DOCUMENT RESUME

ED 469 162 TM 033 831

AUTHOR Edman, Laird R. O.; Robey, Jennifer; Bart, William M.

TITLE Critical Thinking, Belief Bias, Epistemological Assumptions, and the Minnesota Test of Critical Thinking.

PUB DATE 2002-04-00

NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE EDRS Price MF01/PC01 Plus Postage.

DESCRIPTORS *Beliefs; Bias; *College Students; *Critical Thinking; Epistemology; *Evaluation Methods; Higher Education; Objectives; Thinking Skills

ABSTRACT

The Minnesota Test of Critical Thinking-II (MTCT) has been designed to measure both critical thinking (CT) skills and the willingness to evaluate critically arguments that are congruent with one's own goals and beliefs. The MTCT uses a taxonomy of CT skills derived from the American Philosophical Association's definition of critical thinking. Participants were 232 college students at 3 institutions who were administered the MTCT, the Watson-Glaser Critical Thinking Appraisal, the Ennis-Weir Critical Thinking Essay Examination, three subtests of the Multi-Dimensional Aptitude Battery, the Epistemological Questionnaire, and a demographic questionnaire. Results of this study indicate that the MTCT may be a valuable instrument for assessing critical thinking skills. The reliability of the MTCT in this setting with these subjects was high, with the Cronbach's alpha scores for the MTCT higher than those reported for most major tests of CT in the literature. Also promising are the moderately high reliability indices for the subscale scores. Correlations of the scores on the MTCT with scores on the other tests administered are in the ranges hypothesized and support the concurrent validity of the test. However, the failure of the MTCT Bias subscale scores to relate meaningfully to any other measures in this study or the MTCT total score is discouraging, suggesting that this subscale does not measure anything of value. (Contains 3 tables and 39 references.) (SLD)

Critical thinking, belief bias, epistemological assumptions, and the Minnesota Test of Critical Thinking.

Laird R. O. Edman, Jennifer Robey, and William M. Bart

Department of Educational Psychology
The University of Minnesota
Minneapolis, Minnesota

Address all correspondence to:

Laird R. O. Edman
2415 Tamarack Drive
Decorah, IA
52101
edman003@tc.umn.edu or
edman@oneota.net

Paper presented at the 2002 Annual Meeting of the American Educational Research Association
April, 2002
New Orleans, Louisiana

Abstract

The Minnesota Test of Critical Thinking-II (MTCT) has been designed to measure both critical thinking (CT) skills and a key disposition of critical reasoning: the willingness to critically evaluate arguments which are congruent with one's own goals and beliefs. The MTCT uses a taxonomy of CT skills derived from the American Philosophical Association's *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction* (1990). This taxonomy defines critical thinking as "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (p. 3).

Two-hundred thirty two (232) college and university students at three different institutions were administered the MTCT, the Watson-Glaser Critical Thinking Appraisal, the Ennis-Weir Critical Thinking Essay Exam, three subtests of the Multi-Dimensional Aptitude Battery, the Epistemological Questionnaire, and a demographic questionnaire.

The overall Cronbach's alpha for the MTCT with this data is $r\underline{\alpha} = .91$. The subscale reliabilities, using Cronbach's alpha, are: Interpretation: $r\underline{\alpha}=.68$; Analysis: $r\underline{\alpha}=.71$; Inference: $r\underline{\alpha}=.66$; Evaluation: $r\underline{\alpha}=.50$; Explanation: $r\underline{\alpha}=.78$; and Self-Regulation: $r\underline{\alpha}=.71$. Correlations of note are the moderate correlations between the MTCT and the WGCTA ($r=.66$), the Ennis-Weir ($r=.57$), the MAB vocabulary subtest ($r=.61$) and comprehension subtest ($r=.42$), and the Simple Knowledge factor of the EQ ($r=-.51$). Also of note is the virtual lack of significant correlations between the MTCT Bias subscale and any of the other measures in the study, including the EQ subscales.

These data suggest the MTCT may be a useful measure of critical thinking in a variety of college and university settings, with potentially interpretable subscale scores. However, the bias measure on the MTCT failed to yield useful results, and is in need of revision.

Introduction

Over the past two decades, the focus of education has changed from curricular content to curricular outcomes, with a major emphasis on helping students learn to think critically (Edman, 1996; Fisher & Scriven, 1997; Klaczynski, Gordon, & Fauth, 1997; Halpern, 1998; Tucker, 1996). By 1995, most colleges and universities had included critical thinking (CT) skills as an important educational objective in their goal statements, and many accrediting agencies included measurable gains in critical thinking skills into their accreditation criteria (Facione & Facione, 1995).

This emphasis on teaching critical thinking necessarily leads to the need for reliable and valid ways of testing critical thinking. For example, the National League of Nursing has mandated all accredited nursing programs must teach CT to their nursing students and must empirically verify the efficacy of their CT instruction through testing (Rane-Szostak & Robertson, 1996). The assessment of CT is also at the heart of research on CT, for what cannot be measured cannot easily or convincingly be empirically studied. However, the measurement of CT is fraught with difficulty (Ennis, 1993; Tucker, 1996) and has proven to be one of the most difficult aspects of CT research.

