

An Interpretation of a Comparative Study of Japanese and English Versions of the Minnesota Test of Critical Thinking-II

(研究会実施日: 2010 年 2 月 16 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

The focus of this study is the Minnesota Test of Critical Thinking-II (MTCT-II). This test assesses skill at the analysis and evaluation of arguments in contemporary public controversies. The MTCT-II is a test of critical thinking that is based on the conceptual framework for critical thinking formulated by a committee of scholars in the American Philosophical Association. The MTCT-II is one of few tests based on a conceptual framework for critical thinking. This test is a rather difficult test intended to assess critical thinking skills among college students and adults.

The original version of the test was in English. A Japanese version of this test was made and administered to Japanese individuals.

The objectives of this inquiry were the following: (1) to analyze the test data collected in Japan and compare the results with results from an analysis of comparable American test data; and (2) to interpret the results and then offer recommendations regarding the measurement of critical thinking skills.

The Test of Critical Thinking Skills

- The Minnesota Test of Critical Thinking-II (MTCT-II) is a test of critical thinking that entails the analysis and evaluation of arguments regarding contemporary public controversies.

Versions of Test

1. The MTCT-II was originally written in English and used with American students.
2. A Japanese version of the MTCT-II was prepared and used with Japanese individuals.

Objectives of the Comparative Inquiry

1. To analyze Japanese test data and compare its results with results from an analysis of comparable American test data.
2. To interpret the results along with recommendations as to future measurement of critical thinking skills.

Definition of Critical Thinking

- Critical thinking is reasonable reflective thinking that is focused on deciding what to believe or do (Ennis, 1987).

Components of Critical Thinking

Critical thinking has six components according to the American Philosophical Association (Facione, 1990). They are Interpretation, Analysis, Evaluation, Inference, Explanation, and Self-Regulation.

Components of Critical Thinking to the American Philosophical Association (Facione 1990)

1. Interpretation: includes categorization, decoding significance, and clarifying meaning.
2. Analysis: includes examining ideas, identifying arguments, and analyzing arguments.
3. Evaluation: includes assessing claims and arguments.

Slide 1

Components of Critical Thinking to the American Philosophical Association (Facione, 1990)

4. Inference: includes querying evidence, conjecturing alternatives, and drawing conclusions.
5. Explanation: includes stating results, justifying procedures, and presenting arguments.
6. Self-Regulation: includes self-examination and self-correction.

Slide 2

Composition Sample Item of the MTCT-II (revised)

Discussion #1: Logging in national forests

The statement that best describes the main point of Takeshi's argument in paragraph 1 is

- A. Logging companies should pay more for logging on public land.
- B. Commercial logging on public land should be banned completely.
- C. Logging companies should be made to pay for and clean up the damage that they do.
- D. Logging companies should use more environmentally friendly methods.

Answer: B; this item is a measure of Interpretation..

Slide 3

Composition of the MTCT-II (continued)

4. The test involves six controversies:
 - a. Logging in national forests.
 - b. The death penalty.
 - c. The legalization of drugs.
 - d. State sponsorship of lotteries.
 - e. Social promotion.
 - f. State-provided vouchers to attend private school.
5. There are 11 items that relate to each controversy with 2 items measuring each of 5 constructs and 1 item measuring the 6th construct.

Slide 4

Minnesota Test of Critical Thinking-II (MTCT-II)

The MTCT-II consists of multiple-choice items. Examinees read two opposing views on each of six controversies. The examinees then answer items that relate to each controversy. The six controversies relate to four national controversies such as logging in the national forests and the death penalty, but also to two educational controversies related to social promotion and school vouchers.

The MTCT-II has 11 items that relate to each controversy. In addition, the test has 12 items assessing each of the six cognitive constructs except for the construct of Explanation for which there are six items. An examinee receives a score of 1 for an item if the response to the item is correct and a score of 0 for the item if the response to the item is incorrect, excluding Explanation items. An examinee receives a score of 2 for an Explanation item if the response to the item includes an explanation, a score of 1 if the response consists of a restatement of an argument, and a score of 0 if the response provides no explanation. The maximum score for each subscale is 12. The maximum score for the total test is 72.

