [29] specification search procedures to find the noninvariant items with the weak invariance within a group and strong invariance between groups measurement invariance achieved in the present study.

Previous studies on the YCTD [14,21–23] as well as Yoon’s exploratory factor analysis studies [8,12] have reported its content validity and reliability results. Considering that there are few studies verifying the construct validity of YCTD for Korean nursing students, it is hard to compare the findings of the present study with other findings. However, the findings of this study could extend the application of this tool and lead to further validity tests.

Cross-sectional measurement invariance between groups was performed using multigroup CFA to confirm the comparability of the measurement for performance evaluation between groups. Measurement invariance between groups must be confirmed before comparing group performance. To assess correct intervention effects without participants’ perception change (response shift) within a group, the premeasurement and postmeasurement invariance (longitudinal measurement invariance) must be confirmed. According to the findings of the statistical significance measurement invariance test result within and between groups, the YCTD can be used to assess group differences of CT or evaluate the intervention effect within a group. In addition, self-assessed CT instruments such as the YCTD will be useful for motivating and educating nursing students to engage in a metacognitive process to improve their self-monitoring skills.

Limitations

Considering that many professionals lack proficiency in self-assessment [30], measuring only student self-assessment when using the YCTD could be limitations of future application. In addition, this study is limited due to the small sample size for sub-analysis between groups and within group to conduct CFA. This is because of the small student size for each school and limited number of participating schools. Therefore, we should interpret the findings of this study with these limitations.

Conclusion

Effective and unbiased evaluation of students in nursing education requires reliable and valid evaluation instruments. In this study, significant reliability and validity of the YCTD cross-sectionally and longitudinally are established. Therefore, we could use the YCTD to evaluate nursing students’ CT for site comparisons and interventions. The use of sound instrument such as the YCTD in this study is imperative to providing quality of nursing education, by a better understanding of nursing students’ progress as well as mentoring. Further studies that include a comparison between YCTD and other CT tools such as Health Sciences Reasoning Test as well as the further assessment of the measurement variance with the YCTD among different areas of Asian countries and different characteristics are recommended.
Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, and Future Planning (KRF no. 2011-0014034). The authors thank Jon S. Mann, college instructor, Academic Center for Excellence, University of Illinois at Chicago for his editorial assistance.

References


2. KABONE. Nursing education accreditation standards [Internet]. Seoul (Korea): Korean Accreditation Board of Nursing Education; 2012 [cited 2012 Mar 10]. Available from: http://kabon.or.kr/eng/kabon02/index.php


