Project CAT: Assessing Critical Thinking Skills

Barry Stein*, Ada Haynes, Michael Redding - *Tennessee Technological University*

OVERVIEW

This chapter reviews the development of the CAT (Critical thinking Assessment Test) instrument, a new interdisciplinary assessment tool for evaluating students' critical thinking skills. The CAT instrument is designed to directly involve faculty in the assessment of student strengths and weaknesses and thereby support the implementation of pedagogical improvements within and across disciplines. An NSF award allowed Tennessee Technological University (TTU) to partner with six other institutions across the United States (University of Texas, University of Colorado, University of Washington, University of Hawaii, University of Southern Maine, and Howard University) to evaluate and refine the CAT instrument. The CAT instrument has high face validity when evaluated by a broad spectrum of faculty across the United States in STEM and non-STEM disciplines, good criterion validity when compared to other instruments that measure critical thinking and intellectual performance, good reliability, and good construct validity using expert evaluation in the area of learning sciences. The broader impacts and potential benefits for improving undergraduate education are discussed.

IMPORTANCE OF CRITICAL THINKING

Most employers and educational institutions recognize the importance of effective critical thinking skills. For example, in 1990, the U.S. Department of Education stated as a goal that "the proportion of college graduates who demonstrate an advanced ability to think critically, communicate effectively, and solve problems will increase substantially" [1]. This goal became part of the "Goals 2000: Educate America Act" passed by Congress [2]. Many educators have also argued for the importance of preparing people to think critically [3,4,5,6,7,8,9,10]. In the National Assessment of Educational Progress (NAEP), "results suggest that although basic skills have their place in pedagogy, critical thinking skills are essential" [11]. Similarly, a report from the American Association of Universities indicated critical thinking and problem solving skills were essential for college success [12]. According to Derek Bok, president of Harvard University, national studies have found that more than 90 percent of faculty members in the United States consider critical thinking the most important goal of an undergraduate education [13].

Increasingly, the importance of critical thinking/problem solving skills in the workplace is also being recognized. For example, Halpern [14] argues, "virtually every business or industry position that involves responsibility and action in the face of uncertainty would benefit if the people filling that position obtained a high level of the ability to think critically" (see also [15]). A TTU survey of employers revealed that skills typically associated with critical thinking represented four out of the top five skills considered most important [16]. Similarly, a survey of New Jersey employers by the John J. Heldrich Center at Rutgers also found that skills typically associated with critical thinking represented three out of the top six skills considered most important. Those same employers also considered less than half of the graduates of two-year programs prepared for critical thinking, and only 56 percent of the graduates of four-year programs prepared for critical thinking [17]. A recent CNN poll of employers also found that critical thinking is one of the top five skills employers felt both critical to their businesses and most important in potential job candidates [18].

Although the importance of critical thinking is widely recognized, there is a gap between that recognition and the reality of what is taught and learned in the classroom. In order to bridge this gap, valid and reliable assessment tools are needed to measure these higher order thinking skills. With increasing pressure for accountability in education, "What gets measured gets taught... We must measure what we value or it won't get taught" [19].

DEVELOPMENT PROCESS

Our initial efforts to identify and then develop an effective assessment tool for critical thinking were stimulated by a state-wide "Performance Funding Initiative" to pilot test instruments designed to assess critical thinking beginning in 2000. TTU approached this task with the understanding that assessment would ultimately need to be linked to improvement initiatives at some point in the future.

There are three important characteristics of assessments that can foster genuine efforts to improve the quality of student learning:

1) The assessment must be an authentic and valid measure of progress toward the underlying goal.

2) The assessment should promote the motivation from within to change - the more engaged faculty are in the evaluation process the better.

3) The assessment should evaluate skills that are considered important within the framework of contemporary learning sciences.

Assessment approaches that violate these criteria can actually be detrimental to the quality of student learning or at best point out weaknesses that will never be addressed.