Just as in the arena of intelligence testing where there is controversy over definitions, operationalizations, and thus over test construction, so also with CT testing. Because there is no standard definition of CT, the type of test one develops to test for CT depends heavily upon one's definition of the construct. If CT is defined as a set of reasoning competencies, then a measure of those competencies should suffice. However, most theorist and practitioners see CT as more than a set of reasoning competencies. The complex, probably multi-dimensional nature of CT makes simple tests of inductive and deductive logic unsatisfactory.

In order to inform pedagogy, research, and assessment, several CT theorists have proposed taxonomies of CT skills which elaborate the skills and aspects included in the term "critical thinking" (Dick, 1991; Ennis, 1987; Glaser, 1941; Paul, 1993). These taxonomies contain a great deal of overlap in their conceptual presentation of CT, but as of yet there has not been any empirical verification of the elements of CT. However, in 1990 the American Philosophical Association proposed a taxonomy of CT skills which was the result of a two-year Delphi study which included the input of 46 leading theorists and researchers in the field of CT pedagogy and assessment (American Philosophical Association, 1990). This panel defines CT as "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based" (p. 3). The taxonomy of CT skills and subskills devised by this panel has the advantage of the combined expertise of the theorists on the panel, and as such is the most authoritative taxonomy of CT skills available.

The skills and subskills of CT, as delineated by the APA Delphi Study, are:

1. Interpretation
    - Categorization
    - Decoding Significance
    - Clarifying Meaning
2. Analysis
    - Examining Ideas
    - Identifying Arguments
    - Analyzing Arguments
3. Evaluation
    - Assessing Claims

- Assessing Arguments

4. Inference
   - Querying Evidence
   - Conjecturing Alternatives
   - Drawing Conclusions
5. Explanation
   - Stating Results
   - Justifying Procedures
   - Presenting Arguments
6. Self-Regulation
   - Self-Examination
   - Self-Correction

There is widespread theoretical and empirical agreement, however, that critical thinking ability cannot be separated from a person's disposition to use that ability (Facione & Facione, 1995; Halpern, 1998; Klaczynski, Gordon, & Fauth, 1997; Paul, 1993; Perkins, Jay, & Tishman, 1994; Sa, W. C., West, R. F., & Stanovich, K. E., 1999). The relationship between thinking skills and the disposition or propensity to use them has been extensively examined, and several theorists posit that effective critical thinking is a function of two components: the competencies to perform specific cognitive operations, and the metacognitive skill and propensity to evaluate evidence independently of one's own goals and beliefs--to be open minded and objective (Kardash & Scholes, 1996; Klaczynski, Gordon, & Fauth, 1997; Stanovich & West, 1997). It is not enough for the critical thinker to have the skills to use reason when considering ill-defined problems. The critical thinker must also desire to use the skills even in situations in which reasonable reflection may lead to discomfort or difficult decisions on the part of the thinker. That is, the thinker must be willing to use critical thinking skills "against" even her or his own opinions and biases. This is what it means to be intellectually honest or to have intellectual integrity, oft-cited CT dispositional traits (Ennis, 1987; Facione, 1990; Paul, 1993).

If the disposition to use CT skills is an essential component of CT, a test of CT should incorporate assessing this dispositional element into its design. However, the currently available standardized tests of CT measure the construct primarily as a set of reasoning skills divorced from the disposition to use the skills, and they have had only limited success in assessing CT. The current widely used tests of CT have been critiqued as having poor psychometric properties, of relying on limited conceptions of CT, of including confusing or ambiguous questions, and of lacking adequate empirically-based construct validity (Behrens, 1996; Fisher & Scriven, 1997; Follman, 1993; Harris & Clemmons, 1996; Jacobs, 1995; Moss & Koziol, 1991; Rane-Szostak & Robertson, 1996; Tucker, 1996). Many educators and theorists have called for new and better instruments for assessing CT ability (Ennis, 1993; Fisher & Scriven, 1997; Tucker, 1996).

The Minnesota Test of Critical Thinking II (MTCT II) has been designed to measure both CT skills (as proposed by the APA Delphi taxonomy) and a key disposition of critical reasoning: the willingness to critically evaluate arguments that are congruent with one's own goals and beliefs. The purpose of this study is threefold: 1) to examine the reliability (in this research setting with this population) and concurrent validity of the MTCT II with several other measures of critical thinking and general cognitive functioning; 2) to examine the relationship between subjects' critical thinking ability, their general cognitive ability, and their epistemological stance (their beliefs about the nature of knowledge and learning and the development of knowledge); and 3) to examine the approach to measuring belief-bias used in the MTCT II.

Methods

This study involved individual participants completing several different measures in order to examine the relationships among them. These measures included the Minnesota Test of Critical Thinking II (MTCT), the Epistemological Questionnaire (EQ), three subtests of the Multidimensional Aptitude Battery-II (MAB), the Watson-Glaser Critical Thinking Appraisal, Short Form (WGCTA-S), the Ennis-Weir Critical Thinking Essay (EW), and a demographic information sheet. Volunteers for the study were recruited from three different institutions of higher education.

