DOCUMENT RESUME

ED 413 808 HE 030 711

AUTHOR	Cook, Patricia; Johnson, Reid; Moore, Phil; Myers, Phyllis; Pauly, Susan; Pendarvis, Faye; Prus, Joe; Ulmer-Sottong, Lovely
TITLE	Critical Thinking Assessment: Measuring a Moving Target. Report & Recommendations of the South Carolina Higher Education Assessment Network Critical Thinking Task Force.
INSTITUTION	South Carolina Higher Education Assessment Network, Rock Hill.
PUB DATE	1996-06-00
NOTE	94p.
PUB TYPE	Opinion Papers (120) -- Reports - Evaluative (142) -- Tests/Questionnaires (160)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Accountability; Change Strategies; *College Outcomes Assessment; *Critical Thinking; Educational Change; *Evaluation Methods; Feasibility Studies; Higher Education; Needs Assessment; Performance Tests; State Aid; Test Validity
IDENTIFIERS	*Performance Indicators; *South Carolina

ABSTRACT

This report is part of South Carolina's effort to move toward "100 percent performance funding" for the state's public colleges and universities and results from a task force's investigation of ways to assess critical thinking. The following eight major findings are reported: (1) policy makers must determine priorities; (2) critical thinking lacks a common definition; (3) faculty involvement positively impacts change; (4) high test validity correlates with low test feasibility; (5) performance tests are favored over objective, multiple-choice tests; (6) performance tests, however, have significant feasibility limitations; (7) external factors affect test scores; and (8) collaboration, not competition, is essential. The task force was unable to recommend a single test as ranking high on both validity and feasibility, but did recommend the development of a test or series of tests to better measure critical thinking. Specific recommendations are offered to the Educational Board. The bulk of the report consists of seven appendixes which include a literature review, a copy of the survey used and survey results, the criteria used for evaluating measures, reviews of nine specific tests, guidelines for choosing commercial tests, suggestions for future test development, and charts of data on commercial tests. (Contains 65 references.) (DB)

CRITICAL THINKING ASSESSMENT: MEASURING A MOVING TARGET

Report & Recommendations
of the
South Carolina Higher Education Assessment Network
Critical Thinking Task Force
June 1996

Task Force Committee:

Patricia Cook (University of South Carolina at Aiken)
Reid Johnson (Francis Marion University)
Phil Moore (University of South Carolina at Columbia)
Phyllis Myers (Trident Technical College)
Susan Pauly (University of South Carolina at Lancaster)
Faye Pendarvis (Orangeburg-Calhoun Technical College), Task Force Chair
Joe Prus (Winthrop University)
Lovely Ulmer-Sottong (South Carolina Commission on Higher Education)

TABLE OF CONTENTS

➤THE TREND TOWARD ACCOUNTABILITY

In March 1996, South Carolina's legislature became the first in the nation to pass a bill that by 1999 would create "100% performance funding" for South Carolina's public colleges and universities. Instead of traditional formula funding based primarily on the number of full-time students attending a college or university, an institution's state funding will be dependent on a series of "performance indicators." This new performance-based funding will "reward for quality and have consequences for failure" (*Act 1195*, SC Legislature, 1996). The legislation had strong support from the Legislature, the Council of College Presidents, the Governor's office, the State Board for Technical and Comprehensive Education, and the Commission on Higher Education. The new law was based on a February report from the Joint Legislative Study Committee on Higher Education which consisted of several selected legislators, business persons, and one member each from the Commission on Higher Education and the State Board of Technical Education. This Committee, appointed by the Governor, also created a new, ambitious mission statement for higher education in South Carolina:

> "To be a global leader in providing a coordinated, comprehensive system of excellence in education by providing instruction, research and lifelong learning opportunities which are focused on economic development and benefit the State of South Carolina." (*Report of the Joint Legislative Committee to Study the Governance and Operation of Higher Education in South Carolina*, Feb. 1, 1996).

On the national scene conversations concerning education reflect equally strong accountability movements. At the Education Summit in March 1996, President Clinton called for an end to "free passes" in education; that is, an end to the promotion of students from grade to grade or institution to institution using the practice of social promotion. He recommended national standardized tests, one upon entering middle school and high school, and one upon graduating from high school. The forty governors and forty-nine business leaders present went even further and recommended that states voluntarily adopt national standards and publish the results of scores in the form of comparisons. "We're willing to be compared with each other," said Nevada's Governor Bob Miller, Vice Chairman of the Governor's Association. "That comparison will lead to competition and that competition will lead to excellence" (*Governors' Summit Proceedings*, March 1996).

Around the world, business leaders have expressed what they want colleges to produce in their graduates. John Abbott, Director of The Education 2000 Trust, a British not-for-profit entity that links leaders from education, industry, and the social sector, comments that business leaders in a global marketplace . . . want college graduates who have "creativity, enterprise, purposefulness, a good sense of community responsibility and collaborative work" (*AAHE Bulletin*, March 1996).

On the face of it, the three viewpoints outlined above from state legislators, governors and global business leaders have little in common. The first centers on leadership, the second on competition and the last on needs. However, all three turn to the educational community to provide changes that are basically different from the current "norm" for education, and all three are the result of an underlying distrust regarding the perceived state of education today. In addition, there is an underlying hope that the leadership for solving the perceived problems in education will come from within the educational community itself.

Given the vital role that educational leadership will play in continuing to set a growing and diverse agenda for education in the United States, it is not surprising that a wide variety of educational consortia, groups, and organizations continue to study the strengths and weaknesses of our current system, particularly in the area of assessing skill levels of the students who attend and graduate from our schools, colleges and universities.

➤THE CHARGE: SOUTH CAROLINA APPROACHES A MOVING TARGET

The Task Force Takes Shape

It is against the above backdrop that the State of South Carolina's Commissioner of Higher Education, Fred Sheheen, responded to the Southern Regional Educational Board's (SREB) request "to look at the assessment of critical thinking," a skill that forms an important base of higher learning outcomes, particularly at the college level. Specifically, Commissioner Sheheen offered the expertise of a team from the South Carolina Higher Education Network (SCHEA) in issuing a report. SCHEA is a consortium of nearly fifty South Carolina public and private colleges, universities, and higher education agencies working together to plan, implement, and evaluate methods and means of assessment and institutional effectiveness. The SCHEA Task Force agreed to review the current major critical thinking assessment instruments available on the market, cite the strengths and weaknesses of each instrument, rank the instruments relative to their strengths and weaknesses, and recommend whether any of the instruments should be explored for large scale administration (e.g., throughout a state, a consortium of states, or a region).

➤THE UNDERLYING PREMISES OF THE TASK FORCE

Results Should Unify

One of the first steps taken by the Task Force was to identify common beliefs and premises. The first premise on which the Task Force agreed was that the results of their work should not be seen in any way as definitive. Rather, the results would be used to continue conversations regarding critical thinking and its assessment at the collegiate level, broadening the knowledge base about college level skills so better informed decisions could be made about improving student achievement in those skill areas. The Task Force hopes that conversations about the findings and recommendations will unify and not further divide the educational, political and business communities since the support of all are vital to the entire enterprise of education.

Education Is Both A Process and A Product

Another premise of the Task Force is that education is both a process and a product. One separated from the other produces neither. Both must support the other to produce a whole often greater than the sum of any of their individual parts. Just as the skills that are "basic" to education (such as reading, writing and computation) support those that are more "advanced" (such as critical thinking, creativity and discovery), so too does each product of education support the learning process. For example, learning to read--a product of education--often supports the desire to read and comprehend longer and more complex passages--a process of education. Reading can be measured; the desire to learn is much more difficult to quantify, but both are necessary for becoming a successful, "educated" college student.

Higher Order Skills Are Contextual

One difficulty in measuring higher order educational skills, which are both product and process, is that the individual skills themselves are often interwoven. A student who reads better often writes better, and a student who has strong communication skills often has strong adaptation skills. Therefore, another premise of the Task Force was that the acquisition of a higher order skill such as critical thinking is not a discrete process, but often depends on many factors in the learner's history and environment, some of which occur in the formal higher education setting and some of which do not.

Since the evaluation procedures used for critical thinking (e.g., tests, essays, speeches, demonstrations) depend on the mastery of basic skills (i.e., reading, writing, mathematics, and oral communication), it is vital that they be strong before critical thinking is assessed. Individual proficiency levels for these basic skills can vary considerably for individual college students upon college entry . It is folly to believe that weak basic skills will produce a strong critical thinker. For any evaluation of a higher order skill, it is always important to keep in mind that the strength of a student's basic skill proficiencies will certainly affect the assessment outcome of the higher order skill, in this case critical thinking. The basic and higher order skills are not unrelated.

