

The Measurement and Teaching of Critical Thinking Skills

(研究会実施日: 2010 年 2 月 15 日)

William M. Bart [Professor, Educational Psychology, University of Minnesota]

Education	1965, B.S., (Mathematics, Philosophy), Loyola University Chicago 1967, M.A., (Education), Committee on Measurement, Evaluation, and Statistical Analysis, The University of Chicago 1969, Ph.D., (Education), Committee on Measurement, Evaluation, and Statistical Analysis, The University of Chicago
Experience	1969-72, Assistant Professor, Department of Psychological Foundations of Education, University of Minnesota. 1972-80, Associate Professor, Department of Psychological Foundations of Education, University of Minnesota 1974-75, Fulbright Research Scholar, Max Planck Institute for Psychiatry, Munich, West Germany. 1977-78, Senior Fellow, Laboratory for the Study of Adolescence, Michael Reese Hospital and Medical Center. 1980-Present, Professor, Department of Educational Psychology, University of Minnesota. 1984-85, Fulbright Research Scholar, Max Planck Institute for Psychological Research, Munich, West Germany. 1998, Visiting Research Scholar and Professor, Faculty of Education, Hiroshima University, Higashi-Hiroshima, Japan.
Selected Publications	Bart, W., & Williams-Morris, R. (1990). A refined item digraph analysis of a proportional reasoning test. <i>Applied Measurement in Education</i> , 3, 143-165. Bart, W., & Orton, R. (1991). The cognitive effects of a mathematics in-service workshop on elementary school teachers. <i>Instructional Science</i> , 20, 267-288. Bart, W., Post, T., Behr, M., & Lesh, R. (1994). A diagnostic analysis of a proportional reasoning test item: An introduction to the properties of a semi-dense item. <i>FOCUS on Learning Problems in Mathematics</i> 16, 1-11. Bart, W. (2000). Connecting cognitive theory to instruction and testing: Learning and understanding the concept of torque. <i>Journal of Learning and Curriculum Development</i> , 1, 107-110. Edman, L., Bart, W., Robey, J., & Silverman, J. (2004). Psychometric analysis of the Minnesota Test of Critical Thinking. <i>Psychological Reports</i> , 95, 3-9. Yuzawa, M., Bart, W., Yuzawa, M., & Junko, I. (2005). Young children's knowledge and strategies for comparing space. <i>Early Childhood Research Quarterly</i> , 20, 239-253. Bart, W., Hong, S., & Shin, T. (2007). Inquiry on the role of contemporary chess software to enrich human learning and cognition. In O'Neil, H., & Perez, R. (Eds.). <i>Computer games and human learning</i> (pp. 247-268). New York, NY: Erlbaum. Bart, W., Yuzawa, M., & Yuzawa, M. (2008). The development of mathematical reasoning among young children: How do children understand

area and length? In Saracho, O., & Spodek, B. (Eds.). *Contemporary perspectives on mathematics in early childhood education*. (pp. 157-186). Durham, NC: Information Age Publishers.

Kato, K., & Bart, W. (2010). Poisson distribution. In Salkind, N. J., Dougherty, D., & Frey, B. (Eds.). *Encyclopedia of Research Design*. Thousand Oaks, CA: SAGE Publications.

Professional tasks include teaching courses in learning, cognition, intelligence, creativity, and measurement. Research foci are cognitive diagnostic testing and the improvement of basic cognitive abilities and higher reasoning skills, especially critical thinking skills, among learners.

Introduction

Critical thinking skills are among the most important classes of student outcomes in the 21st century, according to the Partnership for 21st Century Skills (2009). These skills are often in service to other cognitive skills such as problem solving and decision making. Using the vernacular, critical thinking skills permit us “to separate the wheat from the chaff.”

1. Importance of Critical Thinking

1. Critical thinking skills are one class of 21st Century Student Outcomes (Partnership for 21st Century Skills, 2009).
2. Critical thinking is one of the major goals of an undergraduate education (Bok, 2006).
3. Critical thinking is a major learner outcome for the 21st century.