Participants

The participants in this study were 232 students from a variety of academic disciplines and from three different types of post-secondary settings. One group of participants included 70 students who had completed their bachelor's degrees and were enrolled in a teacher-training program at a large Midwestern research university. They were recruited from an educational psychology course and offered extra credit for their participation in this study. This group included 49 females and 21 males with a mean age of 27.18 and ages ranging from 19 to 52. The mean of the ACT scores for this group of students was 25.21

Seventy-seven (77) students were recruited from an undergraduate educational psychology course and an introductory psychology course at a medium-sized, selective, private Midwestern liberal arts college. They were also offered extra credit in the course for their participation. This group included 58 females and 19 males, with a mean age of 19.42 and ages ranging from 18 to 41. The mean of the ACT scores for this group was 24.46.

Eighty-five (85) students were recruited from an undergraduate educational psychology course, an introductory statistics course, and an introductory psychology course at a small, rural, church-affiliated college in the Midwest. They also received extra credit for their participation. Thirty-eight (38) of these subjects were female, 47 male. The mean age of this group was 19.65, with ages ranging from 17 to 34. The mean of their ACT scores was 21.15.

Overall, 145 females and 87 males participated in this study. The mean age over all participants was 21.81 and the mean ACT score for the entire group was 23.64. Nine (9) participants had been U.S. residents for less than 5 years.

Instruments

Background to the Minnesota Test of Critical Thinking

The Minnesota Test of Critical Thinking II (MTCT) is a measure of the critical thinking abilities described by the Delphi Study published by the American Philosophical Association (1990). The APA Delphi study was a long-term, interactive process by which a group of experts from philosophy, education, critical thinking assessment, and a variety of other disciplines formed a consensus opinion about a definition for critical thinking and a list of skills essential to good critical thinking. The Delphi process involves several (in this case, six) rounds of questions to which the participants respond in a thoughtful and detailed way, responding to suggestions and comments made by other participants in earlier rounds to ultimately form a consensus opinion. Their final conceptualization of critical thinking (CT) includes two dimensions: cognitive skills and affective dispositions. Specifically, the group's final Consensus Statement Regarding Critical Thinking and the Ideal Critical Thinker states:

> We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential,

> conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based. CT is essential as a tool of inquiry. As such, CT is a liberating force in education and a powerful resource in one's personal and civic life. While not synonymous with good thinking, CT is a pervasive and self-rectifying human phenomenon. The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results, which are as precise as the subject and the circumstances of inquiry permit. Thus, educating good critical thinkers means working toward this ideal. It combines developing CT skills with nurturing those dispositions which consistently yield useful insights and which are the basis of a rational and democratic society. (APA, 1990, p. 3).

This definition and the six critical thinking skills that emerged from it provided the basis for the development of the MTCT.

Development of the Minnesota Test of Critical Thinking

The version of the MTCT used in this study is the second version of the test. After an initial piloting study, adjustments in the focus and format of the instrument were made. For example, the decision was made to include two points of view for each controversy, rather than the single point of view used in the original version, to allow for maximum presentation of arguments and supporting evidence and to offer an opportunity for rebuttal and reply within the discussions. In addition, a variety of strong and weak arguments were put into the mouths of the discussants to provide an opportunity to examine whether subjects who have strong opinions on one side of the argument are able to detect poor reasoning in the arguments presented by the discussant with whom the subject agrees. (This measure of belief bias, or the tendency to be very critical of arguments with which one disagrees and very favorable toward arguments with which one agrees, is discussed in greater detail below.) This manipulation was incorporated on both sides of the issue under discussion in order to maintain balance in the arguments. Furthermore, it was decided that both discussants should be of the same gender, to prevent subjects from being influenced by the sex of the discussants in their evaluation of the arguments. There are three discussions where the discussants are male and three where the discussants are female. This balancing of gender across the instrument is also a change made after the initial pilot study.

The discussions in the MTCT involve controversies of general interest. Two of the controversies presented in the MTCT deal with issues of concern in education: social promotion and the use of state-provided vouchers to attend private schools. The other 4 controversies concern logging in national forests, the death penalty, the legalization of drugs, and state sponsorship of lotteries.

The process of writing the discussions and items (questions) for each discussion was recursive, and as items evolved so did the discussions to better suit the testing of the particular CT skill being considered. For each discussion, there are two items addressing each of five of the skills described above: Interpretation, Analysis, Evaluation, Inference, and Self-Regulation. The two items in each discussion addressing a particular skill are written in such a way so that one item addresses an error or issue on one side of the controversy at hand, and the other item addresses the other side of the controversy in much the same manner. The remaining skill, Explanation, was measured using an open-ended item of the form "Which of the discussants presented a better argument? Explain the reasons for your choice in three to five sentences." In all, 60 multiple choice items and 6 open-ended response items were used in the MTCT. Five of the skills identified by the Delphi study (Interpretation, Analysis, Evaluation, Inference, and Self-regulation) were assessed by 12 multiple

choice questions each. The remaining skill, Explanation, was assessed by the 6 open-ended questions. The readability analysis of the discussions resulted in a Flesch reading ease score of 62.8, a Flesch-Kincaid grade level of 9.0, and a Bormuth grade level of 10.9 (Bormuth, 1966; Flesch, 1948).