Although we explored a variety of assessment tools and pilot tested several existing instruments, none of the available tools satisfied all of our criteria. Many instruments assessed very narrow definitions of critical thinking (i.e., logical reasoning/reading comprehension) and/or did not sufficiently involve faculty in the evaluation of student performance. Consequently, we assembled an interdisciplinary team of faculty with an expert in learning sciences to identify a core set of skills associated with critical thinking across disciplines and then developed questions to assess those skills. TTU spent three years refining the questions and accompanying scoring guide for this test, which involved mostly short answer essay-type questions. These efforts were guided by the following principles:

1) Identify critical thinking skills across disciplines that faculty genuinely believe underlie critical thinking.

2) Develop an instrument that involves faculty and students in activities that reveal weaknesses and encourages quality improvement initiatives.

3) Develop a reliable instrument that students find intrinsically interesting.

4) Develop an instrument based upon contemporary theory in learning sciences.

Once the instrument was developed at our university, NSF funding supported collaboration with six other institutions across the country (University of Texas, University of Colorado, University of Washington, University of Hawaii, University of Southern Maine, and Howard University) to administer the test, conduct scoring workshops, evaluate student performance, and gather faculty input to evaluate and refine the instrument. In addition, NSF funding allowed us to work with external consultants in learning science and education to refine the instrument and collect information to evaluate validity and reliability.

GENERAL FEATURES OF THE CAT INSTRUMENT

The CAT instrument is currently administered in paper form, although there are plans to develop an electronic version. The test consists of 15 questions, with the majority requiring short-answer essay responses to evaluate the 12 skill areas listed in Table 1. Student responses to these essay questions provide a better understanding of students' thought processes and their ability to think critically and creatively when confronted with real-world problems than would be provided by multiple-choice questions. Essay assessments are purported to 1) have higher construct validity; 2) exhibit less racial bias; 3) foster more faculty involvement; and 4) flexibly assess a wider range of skills than multiple-choice questions [20].

Another feature of the CAT instrument is that it provides numerous opportunities for students to learn during the assessment, a process known as "dynamic assessment" [21,22,23,24, 25]. Dynamic assessment techniques are needed to measure the extent to which people can understand new information and apply that information to a novel situation. The CAT instrument uses dynamic assessment together with problems that are intrinsically interesting to students and representative of real-world problems. These features contribute to the validity of the test and students' motivation to perform well on the test regardless of their discipline. Anecdotal reports from a variety of institutions indicate that students do find the test interesting and engaging. The latter observation is particularly important for institutions that are trying to accurately assess students' performance in situations where test performance does not directly impact a student's course grade.

The CAT instrument can be administered in one hour and most students complete the test in about 45 minutes. A detailed scoring guide has been developed and refined to help faculty evaluate student responses.

FINDINGS RELATED TO PROJECT CAT

A major challenge in developing any instrument designed to evaluate critical thinking is to find agreement among faculty across disciplines and institutions about what skills underlie critical thinking. These skill areas must also link to current theory in learning sciences and cognition for construct validity. The skill areas assessed by the CAT instrument correspond to the higher order cognitive skills in Bloom's Taxonomy (comprehension, application, analysis, synthesis, and evaluation) [26].

