

## Measuring Critical Thinking One Step Forward, One Step Back

Catherine M. Walsh, PhD, RN  
Lisa A. Seldomridge, PhD, RN

*When critical thinking was identified as an explicit program outcome by the National League for Nursing in 1991, nursing programs were thrust into frenzied activity defining critical thinking and selecting instruments to evaluate it. This drove research on the measurement of critical thinking among nursing programs and the concomitant search for reliable and valid methodologies to systematically assess changes in critical thinking in their students. Although the ultimate goal of this process was curricular improvement, nearly 15 years later, faculty struggle to make sense of the data they have.*

Before the publication of a Delphi study on critical thinking (CT) in nursing,<sup>1</sup> nursing programs had been challenged by accrediting organizations to develop their own definitions of CT and to assess how students were meeting the outcome. As a result, the published research reflects a variety of definitions of CT, assessment strategies, and findings on the influence of curricula on the development of CT skills and dispositions.

Assorted definitions of CT appeared in the literature from the mid-1980s through mid-1990s. Definitions focused on complex cognitive processes including decision making and clinical problem solving<sup>2</sup>; “rational explanation of ideas, inferences, assumptions, principles, arguments, conclusions, issues, statements, beliefs, and actions”<sup>3(p5)</sup>; evaluation; assertion of opinion based on specialized knowledge<sup>4</sup>; purposeful, self-regulatory judgment<sup>5</sup>; rational-linear problem solving<sup>6</sup>; reflective, reasonable thinking<sup>7</sup>; and clinical problem solving.<sup>8</sup>

As programs refined the definitions of CT, research attention shifted to issues of measurement. Standardized instruments gained popularity because of the established reliability and validity and ease of administration. CT skills have been assessed using the Watson-Glaser Critical Thinking Appraisal (WGCTA)<sup>9</sup> and the California Critical Thinking Skills Test (CCTST),<sup>10</sup> whereas CT dispositions have been measured by the California Critical Thinking Dispositions Inventory (CCTDI).<sup>11</sup>

Findings have been inconclusive regarding the influence of undergraduate nursing curricula on CT as measured by the WGCTA<sup>9</sup> or the CCTST.<sup>10</sup>

**Authors' Affiliations:** Associate Professor (Dr Walsh); Professor (Dr Seldomridge), Department of Nursing, Salisbury University, Salisbury, Md.

**Corresponding Author:** Dr Walsh, Department of Nursing, Salisbury University, 1101 Camden Avenue, Salisbury, MD 21801 (cmwalsh@salisbury.edu).

Although the timing of assessments varied, most studies used a pre-post design. In Adams<sup>12</sup> integrative review, 19 research studies were published between 1977 and 1992 using the WGCTA to measure CT of nursing students. Nine studies reported a significant increase in CT, 6 reported no change, and 4 reported mixed results. More recently, improved WGCTA scores were found in traditional undergraduate and RN-BSN students who completed the instrument at the beginning and end of their nursing course sequence.<sup>13</sup> In the United Kingdom, no differences were found in student performance on the WGCTA before, during, and at the completion of their academic program.<sup>14</sup>

A longitudinal study using the CCTST<sup>15</sup> found that students enrolled in 3 different undergraduate nursing tracks improved overall scores for all 3 tracks. However, in another longitudinal study of baccalaureate nursing students,<sup>16</sup> gains were found in CCTST scores in only 1 of 3 cohorts.

In a cross-sectional study of freshman through senior Canadian baccalaureate nursing students,<sup>17</sup> changes were measured in scores on the CCTST and CCTDI by Profetto-McGrath. Mean CCTST scores increased from years 1 to 4 with the exception of year 3. There were no changes in overall CCTDI

scores, although significant differences were found on 1 subscale (systematicity). McCarthy et al, in a similar study in the United States,<sup>18</sup> found significant changes in CCTST and CCTDI scores from sophomore to senior years. The groups differed significantly on the truth seeking, confidence, analyticity, and inquisitiveness subscales. Like Profetto-McGrath,<sup>17</sup> McCarthy et al<sup>18</sup> found a strong positive relationship between the CCTST and CCTDI. In contrast, Stone et al<sup>19</sup> found no correlation between the CCTST and CCTDI in senior baccalaureate nursing students in the United States. Recently, Stewart and Dempsey<sup>20</sup> also found no significant differences in CCTDI scores from sophomore to senior years in a small study (n = 34) of baccalaureate nursing students. No studies examining a relationship between the WGCTA and the CCTDI were found. With inconsistent results, many questions remain regarding the best way to measure CT skills and dispositions.