In the version of the test used in this study, participants are first asked to respond to six items indicating their own position on each of the controversial issues covered in the main body of the instrument. These items are followed by the six discussions of controversial issues between two individuals, each of which is accompanied by eleven items that ask the participant to evaluate the arguments presented in the discussion. Ten of the items accompanying each discussion are multiple choice items and one item asks the participants to explain in their own words which of the two individuals who were involved in the discussion presented the better argument and why.

Scoring the Minnesota Test of Critical Thinking

There is one correct answer for each multiple choice question. The maximum score for each CT skill tested using multiple choice items (Interpretation, Analysis, Evaluation, Inference, and Self-regulation) is 12, with a maximum score of 60 for all the multiple choice items. Scoring of the open-ended Explanation items involved examining the responses provided by the participants for the quality of the reasons they supplied for preferring one discussant's arguments to the other's arguments. Each response was compared to *a priori* categories of possible responses to determine whether it would receive a score or 0, 1, or 2. A score of 0 was given if the respondent wrote, "X's argument is more convincing," but did not provide any explanation for his or her preference. A score of 1 was given if the respondent indicated that he or she preferred one of the arguments only, or primarily, because it mirrored his or her own opinion. A score of 1 was also given if the respondent supported his or her preference merely by parroting or repeating the arguments presented by one of the discussants. A score of 2 was given if the reasons offered by the respondent in support of his or her preference for one discussant's position over the other included some mention of issues of source credibility, the presence or absence of supporting evidence used by the discussant, identification of appeals to emotion or *ad hominem* attacks, recognition of bias in the arguments or in the rebuttals offered by the opponent, or some other evidence suggesting that the respondent is aware of and can recognize the correct or incorrect implementation of critical thinking skills. The maximum possible score a participant could receive on the Explanation items is 12. In combination with the multiple choice items, the maximum possible score one could receive on the entire test is 72.

A belief bias score was calculated for each participant. The concept of belief bias suggests that people tend to be better at utilizing the critical thinking skills they possess when faced with arguments that are in opposition to the position they hold on an issue or controversy (Klaczynski, Gordon, & Fauth, 1997). In terms of calculating a belief bias score on the MTCT, this means that a participant is expected to be better at identifying weaknesses and biases in the arguments presented by the discussant who is arguing the case that is in opposition to the participant's point of view. It should be more difficult for the participant to identify the weaknesses and biases present in the argument with which they agree. Therefore, the bias score for each controversy is calculated based on the participant's indication of his or her own position on the issue.

As indicated above, the five skills of Interpretation, Analysis, Evaluation, Inference, and Self-regulation were each tested with two items per controversy, one for each side of the issue. Each of these items was labeled as "Pro" or "Anti" depending on the position of the discussant addressed by that particular item. For example, in the logging controversy, Nicole is arguing that logging should be allowed in National Forests and Sarah is arguing that logging should be banned in the National Forests. Items dealing with the arguments presented by Nicole are therefore labeled as "Pro Logging" and items addressing the arguments presented by Sarah are "Anti Logging" items. In a corresponding

fashion, the opinion items ask that the respondent place him- or herself on a 4-point continuum describing his or her stand on the issue. For the logging issue, if the respondent circled a 1 or 2, he or she is said to be "Pro Logging." A response of 3 or 4 indicates the respondent is "Anti Logging." Since it is assumed that respondents will produce better scores (that is, to identify more thinking errors) on the items pertaining to the point of view with which they disagree, the belief bias score for each controversy was calculated by subtracting the respondent's score on the items pertaining to the position held by the respondent from the scores on the items pertaining to the position opposite the respondent's own. For example, if an individual indicated that she is "Anti Logging" on the opinion item, the "Anti Logging" items were subtracted from the "Pro Logging" items. This is the process used to calculate bias scores for all six controversies. A "Total Bias" score was calculated by adding the bias scores from the six controversies together. Using this method, a positive belief bias score indicates that the participant received a better score on those items related to positions with which he or she disagrees with than on the items with which he or she agrees, evidence that the participant is biased in his or her application of critical thinking skills. A belief bias score near zero indicates that the participant is equally good at analyzing arguments on both sides of a controversial issue. A negative belief bias score suggests that the participant was better able to correctly identify thinking errors in the arguments with which he or she agrees than in arguments with which he or she disagrees, a result that is problematic in the present theoretical context.

The Epistemological Questionnaire

The Epistemological Questionnaire (EQ) is the instrument developed by Marlene Schommer (1998) to examine epistemological beliefs in adults. It consists of 63 statements that address various aspects of personal epistemology such as knowledge is certain, success is unrelated to hard work, individuals can learn how to learn, the ability to learn is innate, the process of learning is quick, learning occurs with the first effort, and concentrated effort is a waste of time. The EQ also measures aspects of learner behavior such as the learner should avoid integrating material, seek single answers, avoid ambiguity, depend on authority, and avoid criticizing authority. The participant is to indicate his or her level of agreement with each of the 63 items on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). Four factors have consistently emerged from repeated testing using the EQ: (1) Fixed Ability (the malleability of learning, ranging from the belief that ability to learn is fixed at birth to the belief that the ability to learn can be improved), (2) Simple Knowledge (the structure of knowledge, ranging from the belief that knowledge is best characterized as isolated bits and pieces to the belief that knowledge is best characterized as complex interrelated networks), (2) Quick Learning (speed of learning, ranging from the belief that learning is quick or not-at-all to the belief that learning is gradual), and (4) Certain Knowledge (the stability of knowledge, ranging from the belief that knowledge is unchanging to the belief that knowledge is evolving) (Schommer, 1998; Schommer, 1990; Schommer, 1993; Schommer, Calvert, Gariglietti, & Bajaj, 1997; Schommer & Hutter, 1995).