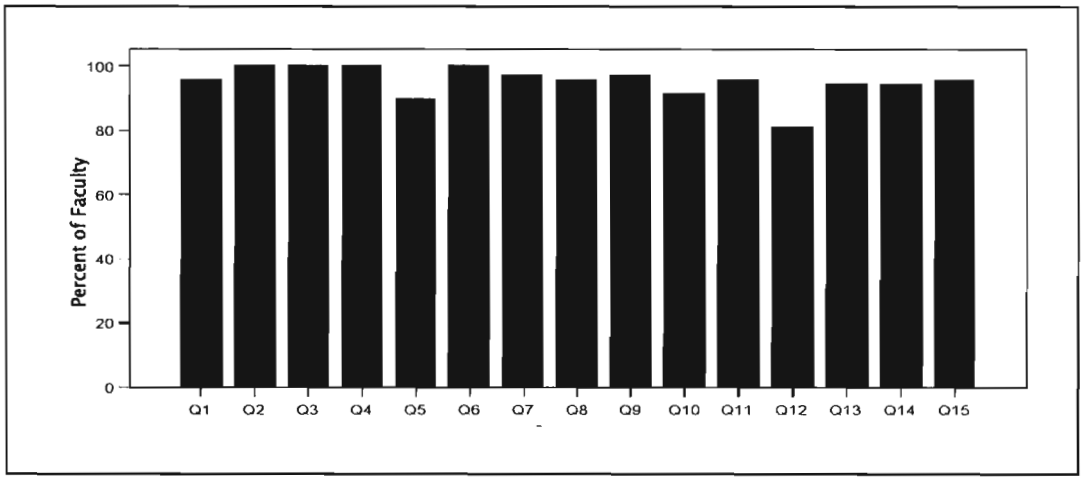
Faculty participants in scoring workshops at each participating university were asked to indicate which of the skill areas targeted by the CAT Instrument (see Table 1) they considered to be important components of critical thinking. The findings indicate that all of the skill areas targeted by the CAT instrument were generally perceived as important components of critical thinking by most faculty who participated in the scoring workshops. The area with least agreement (79.4 percent) concerned using mathematical skills to solve a complex real-world problem.

Table 1. Skill Areas Assessed by the CAT Instrument

Separate factual information from inferences that might be used to interpret those facts.
Identify inappropriate conclusions.
Understand the limitations of correlational data.
Identify evidence that might support or contradict a hypothesis.
Identify new information that is needed to draw conclusions.
Separate relevant from irrelevant information when solving a problem.
Learn and understand complex relationships in an unfamiliar domain.
Interpret numerical relationships in graphs and separate those relationships from inferences.
Use mathematical skills in the context of solving a larger real-world problem.
Analyze and integrate information from separate sources to solve a complex problem.
Recognize how new information might change the solution to a problem.
Communicate critical analyses and problem solutions effectively.

The faculty who participated in the scoring workshops were also asked to evaluate the face validity of each question contained in the CAT instrument. Most faculty felt that the questions included on the CAT instrument were valid measures of critical thinking (see Figure 1). The question with the lowest overall support (question 12 = 81 percent) involved using a mathematical calculation to help solve a complex real-world problem. The percent of faculty who considered this question a valid measure has risen as a result of improvements in the question. STEM faculty rate the face validity of questions slightly higher than non-STEM faculty in every case, although the difference between STEM and non-STEM faculty was only significant ($p < .05$) on question 12. These findings provide strong evidence for the face validity of the test. The external consultant and evaluators have also positively evaluated the construct validity of the instrument.

Figure 1. Percent of Faculty Indicating Each Question Measures a Valid Component of Critical Thinking



Criterion validity for a test of this type is rather difficult to establish since there are no clearly accepted measures that could be used as a standard for comparison. The approach we have taken is to look for reasonable but moderate correlations with other (more narrow) measures of critical thinking and other general measures of academic performance. The rationale underlying these comparisons is that the critical thinking skills measured by the CAT instrument should correlate at a moderate level with other measures of critical thinking and academic

performance. The findings support these aims, with the highest correlation between the CAT and the California Critical Thinking Skills Test (CCTST), indicating that only about 42 percent of the variability in the CAT instrument is explained by the CCTST (see Table 2).

Table 2. CAT Correlations with other Performance Measures

	ACT	SAT	Academic Profile	Grade Point Average	CCTST
CAT	0.599*	0.527*	0.558*	0.345*	0.645*

* correlations significant, $p < .01$

The National Survey of Student Engagement (NSSE) is a widely used instrument designed to assess the types of activities students are engaged in as well as their perceptions of institutional emphasis and the institution's contribution to their learning [27]. The NSSE was administered together with the CAT instrument to a stratified random sample of seniors at TTU. Misty Cecil, one of our doctoral students working on the project, examined the relationship between relevant NSSE questions and student performance on the CAT instrument. Five items on the NSSE were significant predictors of performance on the CAT instrument (multiple $R = .49$, $p < .01$). These five items are listed in the table below. 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 [28].