➤THE COMMON DEFINITION GUIDING THE PROCESS

One Definition From Many

There are many definitions in the literature on critical thinking (See Literature Review, Appendix A). In the process of reaching consensus, the Task Force listened to both in-state and out-of-state speakers regarding critical thinking, conducted outside research, and engaged in a variety of other activities (see The Process). The Task Force unanimously agreed that the following definition of critical thinking would guide its deliberations.

> **"Critical thinking is a reflective, systematic, rational, and skeptical use of cognitive representations, processes, and strategies to make decisions about beliefs, problems, and/or courses of action."** Adapted from Beyer, Ennis, Facione, McPeck, Sternburg, et al.

This definition was chosen because the Task Force believed it best fit this project and could be used as a common base for higher education institutions in South Carolina. However, it is important for other groups to devote serious thought and discussion to their own definition of critical thinking since the instruments chosen to measure the skill must match the definition any group chooses.

Teaching Must Match Testing

Faculty teach what they believe important for students to learn, and faculty test what they teach. If the test does not closely match the types of things that faculty believe should or are being taught in a curriculum, then a resultant low score will mean little. The instrument will have tested something faculty did not believe important enough to teach, or the test will have measured the skill in a very different way from the way in which it was taught or learned. For example, a test of critical thinking might evaluate a student by having him/her perform computer simulations in a high tech laboratory. However, if the student has not encountered computerized simulations before, then the test score will reflect that lack of familiarity rather than a student's ability to think or to react critically to the scenarios presented.

Sum Of Many Parts

Critical thinking has many parts. The ability to effectively present opposing arguments, to comprehend and analyze data, to judge wisely between two possible solutions, to generate multiple approaches to problems, and to utilize information given or implied are but a few possible components of critical thinking. It is imperative that groups determine their own definition for critical thinking and then look for measurements that not only best match that definition but also best match how the faculty at an institution teaches critical thinking.

For example, is critical thinking taught predominately in a "contrast and comparison mode," an "application mode," a "problem solving mode," a "demonstration mode?" Must it be demonstrated by an individual, by a whole team, by only certain members of a team in a given situation? Is it exhibited in writing, in speaking, or by some type of "changed" behavior -- in general education, in the major, in new situations or in familiar situations? These differences affect what instruments are chosen to measure critical thinking.

During an entire collegiate experience, a student will learn many aspects of "critical thinking." **No single assessment instrument available tests all of the possible aspects of critical thinking.** Therefore, after deciding on a definition for critical thinking, groups must actively search for assessment instruments which match the various parts of the definition chosen.

➤ MEASUREMENT IS A MOVING TARGET

One Measure Is Not Enough

Most institutions of higher learning have as a goal that a student exiting their institution should be able to "think critically." **However, depending on the faculty, the curriculum, and the institution, the definition of critical thinking and the ways it is taught and measured are likely to be quite different**. Measurement of critical thinking is a moving target. That doesn't mean critical thinking should not be measured, but it does mean that it can't be measured by one simple means.

Because of this, the Task Force strongly feels that one measurement of critical thinking--one score, one task, one method of measuring this advanced skill--is not enough. To make true formative or summative judgements at the student level, a policy maker needs multiple measures (at least three or more indicators) to assess something as complicated as the collegiate level skill of critical thinking. Whether one of the instruments reviewed by the Task Force is to be used as an outcomes indicator or not is not as important as understanding that there should be at least three *different* types of critical thinking achievement indicators, including grades, used to make sound educational judgements about the teaching and learning of critical thinking.

In addition, and just as importantly, the indicators must be connected to what will be done with them. The assessment of any skill without a well-defined problem to investigate and goals to achieve may create an impression of action but lead nowhere. As noted by Steele (1995), we must use assessment to clarify where we are and where we are going. We must understand how we want to use a test before choosing what the test will be. It is imperative to know what to do with an indicator once the work of getting it has been done. In Fullan's book, *Change Forces: Probing the Depth of Educational Reform* (1993), his first stated lesson is "You can't mandate what matters." In other words, the more complex the change being called for, the less it can be forced.

➤PROCESS OF THE STUDY

Identifying A Plan

Early in its study process, the Task Force developed a plan to accomplish its charge and goal. The plan included a review of relevant literature (see Appendix A); a survey of South Carolina's higher education institutions (see Appendix B); the development of a definition and criteria for evaluating critical thinking instruments (see Appendix C); a review/evaluation of critical thinking instruments (see Appendix D); presentation/interviews with two in-state practitioners of critical thinking assessment; presentations and discussion with three publishers of national instruments designed in whole or part to assess critical thinking; and lengthy deliberations as a Task Force over a five month period. The outcomes of each of these planning components are reproduced in full in the Appendices of this document and will be described only briefly here.

Review of Literature

The SCHEA office provided a list of more than fifty references to critical thinking literature. Each member of the task force was assigned copies of publications to read and review. The SCHEA lending library provided the publications. Readers agreed to identify criteria describing excellence in instruments designed to assess critical thinking. The individual reviews were presented and discussed during Task Force meetings. Outside speakers included South Carolina campus administrators and representatives from three national testing companies: American College Testing, Educational Testing Service, and Riverside Corporation including a University of Missouri representative. They spoke to the Task Force regarding critical thinking concepts, definitions, and measurements (see Appendix A).

South Carolina Critical Thinking Survey

In order to determine the current status of critical thinking instruction and assessment in the state's colleges and universities, the Task Force conducted a survey of 60 public and private South Carolina institutions of higher education. A written, open-ended survey was designed and mailed to the SCHEA representative or other appropriate person at each of the two-year technical colleges, four-year colleges, comprehensive teaching universities and research universities including regional campuses. The survey instrument was accompanied by a memo describing the purpose of the survey and directions for completing the instrument (see Appendix B--SC Survey). Follow-up phone calls were made by Task Force members to elicit telephone or written responses to survey questions.

Forty-two percent of the institutions in South Carolina responded to the survey. In brief, the survey showed that about 88% of the responding institutions cited critical thinking as an expected educational outcome. However, only 46% of the respondents have actually defined critical thinking at the institutional level (that is, in general education requirements). A total of 64% of the respondents indicated that they assess critical thinking. Surveys and individual classroom assessments were the most commonly reported methods. A total of 36% of the respondents reported that they assess critical thinking with institution-wide standardized tests (see Appendix B for a complete discussion of the survey findings).

Criteria For Evaluation Of Instruments

Criteria for describing and evaluating the essential attributes of critical thinking tests were developed, reviewed and revised by the Task Force. Once the final criteria were agreed upon, a numerical score sheet corresponding with the criteria was constructed. The completed criteria covered four general areas: relevance, accuracy, diagnostic utility, and feasibility. Scores for relevance were based on six indicators, while accuracy and diagnostic utility were based on three indicators each. One hundred points were assigned in varying weights to the first three criteria (relevance, accuracy, and diagnostics). In addition to using various weights, 100 points were assigned to the four parts of feasibility. After all members reviewed their assigned instruments, points were collapsed into a Likert-type "high," "average" and "low" symbol-scale. The scale provides a quick and easy reference guide for prospective users of these assessments (see Appendix C--Criteria for a complete list of the criteria and the summary scoring sheet).

Evaluation Of Commercially-Available Tests

Each of ten commercial instruments was independently evaluated by at least two task force members using the common numerical scale described above. It is important to refer to the Task Force definition of critical thinking to understand that assessment instruments which ranked high in "validity" were generally those in which the test was more performance-oriented and had more detailed scoring. These instruments were often more diagnostically useful. Instruments which ranked high on "feasibility," however, were those that could be incorporated more easily into a program, were economical and were easier to administer and score. The more feasible tests were usually more objective in nature (see Appendix C--Criteria).

As previously stated, it is consistent with current measurement practice to find that tests are high in one area and also low in the other. Given those reminders, the following summary serves only as an overall guide.