Scoring scheme for the MTCT-II

1. There are 12 items for each of the six constructs, except for the construct of Explanation.
2. There are 6 items that measure Explanation.
3. An examinee receives a score of 1 for an item if the response to the item is correct and a score of 0 for the item if the response to the item is incorrect, excluding Explanation items.

Slide 5

Scoring scheme for the MTCT-II (continued)

4. An examinee receives a score of 2 for an Explanation item if the response to the item includes an explanation, a score of 1 if the response consists of a restatement of an argument, and a score of 0 if the response provides no explanation.
5. The maximum score for each subscale is 12.
6. The maximum score for the total test is 72.

Slide 6

Scoring scheme for the MTCT-II (continued)

7. The total Belief Bias score is the respondent's score on items pertaining to a position opposite to that of the respondent minus the respondent's score on items pertaining to a position held by the respondent.
8. A positive Total Belief Bias score indicates bias in critical thinking.
9. A Total Belief Bias score near zero indicates objectivity in critical thinking.
10. A negative Total Belief Bias score is not readily explained within the present theoretical context.

Slide 7

Psychometric Properties of the MTCT-II

The reliability of the MTCT-II is .91 and the reliabilities of the six subtests range from .50 for the Evaluation subtest to .78 for the Explanation subtest. A factor analysis of the subtest scores lent support for a single factor structure underlying the subtests.

The MTCT-II has fine concurrent validity with significant correlations with other measures of critical thinking. In addition, it is significantly correlated with participant GPA (.54) and years of

education (.42). There is clear evidence that the MTCT-II is a valid test of critical thinking.

Reliability for the MTCT-II

1. The Cronbach's alpha reliabilities were the following:

a. Total test	.91
b. Interpretation	.68
c. Analysis	.71
d. Evaluation	.50
e. Inference	.66
f. Explanation	.78
g. Self-Regulation	.71

Slide 8

Reliability for the MTCT-II (continued)

2. A maximum-likelihood factor analysis was used on the six subscale scores for 150 participants.
3. The goodness of fit test for the one factor model was statistically significant.
4. The one factor model accounted for 40.71% of the test variance. (The Cronbach's alpha reliability was .78 for the total test.)
5. The factor loadings for the Interpretation and Explanation subscales were .25 and .21 respectively. Those values reflect their low reliabilities.

Slide 9

Validity of the MTCT-II

1. The MTCT-II has fine concurrent validity, as indicated by the correlations between the MTCT-II and WGCTA (.66), Ennis-Weir (57), the Multidimensional Aptitude Battery (MAB) vocabulary test (.61), and MAB comprehension test (.51).
2. Correlations between the MTCT-II and participant GPA (.54) and ACT scores (.69) were also significant.
3. MTCT-II scores were correlated with participant age (.32) and years of education (.42).

Slide 10

Validity of the MTCT-II

4. The MTCT-II was more highly correlated with age than other tests of critical thinking (i.e., WGCTA, Ennis-Weir) with .32 vs. .28 and .19 respectively.
5. The MTCT-II was more highly correlated with years of post-secondary education completed than either WGCTA or Ennis-Weir with .42 vs. .31 and .21 respectively.

Slide 11

Features of the Japanese Version of the MTCT-II

The Japanese students had a slightly lower total mean score (31.33) than the U.S. students (33.39) and their scores had less variability than the American scores. The cognitive construct subtest mean scores for the two groups were quite similar, as were the controversy subtest mean scores.