Table 3. NSSE Questions Related to CAT Performance

NSSE Question	Beta 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	
(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

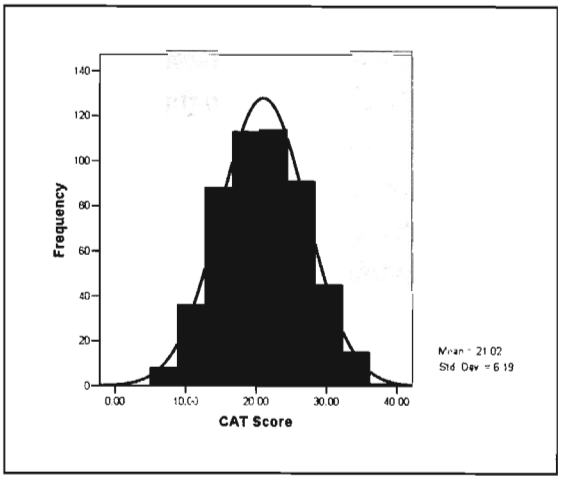
Several other key measures are reported in Table 4 below. Scoring reliability and internal consistency (calculated using Cronbach's alpha) are quite good for a test of this type and have increased as a result of test refinement. In addition, our latest measure of test-retest reliability has risen as a result of instrument improvements. The CAT instrument has also been found to be sensitive enough to assess changes between freshmen and seniors and to reveal the effects of a single course that includes components designed to improve critical thinking [29].

Table 4. Other CAT Statistical Findings

Scoring Reliability = 0.82
Internal Consistency = 0.695
Test–Retest Reliability > 0.80

Figure 2 shows the distribution of student scores (raw) on the CAT (version 4) instrument against the normal curve. Scores ranged from a low of 6 to a high of 36.3. There was no evidence of a floor effect or a ceiling effect (lowest possible score = 0, highest possible score = 40).

Figure 2. Distribution of Student Scores



Although more extensive analyses of any possible ethnic/racial/gender bias in the CAT instrument are currently being conducted, the preliminary analysis of available data is quite encouraging. A multiple regression analysis revealed that once the effects of entering SAT score and GPA, and whether English was the primary language (evaluated during year 2 testing) were taken into account, neither gender, race, nor ethnic background were significant predictors of overall CAT performance.

In addition to the quantitative survey data discussed above, qualitative data were collected from the local testing coordinators and the faculty scorers. Overall, the comments from both these groups were overwhelmingly positive. Many of the faculty members who participated in the CAT scoring workshops also exhibited increased interest in exploring methods for improving their students' critical thinking skills.

EXPLORATORY STUDY: PREDICTIVE VALIDITY

At the suggestion of our external consultant, John Bransford, from the LIFE Center at the University of Washington, Seattle, we began to examine ways in which the CAT instrument might be combined with training to improve students' critical thinking skills. While the benefits of such work go far beyond the goals of the current grant, we began to explore this issue as an additional means to help establish the validity of the CAT instrument. Specifically, the predictive validity of the test could be supported if students could be trained to improve performance on the CAT instrument, and if this training improved their performance on other relevant critical thinking tasks.

