

# **Project CAT: Assessing Critical Thinking Skills**

## **Final Report**

**Tennessee Technological University**



Barry S. Stein, Principal Investigator  
Ada F. Haynes, Co-Principal Investigator  
J. Michael Redding, Co-Principal Investigator



---

Partial support for this work was provided by the National Science Foundation's CCLI Program under grant 0404911.

# **Project CAT**

## **2007 -2008 Activities & Findings During No Cost Extension**

### **Tennessee Technological University**

#### **Overview of Recent Activities**

##### **Testing & Scoring**

The CAT instrument was administered and scored for a second time at the University of Hawaii – Hilo. This institution was the first institution to participate in the project and we were interested in finding out if the improvements to the test and scoring guide resolved issues such as cultural fairness that had been previously identified. We also wanted to see if improvements to the scoring guide increased scoring reliability since the first scoring session. In addition, several outreach and dissemination efforts were undertaken that involved testing students at other institutions. Eastern Kentucky University requested our permission to use the CAT instrument as part of their QEP evaluation plan and we helped them conduct their first scoring workshop. Madisonville Community College continued to use the CAT instrument to evaluate their students as part of their Quality Enhancement Plan. They became the first institution to conduct their own scoring session, and this provided useful information about other institution's capability to conduct reliable scoring workshops on their own.

We also conducted a scoring session at TTU with experienced graders to rescore a sample of tests that were originally scored at other institutions in order to evaluate the accuracy and consistency of scoring at other institutions. A random sample of twenty tests from each institution were rescored at TTU. This rescoring included tests scored at two separate Madisonville Community College scoring sessions (one session led by us and the other led by their own team).

##### **Evaluation**

The PI and one Co-PI met with our external consultant, Dr. John Bransford, to discuss project findings, question weighting, and strategies for dissemination. Several additional statistical analyses were suggested that provided useful information pertaining to question weighting and test validity.

##### **Dissemination**

Information about the project was presented at a concurrent session of the annual SACS/COC meeting in Atlanta in December of 2007. This session was well attended by approximately 200 university administrators/faculty and received excellent evaluations. A proposal for a preconference workshop at the December 2008 annual meeting that would give 35 participants more in-depth knowledge of the CAT instrument with hands-on experience was accepted by the SACS/COC program committee. We will also be presenting a poster at the August 2008 NSF CCLI PI conference. Various other types of dissemination activities are planned as part of the new national dissemination grant.

Our website was updated to better reflect the project accomplishments and usefulness of the test and has 5753 hits as of May 21, 2008 ([www.tntech.edu/cat](http://www.tntech.edu/cat)). We have received numerous requests for information based on visits to this website.

We also continue to receive requests to use the CAT instrument from participants in our May 2007 dissemination conference held at TTU.

### **Instrument Refinement**

The CAT test and scoring guide have now been finalized for use by other institutions. We have used a variety of information to refine the CAT instrument including faculty ratings and comments, statistical analyses of reliability and validity, input from our external evaluators, and input from our external consultant/learning sciences expert. The various validity measures we have obtained for the CAT instrument indicate that it is an excellent tool for evaluating critical thinking/real-world problem solving.

### **National Dissemination Grant**

In January of 2007 we submitted a Phase III proposal to NSF (CCLI) for national dissemination of the CAT Instrument. That grant will focus on the development of regional train-the-trainer workshops to prepare representatives from 20 other institutions to lead CAT scoring workshops on their own campus and the collection of national user norms. That proposal has been funded by NSF and we have held our first train-the-trainer workshop. We have had to add additional self-supporting workshop sessions to our schedule to accommodate the wide-spread institutional interest in using the CAT instrument.

## **Project Findings**

### **General Interest in the CAT Instrument**

Interest in the CAT Instrument continues to be strong with many institutions wanting to know when the instrument will be available for general institutional use. Our success with Madisonville Community College has expanded interest among community colleges for using the instrument. The findings indicate that the test is appropriate for community college students, but that separate user norms will be needed for community college institutions to evaluate overall performance.

