

Faculty Driven Assessment of Critical Thinking: National Dissemination of the CAT Instrument

Barry Stein, Ada Haynes, Michael Redding, Kevin Harris, Mark Tylka, and Elizabeth Lisic
Tennessee Technological University

Abstract- This paper reports the latest findings of a project to nationally disseminate the CAT© instrument, a unique interdisciplinary assessment tool for evaluating students' critical thinking skills. Tennessee Technological University originally partnered with six other institutions across the U.S. to evaluate and refine the CAT instrument beginning in 2004. Through these efforts a test with high face validity, high construct validity, high reliability, and that is culturally fair was developed. In the current national dissemination project, we are collaborating with over 40 institutions across the country to disseminate this unique instrument and support the creation of assessment based efforts to improve student learning. These dissemination efforts involved training representatives from other institutions to lead scoring workshops on their own campuses as part of a broader effort to improve student learning. A variety of findings indicate these dissemination efforts are successful. National user norms that are being collected from this project indicate that largest gains in undergraduate critical thinking appear to occur in the junior and senior years.¹

INTRODUCTION

In our increasingly technological and information driven society, the ability to think critically is a cornerstone to both workplace development and effective educational programs. The importance of critical thinking skills in higher education has become increasingly apparent. Accrediting agencies such as Accreditation Board of Engineering and Technology (ABET) recognize the need for higher order thinking skills and real world problem solving in their accreditation standards [1]. Faculty also recognize the importance of critical thinking skills. According to Derek Bok [2], over ninety percent of the faculty in the U.S. feel that critical thinking is the most important goal of an undergraduate education. Despite the recognized importance of critical thinking skills, it is difficult to find effective tools for assessing these skills. Many of the assessment tools are plagued by problems related to validity, reliability, and cultural fairness [3]. According to Bransford, Brown, and Cocking [4] "a challenge for the learning sciences is to provide a theoretical framework that links assessment practices to learning theory." Pellegrino, Chudowsky and Glaser also note that assessment instruments need to be developed based upon principles of learning and cognition [5].

¹ Partial support for this work was provided by the National Science Foundation's CCLI Program under grants 0404911 and 0717654. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Tennessee Technological University (TTU) has been engaged in an extended effort during the last 9 years to develop and refine an instrument to assess critical thinking that overcomes many of the weaknesses of other existing tools. Preeminent theoreticians and educators in the area of learning sciences and assessment participated in the project. Unlike many other available assessment tools, the Critical thinking Assessment Test (CAT) instrument uses short answer essay responses to assess critical thinking. Although essay questions can be harder to score than multiple choice questions, they provide a better understanding of students' thought processes and ability to think critically and creatively when confronted with real world problems [6]. In addition, the CAT instrument is unique in that it utilizes a campus's own faculty to evaluate student responses. Ewell [7] has noted that one problem with many existing assessments is that "we let assessment get excessively distanced from the day-to-day business of teaching and learning." The CAT instrument allows faculty to directly observe students' critical thinking and understand their students' weaknesses. This activity helps motivate faculty to consider changes in pedagogy that might improve students' critical thinking skills [8]. This becomes increasingly important as accrediting agencies such as ABET increase their focus on efforts to improve students' critical thinking [1].

TTU's approach to developing the CAT instrument has been to involve faculty in the identification of a core set of skills they believe to be an important part of critical thinking across many disciplines (see Table 1). TTU worked with a diverse group of institutions nationwide (University of Texas, University of Colorado, University of Washington, University of

TABLE 1
SKILL AREAS ASSESSED BY THE CAT INSTRUMENT

| |
|---|
| 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. |

Hawaii, University of Southern Maine, and Howard University) to further develop, test, and refine this instrument. The results of that project revealed the CAT instrument had high face validity when evaluated by a broad spectrum of faculty across the U.S. in STEM and non-STEM disciplines, had good criterion validity when compared to other instruments that measure critical thinking and intellectual performance, had good scoring reliability, good test/re-test reliability, and good construct validity using expert evaluation in the area of learning sciences [8].

Student response to the CAT instrument has also been very positive. Many students indicate appreciation for the real world problem solving tasks that make up the test [9]. Using questions that are interesting and engaging helps motivate students to perform their best.

