of Colorado, the University of Washington, the University of Hawaii, the University of Southern Maine, and Howard University). Each institution administered the CAT instrument to a sample of approximately 100 upper division students across STEM and non-STEM disciplines. Students were paid to participate in the testing.

Materials

The CAT instrument is a 15 question test with mostly short answer essay questions. The questions were specifically designed to evaluate the skill areas identified in Table 1. The questions explore various real-world situations that students find intrinsically interesting. A significant part of the test is devoted to solving a complex real-world problem that involves learning and applying new information – a testing practice sometimes referred to as dynamic assessment. The complex problem is designed to mimic many aspects of real-world problem solving. For example, in addition to solving a complex problem, decisions must be made about the about the types of articles that would be relevant to the solution of the problem as well as deciding how the solution to the problem might change if certain aspects of the problem situation were altered.

Testing Procedure

Students were tested in groups and the testing instrument was provided in booklet form with a package of additional readings. Each student was asked to complete all questions in the instrument and was given up to 60 minutes to complete the task. There was seldom a situation where any student did not have sufficient time to complete the test.

Scoring Procedure

A full day scoring workshop was held at each institution to assess student performance using a detailed scoring guide developed for the CAT instrument. Approximately 10-12 faculty from both STEM and non-STEM disciplines participated in the scoring workshop at each institution. Faculty were paid to participate in the scoring process.

During the scoring sessions, faculty received training on using the scoring guide immediately before each question on the test was scored. Each student's response to that question was then was then scored independently by two faculty members. If there was not agreement between the first two scorers, the question was scored by a third faculty member. The final score for each question was either the common score assigned by two graders or, in the case of three different scores, the final score was computed by averaging the three different scores. This process was repeated for each question on the test. Tests were frequently redistributed to insure that each faculty would see as many different tests as possible.

Faculty participants in the scoring workshop also completed several surveys after they finished scoring the tests. These surveys examined the face validity of the questions on the CAT instrument and whether the skills assessed by the CAT instrument were valid components of critical thinking (see Table 1).

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.

RESULTS

Evaluation of Skill Areas Targeted by the CTdMCId[

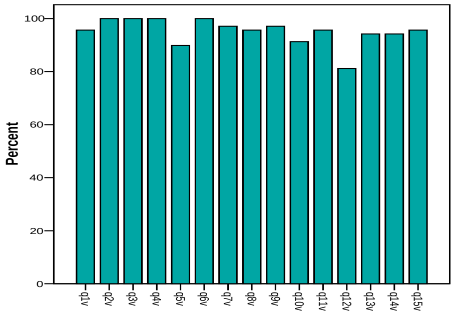

Fig. 2. Percent of Faculty Indicating Question Measures a Valid Component of Critical Thinking

measures of critical thinking (see fig. 2). The question with the lowest overall support (81.2%) involved using a mathematical calculation that was needed to help solve a complex real-world problem.

In addition to the quantitative survey data discussed above, qualitative data were collected from the local testing coordinators and the faculty scorers as well. The comments received from both faculty scorers and the local testing coordinators were overwhelmingly positive. Many faculty felt the test was very useful for identifying student weaknesses and others were interested in using the test within their disciplines to help explore ways of making program improvements.

Distribution of Student Scores and Internal Consistency

Fig. 3 shows the distribution of student scores (raw) on the CAT 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, highother measures of academic performance. Finally, the project has managed to circumvent a major problem for essay type tests by providing relatively high rates of scoring reliability with faculty who have had no prior experience grading this type of exam.

The refinement of the CAT instrument is also significant for another reason. 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' deficiencies and can begin to explore modifications in teaching methods that might address these weaknesses. This becomes increasingly important as accrediting agencies such as the Accreditation Board of Engineering and Technology (ABET) increase their focus on aspects of critical thinking such as life-long learning skills [10].

REFERENCES

[1] S.A. Forawi, "Critical Thinking and the National Science Standards." *Transactions of the Missouri Academy of Science, Annual,* 62.

ade2t7le1 ade2t7le1TJEMC /P6E2 1 uc<- as life-long6( )]TJons o361Me Missou</f3,P]TJ2e-l.1.144 TDTf-13.803 -1.1343Td( )Tj]TJ 2 BDC 843-714(I.256M)6(e 8)4(n)-2