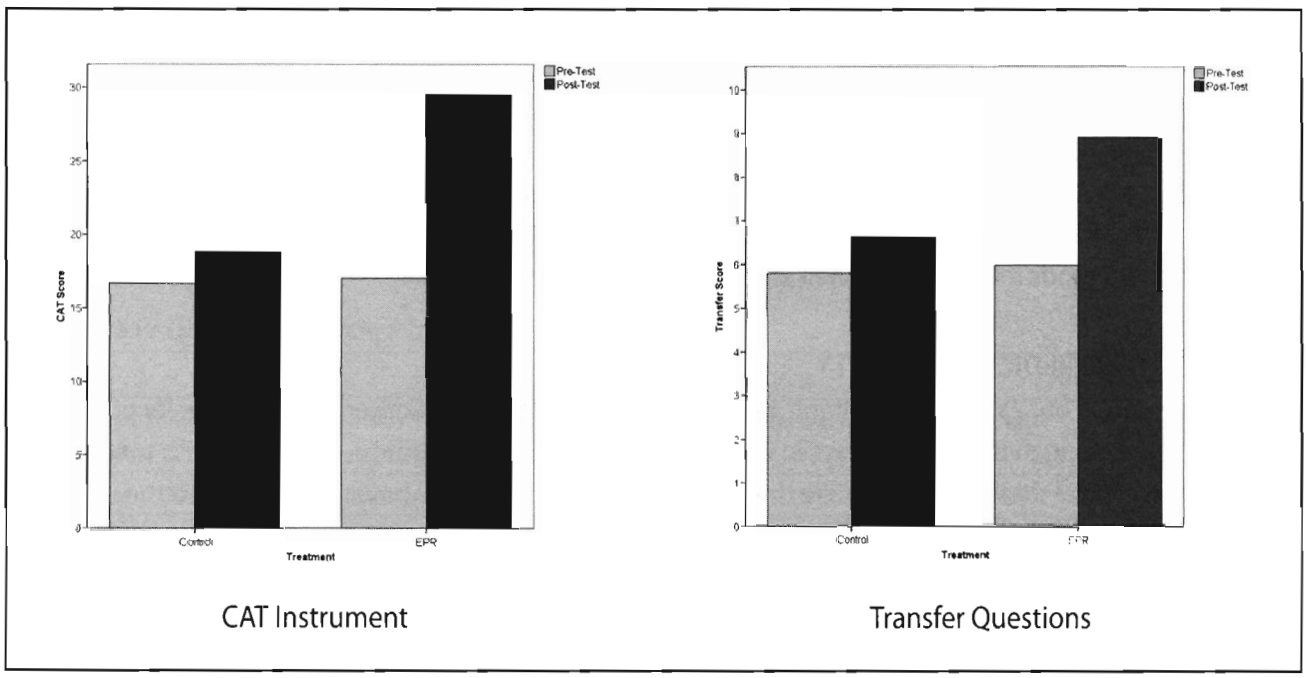
One existing methodology that seemed to hold promise for facilitating training using the CAT instrument was Calibrated Peer Review™ (CPR). Developed at UCLA with NSF support, the CPR system involves a computer network that provides opportunities for students to learn how to evaluate essay-writing assignments through a process of calibrating their evaluations to expert evaluations of the same essays. These exercises afford numerous opportunities to learn and critically evaluate ideas [30].

The constraints surrounding the use of the CAT instrument and the opportunities for implementing special training necessitated our modifying the CPR process and adapting it to this new situation. Training was conducted in group sessions with numerous opportunities for formative feedback. In the pilot study, Theresa Ennis, one of our doctoral students working on the project, trained a small group of students to score the CAT instrument using a procedure we call Enhanced Peer Review (EPR). The training used the same detailed scoring guide that had been developed for faculty scoring workshops. Students were initially given a small sample of test responses that reflected a range of scores for each question on the CAT instrument. During training students calibrated their evaluations of tests with those of the faculty graders. The training afforded numerous opportunities to explore the

rationale for assigning scores to each response on the test and to educate students about a variety of issues related to critical thinking. After two three-hour training sessions, students were given the opportunity to score numerous CAT tests without any further training. The results of the pilot study showed that students could be taught to reliably score the CAT test using the EPR procedure.