The Multidimensional Aptitude Battery

The Multidimensional Aptitude Battery-II (MAB) is a series of paper-and-pencil tests that were developed by Douglas N. Jackson to provide a means for testing large groups of people on the skills tested on an individual basis by the Wechsler Adult Intelligence Scale (Jackson, 1984). Participants in this study took three of the MAB tests: the spatial relations test, the comprehension test, and the vocabulary test. These subscales were chosen because they represent the tests with the highest factor loadings on the Verbal IQ (vocabulary and comprehension) and Performance IQ (spatial) subscales of the MAB, as well as with the overall, Full Scale IQ MAB scores. The spatial relations section has 50 items, the comprehension section has 28 items, and the vocabulary section has 46 items. Participants are given seven minutes to complete each test. The reported reliabilities for these tests for 20 year-olds

are as follows: (1) Spatial relations - .96; (2) Comprehension - .90; and (3) Vocabulary - .88. The reported test-retest reliability for the spatial relations test is .93, .95 for the comprehension test, and .90 for the vocabulary test. For each scale, the score used in all analyses reflects the total number correct on that scale.

The Watson Glaser Critical Thinking Appraisal--Form S

The Watson-Glaser Critical Thinking Appraisal--Form S (Watson & Glaser, 1994) is a shortened version of the Watson Glaser Critical Thinking Appraisal--Form A (Watson & Glaser, 1980). It consists of 16 scenarios and 40 items selected from the 80-item Form A. Examinees are given 30 minutes to complete the test. The measure is comprised of five subtests covering Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. In subtest I, the ability to judge inferences made from a statement of facts is tested. There are five judgments from which to choose: true, probably true, insufficient data, probably false, or false. Subtest II provides five statements followed by a number of proposed assumptions. Examinees are to judge whether a person making a given statement is also making the proposed assumptions. Each assumption is to be judged independently from the others. In subtest III, examinees are to judge whether conclusions follow necessarily from given statements. Subtest IV tests for interpretation abilities. Examinees are to judge whether each of the proposed conclusions follows beyond a reasonable doubt from the information. Finally, in subtest V, examinees are to judge whether given arguments are strong or weak. The manual for the WGCTA-S cautions users against the use of the subtest scores due to their low reliability. It is the overall score that is claimed to be more reliable across a variety of settings. Cronbach's alpha reliability coefficients ($r\underline{\alpha}$, Cronbach, 1970) for the WGCTA-S have ranged from $r\underline{\alpha}$ =.66 to $r\underline{\alpha}$ =.87 across 21 different samples, with an $r\underline{\alpha}$ of .81 for the development sample (N=1,608).

The Ennis-Weir Critical Thinking Essay Test

The Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985) is intended to evaluate the ability to appraise an argument and to formulate a written argument in response. The test begins by asking examinees to read a letter to the editor of a fictional newspaper. In the letter, a proposal is made to end overnight parking on city streets, and a variety of arguments are offered in support of the proposal. Examinees are asked to write a letter evaluating the arguments in each paragraph and in the letter as a whole. The scoring system for each paragraph response is as follows: -1 for judging an argument incorrectly and/or showing bad judgment in justifying; 0 if no response is made; +1 if the argument is judged correctly but not justified; +2 if justified semi adequately; and +3 if justified adequately.

Procedures

Part of the data was collected in supervised testing sessions (including all of the timed tests and demographic questionnaires) and the remainder of the instruments were completed by the participants on their own time and returned to the investigator. Sixty-seven (67) participants completed the entire test packet in the testing sessions. Total testing time was approximately 3 hours.

Results

Means, standard deviations, and score ranges can be found in Table 1. The overall Cronbach's alpha for the MTCT with this data is $r\underline{\alpha}$ = .91. Guttmann Split-half reliability for the MTCT is .85. The subscale reliabilities, using Cronbach's alpha, are: Interpretation: $r\underline{\alpha}$=.68; Analysis: $r\underline{\alpha}$=.71; Inference: $r\underline{\alpha}$=.66; Evaluation: $r\underline{\alpha}$=.50; Explanation: $r\underline{\alpha}$=.78; and Self-Regulation: $r\underline{\alpha}$=.71.

Tables 2 and 3 contain correlation matrices of scores on the measures used in this study. Correlations of note are the moderate correlations between the MTCT and the WGCTA (r=.66), the Ennis-Weir (r=.57), the MAB vocabulary subtest (r=.61) and comprehension subtest (r=.42), and the

Simple Knowledge factor of the EQ (r=-.51). Also of note is the apparent lack of interpretable relationship among the MTCT Bias subscale and any of the other measures in the study, including the EQ subscales.