Slide 1

In his important work on undergraduate education entitled *Our Underachieving Colleges*, Derek Bok, a former President of Harvard University, (2006) provided a similar view by arguing that critical thinking is one of the major goals of an undergraduate education. Critical thinking is indeed a major learner outcome for the 21st century.

But what is critical thinking? What are the critical thinking skills? How is critical thinking measured? How does one teach critical thinking? This paper will attempt to provide answers to these questions through a critical review of a set of measures of critical thinking and their psychometric properties. After that critical review will be an exposition of features of instruction for critical thinking skills. This paper will then end with suggestions for

future measurement and teaching of critical thinking skills.

Critical Review of Selected Critical Thinking Skills

The tests to be reviewed in this section are the following: (1) California Critical Thinking Skills Test (CCTST); (2) Cornell Critical Thinking Test (CCTT); (3) Ennis-Weir Critical Thinking Essay (Ennis-Weir); (4) University of Florida Critical Thinking Tests; and (5) Watson Glaser Critical Thinking Appraisal (WGCTA).

2. Selected Critical Thinking Tests

1. California Critical Thinking Skills Test (CCTST)
2. Cornell Critical Thinking Test (CCTT)
3. Ennis-Weir Critical Thinking Essay (Ennis-Weir)
4. University of Florida Critical Thinking Tests
5. Watson Glaser Critical Thinking Appraisal (WGCTA)

Slide 2

This review is not comprehensive, as not all tests of critical thinking are included in this review. For example, the Minnesota Test of Critical Thinking-II with some fine psychometric properties is not included in this review. However, the tests that are reviewed vary enough to permit a detailed examination of the range of critical thinking skills and their assessment.

California Critical Thinking Skills Test (CCTST)

The first test to be examined is the California Critical Thinking Skills Test. That test is a standardized test of critical thinking skills appropriate for college level

students (Facione, 1990a, 1990d, 1991). The test has 34 multiple choice test items with 4 or 5 answer choices for each item. The examinee reads a short passage in each item and then answers the question that follows the passage. Some items are

similar to reading comprehension questions found in standardized tests such as the LSAT and the SAT. The number of items answered correctly is the test score. The time limit for the test is usually 45 minutes.

2.1. California Critical Thinking Skills Test (CCTST)

1. Standardized test of college level critical thinking skills (Facione, 1990a, 1991).
2. Normed with data collected from 1,673 undergraduate students during the 1989-90 school year at California State University Fullerton (Facione, 1990d).

2.1. Uses of the CCTST

1. To measure college level critical thinking.
2. To screen for positions that require problem solving and decision making.
3. To assess critical thinking abilities of nursing students.

2.1. Composition of the CCTST

1. Test is composed of 34 multiple-choice test items with 4 or 5 answer choices.
2. For each item, the examinee reads a short passage and then answers the question that follows the passage.
3. Test score is the number of items answered correctly.

2.1. Composition of the CCTST (continued)

4. The time limit for the test is typically 45 minutes.
5. Some items resemble reading comprehension questions found in the widely-used tests such as the LSAT and the SAT.

2.1. CCTST Subtests

1. Analysis (9 items)
2. Inference (14 items)
3. Evaluation (11 items)

Secondary Subtests

1. Deductive Reasoning
2. Inductive Reasoning

2.1. CCTST Sample Item

Jeanne weighs 120 pounds. Her brother, Frank, weighs 20 pounds more than Jeanne. Their sister, Suzanna, weighs less than Frank.

- A) Suzanna weighs more than Jeanne.
- B) Suzanna weighs less than Jeanne.
- C) Frank weighs more than Jeanne and Suzanna.
- D) Jeanne weighs less than Frank and Suzanna.

Answer: C

The test has a moderate level of reliability, but with lower subtest reliability.

validity, but lacks any evidence of construct validity. In summary, this test is an adequate measure of critical thinking.

The CCTST has content and concurrent

2.1. CCTST Reliability

1. Using KR-20, the internal consistency of the test is .68-.69, a moderate level of reliability.
2. The subtest reliability (Cronbach's alpha) is lower ranging from .14 for the Analysis subtest to .68 for the Deduction subtest.