SUMMARY OF RANKINGS

VALIDITY RANKING

High Validity

ACT Assessment of Reasoning and Communicating

Moderately High Validity

California Critical Thinking Skills Test

Moderate Validity

ACT CAAP Critical Thinking Module
ACT COMP Objective
Cornell Critical Thinking Test

Moderately Low Validity

College Base (Matrix Form)
College Base (Long Form)
Ennis-Weir Critical Thinking Test
ETS Tasks in Critical Thinking
Watson-Glaser Critical Thinking

Low Validity

None

FEASIBILITY RANKINGS

High Feasibility

ACT CAAP Critical Thinking Module
California Critical Thinking Skills Tests
College Base (Matrix Form)
Cornell Critical Thinking Test
Ennis-Weir Critical Thinking Test
Watson-Glaser Critical Thinking Test

Moderately High Feasibility

None

Moderate Feasibility

None

Moderately Low Feasibility

ACT COMP (Objective Form)
College Base (Long Form)

Low Feasibility

ACT Assessment of Reasoning and Communicating
ETS Tasks in Critical Thinking

CRITERIA With numerical weights		ACT ASSESSMENT OF REASONING & COMMUNICATIN G (COMP -ARC)	ACT CAAP CRITICAL THINKING MODULE	ACT COL. OUTCOME MEAS. PROGRAM (COMP-OBJ. FORM)	CALIF. CRITICAL THINKING SKILLS TEST	COLLEGE BASE MATRIX FORM	COLLEGE BASE LONG FORM	CORNELL CRITICAL THINKING TEST	ENNIS - WEIR CRITICAL THINKING TEST	ETS TASKS IN CRITICAL THINKING	WATSON GLASER-CRITICAL THINKING APPRAISAL
VALIDITY CRITERIA		**Evaluation Key : ★ = High ✔ =Average ○ = Low**									
Directly Measures CT Knowledge	10	★	○	○	○	○	○	○	★	★	○
Requires Direct Demo. of CT Skill	10	★	★	★	✔	○	○	★	★	○	✔
Appr. Level and Range for College	5	★	✔	★	★	✔	✔	✔	○	✔	○
Relevant to Curriculum Taught	5	✔	✔	★	✔	✔	✔	★	★	✔	○
Appro. Results Comparisons	5	★	★	★	★	✔	✔	✔	○	○	✔
Adequate Technical Characteristics	5	★	★	★	★	★	★	✔	○	○	★
Analytical and Global Scores	5	★	○	○	★	✔	✔	★	✔	✔	★
Formative and Summative Scoring	10	★	○	○	○	✔	✔	○	★	○	○
Specificity in Score Reporting	5	★	★	★	✔	✔	✔	✔	○	✔	✔
Student Strengths and Weaknesses	10	★	✔	○	★	○	✔	★	★	○	✔
Program Strengths and Weaknesses	20	★	★	★	★	✔	✔	★	★	○	○
Maximum Generalizability	10	★	★	★	★	★	★	○	○	○	○
FEASIBILITY CRITERIA		**Evaluation Key : ★ = High ✔ =Average ○ = Low**									
Easily Incorporated into Higher Ed.	40	○	★	○	★	★	○	★	★	○	★
Low Cost and Low Time Allocation	30	○	★	○	★	★	○	★	★	○	★
Ease of Adm. and Scoring	20	○	★	✔	★	★	★	★	✔	○	★
Interesting to Students	10	✔	✔	★	✔	★	★	✔	✔	○	○

DEFINITIONS OF CRITERIA USED TO ASSIGN HIGH (★), AVERAGE (✔) AND LOW (O) RANKINGS:

VALIDITY (Primary Criteria):

Relevant:

1. Directly measures critical thinking knowledge by requiring student production of answers
2. Requires demonstration of critical thinking skills by performance or performance simulation
3. At an appropriate level and range for college students' skills
4. Reflects a continuum of knowledge and skills taught in most S.C. college curricula
5. Includes appropriate results' comparisons
6. Clearly describes adequate technical characteristics

Accurate

1. Provides analytical as well as global scores
2. Affords formative as well as summative scores
3. Has sufficient specificity of score reporting to show significant differences in score levels

Diagnostically Useful

1. Shows student level strengths and weaknesses
2. Relates directly to an educational program's strengths and weaknesses
3. Affords maximum generalizability of results (results can be used more than one way)

FEASIBILITY (Secondary Criteria):

1. Easily incorporated into higher education programs
2. Economical in resource requirements (i.e., costs in time, effort, money, for students, for faculty, and institutions. especially in "frontloading," time and effort to prepare items and administer the test, and "backloading," the time and effort to score, interpret, and get useful results from the test)
3. Quickly and/or locally scoreable
4. Clear and easy to administer, score and interpret
5. Interesting to students (to improve motivation)

➤ PRIMARY FINDINGS OF TASK FORCE

The Task Force on Critical Thinking hopes that the following findings will guide others in their study of the measurement of critical thinking as a college level skill.

F① POLICY MAKERS MUST DETERMINE PRIORITIES

There is little doubt that critical thinking is important to college graduates, employers, and citizens and should be a stated goal of general education and virtually all major fields in every college and university. However, higher education institutions and the bodies that oversee and advise them *must determine the relative priority to give critical thinking and its assessment, versus other important aspects of general education (e.g., oral and written communication, reading, mathematics, computer literacy, etc.).* Allocation of resources and effort should then correspond to these priorities.

F② CRITICAL THINKING LACKS A COMMON DEFINITION

Despite the fact that higher education institutions in South Carolina as a whole have a longer history of comprehensive assessment than is the case in many states, most institutions in the state have not defined and assessed critical thinking in a substantive manner. This, in part, is probably reflective of the complexity of pedagogical, assessment, and feasibility issues involved.

F③ FACULTY INVOLVEMENT POSITIVELY IMPACTS CHANGE

South Carolina colleges or universities which have used national standardized tests of general education, including critical thinking, over a multi-year period report no substantial impact of results on programs, instruction, or curriculum. On the other hand, the few institutions which have implemented locally-developed, performance-based measures are more likely to report subsequent improvements in how students are taught.

F④ HIGH TEST VALIDITY EQUALS LOW TEST FEASIBILITY

An inverse relationship often exists between validity and feasibility. Task Force evaluations of instruments clearly show this relationship. Instruments which provide more detailed, performance based measures of critical thinking require more time, effort, and expense and are thus less feasible for administration on a wide scale basis in most higher education institutions.

F⑤ PERFORMANCE TESTS ARE FAVORED OVER OBJECTIVE, MULTIPLE-CHOICE TESTS

The Task Force found that, particularly for critical thinking skills, performance-based tests are favored over objective tests. From a philosophical standpoint, since there is usually no "one right answer" when critical thinking is applied in the real world, nationally-normed, multiple-choice tests that continue to provide one right answer only tend to divide the real world from the classroom. If the public wants the classroom to more realistically mirror the real world, then "one right answer" solutions in teaching or testing critical thinking must be eliminated.

F⑥ FEASIBILITY LIMITATIONS EXIST FOR PERFORMANCE-BASED TESTS

Although performance-based tests are often preferred by educators, feasibility issues keep the use of performance tests low. Feasibility issues include such things as instrument costs, staff time required for secure test administration and make-up test administrations, staff time for scoring and analyzing results, special facilities and equipment required for test administration, and results that are often not able to be tied to specific courses taught in the curriculum. All of these issues lower the use of performance-based tests, even though that type of test often gives the most diagnostically useful information about a student's skill levels and most directly measures a skill.

F⑦ EXTERNAL FACTORS AFFECT TEST SCORES

How and when student critical thinking assessment is measured are just as important as what instrument is used. The results of assessment are considerably affected, for example, by such factors as student motivation, sample selection, standardized administration procedures, and the timing and circumstances surrounding the assessment activity. These factors need to be carefully considered by institutions seeking to implement comprehensive assessment programs.

F⑧ COLLABORATION, NOT COMPETITION, IS ESSENTIAL

South Carolina legislation, Act 1195, designates performance funding on a variety of indicators, among which is "Institutional Cooperation and Collaboration." Any approach that relies solely or primarily on inter-institutional comparisons based on a single mandated measure will likely be to the detriment of cooperation and collaboration among and between institutions in the area of assessment. Although this Task Force is not insensitive to the benefits of healthy competition which supports collaboration and assessment, comparative approaches which ignore key differences in mission characteristics and the needs of students entering different types of institutions must be avoided.

➤ MAJOR CONCERN OF TASK FORCE

The most basic and underlying concern of the Task Force is that a single indicator of critical thinking, for example a "test score," will be used to make absolute judgements about quality, funding, and curriculum in the absence of other indicators instead of in addition to them. Critical thinking is both a complicated product and a complicated process; there is no single skill to be measured in any single way. Critical thinking is not one discrete skill but many, including a desire to use them all.