The reliabilities of the total test and the subtests tended to be lower for the Japanese students than for the American students. The correlations among the subtests for the Japanese students (range, .21-.51) were lower than the correlations among the subtests for the American students (.54-.75)

Comparisons of the Japanese and USA Versions

1. Japanese students received a slightly lower total test score mean than the U.S. students (31.33 vs. 33.39). This may be due to the lack of familiarity of some of the controversies to the Japanese students.
2. Japanese student scores had less variability than U.S. student scores. This may be due to the greater variability among American students.

Comparisons of the Japanese and USA Versions

3. Subscale means were similar between the two groups for the Analysis and Interpretation subscales.
4. Subscale means for the American students were higher for the other subscales but the differences do not seem significant.

Comparisons of the Japanese and USA Versions

5. Controversy means were similar for the two groups.
6. The testlet analysis with the Japanese data suggested that certain examinees gave benefits to some examinees separate from their critical thinking skills.
7. One unanswered question is the following: why are some controversy subtests more difficult than others?

Comparisons of the Japanese and USA Versions

8. Controversy means	Japan	U.S.A.
Logging in natural forests	6.07	6.15
School vouchers	4.92	5.33
Legalizing drugs	4.97	6.22
The death penalty	5.88	5.97
Grade retention and promotion	5.51	5.22
State sponsored lotteries	3.78	4.50

Comparisons of the Japanese and USA Versions

9. The reliabilities of the total test and the subscales tended to be lower for the Japanese students than the American students.
10. The Evaluation subscale had the lowest subscale reliability for each group.
11. Several items in the Japanese version had negative discriminations and that likely accounts for the lower reliabilities for the Japanese version.

Comparisons of the Japanese and USA Versions

12. The correlations among subscale scores were large and positive for the American students (range, .54-.75) but less for the Japanese students (.21-.51).
13. As a result, the use of subscale scores makes less practical sense with American students than with Japanese students.

Factor analyses of both data sets indicate that the MTCT-II is unidimensional. Some items were more difficult for one group than for the other group.

There were no significant differences for gender and academic major for Japanese students. Also, age, grade, and achievement level were moderately correlated with total test score with correlations being .18, .22, and .29 respectively.

Comparisons of the Japanese and USA Versions

14. Factor analysis of test data revealed relatively large first factors for both groups with the first factor for the American data accounting for 16% of that test data and the first factor for the Japanese data accounting for 12% of that test data.
15. The subsequent five factors for each version accounted for approximately 30% of the total data.
16. These analyses suggest that the MTCT-II is unidimensional.

Slide 12

Comparisons of the Japanese and USA Versions

17. Certain items were more difficult for one group than the other group. For example, items 13, 25, 26, and 29 were more difficult for Japanese students and item 44 was more difficult for American students.
18. Certain items had negative or zero discriminations. For example, items 12, 18, 26, 29, and 33 in the Japanese version had negative or zero discrimination.
19. Those and other discrepancies among the items if addressed would likely engender improvements of the two test versions.

Slide 13

Conclusions

The similarities between the psychometric properties of the Japanese and the American versions of the MTCT-II are striking. The test has many positive psychometric properties, but improvements

are possible. One path for the improvement of the test is the introduction of at least one new controversy known by all examinees.

Comparisons of the Japanese and USA Versions - Conclusions

1. The CRET Report termed "The Comparison of the U.S. and Japanese Versions of the Minnesota Test of Critical Thinking-II" and dated October 20, 2009 is well done, extensive, and informative.
2. The similarities between the psychometric properties of the Japanese and American versions of the MTCT-II are striking.
3. This critical review and the report have identified ways to improve the test.

Slide 14

Comparisons of the Japanese and USA Versions - Conclusions

4. Additional ways to improve the test emanate from a reflection of the weaknesses of the test.
5. One weakness of the test is the difficulty of the test. It is preferable that the test items have difficulty of approximately .50.
6. A related weakness is the set of controversies used in the test. The controversies in test ideally should all be known by the examinees.
7. An additional weakness is the lack of specific diagnostic value of the test. Measurement should inform instruction for critical thinking skills.