2.1. CCTST Validity

1. Content validity supported by a panel of experts.
2. Concurrent validity supported by significant correlations with GPA, SAT Math and Verbal, and Nelson-Denny Reading Test (Facione, 1990b).

2.1. CCTST Bias

1. Lack of gender bias
2. Lack of ethnic bias
3. Lack of racial bias
4. Lack of bias based on academic major
5. Students who are English language learners or students with poor reading comprehension ability performed worse than their peers

2.1. CCTST and Reading Comprehension Ability

1. Students who are English language learners or students with poor reading comprehension ability performed worse than their peers.
2. Students who were non-native English speakers did not improve in their CCTST scores as a result of a semester course in critical thinking (Facione, 1990c).

2.1. Conclusions - CCTST

1. The CCTST is an adequate measure of critical thinking ability.
2. However, subtest reliability is lacking.
3. Verbal ability may impact accurate measurement of critical thinking.

Cornell Critical Thinking Test (CCTT)

Robert Ennis and Jason Millman developed the Cornell Critical Thinking Test, composed of multiple choice test

items.

This test has relatively weak psychometric properties and lacks construct validity

2.2. Cornell Critical Thinking Test (CCTT)

1. Robert H. Ennis and Jason Millman developed the Cornell Critical Thinking Test (CCTT).
2. This test has two levels:
 - A) Level X for students in grades 4-14.
 - B) Level Z for gifted secondary school students, college students, and adults.

2.2. Composition of the CCTT

1. Level X has 71 multiple-choice items.
2. Level Z has 52 multiple-choice items.

2.2. Subscales of the CCTT

1. Deduction
2. Semantics
3. Credibility
4. Induction-judging conclusions
5. Induction-planning experiments
6. Definition
7. Assumption identification

2.2. CCTT Sample Item

Suppose you know that
All cats can jump.
All animals that can jump are brown.
Then would this be true?
All cats are brown.

Answers:
1. True
2. False
3. Maybe
Answer: 1. True

2.2. Reliability of the CCTT

1. The internal consistency ranges from .67 to .90 for Level X.
2. The internal consistency ranges from .52 to .77 for Level Z.
3. The split-half reliability ranges from .76 to .87 for Level X.
4. The split-half reliability ranges from .55 to .76 for Level Z.

2.2. Validity of the CCTT

1. Hughes (1992) reported that the concurrent validity of the CCTT to be relatively low.
2. The correlations of the CCTT with student grades were in the .15-.17 range, but the correlations of the CCTT with scholastic aptitude and intelligence measures were approximately .50.

2.2. Conclusions - CCTT

1. The CCTT has relatively weak psychometric properties – especially, validity.
2. There is a need for empirical research on the construct of critical thinking and on how that construct is different from intelligence and reading comprehension ability.
3. Such research may contribute to research on the construct validity of the CCTT.

Ennis-Weir Critical Thinking Essay Test (Ennis-Weir)

This is an essay test in which examinees evaluate arguments presented in fictitious letters to the editor.

Although the test items have more

ecological validity than many other test items assessing critical thinking, there is a need for more research on the reliability and validity of the test. In addition, there is a need for greater precision in the scoring system to reduce the subjectivity of the test.

2.3. Ennis-Weir Critical Thinking Essay Test (Ennis-Weir)

1. Robert Ennis and Eric Weir developed the Ennis-Weir Critical Thinking Essay Test (Ennis-Weir).
2. This is an essay test in which the examinee evaluates arguments presented in fictitious letters to the editor.

2.3. Ennis-Weir Scoring Scheme - 1

1. After reading a fictitious letter to the editor, the examinee evaluates nine arguments presented in the letter.
2. The examinee then writes an evaluative essay for each of the nine arguments.
3. A qualified examiner assesses each essay.
4. The maximum score for the test is 29 points.

2.3. Ennis-Weir Scoring Scheme - 2

5. The test requires approximately 40 minutes to complete.
6. The test has external validity in that the item responses are similar to typical instances of critical thinking.
7. However, the test scoring scheme is subjective.

