

Conference on Quality in Higher Education



A Collection of Papers on Self-Study
and Institutional Improvement, 2011



Higher
Learning
Commission

A commission of the
North Central Association

2011



Getting Faculty Involved in Assessing and Improving Students' Critical Thinking

Barry Stein, Ada Haynes, and Denise Drane

Although the acquisition of factual knowledge is important, most educators would argue that an undergraduate education should prepare individuals to do more than recite facts, that it should prepare students to think critically, solve problems, and effectively communicate ideas. Derek Bok (2006), president emeritus of Harvard University, notes that “ninety-five percent of all American faculty in American universities believe that developing the powers of critical thinking of their students is not just a, but *the*, most important objective of a college education.” Consequently, it is not surprising that many colleges and universities are looking for better ways to evaluate students' critical thinking in response to state and national mandates for accountability.

Tennessee Tech University (TTU), along with a broad range of institutions across the country, has been exploring existing methods and developing new tools to assess critical thinking skills for the past ten years with support from the National Science Foundation (NSF). These efforts have focused on (1) developing an instrument to assess critical thinking that has high validity for both faculty and experts in learning science, (2) creating an assessment tool that involves faculty in the scoring process to help them understand student weaknesses, and (3) creating an assessment system that supports broad-based efforts to improve student learning. The success of the project has culminated in the national dissemination of the Critical Thinking Assessment Test (CAT). More than fifty institutions across the country are now using this instrument, and it has become an integral component of many institutional efforts to improve student learning.

Historical Context

Eleven years ago TTU began the process of exploring methods to evaluate critical thinking skills as part of a statewide higher education mandate. We explored a variety of tests that were being marketed for that purpose. We searched for a test that would have high face validity in the eyes of our faculty and that would directly involve our faculty in the scoring, since we believed that any initiative to evaluate critical thinking might ultimately require efforts to improve critical thinking performance. Our search for an appropriate assessment tool revealed that the options were fairly limited. We chose a test that involved short-answer essay questions that could be graded by our own faculty. Although selecting an essay test that must be scored by one's own faculty might seem like an unnecessary burden, it has the advantage of allowing faculty to directly experience student weaknesses. This type of faculty engagement is particularly important if an institution is interested in improving student learning.

Our experiences with that test (Tasks in Critical Thinking) were encouraging, although our subsequent analysis of results showed poor criterion validity when related to students' entering ACT scores. Our faculty also had some concerns about the face validity of the instrument and the reliability of the scoring. These concerns coupled with the subsequent removal of the test from the market necessitated an alternative approach. We decided to pursue a somewhat radical and ambitious alternative—the development of a



new test to assess critical thinking skills. Our goal was to assess what faculty considered to be important and relevant critical thinking skills. It was acknowledged from the outset that this test would probably not evaluate all aspects of critical thinking but would instead focus on a core set of critical thinking skills relevant to most disciplines and contributing to successful careers in most fields.

Faculty members worked in teams and as members of a larger group to identify important critical thinking skills and develop questions and materials that could measure those skills. The test would involve mostly essay answers to help assess communication skills and leave opportunities for creative answers to questions that do not always have a single correct response. The essay format would also involve faculty in the scoring of exams and hence promote more interest in improving critical thinking skills. In addition, the test would be based on topics that would be interesting and engaging for students to help motivate them to perform to their maximum potential. The skills assessed by the CAT instrument were as follows:

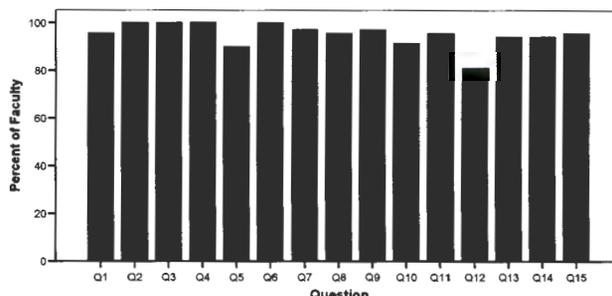
- Evaluating information
 - Separate factual information from inferences.
 - Interpret numerical relationships in graphs.
 - Understand the limitations of correlational data.
 - Evaluate evidence and identify inappropriate conclusions.
- Creative thinking
 - Identify alternative interpretations for data or observations.
 - Identify new information that might support or contradict a hypothesis.
 - Explain how new information can change a problem.
- Learning and problem solving
 - Separate relevant from irrelevant information.
 - Integrate information to solve problems.
 - Learn and apply new information.
 - Use mathematical skills to solve real-world problems.
- Communication
 - Communicate ideas effectively.

The CAT test and scoring guide have been tested and refined through repeated administrations to students and scoring sessions over an eleven-year period. In 2004 NSF provided funding to work with seven other institutions across the country (the University of Hawaii, the University of Texas, the University of Southern Maine, the University of Colorado, the University of Washington, Howard University, and Madisonville Community College) to use, evaluate, and help refine the CAT instrument. The success of this work led to additional funding in 2007 to train representatives from other institutions to lead CAT scoring workshops on their own campuses and to evaluate the effectiveness of these regional training workshops. The dissemination model has been successful, and in 2010 NSF provided additional funding to expand national dissemination and to expand the use of the CAT instrument for the assessment of project outcomes in other NSF-funded projects.



Face Validity

The CAT instrument was designed to measure the components of critical thinking and problem solving that faculty across disciplines think are most important. The graph below shows the percentage of faculty that think each question is a valid measure of critical thinking. These evaluations include a wide variety of disciplines from six institutions involved in a recent NSF project to evaluate and refine the instrument.