Unweighted least squares, unrotated factor analysis of the MTCT failed to reproduce the subscale structure. The first factor to emerge from the analysis, with an eigen value of 11.93, accounts for 18% of the variance in scores. The remaining factors are not interpretable.

Discussion

Assessing critical thinking is an important, and in some cases, high stakes undertaking. As more and more secondary and post-secondary institutions look to teaching and testing the critical thinking skills of students, assessments that are reliable in many settings and that do indeed accurately test critical thinking skills are needed. The results of this study indicate that the Minnesota Test of Critical Thinking may be a valuable instrument for assessing critical thinking skills.

The reliability of the MTCT in this setting, with these subjects, is pleasantly high. Reported reliabilities of CT tests tend to be low (Ennis & Norris, 1990; Loo & Thorpe, 1999; Norris, 1995; Norris & Ennis, 1989; Watson & Glaser, 1994), with Cronbach's Alpha scores ranging from .37 to .87. The Cronbach's Alpha scores reported for the MTCT in this study are higher than those reported for most other major tests of CT in the research literature. Obviously, further research on the reliability of the MTCT is required for us to know more about how the test functions with different populations and in different settings. However, the reliability indices of these data are promising.

Also promising are the moderately high reliability indices for the subscale scores. The authors of the Watson-Glaser Critical Thinking Appraisal caution test users against interpreting individual subscale scores on the WGCTA because of the instability of such scores (Watson & Glaser, 1994). The MTCT, in this setting at least, appears to have much more stable subscale scores. This may allow for subscale scores that lend themselves to more detailed, diagnostic interpretation. A great deal more research on the correlates and predictive power of the subscales is required, however, before more can be said of their usefulness.

The correlations of the scores on the MTCT with the scores on the three subtests of the MAB, the WGCTA-S, the Ennis-Weir, and ACT scores are in the ranges hypothesized, and support the concurrent validity of the test. That the MTCT correlates most highly with the WGCTA-S is exactly as hypothesized, since both tests purport to test critical thinking abilities and both use multiple-choice methodology to do so. That the MTCT correlates more highly with the vocabulary subtest of the MAB than it does with the Ennis-Weir is slightly problematic. However, that CT skills are correlated with general intellectual functioning is assumed, and given the verbal nature of the MTCT, a moderate correlation with verbal intelligence would also be assumed. The difference in the correlations between the MTCT, the Ennis-Weir, and the MAB-vocabulary subtest are so small as to be statistically insignificant.

The correlations of the MTCT with the factors of the EQ are encouraging. There is good evidence that epistemological beliefs are related to what are commonly referred to as critical thinking dispositions and skills. Epistemological beliefs may influence how students interpret information (Schommer, 1990), comprehend written text (Kardash & Scholes, 1996, Schommer, 1990), monitor their comprehension (an important meta-cognitive skill related to CT) (Ryan, 1984), and persist in the face of a difficult task (Dweck & Leggett, 1988). As mentioned above, since CT skills include essential dispositional elements, a valid measure of CT should tap those dispositional elements as well as reasoning competencies. The correlations of the MTCT with the 4 factors of the EQ indicate the MTCT is touching on something in common with the EQ, and doing so in a more effective way than

either the WGCTA-S or the Ennis-Weir.

However, the complete failure of the MTCT Bias subscale scores to meaningfully relate to any other measures in this study, or even with the MTCT total score, is discouraging. This subscale does not seem to be effectively measuring anything of value. This leads the authors of the test to conclude the methodology of measuring belief bias used in the subscale is not effective. It may be that subjects were not sufficiently opinionated about the topics in the test for belief bias to be a factor; or it may be that a number of the items are not clearly "pro" or "anti" items—that is, they are not designed well enough so that bias would be likely to influence an examinee's answer. Whatever the reason, clearly the bias subscale needs to be recast, or the items need to be refocused. The issue of belief bias and objectivity of judgment is an important issue, and one that needs to be further refined and tested.

As mentioned above, least-squares factor analysis failed to reproduce the subscale structure of the instrument, an expected result that is consistent with other research on tests of CT instruments (Ennis & Norris, 1990; Loo & Thorpe, 1999; Norris, 1995; Norris & Ennis, 1989; Tucker, 1996). The analysis did seem to reveal a general factor that accounted for a significant amount of score variance. This is a result that is not surprising, and is consistent with Norris (1995) who in a study that examined 15 different CT measures also posits a general critical thinking factor emerging from the results of confirmatory factor analysis. The underlying structure of CT is still an issue of debate, and the Delphi structure has not been verified in any study. This supports the conclusions above that further research on the subscales is needed, as well as further analysis of the structure and components of CT.

In a world of accelerating change and ever increasing amounts of easily available information, the need for people to develop good critical thinking skills is profound. It is thus a primary educational imperative to teach students to become better critical thinkers. Testing critical thinking skills is a difficult task, however. Issues which need to be resolved include the extent to which critical thinking is a domain--specific or general competency, and the extent to which critical thinking comprises discrete skills (such as identifying assumptions, evaluating credibility, deduction, induction, and metacognitive elements such as self-monitoring and self-awareness of cognitive strategies) which can be taught and tested individually or interdependent aspects of a complex concept that cannot be disassembled without altering its nature (Moss & Koziol, 1991). It is also essential that we understand the nature and importance of critical thinking dispositions to the teaching, assessment, and practice of critical thinking. These issues require us to engage much more fully in research on CT. The initial evidence indicates that the MTCT may be a useful instrument for testing CT abilities and researching the questions raised above.