Performance on the CAT instrument is significantly correlated with other measures of student performance including cumulative college grade-point average ($r = 0.295$), entering SAT or ACT scores ($r = 0.528$, and $r = 0.560$, respectively), and performance on several other tests that are thought to measure certain skills associated with critical thinking such as the California Critical Thinking Skills Test ($r = 0.645$) and the CAAP Critical Thinking Module ($r = 0.691$). These correlations provide support for the criterion validity of the CAT instrument. However, the magnitude of the correlations also indicate that the CAT instrument measures something different than the other assessment tools [9]. Performance on the CAT instrument has also been found to significantly correlate with several key questions on the National Survey of Student Engagement (NSSE). One NSSE question that is significantly negatively correlated with students' performance on the CAT instrument is the extent to which students think their courses emphasized rote retention [10]. In contrast, CAT performance is significantly positively correlated with NSSE questions that reflect the extent to which students' think their college experience emphasized critical thinking and real-world problem solving.

Extensive work has been done to refine the scoring process for the CAT instrument to improve scoring accuracy and reliability. For example, the scoring process uses multiple scorers for each question and uses a detailed scoring guide that has been refined with extensive faculty input to ensure scoring consistency. Although scoring accuracy is typically a problem for essay type tests, the scoring process for the CAT instrument overcomes many of the problems typically associated with scoring essay tests.

The major focus of the new work reported in this paper involves efforts to disseminate the instrument to a broad range of institutions across the U.S. These efforts involved training representatives from other institutions to lead scoring workshops on their own campuses as part of a broader effort to improve student learning.

METHOD

Funding from the National Science Foundation supported collaboration with 20 institutions across the country to admi-

nister the CAT instrument to their students and to conduct scoring workshops with their own faculty to score their students' responses on the CAT instrument. Since institutional interest in the CAT instrument greatly exceeded that level of participation supported by NSF, we created an alternative method of institutional participation that was self-supporting. We are currently collaborating with over 40 institutions across the country. Grant funded participants were selected to insure diversity in mission, institutional characteristics, ethnicity/racial demographics, and geographic location. The collaborating institutions range from community colleges to large research institutions and include private, as well as, publicly funded institutions.

Materials

A variety of training materials were developed to support the training of new scoring workshop leaders and to help those leaders train their own faculty to effectively score the CAT instrument on their respective campuses. These materials included a narrated overview of the development process for the CAT instrument that streams from our website (www.CriticalThinkingTest.org), a detailed training manual that covers issues from sampling and obtaining institutional review board approval to administering the test, recruiting faculty scorers, conducting the scoring workshop, and interpreting institutional results. A detailed scoring guide and multi-media computer based training modules were also developed to explain question scoring.

Train-the-Trainer Workshops

Participating institutions sent two to three representatives to a two-day training workshop where they received extensive training in scoring the CAT instrument and preparation to lead a scoring workshop on their own campus. The workshop also covered other topics related to understanding and using different assessment designs, sampling, and using the scoring session as an opportunity for faculty development. One of the advantages of the CAT instrument is that it gives faculty a good understanding of student weaknesses. Coupling this experience with a discussion of ways to improve student learning creates an effective faculty development experience that can lead to real improvements in instruction and student learning.

RESULTS

The effectiveness of the train-the-trainer workshops were evaluated in a number of ways including; participant surveys, direct observations by our external evaluator, follow-up onsite observations of campus scoring sessions, and scoring accuracy checks of tests scored by each institution.

Training Workshop Surveys

At the conclusion of each of the training workshops participants were asked to complete an anonymous survey to evaluate various components of the workshop training. The survey used a 6 point Likert type scale: 6 – Strongly Agree, 5 –

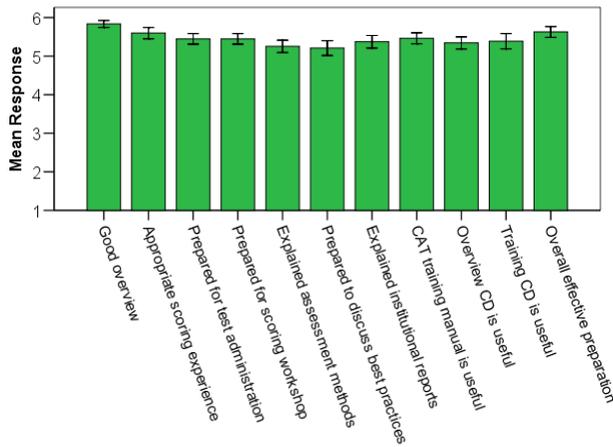


Fig. 1
Average Training Workshop Evaluation Responses (with 95% CI)

Agree, 4 – Slightly Agree, 3 – Slightly Disagree, 2 – Disagree, 1 – Strongly Disagree. Eighty-five participants were included in this analysis.

