

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

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 $r = 0.2668$), and

measure certain skills associated with critical thinking such as

the California Critical Thinking Skills Test ($r = 0.645$), and the

University of California, San Diego Critical Thinking Test ($r = 0.0953$),

indicating that the CAT instrument measures something different

from other instruments. The CAT instrument has been found to significantly correlate with

student engagement (NSSE, $r = 0.16$), student learning (NSSE, $r = 0.16$), and student problem solving (NSSE, $r = 0.15$).

Extensive work has been done to refine the scoring process to improve student learning.

Minister the CAT instrument to their students and to conduct scoring workshops with their own faculty to score their students.

The 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 United States. Institutions were selected to insure that

the materials were developed by a variety of institutions, including private as well as public institutions. These materials included a narrated overview of the development process for the CAT instrument that streams from our website (www.CriticalThinking.org).

These materials included a narrated overview of the development process for the CAT instrument that streams from our website (www.CriticalThinking.org).

www.CriticalThinking.org

that covers issues from sampling and scoring

faculty scorers, conducting the scoring workshop, and interpreting institutional results. A detailed multi-media computer based training manual was developed for the training workshop.

8

a two-day training workshop where they receive training in scoring the CAT instrument.

METHOD

Funding from the National Science Foundation supported

the research.

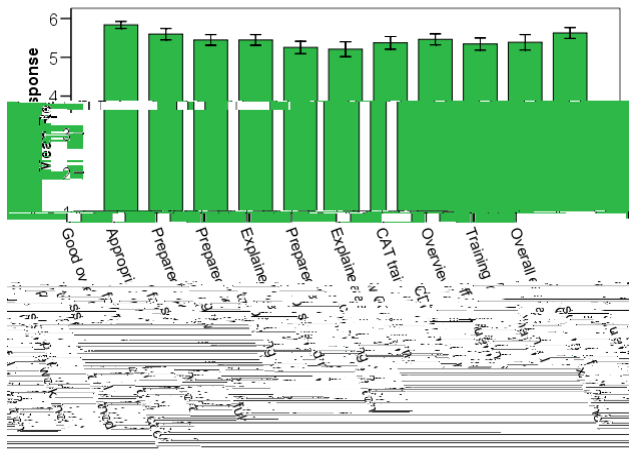


Fig. 1
Average Training Workshop Evaluation Responses (with 95% CI)

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

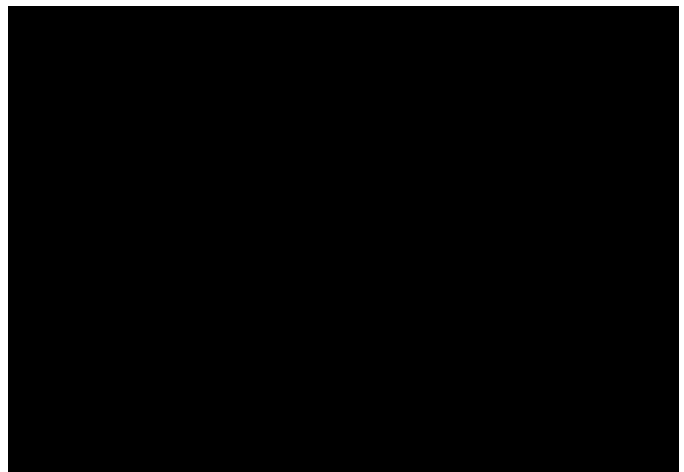


Fig. 2
Results of Scoring Accuracy Checks Across Institutions as Percent Error

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

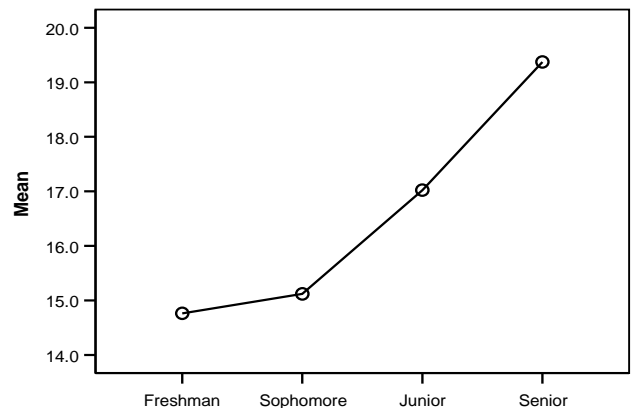


Fig. 3
Mean Scores on the CAT Instrument for Students at 4-year Institutions

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.

and can begin to explore modifications in teaching methods that might address these weaknesses. Many of the participants in our regional training works

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