References

American Philosophical Association. (1990). 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. (ERIC Document Reproduction Service No. ED 315-423).

Behrens, P. (1996). The Watson-Glaser Critical Thinking Appraisal and academic performance of diploma school students. Journal of Nursing Education, 35, 34-36.

Bormuth, J. R. (1966). Readability: A new approach. Reading Research Quarterly, 1, 79-132.

Dick, R. D. (1991). An empirical taxonomy of critical thinking. Journal of Instructional Psychology, 18, 79-92.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. Psychological Review, 95, 256-273.

Edman, L. (1996). Teaching teachers to teach thinking. The National Honors Report, 16, 8-12.

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

Ennis, R. H. (1993). Critical thinking assessment. Theory into Practice, 32, 179-186.

Ennis, R. H., & Norris, S. P. (1990). Critical thinking assessment: Status, issues, needs. In S. Legg & J. Algina (Eds.), Cognitive assessment of language and math outcomes (pp. 1-42). Norwood, NJ: Ablex.

Ennis, R. H., & Weir, E. (1985). The Ennis-Weir critical thinking essay test. Pacific Grove, CA: Midwest.

Facione, P. A. (1990). The California Critical Thinking Skills Test (CCTST): Forms A and B; and the CCTST Test manual. Millbrae, CA: California Academic Press.

Facione, P.A. & Facione, N. (1995). The disposition toward critical thinking. The Journal of General Education, 44, 25-50.

Fisher, A, & Scriven, M. (1997). Critical thinking: Its definition and assessment. Point Reyes, CA: Edgepress.

Flesch, R. (1948). A new readability yardstick. Journal of Applied Psychology, 32, 221-233.

Follman, J. (1993). Critical thinking and verbal ability. ACEHI Journal, 19, 71-76.

Glaser, E.M. (1941). An experiment in the development of critical thinking. AMS reprint, New York: AMS Press.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. American Psychologist, 53, 449-455.

Harris, J. C., & Clemmons, S. (1996). Utilization of standardized critical thinking tests with developmental freshman. Paper presented at the National Conference on Research in Developmental Education, October, 1996.

Jacobs, S. S. (1995). Technical characteristics and some correlates of the California Critical Thinking Skills Test Forms A and B. Research in Higher Education, 36 89-108.

Jackson, D. N. (1984). The Multidimensional Aptitude Battery. Port Huron, MI: Sigma Assessment Systems, Inc.

Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. Journal of Educational Psychology, 88, 260-71.

Klaczynski, P., Gordon, D., & Fauth, J. (1997). Goal-oriented critical reasoning and individual

differences in critical reasoning biases. Journal of Educational Psychology, 89, 470-485.

Loo, R., & Thorpe, K. (1999). A psychometric investigation of scores on the Watson-Glaser Critical Thinking Appraisal new form S. Educational and Psychological Measurement, 59, 995-1003.

Moss, P. A., & Koziol, S. M. (1991). Investigating the validity of a locally developed critical thinking test. Educational Measurement: Issues and Practices, 17-22.

Norris, S. P. (1995). Format effects on critical thinking test performance. The Alberta Journal of Educational Research, 41, 378-406.

Norris, S. P., & Ennis, R. H. (1989). Evaluating Critical Thinking. Pacific Grove, CA: Critical Thinking Press and Software.

Paul, R. (1993). Critical thinking: What every person needs to know to survive in a rapidly changing world. 3rd ed. Santa Rosa, CA: Foundation for Critical Thinking.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. Merrill-Palmer Quarterly, 39, 1-21.

Rane-Szostak, D. & Robertson, J.F. (1996). Issues in measuring critical thinking: meeting the challenge. Journal of Nursing Education, 35, 5-11.

Ryan, M. P. (1984). Monitoring text comprehension: Individual differences in epistemological standards. Journal of Educational Psychology, 76, 248-258.

Sa, W. C., Stanovich, K. E., & West, R. F. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. Journal of Educational Psychology, 91, 497-510.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. Journal of Educational Psychology, 82, 498-504.

Schommer, M. (1998). The influence of age and education on epistemological beliefs. British Journal of Educational Psychology, 68, 551-562.

Schommer, M. (1993). Epistemological development and academic performance among secondary students. Journal of Educational Psychology, 85, 406-411.

Schommer, M., Calvert, C., Gariglietti, G., & Bajaj, A. (1997). Journal of Educational Psychology, 89, 37-40.

Schommer, M. & Hutter, R. (1995). The relationship between epistemological beliefs and controversial day-to-day issues. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. Journal of Educational Psychology, 89, 342-357.

Tucker, R. W. (1996). Less than critical thinking. Adult Assessment Forum Archives, 6, 1-6

Watson, G. B., & Glaser, E. M. (1994). Watson-Glaser critical thinking appraisal Form S manual. San Antonio, TX: Harcourt Brace.