### **A Comparison of relationships between the NSSE and the CAT versus CCTST**

At the suggestion of our external consultant, John Bransford, we examined the relationship between relevant items on the National Survey of Student Engagement and Performance on the CAT instrument versus that between the NSSE and the California Critical Thinking Skills Test (CCTST). Table 1 shows that the CAT instrument has a strong negative relationship with an item on the NSSE that reflects the extent to which students thought their classes emphasized rote retention. The CCTST score also had a negative relationship with courses emphasizing rote retention, but it was smaller in magnitude than that found with the CAT instrument. In addition, the CAT instrument had a significant positive relationship with other areas of the NSSE that relate to critical thinking. In contrast, the CCTST was not significantly related to these other areas.

Table 1

NSSE Question	CAT Coefficient	CCTST Coefficient
(2a) Memorizing facts, ideas, or methods from your courses and readings so you can repeat them in pretty much the same form. (negative relationship)	-.341 *	-.212*
(3b) Number of books read on your own (not assigned) for personal enjoyment or academic enrichment.	.277 *	-.124
(11e) Thinking critically and analytically & (11m) Solving complex real-world problems	.244 *	.144
(7h) Culminating Senior Experience (thesis, capstone course, project, comprehensive exam, etc.)	.231 *	.025

\* Significant at .01 level

These findings provide additional support for the validity of the CAT instrument and demonstrate its relationship to other indirect measures that reflect student’s perceptions of college experiences that should impact their critical thinking. The findings also indicate that the CAT instrument may provide a better measure of critical thinking than the CCTST.

**Findings From Retesting at the University of Hawaii**

Informal discussions with faculty during the scoring session indicate that we had corrected any problems that might affect the cultural fairness of the test that been raised during the first test scoring session there three years earlier. Faculty were also impressed by the substantial improvements made to the scoring guide. Those improvements probably contributed to the increase in average scoring reliability observed between the first and second scorers. The improvement in scoring reliability is shown in Table 2.

Table 2  
Average Scoring Reliability in Hawaii

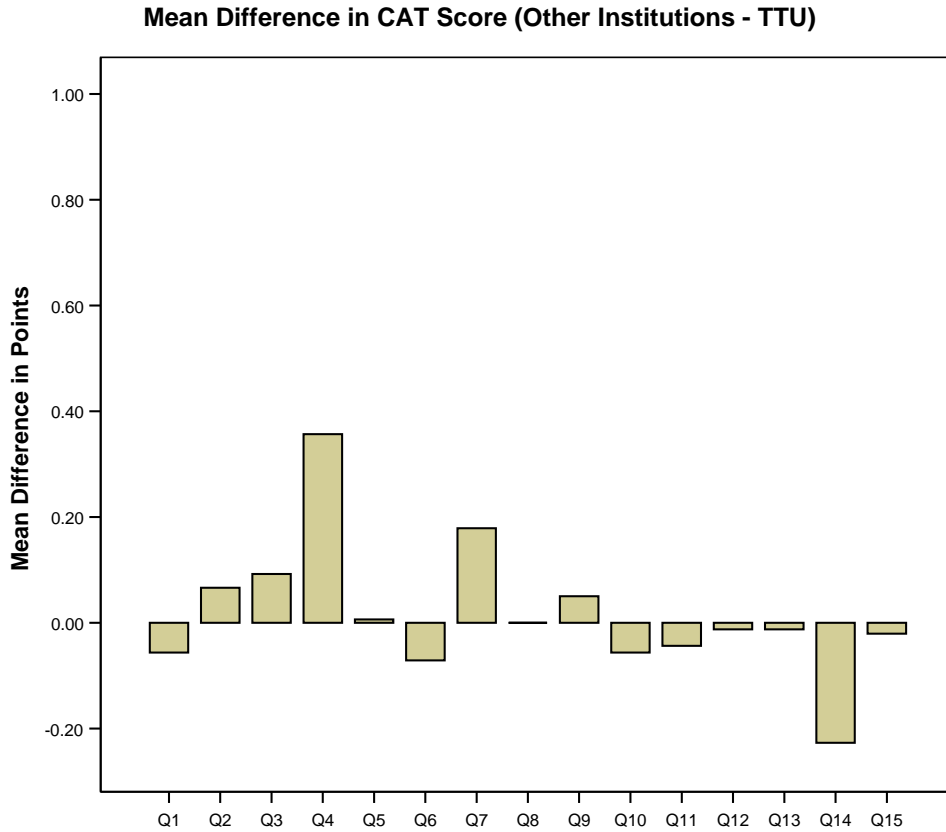
Original Scoring Session	New Scoring Session
.795	0.848