2.3. Ennis-Weir Scoring Scheme - 3

8. Properties of item essays:
 - a. Getting the point.
 - b. Seeing the reasons and assumptions.
 - c. Stating one's point.
 - d. Offering good reasons.
 - e. Seeing other possibilities, and
 - f. Responding appropriately to and/or avoiding the following argument weaknesses such as use of emotive language to persuade.

2.3. Reliability of the Ennis-Weir

1. Inter-rater reliability for 27 college students is .86.
2. Inter-rater reliability for 28 gifted 8th grade students is .82. (Ennis & Weir, 1985).

2.3. Validity of the Ennis-Weir

1. The test seems to have content validity.
2. No concurrent validity of the test is available in the test manual.

2.3. Conclusions - Ennis-Weir

1. There is a need for research on the validity and reliability of the test.
2. There is also a need for more precise scoring criteria to decrease test subjectivity.

University of Florida Critical Thinking Tests

The tests of critical thinking developed at the University of Florida are the following: (1) the Critical Thinking Disposition Assessment Instrument (EMI); (2) the Quantitative Florida Test for Critical Thinking Skills (UFCTS); and (3) the Qualitative Florida Test for Critical

Thinking Skills (UFQCTS). The tests were measures of critical thinking applied to food biotechnology.

These tests have limited utility, because they are set in the context of food biotechnology. The subtests tend to have fine reliabilities, but these tests lack validity information.

2.4. University of Florida Critical Thinking Tests

- There are three tests:
 - a. Critical Thinking Disposition Assessment Instrument.
 - b. Quantitative Florida Test for Critical Thinking Skills.
 - c. Qualitative Florida Scale for Critical Thinking Skills.

2.4.a. Critical Thinking Disposition Assessment Instrument (EMI)

- This test assesses three constructs:
 - a. Engagement -- confidence in thinking and communication.
 - b. Cognitive Maturity -- self-awareness and objectivity.
 - c. Innovativeness -- intellectual curiosity.

2.4.a. Scoring Scheme for the EMI

1. The test has 26 items making use of a 5-point Likert scale with 1 representing "Strongly Agree" and "5" representing "Strongly Disagree".
2. Maximal scores were 55 for Engagement (11 items), 40 for Maturity (8 items), 35 for Innovativeness (7 items), and 130 for the total (26 items).

2.4.a. Sample Item for the EMI

"I am interested in many topics."

1	2	3	4	5
Strongly			Strongly	
	Agree		Disagree	

- This item measures Engagement.

2.4.b. Quantitative Florida Test for Critical Thinking Skills (UF/CTS)

- This test assesses three constructs:
 - a. Analysis
 - b. Inference
 - c. Evaluation

2.4.b. Scoring Scheme for the UF/CTS

1. Items are multiple choice items with 5-8 items assessing each construct.
2. All items involve a food biotechnology context.
3. Maximal scores are 100 for Analysis, 100 for Evaluation, 100 for Inference, and 300 for the Total.

2.4.c. Qualitative Florida Test for Critical Thinking Skills (UF/QCTS)

1. This test assesses six constructs:
 - a. Interpretation
 - b. Analysis
 - c. Evaluation
 - d. Inference
 - e. Explanation
 - f. Self-regulation
2. Items entailed evaluations of editorials.

2.4.c. Scoring Scheme for the UF/QCTS

1. This test has 24 items with 4 items relating to each construct.
2. Each item involves the evaluation of an editorial.
3. Each item response is measured on a 5-point scale with 1 representing "Poor Critical Thinking" and 5 representing "Excellent Critical Thinking".
4. There are three raters for each test response.

2.4. Reliability of the UF Tests

1. EMI alphas are .89 for Engagement, .73 for Cognitive Maturity, and .82 for Innovativeness.
2. UF-CTS alphas are .68 for Analysis, .85 for Inference, and .72 for Evaluation.

2.4. Validity of the UF Tests

- A panel of experts in critical thinking and food biotechnology attested to the content and face validity of the tests.