➤ RECOMMENDATIONS OF THE TASK FORCE

No one test can be recommended by the Task Force that ranks equally high on both validity and feasibility. However, the Task Force does recommend that a test or series of tests be created that could better measure critical thinking. We do not believe that this would be either inordinately expensive or difficult, but it would take the combined efforts of a variety of people (see the SREB Recommendation).

We recommend that such an endeavor begin as soon as possible. While this process is occurring, we recommend that institutions in South Carolina begin a collaborative study to determine in what courses on each campus the teaching and learning of critical thinking takes place. If and when pilot testing occurs, the results could be directly applied to the curricula which the institutions determine are most appropriate to their missions, and student bodies.

The Task Force on Critical Thinking further recommends the following actions for consideration by South Carolina and the Southern Regional Educational Board:

R❶ SET PRIORITIES AND DEDICATE RESOURCES

More needs to be done in South Carolina to encourage and support the teaching and assessment of critical thinking in higher education. There exists within the state substantial expertise to address this issue but additional support and resources are needed. *In addition, the State must determine the relative priority to give critical thinking and its assessment, versus other important aspects of general education (e.g.. writing, oral and written communication, mathematics, computer literacy, etc.).*

R❷ INVOLVE ALL STAKEHOLDERS IN THE PROCESS

Efforts to address critical thinking must be multifaceted and must involve faculty and students in significant ways in order to produce meaningful change. Pilot projects or other multi-institutional efforts related to critical thinking should not be limited to assessment alone but also need to address curricula, instruction, and/or student learning. Research has shown repeatedly that educational change does not occur in higher education until the faculty support and promote it. Faculty and students are the heart of every institution of higher education. The relationship of those faculty and students to any assessment procedure or instrument isn't just desirable; it is essential. Research at institutions which use either commercially based instruments or locally-developed instruments has shown that faculty must support any test chosen to measure critical thinking, or that testing experience will fail. If it fails, it will only increase the general mistrust of the public, the legislature, faculty and students, and it will mean that succeeding will be that much harder the next time.

R❸ CONTINUE TO ENCOURAGE LOCAL ASSESSMENT

Locally-developed approaches to assessment within individual institutions, departments, and programs need to be strongly encouraged and not inadvertently discouraged by focusing too extensively on national measures. Multiple measures must be used and local measures must be a major part of critical thinking assessment.

R❹ CHOOSE INSTRUMENTS AND PILOT PROJECTS CAREFULLY

Since no one instrument available measures all aspects of critical thinking, make sure that the instruments chosen most closely match the definition of critical thinking on which the stakeholders agree. If commercially prepared tests are chosen for pilot projects, carefully consider the following (see Appendix E for full guideline explanations):

1. Define critical thinking. What definition encompasses what is being or should be taught? Which classes are addressing critical thinking according to the definition chosen?

2. Know how the results will be used before the assessment is administered. Will individual student scores be necessary for long-range curricular change, high student testing motivation, student level improvement? How and to whom will the results be reported?

3. Determine which testing criteria are important based on a philosophy of measurement. Will your institution use a performance test or a multiple-choice objective test to evaluate critical thinking? Review the criteria shown in Appendix E in reference to your needs.

4. Fit the test into your overall assessment plans. Make sure that basic skills are being assessed before making judgements on the results of advanced skill tests such as critical thinking. What are the other criteria being used to evaluate critical thinking (there should be at least three criteria including grades) ?

5. Test feasibility must be determined before test adoption. Does the institution have adequate facilities for administration of the test-laboratories for taped responses; computers for computerized responses; faculty for grading responses?

6. *Carefully determine how and to whom the test will be administered.* Will the test be given to freshmen and seniors to determine differences or just to seniors? How will students be motivated to do well? Will all students take the test or only selected students? Will the test be voluntary?

7. Take the test and have all stakeholders take the test. It is unfair to ask students to be graded on an assessment that the major stakeholders have never seen or taken themselves. It is essential that stakeholders understand what the instrument does and does not measure (scores can be kept confidential).

➤ REGIONAL RECOMMENDATION FOR SREB CONSIDERATION

Testing companies say that tests are only as good as how they are used--that tests are just "pieces of paper with questions on them" (Saterfiel, 1992). That is only partially true. Testing companies have a responsibility not just to produce valid tests, but to use all means available to solve the problems of wide scale testing. There is a known problem with current tests of critical thinking. Performance-based tests mean high cost and low feasibility. Current measurement design says there is no solution to this problem. This Task Force simply doesn't believe that.

What is needed is an instrument(s) with acceptable levels of both validity (i.e., assurance that students' critical thinking skills are actually being measured in ways that can be directly related to curricular change) and feasibility (in terms of the time, effort, and expense required to administer, serve, interpret, and utilize the data). While no currently available instruments fit our needs or the needs of other individuals who have also studied the assessment of critical thinking (Blai, 1992, Rock, 1991, Nummedal, and others), the Task Force offers several ideas to be explored regarding current instrument combinations, changes, or future instrument development.

1. *Increase the validity of current multiple choice tests by using enhanced multiple choice answers and scoring.* For example, have the number of "correct" responses vary, requiring the marking of an answer as "fully correct or incorrect" or "partially correct." Supplement multiple choice responses with short open-ended narratives explaining why correct answers are correct. (See Appendix E)

2. *Increase the feasibility of performance-based approaches by using objectively scoreable, student produced responses.* Scanable score sheets for short answers, for example, might be explored.

3. *Combine the best features of several tests by exploring computer adaptive testing having the student generate answers.* (More extensive ideas are outlined in Appendix E)

Given that educational stakeholders know the problems of large scale testing, we recommend that a SREB study involving faculty from more than one SREB state and from all types of institutions, working with whatever testing firms are interested, begin to solve these problems. We further recommend that a meeting, or series of meetings be held to explore new testing designs, administrative procedures and usage. At the least, these meetings should include selected academic faculty from two and four year colleges and universities in the SREB region. The group should include at least four groups of people: 1) selected faculty who are knowledgeable in undergraduate curriculum design and the teaching of critical thinking,

2) computer programmers from both education and business who are knowledgeable about software design, game theory, simulations, and cost-effective product development, 3) selected staff from a testing company's development division who are knowledgeable about large scale administration and measurement design, testing theory, and test security, and 4) experts in higher education assessment to facilitate communication and bring to the table the experience of implementing good assessment practices in a higher education setting.

The meetings don't have to be large or expensive. They just need to involve creative people from a variety of backgrounds with a common goal which they all believe is possible. If the meetings prove fruitful, this committee then recommends that a grant be submitted to FIPSE and/or other outside sources to explore further funding possibilities since assessment often drives educational change.

The goal is to change the way we test, particularly in advanced skills such as critical thinking. A change is needed to more curriculum friendly, performance based, easily scored, low cost and diagnostically useful tests--tests that are marketable for testing companies and good for education. We recommend that under the excellent leadership of SREB, testing companies, businesses with the expertise and interest, and educators take the first step by bringing together creative people with a common goal. New solutions will follow.

REFERENCES

*American College Testing. (1995). *CAAP: Assuring Academic Excellence in General Education Skills*. Iowa City, IA.

*American College Testing. (1995). *CAAP Technical Handbook.* Iowa City, IA.

*American College Testing. (1995). *COMP: Using Today's Knowledge to Build a Brighter Future.* Iowa City, IA.

American Philosophical Association. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. The Delphi Report: Research Findings and Recommendations Prepared for the Committee on Pre-college Philosophy*. (ERIC Document Reproduction Service No. ED 315-423).

Bandman, E. L. and B. Bandman. (1988). *Critical Thinking in Nursing*. Norwalk, CT: Appleton and Lange.

Baron, Paul E. and Archie Lapointe. (1995). "Learning by Degrees: Indicators of Performance in Higher Education." Princeton, NJ: ETS Policy Information Center. Sponsored by Pew Charitable Trusts.

Beyer, B.K. (1987). *Practical strategies for the teaching of thinking.* Boston: Allyn and Bacon, Inc.

*Bill 1195. (1996). Higher Education Commission and institutions. South Carolina General Assembly.

*Blai, Boris Jr. (1992). "Assessment of Critical Thinking in Postsecondary Education." ED 351954.

*Brantford, John et. Al. (1986). "Teaching Thinking and Problem Solving." *American Psychologist*, 41.10, 1078-1089.