Recently, interest has shifted to understanding what facilitates and stifles the development of CT. Allen et al<sup>21</sup> recommended teaching students how to think critically in nursing. Others<sup>22-24</sup> urged faculties to debate *how* content could be taught to foster the development of CT rather

than *what* should be taught. Several barriers to teaching for CT were found in baccalaureate programs including student resistance to active learning, inadequate class time, insufficient time to prepare CT activities, and need to cover content.<sup>25</sup>

Published research on teaching strategies and their effectiveness in improving CT has been reviewed.<sup>26</sup> These strategies included reflective writing/journaling, use of concept maps, case studies/vignettes, role-playing, CT rounds, and computer-assisted instruction. With a wide range of dependent measures, many of which were anecdotal reports, the research is largely inconclusive.

## Purpose

Although nursing education has made strides in defining CT, much work remains. Ongoing research on CT measurement, construction and evaluation of teaching strategies to develop CT are imperative if the desired outcome is a graduation of self-directed critically thinking nurses. The purpose of this article is to share the results of CT measurement and the experiences and insights of 1 baccalaureate program using 2 standardized instruments over an 8-year period.

## Instruments Used

Critical thinking was conceptually defined in our nursing program as “knowing what to believe or do.”<sup>27</sup> Instruments measuring both the dispositional nature of this attribute and the skills were sought because the notion of the 2-pronged nature of CT was well supported in the literature.<sup>28-30</sup> The CCTDI<sup>10,11</sup> was chosen to track CT dispositions based on psychometrics and ease of administration. It consists of 75 forced-choice items (Likert scale 1-6), which can be clustered in 7 subscales (truth seeking, open-mindedness, analyticity, systematicity, confidence, inquisitiveness, and maturity). Subscales can be scored (ranging from the low 20s to a maximum of 60), and a total overall score can be obtained (maximum 420). Overall scores of less than 270 were deemed to be “weak,” and scores greater than 350 were said to show a “solid indication of across

the board strength.”<sup>30(p16)</sup> Reliability studies<sup>12,30</sup> showed overall CCTDI reliability of .91 and .89, respectively. The instrument was administered to students who were beginning students in upper-division nursing courses and again, 2 years later, at the time of graduation to ascertain changes in students’ CT dispositions between 1995 and 2002 (n = 163).

Unsuccessful attempts to find a satisfactory nursing-specific CT skills test delayed the start-up date of the administration of a CT skills test, but ultimately, the WGCTA<sup>9</sup> was adopted in 2000 because it was widely used by many nursing programs for CT assessment. The instrument is composed of 40 items, grouped into 5 subscales of 5 to 9 items. The maximum score is 40. Reliability was .81 for a development sample of participants in various occupations (n = 4,732) and was .74 for 111 of those participants who were nurse managers and educators.<sup>9(p34)</sup> Beginning in 2000, it was administered, along with the CCTDI, as described above. The final study sample consisted of 93 participants.

## Data Analysis and Interpretation

Analyses of CCTDI and WGCTA data were done using SPSS for Windows, Version 11<sup>31</sup> at an alpha level of .05. The results are presented in Table 1. The top half of the table presents aggregated CCTDI pretest and posttest scores of students graduating from 1997 through 2002 (n = 163). The results of pre-post tests on the CCTDI were mixed, with no consistent pattern from year to year over the 6 years for which data were collected. Al-

though paired *t* test analyses revealed an overall gain in scores, statistical differences were found only for the Classes of 1999 ( $t = 5.858$ ,  $df 21$ ,  $P = .000$ ) and 2002 ( $t = 2.646$ ,  $df 29$ ,  $P = .013$ ). No differences were found for the Classes of 1997, 1998, and 2000. The lack of consistent direction in an upward (or downward) trend was puzzling and became a challenge to explain. A fairly substantial gain was seen in the first cohort (1999) and a moderate gain was seen in last cohort (2002), making the overall picture unclear at best. One explanation might be that the CCTDI does not measure dispositional attributes reliably. Gains for any 1 cohort may not thus be a reliable indicator of CT. Further, there is some evidence in the literature to support the lack of stability in some of the subscales of the CCTDI.<sup>32-34</sup>

The bottom half of Table 1 presents the means, SDs, and paired *t* tests for the WGCTA for the Classes of 2001 and 2002 (n = 93). This analysis revealed a small, but significant mean decline in scores ( $t = -1.988$ ,  $df 92$ ,  $P = .050$ ). The results should be interpreted with obvious caution, as the data represented only 2 cohorts. These results, although showing undesirable direction, may be easier to explain, in that the particular CT skills tested by the WGCTA are not emphasized in the classroom or in clinical settings.