2.4. Conclusions - UF Tests

1. The tests provide measures of critical thinking in the context of food biotechnology.
2. The subtests tended to have fine reliabilities.
3. The UF-QCTS is difficult to score.
4. Validity research on the tests is sparse.

Slide Series 4-2

The Watson-Glaser Critical Thinking Assessment (WGCLA)

This test is the classic measure of critical thinking. Its items form five cognitive classes: inference, recognition of assumptions, deduction, interpretation,

and evaluation of arguments.

The Watson-Glaser Critical Thinking Assessment has moderate levels of reliability. In addition, there is evidence that the test has content validity and concurrent validity.

<p>2.5. Watson-Glaser Critical Thinking Appraisal (WGCTA)</p> <ol style="list-style-type: none"> 1. Classic measure of critical thinking 2. Uses <ol style="list-style-type: none"> a. To predict success in many professional occupations. b. To assess critical thinking as an outcome measure. 	<p>2.5. WGCTA - Types of Items</p> <ol style="list-style-type: none"> 1. Inference 2. Recognition of Assumptions 3. Deduction 4. Interpretation 5. Evaluation of Arguments.
<p>2.5. WGCTA - Forms</p> <ol style="list-style-type: none"> 1. There are various forms of this test, including Form S, Form A, and Form B. 2. Form S has 16 scenarios and 40 items, all selected from the 80-item Form A. 	<p>2.5. WGCTA Sample Item</p> <ul style="list-style-type: none"> • Evaluation of Arguments Item Should all young adults learn a trade? <p>No; excessive studying permanently warps an individual's character. (This argument is not directly related to the question, because learning a trade does not necessarily require excessive studying.)</p>
<p>2.5. WCGTA - Reliability</p> <ol style="list-style-type: none"> 1. The alpha reliability tends to be in the .73-.86 range. 2. The alpha reliability of the subscales tends to be lower in the .31-.83 range. 3. The split-half reliability is .75. 4. The reliabilities of the WCGTA and its subscales tend to be moderate. 	<p>2.5. WCGTA - Validity</p> <ol style="list-style-type: none"> 1. The scores on the Inference and Deduction subscales significantly predicted course grades. 2. Correlation between course grades and Total test scores was .41, $p < .01$. 3. There is evidence in support of the face and concurrent validity of the test.

Slide Series 5

Certain conclusions result from this review. First, test subscales lack adequate levels of reliability. Second, the tests lack evidence of construct validity. Third, there is a lack of systematic inquiry on the factor structure of critical thinking tests. Fourth, there is a lack of inquiry on the diagnostic and predictive features of critical thinking measures.

2.6. General Conclusions regarding the Selected Tests of Critical Thinking

- Needed are the following:
 - a. Improved reliability of subtests.
 - b. Inquiry on the construct validity of the tests.
 - c. Careful investigation of the factor structure of critical thinking.
 - d. Development of a new generation critical thinking test with predictive and diagnostic properties.

Slide 3

Features of Critical Thinking Skill Instruction

Critical thinking skills can be taught. One

approach involves the evaluation of arguments through answering a set of questions such as “what is the conclusion?”

3. Instruction for Critical Thinking Skills

1. One approach involves the critical evaluation of written arguments (e.g., Browne and Keeley, 2004).
2. The student learns to answer specific questions when reading text.
3. Students learn to evaluate critically letters to the editor, editorials, and newspaper opinion pieces, and essays.

3. Questions used in critical reading of text (Browne & Keeley, 2004).

1. What is the issue?
2. What is the conclusion?
3. What are the reasons?
4. What words or phrases are ambiguous?
5. What are the value conflicts and assumptions?
6. Are there any fallacies in the reasoning?

3. Questions used in critical reading of text (Browne & Keeley, 2004).

7. How good is the evidence?
8. Are there rival causes?
9. Are the statistics deceptive?
10. What significant information is omitted?
11. What reasonable conclusions are possible?

3. Example of a Fallacy

- 1 Argumentum ad baculum (i.e., argument of the staff, appeal to force)
 - a Fallacy occurs when the listener is coerced to accept a conclusion of an argument.
 - b “Please accept the conclusion of this argument; otherwise, seek employment elsewhere.”