Brookfield, S. (1987). *Developing Critical Thinkers: Challenging Adults to Explore Alternative Ways of Thinking and Acting*. San Francisco, CA: Jossey-Bass.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Chicago: Regnery.

Dressell, P. L. and L. B. Mayhew. (1954). *General Education: Explorations in Evaluation.* Washington, D.C.: American Council of Education.

Ennis, R. H. (1985). "A Logical Basis for Measuring Critical Thinking Skills." *Educational Leadership*, 43(2), 44-48.

Ennis, R. H. (1987). "A Taxonomy of Critical Thinking Dispositions and Abilities." In J. B. Baron and R. J. Sternberg, (eds.), *Teaching Thinking Skills: Theory and Practice* (9-26). New York: W. H. Freeman & Company.

Ennis, R.H. (1993). "Critical Thinking Assessment." *Theory into Practice* 32.3 (Summer), 179-186. EJ 473735.

Ennis, R.H. and J. Millman. (1985). *Cornell Critical Thinking Tests. Level X and Level Z.* Pacific Grove, CA: Midwest Publications.

Facione, P.A. (1989). A Critical Thinking Bibliography with Emphasis on Assessment. Placentia, CA: California Academic Press.

*Facione, P.A. (1990). "Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction." Newark, DL: American Philosophical Association. ED 315 423.

Facione, P.A. (1990). *The California Critical Thinking Skills Test - College Level. Technical Report #1. Experimental Validation and Content Validity; #3 Gender, Ethnicity, Major, CT Self-Esteem, and the CCTST.* Millbrae, CA: California Academic Press.

Facione, P.A. (1991). *Using the California Critical Thinking Skills Test in Research, Evaluation, and Assessment.* Millbrae, CA: California Academic Press. ED 337498.

Facione, N., Facione, P. A. and Sanches, C.A. (1994). "Critical Thinking Disposition as a Measure of Competent Clinical Judgment: The Development of the California Critical Thinking Disposition Inventory." Journal of Nursing Education, 33(8), 345-350.

Fisher, Alec E. (1988). *The Higher Studies Test: Report Prepared for the University of Cambridge Local Examinations Syndicate.*

Fullan, Michael. (1993). *Change Forces: Probing the Depths of Educational Reform.* The Falmer Press.

Halpern, Diane F. (1992). "A Cognitive Approach to Improving Thinking Skills in the Sciences and Mathematics." In Diane F. Halpern (ed.), *Enhancing Thinking Skills in the Sciences and Mathematics*, Hillsdale, NJ: Lawrence Erlbaum Associates.

"Governor's Summit May Agree to State Comparisons Competition." (1996, April). Vocational *Education Weekly.* 1, 3.

Halpern, Diane F. (1993). "Assessing the Effectiveness of Critical-Thinking Instruction." Journal of General Education 42.4: 239-254. EJ 476384.

Hart, K. A. and M. K. Joscelyn. (1989). "Assessing Growth in Thinking in College Courses: A Caveat." *Accent on Improving College Teaching and Learning #4.* Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

Hickman, J. S. (1993). "A Critical Assessment of Critical Thinking in Nursing Education." *Holistic Nursing Practice*, 7(3), 36-47.

Jacobs, Stanley S. (1994). "Technical Characteristics and Some Correlates of the California Critical Thinking Skills Test, Forms A and B." AIR Annual Forum Paper. ED 373631.

Klassens, E. L. (1988). "Improving Teaching for Thinking." *Nurse Educator*, 13(6), 15-29.

Lipman, Matthew. (1989). "Critical Thinking: Some Differences of Approach." In Robert Michael Esformes (ed.), *Inquiry: Critical Thinking Across the Disciplines*. Volume III:2, March 1989. Institute for Critical Thinking, Montclair State College, Upper Montclair, NJ.

Marchese, Ted. (1996). "The Search for Next-Century Learning." American *Association of Higher Education Bulletin*, (48) 7, 3-5.

McMillan, J. H. (1987). "Enhancing College Students' Critical Thinking: A Review of Studies." *Research in Higher Education*, 26, 3-29.

McPeck, J. E. (1981). "Critical Thinking and Education." New York: St. Martin's Press.

Miller, Deborah A. et. al. (1993). "Critical Thinking in Preclinical Course Examinations." Academic Medicine 68.4 (Apr.): 303-305. EJ 462752.

Miller, M. A. and Malcolm, N. S. (1991). "Critical Thinking in the Nursing Curriculum." *Nursing and Health Care*, 11(2), 67-73.

National Council for Excellence in Critical Thinking Instructions. "Draft Statements on Critical Thinking." Santa Rosa, CA: Foundation for Critical Thinking.

Norris, S.P. (1985). Synthesis of research on critical thinking. *Educational Leadership*, 42(8), 40-45.

Norris, Stephen P, and Robert H. Ennis. (1989). *Evaluating Critical Thinking*. Pacific Grove, CA: Midwest Publications.

*Nummedal, Susan G. "Designing a Process To Assess Higher Order Thinking Sills in College Graduates: Issues of Concern." Washington, DC.: National Center for Educational Statistics. ED 340 761

Olsen, Scott A. (1990). "Examining the Relationship between the CAAP Critical Thinking Test and the COMP." Paper presented at American Educational Research Association. ED 326576.

*Osterlind, S. & Merz, W. (1990). *College BASE Technical Manual*. Columbia, MO: University of Missouri Center for Educational Assessment.

*Paul, R. and Gerald M. Nosich. (1991). "A Proposal for the National Assessment of Higher-Order Thinking at the Community College, College, and University Levels." Washington, DC.: National Center for Educational Statistics. ED 340 762.

Paul, R. W. (1993). "Critical Thinking: What Every Person Needs to Know to Survive in a Rapidly Changing World." Rohnert Park, CA: Center for Critical Thinking.

Paul, Richard. (1992). "Critical Thinking: What, Why, and How." In Cynthia A. Barnes (ed.), *Critical Thinking: Educational Imperative*, New Directions for Community Colleges, No. 77. San Francisco: Jossey-Bass.

*Pike, G. (1996, March-April). "Assessing the Critical Thinking Abilities of College Students." In Banta, T. (ed.) *Assessment Update*, 8(2), 10-11.

*Riverside Publishing Company. *College Base Guide to Test Content.* Norcross, GA.

*Riverside Publishing Company. (1989). *Presenting the College Base.* Norcross, GA.

*Rock, D. A. (1991). "Development of a Process To Assess Higher Order Thinking Skills for College Graduates." Washington, DC.: National Center for Educational Statistics. ED 340 765.

Saterfiel, T. (1992). "Address to Tennessee Higher Education Commission". Knoxville, TN.

Scriven, Michael. (1989). "Critical Thinking and the Concept of Literacy." *Informal Logic*, Spring.

Scriven, Michael. (in press). *Defining and Assessing Critical Thinking.*

Smith, F. (1993). *To Think.* New York: Teachers College Press.

Sormunen, Carolee, and Marilyn Cahlupa. (1994). "Critical Thinking Skills Research: Developing Evaluation Techniques." *Journal of Education for Business* 69.3 (Jan.-Feb.): 172-177. EJ478876.

Steele, J. M. (1994). *Comparing Measures of College Outcomes of General Education.* Iowa City: American College Testing.

*Steele, J. M. (1995). Tasks for Critical Thinking: When is a "Problem" Just an "Exercise?" Workshops at the 15th Annual Critical Thinking Conference at Sonoma State University, Iowa City, Iowa.

*Steele, J. M. (1995). *Postsecondary Measures of Reasoning and Critical Thinking.* Iowa City: American College Testing.

*Sternberg, Robert J. (1986). *Critical Thinking: Its Nature, Measurement, and Improvement.* Washington, DC: National Institute of Education.

"Testing Thinking (Open to Suggestion)." (1990). *Journal of Reading* 33.5 (Feb.): 380-81. EJ403706.

*Venezky, Richard L. (1991). "Assessing Higher Order Thinking and Communication Skills: Literacy." Washington, DC.: National Center for Educational Statistics.

Watson, G, and E.M. Glaser. (1980). *Watson-Glaser Critical Thinking Appraisal.* Dallas, TX: Psychological Corporation.

Werner, Patricia Holden. (1991). "The Ennis-Weir Critical Thinking Essay Test: An Instrument for Testing and Teaching (Test Review). *Journal of Reading* 34.6 (Mar): 94-95. EJ422614.

White, Edward M. (1991). "Assessing Higher Order Thinking and Communication Skills in College Graduates Through Writing." Washington, DC.: National Center for Educational Statistics. ED 340 767.

