

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

William M. Bart, Ph.D.
The University of Minnesota
Invited Colloquium
Benesse/CRET
Tokyo, Japan
February 16, 2010

1

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.

2

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.

3

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.

4

Definition of Critical Thinking

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

5

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.

6

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.

7

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..

8

Composition of the MTCT-II

1. There are 12 multiple choice items assessing each of five constructs:
 1. Interpretation.
 2. Analysis.
 3. Evaluation.
 4. Inference.
 5. Self-Regulation.
2. There are six open-ended items assessing Explanation.
3. There are six items in which a participant indicates which viewpoint that the participant tends to share.

9

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.

10

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.

11

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.

12

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.

13

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

14

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.

15

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).

16

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.

17

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.

18

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.

19

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?

20

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

21

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.

22

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.

23

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.

24

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.

25

Comparisons of the Japanese and USA Versions

20. The analysis of prior opinions and controversy scores may benefit from an examination of the bias analysis that was done with the American version by its developers.
21. That Japanese examinees who agreed with Statement B in controversy 5 tended to score higher on controversy #5 than examinees who agreed with Statement A.
22. That result suggests that controversy #5 lacks neutrality.

26

Comparisons of the Japanese and USA Versions

23. There was no significant difference for gender and academic major for Japanese students.
24. Age, grade, and achievement level were moderately correlated with total test score with the correlations being .18, .22, and .29 respectively and the R-squared values being .035, .047, and .084 respectively.

27

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.

28

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.

29

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.

30

Comparisons of the Japanese and USA Versions - Future Directions

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.

31

Comparisons of the Japanese and USA Versions - Future Directions

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.

32

Comparisons of the Japanese and USA Versions - Future Directions

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.

33

Comparisons of the Japanese and USA Versions - Future Directions

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.

34

Comparisons of the Japanese and USA Versions - Future Directions

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.

35

Open Invitation

- Any interested party, who is interested in the further inquiry on the assessment and the development of critical thinking is invited to contact the presenter, William M. Bart, by e-mail or by regular mail.

36

Contact Information

William M. Bart, Ph.D.

Professor, Department of Educational Psychology
University of Minnesota

Regular mail address: 162 Educational Sciences Building
University of Minnesota
56 East River Road
Minneapolis, Minnesota 55455
USA

E-mail address: lindentree10@gmail.com

37

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.

38