We also compared students in the EPR pilot study to a control group. Both groups took a pre-test and post-test that included questions from the CAT instrument together with a new set of analogous transfer questions. Students who participated in the EPR training not only improved significantly more than the control group on the CAT questions they were trained on, they also improved significantly more than the control group on transfer questions (see Figure 3).

Figure 3. Performance on CAT Instrument & Transfer Test Questions



Student comments from a follow-up interview a year after the testing indicate lasting effects of the EPR training (e.g., "I'm a lot more critical about what I'm going to be listening to or believing in"; "I don't just look at one side and say this is definitely the way it is… I explore more sides and try to find alternative answers"; and "It helped me work through a personal serious life episode.").

We are currently conducting further studies of EPR using the CAT instrument and analog transfer questions. The pilot work done with training students to score the CAT instrument has also provided insights into better methods for training faculty to score the CAT instrument that could be incorporated into future workshops to train scoring workshop leaders at other institutions.

BROADER IMPLICATIONS

While there is broad agreement among faculty, administrators, educational experts, governmental officials, and employers that critical thinking skills are very important, efforts to improve students' critical thinking skills are currently hindered by the absence of an effective tool for assessing critical thinking skills that is valid, reliable,

and culturally fair. The CAT instrument is an innovative tool for assessing critical thinking that meets these criteria and, in addition, is designed to maximize faculty engagement in the assessment process and help motivate instructional improvement. These features of the instrument can help close the loop between assessment and the implementation of improvement initiatives. The CAT instrument is appropriate for both STEM and non-STEM disciplines. The interdisciplinary nature of the instrument is compatible with focused or broad-based institutional efforts to improve critical thinking across disciplines.

The CAT instrument may also be a useful assessment tool that can be used to strengthen other research projects designed to assess or improve student learning. We feel that combining the CAT instrument with initiatives that encourage the use of best practices to improve students' critical thinking and real-world problem solving to improve student learning can greatly improve undergraduate education in both STEM and non-STEM disciplines.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Facione, P.A., Facione, N.C., Sanchez, C., and Gainen, J., "The disposition toward critical thinking," *Journal of General Education,* vol. 44, no.1, ERIC No. EJ499944, pp. 1-25, 1995.

[2] U.S. Congress, Goals 2000: *Educate America Act, H.R. 1804*, (U.S. Congress, Washington, D.C.), http://www.ed.gov/legislation/GOALS2000/TheAct/index.html, 1994.

[3] Bok, D., Our Underachieving Colleges: *A candid look at how much students learn and why they should be learning more,* (Princeton University Press, Princeton, N.J.), 2006.

[4] Bransford, J. D., Brown, A., L., and Cocking, R. R., (eds.), *How People Learn: Brain, Mind, Experience, and School*, (National Academy Press, Washington, D.C.), 2000.

[5] Ennis, R., "A Logical Basis for Measuring Critical Thinking Skills," *Educational Leadership*, vol. 43, no.2, pp. 44-48, 1985.

[6] Paul, R., and Nosich, G., *A Model for the National Assessment of Higher Order Thinking*, (National Center for Educational Statistics, Washington, D.C.), 1992.

[7] Pawlowski, D.R., and Danielson, M.A., "Critical Thinking in the Basic Course: Are We Meeting the Needs of the Core, the Mission, and the Students?" Paper Presented at the Annual Meeting of the National Communication Association, 1998.

[8] Resnick, L.B., *Education and Learning to Think,* Committee on Mathematics, Science, and Technology Education, Commission on Behavioral and Social Sciences and Education, National Research Council, (National Academy Press, Washington, D.C.), http://www.nap.edu, 1987.

[9] Siegel, H., *Education Reason: Rationality, Critical Thinking, and Education*, (Routledge, New York), 1988.

[10] Vygotsky, L. S., *Thought and language,* (MIT Press, Cambridge, Mass.), 1986.

[11] Wenglinsky, H., "Facts or critical thinking skills? What NAEP results say," *Educational Leandership*, vol. 62, no.1, pp. 32-35, 2004.