Yinger, R. J. (1980). "Can We Really Teach Them to Think?" In R. E. Young (ed.), *Fostering Critical Thinking*. 11-31. San Francisco, CA: Jossey-Bass, Inc.

* denotes references reviewed in-depth by the SCHEA Task Force on Critical Thinking Assessment

APPENDICES

APPENDIX A

LITERATURE REVIEW

APPENDIX A
LITERATURE REVIEW

The process of critical thinking has been discussed by educators, psychologists, and philosophers. Most of today's definitions continue to embrace the long standing definitions of critical thinking formulated by researchers in the field of education (McPeck, 1981; Ennis, 1985; Norris, 1985; and Beyer, 1987). These early researchers had little difficulty in identifying the higher order thinking and reasoning skills we see today as critical thinking, and their diverse definitions have provided guidance and direction to today's current study of the subject.

In 1933, Dewey studied thinking and distinguished between the product of one's thinking and the process of that thinking. He used the term "reflective thinking" and linked it with mental activities to reach a conclusion. Reflection also was noted in McPeck's (1981) discussion of critical thinking. He noted that critical thinking meant not only raising questions, but also using reflective skepticism. According to McPeck, there is a linkage between knowledge and expertise. Even though critical thinking involves knowledge and skills, a person may have these skills in one area and not another.

A major contribution to the understanding of critical thinking was noted in 1988. A panel of 46 theoreticians was assembled by the American Psychological Association (APA) to formulate a cross-disciplinary definition of critical thinking. This definition focuses more on attitude and mental stances rather than on how this thinking process is conducted. The APA definition provides understanding of the concept of critical thinking from a more philosophical perspective:

> The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. (APA, 1990, p. 3)

Looking at definitions from the literature and discussing their relevance to our purpose, this task force developed its working definition of critical thinking. The task force defined critical thinking as a reflective, systematic, rational, and skeptical use of cognitive representations, processes, and strategies to make decisions about beliefs, problems, and/or courses of action (adapted from Beyer, 1987; Ennis, 1985; Facione, 1990; McPeck, 1981; and Sternburg, 1986). This definition provided the direction for assessing and evaluating tools available for the measurement of critical thinking.

MEASURING CRITICAL THINKING

Today's researchers of critical thinking tend to agree on the importance of this essential thinking process; however, there is little consensus on "how" critical thinking can be taught and measured. A key problem to the critical thinking dilemma is differing conceptualizations of critical thinking. While the APA definition does provide a cross-disciplinary view of critical thinking, many institutions are establishing their individual definitions based upon mission statements and goals.

Is there ONE tool to assess critical thinking? Most experts would say NO. Facione, Facione, and Sanchez (1994) reported that many methods of assessment of critical thinking would be preferable. But, is it realistic to incorporate multiple tools to assess the process of critical thinking?

Today, we have available many tools for the assessment of critical thinking, namely the Cornell Critical Thinking Test, the California Critical Thinking Skills Test, the CAAP Critical Thinking Test, the COMP Assessment of Reasoning and Communicating, the Ennis-Weir Critical Thinking Test, the College Basic Academic Subjects Examination, the ETS Skills Test, the Academic Profile II, and the Watson-Glaser Critical Thinking Appraisal. All of these tools do not market themselves as complete tools for assessing critical thinking, but they all test components or some part of critical thinking. While each of these tools possesses value in its focus on assessment, noted differences do exist. These differences can be seen in pages 9-13 describing the task force's assessment process.

Various authors have looked at assessment of critical thinking. Nummedal (1991) conveyed her reservations about existing tests and noted these tests to be "substantially flawed" (p. 10). She noted that the skills being evaluated in these tools were derived from a definition of critical thinking that was too narrow and defined as specific to a certain discipline.

Rock (1991) had a similar view of measuring critical thinking. He saw the need to develop critical thinking, but this task, he admitted, is difficult since there is little consensus on teaching or assessing critical thinking. Rock continued by identifying the problem of student responses on "extended free response items" on diagnostic tests. He concluded that inadequate student responses did not provide the needed diagnostic information.

Another view of critical thinking is seen in the comments made by Blai (1992). He identified several problems with assessment of critical thinking. He noted that intelligence tests do not address critical thinking in specific areas of knowledge. Also, he noted validated tests do not measure the open-ended problem solving specific to the critical thinking process.

These views mirror the opinion of most of those who are knowledgeable about critical thinking: no ideal tool for assessing critical thinking currently exists.

APPENDIX B

COPY OF SURVEY USED IN SOUTH CAROLINA

AND

SURVEY RESULTS AND SUMMARY

APPENDIX B1

SURVEY ON CRITICAL THINKING
from the SCHEA Critical Thinking Task Force

The SCHEA Critical Thinking Task Force was asked by the Southern Regional Educational Board and the Commission on Higher Education to review several Critical Thinking Evaluation Instruments. In order to do this effectively, we need your help. Please take some time to think about the following questions. A SCHEA Task Force member will be contacting you either by person, phone or E-mail in the next several weeks so you may discuss your answers with him/her. If you have any questions or comments, please do not hesitate to ask the Task Force member. Thank you for your time and thoughts on this issue.

1. Is Critical Thinking one of the expected educational outcomes for your students?

2. Have you defined Critical Thinking and, if so, how is it defined?

3. Do you assess Critical Thinking and, if so, what do you use to measure it? That is, what is your methodology used to measure Critical Thinking?

4. When do you measure Critical Thinking? (Upon entry, exit, junior year . . .)

5. How do you measure Critical Thinking? (Pre-post, capstone courses, outside examiners . . .)

6. What is the length of time your institution has used an instrument (test, survey, activity . . .) or a specific method to measure Critical Thinking?

7. Why did you choose that method or instrument? What were the strengths and weaknesses of the method or instrument? Please be specific. Use the back of this page if you need more space to jot down your notes.

8. What has been the impact of using this method and/or instrument?

9. If you had to do it over again, would you use this same instrument or method? Why?

10. If you could change one or two things about this method or instrument, what would they be and why?

APPENDIX B2
SURVEY RESULTS
AND SUMMARY

Responses to the Task Force's critical thinking assessment survey were received from 25 institutions or 42% of all South Carolina public and private colleges and universities. Critical thinking was cited as an expected educational outcome by eighty-eight percent (88%) of institutions responding to the survey. One additional institution reported that critical thinking was presently being included in proposed general education revisions, leaving only eight percent (8%) of respondents--or two institutions--that do <u>not</u> have critical thinking as an expected educational outcome.

Despite the high percentage of South Carolina colleges and universities that cite critical thinking as an expected outcome for students, most institutions (64%) responding to the survey have not yet defined what exactly is meant by critical thinking. Nine institutions reported having such a definition. Another three reported that a definition was, at the time of the survey, being developed for their institutions. Survey responses also suggest that critical thinking may be defined in particular departments, majors, and/or classes without being addressed on an institution-wide basis.

A total of sixty-four percent (64%) of respondents indicated that they assess student critical thinking, with one additional institution reporting plans to conduct such assessment in the near future. However, nearly one-third of colleges and universities with critical thinking assessment procedures in place reported that assessment is conducted in individual classes or majors but not on an institution-wide basis.

The most common reported means of assessing critical thinking in South Carolina's colleges and universities consists of surveys. Various types of surveys, including those conducted with students, alumni, faculty, and employers were reported.

Institution-wide use of standardized tests to assess student critical thinking was reported by thirty-six percent (36%) of South Carolina institutions responding to the survey (or 56% of those with programs to assess critical thinking). Four institutions reported using the College BASE for this purpose, with another four using the Academic Profile and one using the ACT COMP. Other respondents indicated that standardized tests (e.g., Watson-Glaser) or locally developed examinations are used to assess critical thinking in particular majors or classes.

Seven institutions reported having other locally developed methods of assessing critical thinking. The most common of these methods consists of assessing written student work such as essays and papers and/or collections of written work such as portfolios. In some institutions these products are evaluated using criteria that explicitly address critical thinking, but in most cases critical thinking is an *implicit* part of the required activity and grading. One university indicated that critical thinking is part of some activities included in an institutional classroom assessment project to assess general education. Other institutions reported using a variety of methods such as simulations, individual and group projects, complex problem-solving exercises, and similar means to assess critical thinking in individual courses and/or majors. One college has designed and implemented a model that encourages systematic inclusion and assessment of problem-solving and other skills in individual majors and courses. This approach integrates course content and work-related skills.