3. Another Example of a Fallacy

2. Argumentum ad hominem
 - a. Fallacy occurs when the premises castigate an individual rather than question issues.
 - b. “Hideki’s view of the Tokyo Park System has no value, because he does not like nature. He owns neither plants nor pets.”

3. Anticipated Outcomes

1. To develop critical thinking skills.
2. To facilitate the attainment of superior levels of verbal achievement among participating learners.
3. To develop advanced reading and writing skills.

Among the materials that a teacher could use when teaching critical thinking skills are newspapers. The editorials, letters to the editor, and opinion articles in newspapers are especially useful. Teachers

may ask students to read newspaper editorials or letters to the editor and critically evaluate those passages by writing the answers to the questions associated with critical thinking skills.

3. Importance of Critical Thinking

Critical thinking is the principal aim of undergraduate education (Bok, 2006).

4. Suggestions for Future Measurement and Teaching of Critical Thinking Skills

1. Development of a new and improved test of critical thinking skills that has predictive and diagnostic properties.
2. Development of a new and improved instructional plan for the teaching of critical thinking skills.

4. Suggestions for Future Measurement and Teaching of Critical Thinking Skills

3. Research on how to reconcile conceptual frameworks underlying measures of critical thinking (e.g., the APA Delphi model) with an instructional plan to teach critical thinking skills (e.g., the Browne & Keeley, 2004 approach).

Slide Series 6-2

Critical thinking is a crucial 21st century skill. Three lines of future inquiry are the following: (1) the development of measures of critical thinking that have predictive and diagnostic potential; (2) continued refinement of instruction for critical thinking skills; and (3) the coordination of conceptual frameworks underlying critical thinking skills and instructional plans for teaching critical thinking skills. Pursuit of these three lines of inquiry will likely facilitate the development of critical thinking skills among people.

< References >

Bok, D. (2006). *Underachieving colleges*. Princeton, NJ: Princeton University Press.

Browne, M. N., & Keeley, S. M. (2004). *Asking the right questions: A guide to critical thinking (7th ed.)*. Englewood Cliffs, NJ: Prentice-Hall.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9-26). New York, NY: W. H. Freeman.

- Facione, P. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. The executive summary of the American Philosophical Association Delphi Research Report*. Millbrae, CA: The California Academic Press.
- Facione, P. A. (1990a). *The California Critical Thinking Skills Test – College Level. Technical Report #1: Experimental Validation and Content Validity*. Millbrae, CA: California Academic Press. (ERIC Document Reproduction Service No. TM 015818).
- Facione, P. A. (1990b). *The California Critical Thinking Skills Test – College Level. Technical Report #2: Factors predictive of CT skills*. Millbrae, CA: California Academic Press. (ERIC Document Reproduction Service No. TM 015819).
- Facione, P. A. (1990c). *The California Critical Thinking Skills Test – College Level. Technical Report #3: Gender, Ethnicity, Major, CT Self-esteem, and the CCTST*. Millbrae, CA: California Academic Press. (ERIC Document Reproduction Service No. TM 015962).
- Facione, Peter A. (1990d). *The California Critical Thinking Skills Test – College Level. Technical Report #4: Interpreting the CCTST, Group Norms, and Subscores*. Millbrae, CA: California Academic Press. (ERIC Document Reproduction Service No. TM 015996)
- Facione, Peter A. (1991). *Using the California Critical Thinking Skills Test in Research, Evaluation, and Assessment*. Millbrae, CA: California Academic Press. (ERIC Document Reproduction Service No. TM 017349)
- Hughes, J. N. (1992). Review of the Cornell Critical Thinking Tests. In J. J. Kramer & J. C. Conoley (Eds.), *The Eleventh Mental Measurements Yearbook*. Lincoln, NE: Burros Institute of Mental Measurements.
- Partnership for 21st Century Skills. (2009, April 22). *Framework for 21st Century Learning*. Retrieved October 29, 2009 from http://www.21stcenturyskills.org/documents/P21_framework.pdf