The mean response on each survey item is shown in Figure 1 (error bars indicate 95% confidence interval). The results indicate that the workshop evaluations have been quite positive across the topics surveyed.

Qualitative Evaluation of Training Workshops

Observations by our external evaluators and project staff are used formatively to continually improve the training workshops. For example, we noted early in the project that participants in the training workshops may find it difficult to process information from too many different sources (the workshop leader, the multimedia training modules, and the scoring guide). This led to changes in how the multi-media training module was used in the workshop. It is now used during the second day as a resource for explaining scoring in situations where scorers may need further explanation. Many other changes were also made in response to qualitative evaluations (e.g., extending the discussion of sampling methods and exploring methods for obtaining representative samples at different institutions, and encouraging participants to develop discipline specific analog learning activities that correspond to questions on the CAT instrument).

Onsite Evaluation of Scoring Workshops

Observations by our external evaluators and project staff who have attended scoring workshops at other institutions have been particularly useful. These observations have confirmed that procedures outlined in our training manual must be carefully adhered to in order to efficiently and accurately score tests. Institutions that have deviated in important ways from those guidelines experienced numerous problems. These observations have led to increased emphasis and training support to explain various procedures associated with test scoring.

Scoring Accuracy Checks

One of the most important pieces of evaluative information

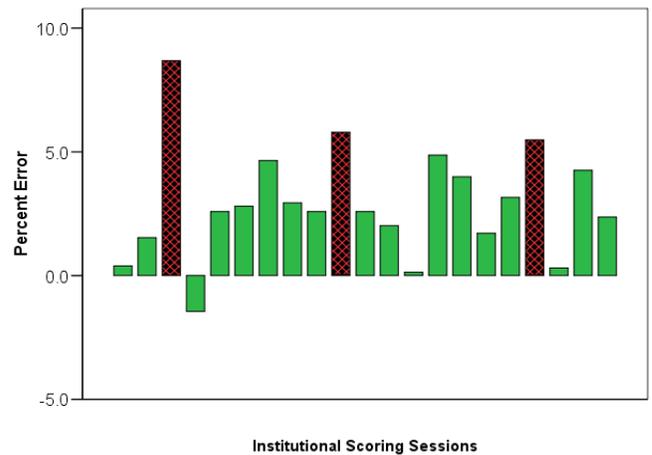


Fig. 2
Results of Scoring Accuracy Checks Across Institutions as Percent Error

about the success of this national dissemination model comes from checking the accuracy of scoring performed by each participating institution. Experienced scorers at our institution rescore randomly selected tests from each scoring session at other institutions. The sample includes about 15% to 20% of the tests scored by each institution. Error rates below 5% are considered acceptable. Thus far, overall accuracy has been very good. Figure 2 shows the overall test scoring error found across 21 institutional scoring sessions that have undergone accuracy checks. About 1 in 7 institutions has been found to deviate more than 5% from our experienced scorers. All institutions receive a scoring accuracy report together with a question by question analysis of accuracy, and suggestions for improving scoring accuracy in subsequent sessions. Recommendations are given to institutions that deviate more than 5% on how to appropriately adjust their scores for comparison to other institutions.

National Norms

Another goal of the current project is to begin assembling national norms for various populations of students including community college students and students at four-year institutions. Although many institutions will track their progress by