Respondents with locally developed assessment methods embedded in courses and/or majors reported considerable impact on teaching, learning, and further assessment. Additional advantages cited by these respondents included greater faculty acceptance of such methods and greater relevance to, as well as increased likelihood of impact on, curricula. Difficulties in generalizing results across courses/disciplines, not having a "standard" to compare results to, and lack of demonstrated validity were mentioned as drawbacks.

Most institutions, including all but one of those using institution-wide, standardized national examinations reported that assessment methods have not been in place consistently or long enough to have had much impact to date. One respondent indicated that results of national examinations" have prompted discussions and increased faculty awareness regarding general education goals but have provided few other benefits."

APPENDIX C
CRITERIA

APPENDIX C1

COMPLETE COPY OF CRITERIA

On February 15, the SCHEA Critical Thinking Task Force adopted the following working definition and statement of purpose, as well as criteria for critiquing and selecting test(s) of critical thinking to recommend to the South Carolina Commission on Higher Education. The criteria have been divided into functional categories, i.e., "primary criteria"--those which directly affect the validity of tests for our purpose, and "secondary criteria"--those which do not directly affect validity, but which might afford additional advantages or disadvantages, once the validity of particular tests is assured at a satisfactory level.

Definition: Critical thinking is a reflective, systematic, rational, and skeptical use of cognitive representations, processes, and strategies to make decisions about beliefs, problems, and/or courses of action. (Adapted from Beyer, Ennis, Facione, McPeck, Sternburg, et.al.)

Purpose: To assess college students' critical thinking knowledge and skills as an outcome of higher education course(s) of study.

RECOMMENDED PRIMARY CRITERIA

A valid critical thinking (CT) measure should be . . .

A. Relevant, i.e., it should . . .

1. directly measure CT knowledge (require student production of answers, not just recognition of correct answers from true-false, matching, or multiple-choice lists); and
2. require demonstration of CT skills (by performance or performance simulation which could still be paper and pencil task); and
3. be at an appropriate level and range for college students' skills; and should
4. reflect the continuum of knowledge and skills taught in most S.C. college curricula; and should
5. include appropriate results comparisons (i.e., a representative national or regional sample for norm-referenced tests and/or authoritative mastery standards for criterion-referenced tests).
6. It should also clearly describe technical characteristics adequate for our task; i.e., reliability (consistency over time and/or inter-rater, as appropriate), criterion validity (predictive to real world CT tasks, concurrent with other CT measures, etc.), and appropriate numbers and types of norm groups, etc. Efficacy studies with college students and curriculum improvements are especially important.

B. Accurate, i.e., it should . . .

7. provide analytical as well as global scores (measure the "core components" of CT); and
8. afford formative as well as summative score(s). (Tell how students attain their answers, not just the correctness/incorrectness of their answers); and
9. have sufficient specificity of score reporting to show significant differences in CT competency levels (e.g., not just stanines or quartiles).

C. Diagnostically useful, i.e., it should . . . .

10. show student strengths and weaknesses referenced to a clear CT model or system which
11. relates directly to educational program strengths and weaknesses (= internal validity); and
12. affords maximum generalizability of results (=external validity).

D. RECOMMENDED SECONDARY CRITERIA

Once valid measure(s) have been identified by the primary criteria, it would be desirable if they were also . . . .

1. easily incorporated into higher education programs
2. economical in resource requirements (i.e., costs in time, effort, and money for students, faculty, and institutions, especially in "front-loading"--the time and effort to prepare items and administer the test, and "back-loading"-- the time and effort to score, interpret, and get useful results from the test).
3. quickly and/or locally scorable,
4. clear and easy to administer, score, and interpret, and
5. interesting to students

In applying these criteria for evaluating the potential appropriateness of specific measures for our purpose, all Task Force members should be consistent in the following:

- Rate each of the four sections of these criteria (i.e., A=Relevance, B=Accuracy, C=Diagnostic Utility, D=Secondary Criteria), citing the subsections as "strength" and "weakness" guidelines.
- Use the following six point rating scale for each section (and subsection, if you're really detailed in your approach); 5=Excellent/Outstanding, 4=Very Good/Above Average, 3=Good/Satisfactory, 2=Fair/Somewhat Unsatisfactory, 1=Poor/Unsatisfactory, 0=Not Addressed/Very Unsatisfactory.
- When you've finished assigning ratings to your four sections (maximum possible score = 20 pts.), get a weighted overall score by multiplying your Section A (Relevance) Score by 8, your Section B (Accuracy) Score by 6, your Section C (Diagnostic Utility) Score by 4, and your Section D (Secondary) Score by 2. This will yield a maximum overall rating of 100, and help discriminate among similar ratings. (If some aspect of this weighting system proves problematic, we will adjust it accordingly.)
- Remember to keep your ratings to yourself until your "secondary rater" has had a chance to complete his/her task independently.

APPENDIX C2
SCORING SHEET

SCHEA CRITICAL THINKING MEASURE

SCORE/COMMENT/SUMMARY SHEET

CRITERIA	POINTS	NAMES OF MEASURES		
1. Direct Knowledge	/10			
2. Skill Demonstration	/10			
3. College Appropriate	/5			
4. Curriculum Relevant	/5			
5. Results Comparison	/5			
6. Technical Chars.	/5			
I. RELEVANCE SUM = Comments:	/40			
7. Analytical Scoring	/5			
8. Formative Scoring	/10			
9. Detailed Scoring	/5			
II. ACCURACY SUM = Comments:	/20			
10. Student S&W Profile	/10			
11. Ed. Prog. S&W Profile	/20			
12. Generalizability	/10			
III. DIAGNOSTIC SUM = Comments:	/40			
TOTAL VALIDITY SCORE =	/100			

A. Ease of Program Use	/40			
B. Low Time & Cost	/30			
C. Ease of Adm./Scoring	/20			
D. Interesting to Students	/10			
FEASIBILITY SCORE =	/100			

Comments:

APPENDIX D
INDIVIDUAL TEST REVIEWS

(Please note that because reviews were done by different committee members there is no standardized review form. However, in all instances the strengths and weaknesses of the instruments are described and reviewed.)

APPENDIX D1

REVIEW OF COLLEGE OUTCOME MEASURES PROGRAM(COMP), ASSESSMENT OF REASONING AND COMMUNICATING EXAMINATION (ARC).

This review will focus primarily on the aspects of the test relevant to assessing critical thinking/reasoning as defined in the Revised Criteria (2/16/96).

Published by: American College Testing

Developed by: American College Testing through a FIPSE grant which involved college faculty in the field who developed the format, criteria, and question types.

Original Publication Date: 1986 (In its current format); Updated forms (every three years)

Forms Available: Except where otherwise noted this review will concentrate on the ARC (Assessment of Reasoning and Communicating Examination). This reviewer reviewed the specimen copy of Form IV which is a non-secure form of the test. Secure forms in current use are Forms X, XI and XII.

Format: The Assessment of Reasoning and Communicating was derived from the COMP Composite Exam (a 4.5 hour administration time instrument). It consists of six simulation activities, three requiring written responses and three requiring oral responses. A basic philosophy underlying ARC is that critical reasoning must be communicated to be observable/measurable and that most communication takes place through writing and speaking.
The questions used for the COMP ARC are based on "real life" rather than on discipline or content-based outcomes.

Administration Time: Writing Skills Assessment is 80 minutes.
Speaking Skills Assessment is 15 minutes in a laboratory.
Total Administration Time is 2 hours.

Scoring: ARC is evaluated using rating scales which have been developed "with the assistance of college instructors." COMP provides rater trainer materials, including prescored writing and speaking samples, and specific criteria for the essays and speeches, to allow institutions to evaluate ARC locally. The scoring is criterion driven; that is, "set" criteria are used to judge if a student should be at Level A through Level E (with E being the lowest rating). Level A might require, for example, that a student state multiple problems with the viewpoint expressed and more solutions on how to solve or face those problems. After training, raters usually take about 45 minutes per student to evaluate the ARC assignments. COMP provides a rescore service for checking rater reliability.