[12] Conley, D., A report from standards for success, Center for Educational Policy Research, http://www.s4s.org/cepr.s4s.php, 2003.

[13] See reference 3.

[14] Halpern, D. E., "Assessing the Effectiveness of Critical-thinking Instruction," *Journal of General Education*, vol. 42, no.4, pp. 238-254, 1993.

[15] Duchesne, R. E., "Critical Thinking, Developmental Learning, and Adaptive Flexibility in Organizational Leaders," paper presented at the Annual Meeting of the American Association for Adult and Continuing Education in Charlotte, N.C., 1996.

[16] Stein, B., Haynes, A., and Unterstein, J., "Assessing Critical Thinking," paper presented at the annual meeting of the Council on Colleges, Southern Association of Colleges and Schools, Atlanta, Ga. http://www.tntech.edu/cat/sacs2004stein.ppt, 2004.

[17] John J. Heldrich Center for Workforce Development, *Survey of New Jersey Employers to Assess the Ability of Higher Education Institutions to Prepare Students for Employment*, ERIC No. ED485290, 2005.

[18] Castellini, R., "Survey: More to see pay increase in 2006," http://www.cnn.com/2006/US/Careers/01/04/cb.aol.survey/index.html, 2006.

[19] Partnership for 21st Century Skills, *Learning for the 21st Century*, http://www.21stcenturyskills.org/images/stories/otherdocs/p21up_Report.pdf, 2002.

[20] U.S. Department of Education, National Center for Education Statistics, *The NPEC Sourcebook on Assessment, Volume 1: Definitions and Assessment Methods for Critical Thinking, Problem Solving, and Writing*, NCES 2000—172, Erwin, T.D. (ed.), for the Council of the National Postsecondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development, (U.S. Government Printing Office, Washington, D.C.), 2000.

[21] Campione, J.C., and Brown, A.L., "Guided learning and transfer: Implications for approaches to assessment," in Frederiksen, N., and Glaser, R., et al. (eds.), *Diagnostic monitoring of skill and knowledge acquisition,* pp. 141-172, (Erlbaum, Hillsdale, N.J.), 1990.

[22] Feuerstein, R., *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques,* (University Park Press, Baltimore), 1979.

[23] Lidz, C. S., *Dynamic assessment: An interactional approach to valuating learning potential,* (Guilford Press, New York), 1987.

[24] Samuels, M.T., "Assessment of post-secondary students with learning difficulties: Using dynamic assessment in a problem-solving process," In Lidz, C.S., and Elliott, J.G. (eds.), *Dynamic assessment: Prevailing models and applications,* pp. 521-542), (JAI/Elsevier Science, Amsterdam), 2000.

[25] Sternberg, R.J., and Grigorenko, E.L., *Dynamic testing: The nature and measurement of learning potential*, (University of Cambridge, Cambridge), 2002.

[26] Bloom, B.S., (ed.), *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain,* (Longmans, New York), 1956.

[27] Indiana University Center for Postsecondary Research, *National Survey of Student Engagement 2005 Annual Report: Exploring different dimensions of student engagement*, http://nsse.iub.edu/NSSE_2005_Annual_Report/index.cfm, 2005.

[28] Cecil, M.J., *The relationship between student responses on the National Survey of Student Engagement (NSSE) and performance on the Critical-thinking Assessment Test (CAT),* doctoral dissertation, Tennessee Technological University, 2006.

[29] Stein, B., Haynes, A., and Ennis, T., "Assessing Critical Thinking," paper presented at the Commission on Colleges of the Southern Association of Colleges and Schools Annual Meeting in Atlanta, Ga. in December 2005, http://www.tntech.edu/cat/sacs2005Stein.ppt, 2005.

[30] Chapman, O.L., and Fiore, M.A., "Calibrated Peer Review™," *Journal of Interactive Instruction Development*, vol. 12, no.3, pp. 1-15, 2000.