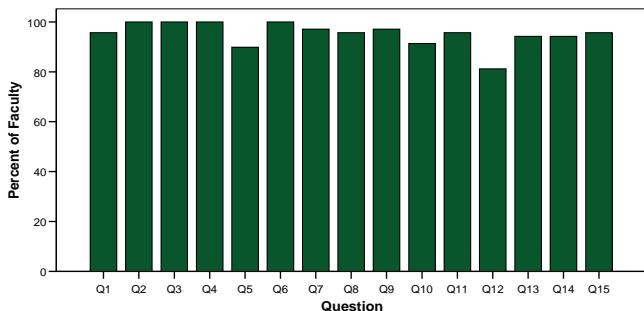


CAT[®] Instrument Technical Information

Face Validity

The CAT instrument was designed to measure those components of critical thinking and problem solving that faculty across disciplines think are most important. The graph below shows the percent 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 a test of this type 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 (more narrow) measures of critical thinking and academic performance.

General Measures of Academic Performance

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

* correlations significant, $p < .01$

Other Measures of Critical Thinking

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

* 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.

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. (<i>negative relationship</i>)	-.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

Scoring Reliability

Since this instrument involves mostly short-answer essay questions, the reliability of scoring is of great importance. Each question is scored by a minimum of two scorers and disagreements are resolved by third scorer. Refinements in the test and the scoring guide have yielded scoring reliability = 0.92 between the first and second scorer.

Internal Consistency

Most of the questions on the CAT instrument are designed to assess more than one component of critical thinking. The internal consistency of questions is reasonably good, $\alpha = 0.699$.

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 entering SAT score and 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.

Score Range

Performance on the CAT instrument reveals neither floor effects nor ceiling effects for any of the participants tested so far. Test takers have included all levels of 4-year undergraduates and community college students. The sensitivity of the test is also sufficient to reveal differences between freshman and seniors and to reveal the effects of single course that emphasizes critical thinking.