## Experience and Insights Using CT Tests

Results of CT measurement have been consistently disappointing to our faculty over the past 10 years. Initially,

**Table 1. CCTDI (n = 163) and WGCTA (n = 93) Means, SD, and Paired *t* Test for Traditional Undergraduates**

	<i>M</i>	SD	<i>t</i>	<i>P</i>
CCTDI (1997-2002)				
Pretest	295.63	34.33	2.34	.021
Posttest	302.72	27.76		
WGCTA (2001-2002)				
Pretest	23.28	8.47	-1.988	.05
Posttest	20.67	11.26		

CCTDI indicates California Critical Thinking Dispositions Inventory; WGCTA, Watson Glaser Critical Thinking Appraisal.

perhaps through naivete, we believed that any measurement of CT would yield clear results; that is, it would consistently go up (or down) for most cohorts. This belief was generated by virtue of having an extremely stable faculty with a tacit understanding of the curriculum and a teaching culture to which members subscribed. Furthermore, our investigation of the available standardized instruments convinced us that using some type of standardized instrument, despite its limitations, would provide meaningful data that could be used to improve our program.

In looking back, there were several strong points on the CT measurement program. Testing conditions at entry were good. Instruments were administered 1 per class period in a supervised situation. Because students were new to the major, we believed that they took the tests seriously, taking care with their answers. Testing conditions were also consistent over the years, with suitable space, lighting, and temperature control.

There were also weaknesses in the CT measurement program. In general, testing conditions were not as good at exit because a variety of end-of-program instruments were administered to exiting seniors on a designated testing day. Seniors did not take these tests as seriously, reporting "senioritis" and a general desire to be "done." Room proctors corroborated that students moved through the instruments hastily.

Many threats to internal validity could not be controlled, for example, maturation and selection, because of the 1-group design. Instrument quality emerged as a problem as we learned more about each of them. The CCTDI was found to lack stability on several factors, creating interference in the interpretation of results.<sup>32-34</sup> More troubling, from a philosophical standpoint, was the use of the WGCTA, an instrument that assessed general reasoning skills rather than discipline-specific thinking skills that would be learned in a nursing program. Although the faculty were aware of its limitations at the outset, we were reluctant to evaluate it until we had amassed a large enough sample for statistical analysis. With no course in CT or formal logic, it became clear

that the WGCTA was a poor match for our program. In addition, like many nursing programs, we were in our relative infancy of thinking about what constituted CT in nursing. Over time, the faculty have progressed in their thinking and now believe that CT is composed of varied skills, not all of which are used by students at novice levels. Some skills are more commonly needed than others (eg, problem solving rather than creativity).

### What's Next?

Assuming that CT in nursing is something that can be measured, and we believe it can, the improvement of the measures would seem to be the path we should explore. In terms of CT dispositions, items on the CCTDI that consistently show factor analytic irrelevance (low loadings) should be discarded. In terms of CT skills, important decisions need to be made regarding which particular CT skills are appropriate for novice nurses and how to best develop and measure them.

Classroom activities and assignments must be structured to build CT skills at increasing levels of sophistication over 2 years. It is time to re-examine how much classroom time is spent to fact feeding and how much is devoted to helping students learn how to think about clinical situations. Clinical teaching must be carried out to develop specific novice CT skills in generalist settings. New approaches to student clinical assignments could be used where the goal is to improve thinking and problem solving about care rather than to actually perform the care. Strategies to externalize the thought processes of experts, faculty "think-aloud," could be used to model CT in the clinical area.

Finally, continued collection of data using measures of general thinking is unlikely to help improve curricula, instruction, or professional practice; therefore, we believe that it is imperative to develop a discipline-specific instrument to measure thinking skills that are fundamental to entry-level nursing. Such an instrument would have to be structured to illustrate thought and reflection regarding varied ways to care for patients. Thus, there would not neces-

sarily be one narrow "right" pathway to a desired patient outcome but a richer approach that would reveal the presence of pertinent CT skills in the students' answers.