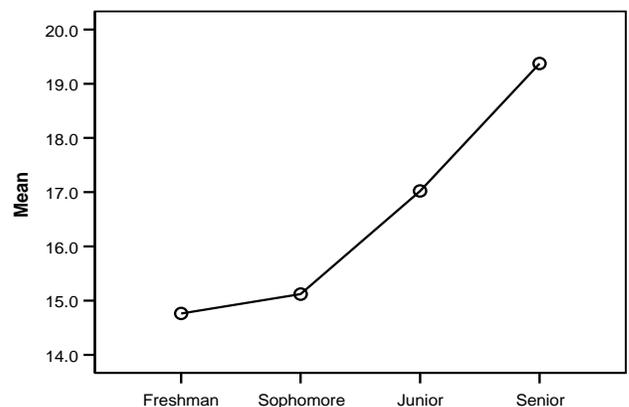


Fig. 3
Mean Scores on the CAT Instrument for Students at 4-year Institutions

comparison to their own test scores over a period of time, others have requested information that would allow comparison to larger populations. Figure 3 shows the mean score on the CAT instrument for approximately 3000 students of different class standings at four-year institutions across the country. The largest gains in critical thinking appear to occur in the junior and senior years of undergraduate education. The average senior score was about 51% of the maximum possible score on the CAT instrument (maximum score = 38 points). These scores indicate there is considerable room for improvement. The fact that no student or has obtained a perfect score or a score of zero at any institution suggests the absence of a ceiling or floor effect.

CONCLUSION

There have been several significant outcomes of the current national dissemination project. We have found that our dissemination model for training representatives at other institutions to use the CAT instrument at regional train-the-trainer workshops is successful. This finding is supported by information from participant surveys, onsite observations, and scoring accuracy checks. We do, however, continue to modify our training methods and supporting materials to improve effective dissemination.

We have found that many institutions are interested in finding ways to engage faculty in quality improvement efforts. Faculty involvement in the scoring of student essay exams greatly facilitates this process. We are also expanding our efforts to provide links to institutional and grant funded projects that have positively impacted student performance on the CAT instrument. This type of information is just beginning to emerge from a broad range of institutional initiatives. We are also collaborating with several other NSF funded projects that are using the CAT instrument to help identify potential student gains in critical thinking and real-world problem solving that may result from innovative educational pedagogies.

The CAT instrument is one of the few interdisciplinary assessment instruments available that also provides an opportunity for faculty development. By participating in the scoring process, faculty become aware of their students' weaknesses

and can begin to explore modifications in teaching methods that might address these weaknesses. Many of the participants in our regional training workshops feel that the CAT instrument is particularly useful because it helps faculty understand how they can develop discipline specific assessments that encourage students to think critically instead of just encouraging rote retention. This becomes increasingly important as accrediting agencies such as the Accreditation Board of Engineering and Technology (ABET) increase their focus on efforts to improve student learning.

REFERENCES

- [1] Accreditation Board of Engineering and Technology, *Criteria for Accrediting Engineering Programs: Effective for Evaluations During the 2006-2007 Accreditation Cycle*. Baltimore, MD: Accreditation Board of Engineering and Technology, 2005
- [2] D. Bok, *Our Underachieving Colleges: A candid look at how much students learn and why they should be learning more*. Princeton: Princeton University Press, 2006.
- [3] 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, prepared by T. Dary Erwin for the Council of the National Post-secondary Education Cooperative Student Outcomes Pilot Working Group: Cognitive and Intellectual Development. Washington, DC: U.S. Government Printing Office, 2000.
- [4] J. Bransford, A. Brown, and R. Cocking, (Eds.) *How People Learn: Brain, Mind, Experience, and School*. Washington, D.C.: National Academy Press, 2000.
- [5] J. Pellegrino, N. Chudowski, and R. Glaser (Eds.) *Knowing What Students Know*. Washington, D.C., 2001.
- [6] B. Stein, A. Haynes, and M. Redding. "Project CAT: Assessing Critical Thinking Skills," in *Proceedings of the National STEM Assessment Conference*, D. Deeds and B. Callen, Eds. NSF, 2007.
- [7] P. Ewell. "The examined life: Assessment and the ends of general education," Association of American Colleges & Universities General Education and Assessment Conference, 2004.
- [8] B. Stein, A. Haynes, M. Redding, T. Ennis, and M. Cecil. "Assessing Critical Thinking in STEM and Beyond," in *Innovations in E-learning, Instruction Technology, Assessment, and Engineering Education*, M. Iskander, Ed. Springer, 2007, pp.79-82.
- [9] B. Stein, A. Haynes, and K. Harris. "Assessing and Improving Critical Thinking Using the CAT Instrument," *ABET Best Assessment Processes Symposium XI*. Indianapolis, IN: Accreditation Board of Engineering and Technology, 2009.
- [10] M. Cecil, "The Relationship Between Student Responses on the National Survey of Student Engagement (NSSE) and Performance of the Critical-Thinking Assessment Test (CAT)," Dissertation, Tennessee Technological University, 2006.