**Accuracy of Scoring at Other Institutions**

A random sample of twenty tests from each of the 7 institutions were rescored at TTU. Overall, the results were quite encouraging. The overall difference in total CAT score assigned to students by other institutions versus TTU scorers was 0.25 points. This is a remarkably small difference in total assigned score when the average score for this sample was about 18 points (individual test scores ranged from 5 to 32).

We also broke down the analysis of scoring consistency by question. Figure 1 shows the average difference by question between tests scored at other institutions and those same tests scored at TTU by experienced graders. In this analysis we see that in the worst case, one question (#4) was scored 0.4 points too high compared to TTU’s experienced graders. Further examination of the data revealed that two institutions were primarily responsible for this question’s scoring difference and these institutions probably misunderstood the scoring rubric for that question. This type of analysis should provide extremely useful information about which questions may need further explanation to assure better scoring at each institution. We plan to conduct this type of analysis for each institution using the CAT to provide corrective feedback.

Figure 1  
Overall Difference in Mean Scores given by Institutions and Follow up TTU Scoring



In general, errors tended to equalize when summed across questions and averaged across the twenty tests sampled from each institution.

**Madisonville Community College Scoring Sessions**

This institution was the first to conduct their own scoring session after having invited us to lead a scoring session at their institution last year. We were particularly interested in whether this institution could lead a scoring workshop themselves after watching members of the project team lead a workshop on their campus. The qualitative reports we received about their own scoring session were very encouraging. We also met with an interdisciplinary group of faculty from that institution to discuss ways of using the CAT instrument as a model for building active learning tasks that would improve students’ critical thinking.

We also examined a random sample of 20 tests from each scoring session to evaluate scoring accuracy. The average test score in these samples was about 15 with a range of 5 to 28. The results of this analysis were also very encouraging. The overall mean score in both sessions differed by less than 0.33 points from the score assigned by experienced TTU scorers. Individual question scores (see figures 2 and 3) differed 0.2 points or less from TTU scores in both sessions. These findings provide strong support for the idea that other institutions can learn to lead reliable scoring workshops at their own institutions after training by our project team.