### References

1. Scheffer BK, Rubenfeld MG. A consensus statement on critical thinking in nursing. *J Nurs Educ.* 2000;39(8):352-360.
2. Tanner CA. Research on clinical judgment. In: Holzemer W ed. *Rev Res Nurs Educ.* Thorofare, NJ: Slack; 1983:1-32.
3. Bandman EL, Bandman G. *Critical Thinking in Nursing.* East Norwalk, Conn: Appleton & Lange; 1998.
4. Itano JK. A comparison of the clinical judgment process in experienced registered nurses and student nurses. *J Nurs Educ.* 1989;28(3):120-126.
5. Facione PA. *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. The Delphi Report.* Millbrae, Calif: California Academic Press; 1990.
6. Jones S, Brown L. Critical thinking: impact on nursing education. *J Adv Nurs.* 1991; 16:529-533.
7. Kataoka-Yahiro M, Saylor C. A critical thinking model for nursing judgment. *J Nurs Educ.* 1994;33:351-356.
8. Oermann MH. Evaluating critical thinking in clinical practice. *Nurs Educ.* 1997;22(5): 25-28.
9. Watson GB, Glaser EM. *WGCTA Watson-Glaser Critical Thinking Appraisal S. Form.* San Antonio, Tex: The Psychological Corporation Harcourt Brace and Company; 1994.
10. Facione PA. *California Critical Thinking Skills Test.* Millbrae, Calif: California Academic Press; 1990.
11. Facione PA, Facione NC. *The CCTDI: A Disposition Inventory.* Millbrae, Calif: California Academic Press; 1992.
12. Adams BL. Nursing education for critical thinking: an integrative review. *J Nurs Educ.* 1999;38(3):111-119.
13. Brown JM, Alverson M, Pepa CA. The influence of a baccalaureate program on traditional, RN-BSN, and accelerated students' critical thinking abilities. *Holist Nurs Pract.* 2001;15(3):4-8.
14. Girot EA. Graduate nurses: critical thinkers or better decision makers? *J Adv Nurs.* 2000; 31(2):288-297.
15. Spelic S, Parsons M, Hercinger M, Andrews A, Parks J, Norris J. Evaluation of critical thinking outcomes of a BSN program. *Holist Nurs Pract.* 2001;15(3):27-34.
16. Beckie T, Lowry L, Barnett S. Assessing critical thinking in baccalaureate nursing students: a longitudinal study. *Holist Nurs Pract.* 2001;15(3):18-26.
17. Profetto-McGrath J. The relationship of critical thinking skills and critical thinking

- dispositions of baccalaureate nursing students. *J Adv Nurs*. 2003;43(6):569-577.
18. McCarthy P, Schuster P, Zehr P, McDougal D. Evaluation of critical thinking in a baccalaureate nursing program. *J Nurs Educ*. 1999;38(3):142-144.
  19. Stone C, Davidson L, Evans J, Hansen M. Validity evidence for using a general critical thinking test to measure nursing students' critical thinking. *Holist Nurs Pract*. 2001;15(4):65-74.
  20. Stewart S, Dempsey L. A longitudinal study of baccalaureate nursing students' critical thinking dispositions. *J Nurs Educ*. 2005;44(2):81-85.
  21. Allen G, Rubenfeld M, Scheffer B. Reliability of assessment of critical thinking. *J Prof Nurs*. 2004;20(1):15-22.
  22. Diekelmann N. "Pitching a lecture" and "reading the faces of students:" learning lecturing and the embodied practices of teaching. *J Nurs Educ*. 2002;41(3):97-100.
  23. Valiga T. Teaching thinking: is it worth the effort? *J Nurs Educ*. 2003;42(11):479-480.
  24. Ironside P. "Covering content" and teaching thinking: deconstructing the additive curriculum. *J Nurs Educ*. 2004;43(1):5-14.
  25. Shell R. Perceived barriers to teaching critical thinking by BSN nursing faculty. *Nurs Health Care Perspect*. 2001;22(6):286-291.
  26. Staib S. Teaching and measuring critical thinking. *J Nurs Educ*. 2003;42(11):498-509.
  27. Ennis RH. A logical basis for measuring CT skills. *Educ Leadersh*. 1985;43:44-48.
  28. Baron JB, Sternberg RJ. *Teaching Thinking Skills: Theory and Practice*. New York: Freeman; 1987.
  29. American Philosophical Association. *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. The Delphi Report: Research Findings and Recommendations Prepared for the Committee on Pre-College Philosophy*. Millbrae, Calif: The California Academic Press; 1988.
  30. Facione PA, Facione NC, Sanchez CA. *The California Critical Thinking Disposition Inventory Test Manual*. Millbrae, Calif: California Academic Press; 1992. rev. 1994.
  31. SPSS for Windows, v. 11. Chicago: SPSS.
  32. Walsh CM. *Critical thinking disposition of university students in practice disciplines (nursing, education, and business) and non-practice disciplines (English, history, and psychology): an exploratory study* [unpublished dissertation]. University of Maryland, College Park; 1996.
  33. Walsh CM, Hardy RC. Factor structure stability of the California Critical Thinking Disposition Inventory across gender and various student majors. *Percept Mot Skills*. 1997;85:1211-1228.
  34. Kakai H. Re-examining the factor structure of the California Critical Thinking Disposition Inventory. *Percept Mot Skills*. 2003;96(2):435-438.