Table 1: Means, Standard Deviations, Minimum scores, and Maximum scores

	N	Mean	Std. Deviation	Minimum	Maximum
MTCT-II	210	40.29	12.14	9	64
WGCTA-S	226	27.38	5.97	14	39
Ennis-Weir CT Essay	203	16.39	6.94	0	28
MAB vocabulary	232	23.07	7.51	9	44
MAB comprehension	232	20.62	3.77	5	28
MAB spatial	232	31	10.81	3	50
EQ--Fixed ability	222	2.33	.42	1.18	3.76
EQ--Simple Knowledge	222	2.86	.37	1.81	3.79
EQ--Certain Knowledge	222	2.16	.40	1.10	3.53
EQ--Quick Learning	222	2.59	.57	1	4
ACT	184	23.64	4.25	14	34

Table 2: Correlations among the MTCT II, the WGCTA-S, the Ennis-Weir, MAB, and ACT

		MTCT II	WGCTA	Ennis-Weir	MAB vocab	MAB comp	MAB spatial	ACT
MTCT II	Pearson Correlation	1.000	.659	.573	.609	.422	.210	.692
	Sig. (2-tailed)	.	.000	.000	.000	.000	.002	.000
	N	210	202	200	209	209	209	167
WGCTA	Pearson Correlation	.659	1.000	.483	.555	.441	.201	.649
	Sig. (2-tailed)	.000	.	.000	.000	.000	.003	.000
	N	202	226	203	224	224	224	179
Ennis-Weir	Pearson Correlation	.573	.483	1.000	.405	.354	.161	.516
	Sig. (2-tailed)	.000	.000	.	.000	.000	.022	.000
	N	200	203	203	202	202	202	161
MAB vocab	Pearson Correlation	.609	.555	.405	1.000	.550	.242	.678
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000
	N	209	224	202	232	232	232	184
MAB comp	Pearson Correlation	.422	.441	.354	.550	1.000	.204	.530
	Sig. (2-tailed)	.000	.000	.000	.000	.	.002	.000
	N	209	224	202	232	232	232	184
MAB spatial	Pearson Correlation	.210	.201	.161	.242	.204	1.000	.374
	Sig. (2-tailed)	.002	.003	.022	.000	.002	.	.000
	N	209	224	202	232	232	232	184
ACT	Pearson Correlation	.692	.649	.516	.678	.530	.374	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.
	N	167	179	161	184	184	184	184

Table 3: Correlations.

		MTCT II	WGCTA	Ennis-Weir	Fixed Ability	Simple Knowledge	Certain Knowledge	Quick Learning	Bias total score
MTCT II	Pearson's r	1.000	.659	.573	-.178	-.513	-.340	-.146	.076
	Sig. (2-tailed)	.	.000	.000	.010	.000	.000	.034	.287
	N	210	202	200	210	210	210	210	196
WGCTA	Pearson's r	.659	1.000	.483	-.099	-.437	-.234	-.134	.033
	Sig. (2-tailed)	.000	.	.000	.150	.000	.001	.051	.657
	N	202	226	203	214	214	214	214	188
Ennis-Weir	Pearson's r	.573	.483	1.000	-.077	-.352	-.207	-.118	-.003
	Sig. (2-tailed)	.000	.000	.	.277	.000	.003	.094	.969
	N	200	203	203	203	203	203	203	186
Fixed Ability	Pearson's r	-.178	-.099	-.077	1.000	.281	.616	.122	-.090
	Sig. (2-tailed)	.010	.150	.277	.	.000	.000	.071	.207
	N	210	214	203	222	222	222	222	196
Simple Knowledge	Pearson's r	-.513	-.437	-.352	.281	1.000	.431	.160	-.086
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.017	.233
	N	210	214	203	222	222	222	222	196
Certain Knowledge	Pearson's r	-.340	-.234	-.207	.616	.431	1.000	.253	-.164
	Sig. (2-tailed)	.000	.001	.003	.000	.000	.	.000	.021
	N	210	214	203	222	222	222	222	196
Quick Learning	Pearson's r	-.146	-.134	-.118	.122	.160	.253	1.000	-.127
	Sig. (2-tailed)	.034	.051	.094	.071	.017	.000	.	.076
	N	210	214	203	222	222	222	222	196
Bias total score	Pearson's r	.076	.033	-.003	-.090	-.086	-.164	-.127	1.000
	Sig. (2-tailed)	.287	.657	.969	.207	.233	.021	.076	.
	N	196	188	186	196	196	196	196	196

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

TM033831

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Critical thinking, belief bias, epistemological assumptions, and the Minnesota Test of Critical Thinking

Author(s): Laird R.O. Edman, Jennifer Robey, & William M. Bart

Corporate Source: University of Minnesota

Publication Date: —

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1 ↑ ☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A ↑ ☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B ↑ ☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature:	Printed Name/Position/Title: Laird R.O. Edman
Organization/Address: 2415 Tamarack Dr, Decorah, IA 52101	Telephone: 563-387-0068 FAX: [illegible]
	E-Mail Address: edman@oneota.net Date: 4/3/02

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: info@ericfac.piccard.csc.com
WWW: http://ericfacility.org