Figure 2  
Scoring Accuracy I

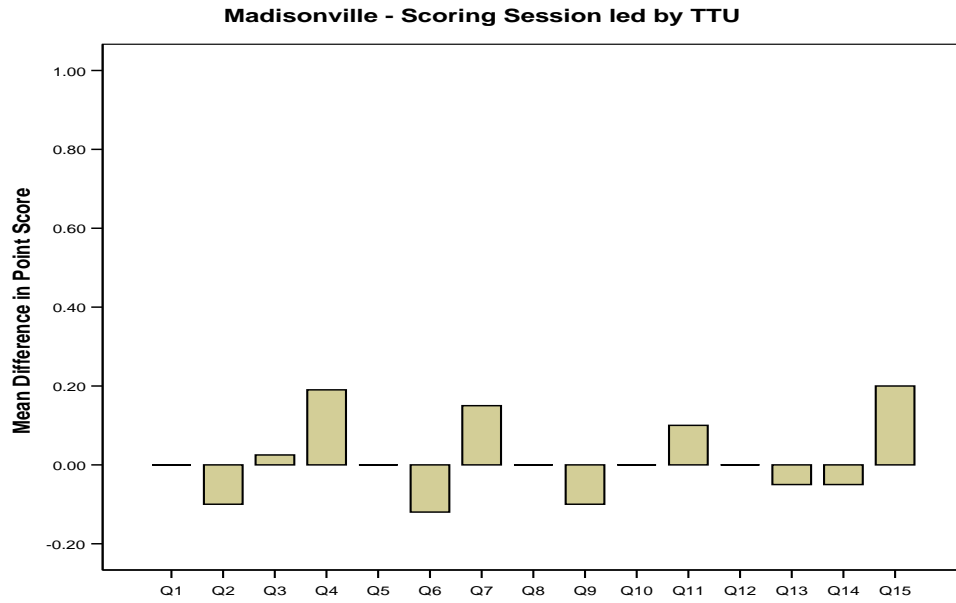
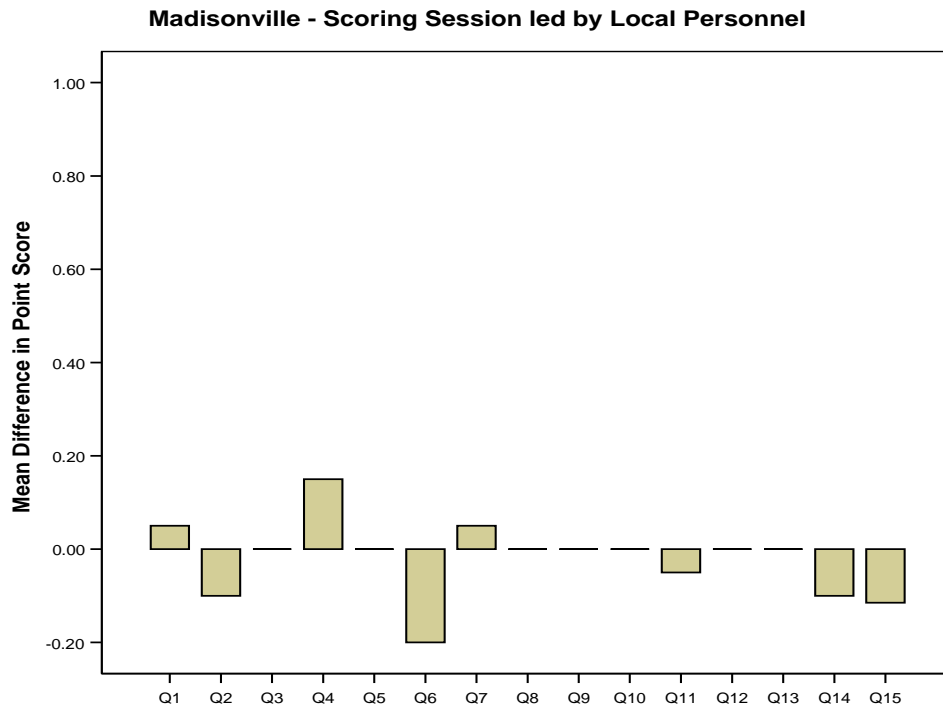


Figure 3  
Scoring Accuracy II



## Summary

The project has met or exceeded all goals. The previous annual reports have described in detail the specific accomplishments in a cumulative fashion. The CAT instrument and scoring guide have been tested and refined with input from faculty and experts across the country. The instrument has been validated in many ways and the reliability of scoring is quite good for this type of test. The instrument is particularly well suited for engaging faculty in improvement initiatives since it gives faculty first hand experience with student weaknesses in critical thinking. The test questions also provide excellent examples of real-world problems that faculty can use as models to develop their own discipline specific analogs to engage students in active learning that will improve their critical thinking. The instrument should be of great value to institutions, programs and researchers interested in assessing and improving critical thinking and real-world problem solving skills.

Our next challenge that is currently being pursued through our national dissemination grant is to develop effective and efficient methods for training other institutions to conduct their own scoring workshops using the CAT instrument. These methods must also insure the accuracy and reliability of test scoring. The current report describes a procedure we have developed for evaluating the accuracy of scoring at other institutions and that will be used to evaluate the effectiveness of our national dissemination efforts.