Slide 15

Future Directions for the Assessment of Critical Thinking

This comparison of the Japanese and English versions of the MTCT-II provide ample evidence that the MTCT-II is a fine test of critical thinking with many desirable psychometric properties that are maintained across distinctly different nations. Despite the many positive features of the MTCT-II, there are still areas of improvement for the assessment of critical thinking. These areas of improvement constitute future directions for the assessment of critical thinking.

Proposed properties of any improved test of critical thinking determine the future directions for inquiry on the assessment of critical thinking. Those proposed properties are the following:

1. Excellent psychometric properties, including construct validity (Ennis, 1987).
2. Well-integrated with models of effectively teaching critical thinking (e.g., Browne & Keeley, 2004).
3. High potential as a fine diagnostic tool.
4. High potential as a predictor of success in the professions.

5. High potential as a fine measure of an important cognitive outcome.
6. Development of different versions for children, adolescents, and adults.
7. Development of versions in different world languages.

Critical thinking skills are among the most important cognitive educational outcomes throughout the world (Bok, 2006). The assessment of critical thinking skills will continue to play a vital role in the development of critical thinking skills in students.

Comparisons of the Japanese and USA Versions - Future Directions

Let me recommend that a new test of critical thinking be developed that has the following properties:

1. Excellent psychometric properties, including construct validity.
2. Well integrated with models of effective instruction for critical thinking.
3. High potential as a diagnostic tool.
 - a. Identification of errors in critical thinking.
 - b. Grouping of errors in terms of level of pre-critical thinking.
 - c. Formation of items with foils that result from errors and misconceptions in critical thinking.

Comparisons of the Japanese and USA Versions - Future Directions

Let me recommend that a new test of critical thinking be developed that has the following properties:

3. High potential as a diagnostic tool.
 - d. Formation of a scoring scheme that will permit qualitative inferences from item response patterns to recommendations for instructional correctives for examinees.
 - e. Possible use of multiple keys for such a scoring scheme such that some keys would identify answer choices that would result from certain types of defects in critical thinking.
 - f. Use of psychometric models that identify levels of critical thinking competence.

Comparisons of the Japanese and USA Versions - Future Directions

Let me recommend that a new test of critical thinking be developed that has the following properties:

3. High potential as a diagnostic tool.
 - g. The measurement of critical thinking should inform instruction for critical thinking skills and instruction for critical thinking skills should inform the measurement of critical thinking.
 - h. More research is needed on how to make a new-generation critical thinking test highly diagnostic.
 - i. This quest should likely best be done in stages with moderate forms of improvement occurring with each new version of the test.

Comparisons of the Japanese and USA Versions - Future Directions

Let me recommend that a new test of critical thinking be developed that has the following properties:

4. High potential as a predictor of success in the professions.
 - a. Total score or subscale scores could be used to predict future professional success.
 - b. Other scoring methods could be used to diagnose weaknesses in critical thinking.
5. High potential as an outcome measure.

Comparisons of the Japanese and USA Versions - Future Directions

Let me recommend that a new test of critical thinking be developed that has the following properties:

6. Possible subsequent development of different versions of the test.
 - a. a version for children to assess novice levels of critical thinking.
 - b. a version for adolescents to assess intermediate levels of critical thinking.
 - c. a version for gifted adolescents and adults to assess advanced levels of critical thinking.

Comparisons of the Japanese and USA Versions - Future Directions

Let me recommend that a new test of critical thinking be developed that has the following properties:

7. Possible use of the test worldwide.
 - a. Critical thinking is a worldwide concern.
 - b. Research on versions of the test in different world languages could provide the basis for the test to be used worldwide.
 - c. In addition, a form of the test could become part of a possible 21st century skills test battery that could also be used worldwide.
 - d. Excellent tests may likely serve the needs of people throughout the world.

Slide Series 3-2

< 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.