Criterion Validity

Criterion validity for this type of test is difficult to establish, since there are no clearly accepted measures that could be used as a standard for comparison. Since the CAT instrument is designed to assess a broad range of skills associated with critical thinking, we looked for reasonable but moderate correlations with other measures of academic performance and (more narrow) measures of critical thinking.

General Measures of Academic Performance

	ACT	SAT	Academic Profile	Grade Point Average
CAT	0.501*	0.516*	0.562*	0.295*

* Correlations significant, $p < .01$

The relationship between student responses on the National Survey of Student Engagement (NSSE) and performance on the CAT instrument has also been examined. Five items on the NSSE were significant predictors of performance on the CAT instrument (multiple $R = .49, p < .01$). The negative relationship between CAT performance and the extent to which students felt that their college courses emphasized rote retention is particularly important and supports both the criterion validity and the construct validity of the CAT instrument.

Other Measures of Critical Thinking

	CCTST (California Critical Thinking Skills Tests)	CAAP Critical Thinking Module
CAT	0.645*	0.691*

* Correlations significant, $p < .01$

NSSE Question	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 **
(3b) Number of books read on your own (not assigned) for personal enjoyment or academic enrichment	.277 **
(11e) Thinking critically and analytically and (11m) Solving complex real-world problems	.244 *
(7h) Culminating Senior Experience (thesis, capstone course, project, comprehensive exam, etc.)	.231 *

* Significant at .01 level ** Significant at .001 level

Cultural Fairness

The cultural fairness of the test has been evaluated in two ways. A multiple regression analysis of CAT performance revealed that once the effects of the entering SAT score and the GPA and whether English was the primary language were taken into account, neither gender, race, nor ethnic background were significant predictors of overall CAT performance. A cultural differential item functioning

(DIF) analysis was also performed to examine question bias. The review of DIF results did not reveal any items with prevalent cultural bias.



Test Sensitivity

Performance on the CAT instrument reveals neither floor effects nor ceiling effects for any of the participant groups tested thus far. Test-takers have included community college students and a very diverse collection of four-year undergraduate institutions (including regional universities, R-1 institutions, and Ivy League institutions). The sensitivity of the test is also sufficient to reveal differences between freshmen and seniors and to reveal the effects of a single course that develops critical thinking skills.

Institutional Use

More than fifty institutions across the country are collaborating on national dissemination of the CAT instrument. The participating institutions range from community colleges to large research institutions and include private and publicly funded institutions. Institutions are using the CAT for a variety of assessment purposes, including: campus-wide outcomes, general education outcomes, specific program outcomes, and specific course outcomes.

Faculty Development and Emerging Success Stories

Regional training workshops prepare participants to lead scoring sessions on their own campuses and to stimulate discussions of how faculty might address student weaknesses through modifications to their teaching and assessment methods. Faculty are encouraged to develop their own discipline-specific analog activities that engage students in critical thinking. This type of faculty development along with appropriate institutional support can transform instruction and student learning.

Institutional findings are beginning to emerge that provide a better understanding of the types of activities leading to significant gains on the CAT instrument. These findings indicate that a single course or project can impact critical thinking and significantly improve students' performance on the CAT instrument. Faculty members at Clemson University, Sam Houston State University, and the University of Wisconsin at Madison, among others, have implemented courses that have significantly improved students' performance on the CAT instrument.

Northwestern University and City Colleges of Chicago

Many NSF projects are beginning to use the CAT instrument to assess program outcomes, and these results will provide useful information to institutions that are looking for successful strategies to improve student learning. A recently funded project involving the Searle Center for Teaching Excellence at Northwestern University and the City Colleges of Chicago to design, pilot, and study a science, technology, engineering, and mathematics (STEM) faculty development program focused on improving higher-order learning outcomes of STEM students. The pilot program will consist of sixteen STEM faculty members, eight from each institution, participating in a linked series of workshops over one-and-a-half academic years. Goals of the project are for faculty to

- Reflect critically on key issues in learning and teaching in higher education
- Develop their learning, teaching, and assessment practices to facilitate higher-order student learning outcomes



- Redesign an existing course or develop a new course that implements these new learning and teaching practices
- Improve STEM students' higher-order learning outcomes

The project will be evaluated in terms of its effectiveness at the two institutions with respect to changes in instructional and assessment practices, faculty approaches and conceptions of teaching, and student learning outcomes. In addition, the effectiveness of this program in changing faculty conceptions of and approaches to teaching will be compared to the effectiveness of other faculty development programs by comparing data from this project to historical data on faculty who participated in standard faculty development programs at Northwestern.

The hypothesis driving the study is that faculty who are confronted with data on their own students' critical thinking using the CAT instrument and who conduct their own experiment to improve critical thinking will make greater changes in their conceptions of and approaches to teaching than faculty who participate in a standard faculty development program.

Notes

Partial support for the development and dissemination of the CAT instrument was provided by the National Science Foundation's TUES program (formerly CCLI) under grants 0404911, 0717654, and 1022789. Additional information and project reports can be found at www.CriticalThinkingTest.org.

The National Science Foundation's TUES program is also providing support for work related to enhancing critical thinking in STEM disciplines at the Searle Center for Teaching Excellence at Northwestern University and the City Colleges of Chicago to design and evaluate a STEM faculty development program for improving higher-order learning outcomes under grant 942404.

Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Reference

Bok, D. 2006. *Our underachieving colleges: A candid look at how much students learn and why they should be learning more*. Princeton, NJ: Princeton University Press.

Barry Stein is Professor and Co-Director, Center for Assessment and Improvement of Learning, and **Ada Haynes** is Professor and Co-Director, Center for Assessment and Improvement of Learning, at Tennessee Tech University in Cookeville; **Denise Drane** is Associate Director, Searle Center for Teaching Excellence, at Northwestern University in Evanston, Illinois.