Types of Scores Produced: There are student scores as well as an institutional composite package. The following is taken from the **COMP GUIDE**: " The ARC yields a **total score and three subscores in the following areas:**

> *Reasoning:* Can analyze social, scientific and artistic problems; generate solutions to problems; and analyze values and implications of resulting decisions. (This subscore is a combination of scores for two COMP process areas: Solving Problems and Clarifying Values). Reasoning scores can be plotted on the ARC scoring graph in the following areas: total reasoning and communicating, total reasoning, social reasoning, scientific reasoning and artistic reasoning.
>
> *Writing:* Can make and sustain contact with a relevant audience; organize a persuasive message that develops a number of relevant ideas; and present ideas clearly using correct and lively language to support an argument. Writing scores can be plotted on the ARC scoring graph in the following areas: total writing score, audience score, organization score and language score.
>
> *Speaking:* "Can make and sustain contact with a relevant audience; organize a persuasive message that develops a number of relevant ideas; and present ideas clearly without hesitation and with energy and variety in voice quality." Speaking can be plotted on the ARC scoring graph in the following areas: total speaking, audience, discourse and delivery.

If ACT scores the ARC, or if trained faculty scores ARC (using ACT criteria) and the scores are submitted to ACT, there are extensive data files that may be used for comparison to similar institutions in the nation on such things as score differences in gender, race and age.

The scoring is criterion based; that is, the criteria and scoring rubrics are set as external to the test or test takers. The ARC renders individual student scores as well as program information. The comparison data which can be used to compare like institutions is on a normative scale.

General Description of the Test: For the writing test, the student is given 20 minutes to write a letter or a memo in response to listening to a 2-3 minute tape which consists of excerpts from public radio, e.g., a taped selection of music, a part of a lecture, or some other radio segment. The excerpts provide a context from which the student responds to critical societal issues. The three written responses call for writing in each of three areas: social sciences, natural sciences and the arts/humanities. The student has 20 minutes to write after the tape ends. The student may take notes during the tape and may use the notes and his own knowledge of the subject to do the writing assignment. The content for one letter is a social science topic; for another, a technological topic; and for the third, a fine arts topic. Tasks measure "whether a person can develop a convincing written argument, elaborated and designed to address the concerns of a particular audience" (from ACT, *COMP Using Today's Knowledge to Build a Brighter Future).*

For the three oral-response questions, the student is given the instructions at least one day prior to taking the test, is told to think about and prepare brief notes (one 3" x 5" card) and go to a lab at her convenience or at a scheduled time to make the speech. The speech format consists of the student role playing a presentation to an audience of one, a second presentation to a small group, and a third presentation to a large group. Written stimulus materials provide a context for the role playing activities. The student is expected to develop a persuasive oral discourse, with examples and elaboration delivered in a way that supports the communication. This assessment is usually administered to groups in a language lab setting so presentations can be taped for later playback to faculty raters. (from ACT, *COMP Using Today's Knowledge to Build a Brighter Future).*

Relevance:

1. Directly measures CT knowledge: The ARC is especially appropriate "where emphasis across the curriculum is on reasoning and communicating. The assessment models the production of ideas rather than selecting the right answers."

In constructing the COMP ARC, fundamental assumptions were made: it should assess higher order cognitive skills using realistic problems rather than discipline or content based outcomes; it should go beyond the assessment of academic information (recall) and assess the ability to apply skills and knowledge useful in civic, volunteer, and on-the-job roles outside of the classroom; problem settings should focus on fundamental social issues such as energy, pollution, waste disposal, housing, transportation, etc.

2. Requires demonstration of CT skills: The ARC directly measures critical thinking through the production of answers which may vary from student to student. A student must write from a listening prompt and speak from a written prompt. COMP ARC "measures the ability to analyze a variety of problems, generating logical and reasonable approaches to solve and implement solutions, that reflect a consistent value orientation and a recognition of conditions and perspectives that would support or conflict with such actions." *Postsecondary Measures of Reasoning and Critical Thinking, Steele, April 7,1995.*

3. *Is at an appropriate level and range:* The ARC has been successfully used to measure growth from freshmen to senior exit in the area of critical thinking; however, most of its use has been in the Midwest and West. Little broad based use has been in the Southeast.

4. Reflects a continuum of knowledge and skills: Although there may be a continuum of knowledge and skills measured by the COMP ARC, there is little evidence that the application of those critical thinking skills are taught across the curriculum or by "real-life" examples in the South Carolina higher education institutions. Critical thinking may be taught from more of a "content or discipline base" in S.C. institutions.

5. Includes appropriate results' comparisons: The original criteria for the COMP were set by a group of over 100 higher education practitioners including Trudy Banta working through a FIPSE grant awarded to ACT to develop " a criterion referenced, application oriented, general education exit exam." There are many studies attesting to both national norming data and institutionally based criterion data.

There are many nation-wide norms for the COMP ARC including those for freshmen, seniors, for age, gender and institutional type. This reviewer did not find any comparative norms based on race.

6. Technical Characteristics e.g., norming groups, efficacy studies, predictive to real-world tasks, etc. COMP ARC fairs well in technical characteristics. However, some institutions have reported a maturation function on the COMP scores. Although COMP national studies do not bear this out, it is extremely difficult to ignore that for adult students their experience with real-life problems before entering the classroom often might affect COMP ARC scores particularly in reasoning. *Also see below: Diagnostically useful*

Accurate:

7. Provides analytical as well as global scores: In the 10 years or so of the COMP's use, there is extensive reliability and validity data on the instrument. In addition, an examination of relationships between communicating and reasoning skills was based on an analysis of data for 1589 freshmen from 13 institutions and 1366 seniors from 10 institutions. The correlation coefficients between the scales of the instruments show that distinctly different skills are measured by the instrument. Clearly the reasoning skills measure something different than the writing skills measure and the speaking skills measure.

8. Affords formative as well as summative score(s): This may be a weak area for the COMP ARC and S.C. institutions as it is often difficult to increase COMP ARC scores without broad based changes in the curriculum as opposed to singular course changes.

In addition, although the writing, speaking and critical reasoning scores can be both formative and summative, the subscores in each of the areas may not be differentiating the individual skill finely enough for an institution to know what to change. For example, in speaking, discourse may be too highly correlated to delivery at a specific institution to differentiate the skills that should be taught. Nonetheless, low scores in speaking (overall) would indicate that more speaking should be integrated into a curriculum.

9. Has sufficient specificity of score reporting: Although the score reporting is more difficult to understand perhaps than a standard, multiple choice score report, it renders both individual and group information in the standard scoring service.

Diagnostically useful:

10. Shows student strengths and weaknesses: Renders student and group level scores. Shows student strengths and weaknesses relative to other students in a chosen peer group of institutions. Therefore, although the actual scoring of the COMP ARC is criterion based, the reporting of the scores is on a normative, comparison scale.

11. Relates directly to educational program: There is much discussion as to whether COMP relates directly to the traditional academic content-oriented curriculum since its items are so distinctly application-oriented using real life examples. Therefore, much discussion is centered around what happens when a specific COMP ARC score is low. The question usually asked by faculty is which course(s) should we change to improve a low score on the ARC? COMP does not directly answer that question. Rather, low COMP ARC scores would indicate that an entire weak area should be attended to across the curriculum as a whole; that is, all course(s) would systematically increase their use of application problems in classes, on in class tests, and in exercises, thereby increasing the teaching and learning of the weak area, and as a by-product higher COMP ARC scores should result. This is often far more difficult for a faculty to do in a large, "open" general education curriculum than in one in which there is a set progression of courses.

12. Affords maximum generalizability of results: Although ACT has information on how institutions have used COMP ARC scores to strengthen their curricular offerings, the examples tend to be from small four year teaching oriented institutions where there has been high faculty buy-in and involvement in using the scores for improvement. The ARC can show both longitudinal growth information and point-in-time evaluation information.

D. RECOMMENDED SECONDARY CRITERIA

1. EASILY INCORPORATED INTO HIGHER EDUCATION PROGRAMS: The results of the COMP ARC are not easily used in many higher education curricula because most curricula have not identified which courses are teaching their critical thinking skills and which are not. Therefore, to increase scores on the COMP ARC one is left with the two solutions: 1) increase the teaching of all critical thinking in all general education courses making it an "across the curriculum" strategy, or 2) identify specifically by score differentiation of students exactly which courses are best teaching critical thinking skills and a) requiring those courses of all students or b) increasing the teaching of critical thinking in other courses in which the majority of the students enroll, with the subsequent result of teaching across a "specified" curriculum. In either case, low COMP ARC scores require a genuine willingness on the part of the faculty to make broad based changes in the curriculum.

2. ECONOMICAL IN RESOURCE REQUIREMENTS: The COMP ARC's greatest weaknesses for broad-base application in S.C. institutions of higher education are (as evaluated by this rater) in this area. COMP ARC is not a simple test to administer since it involves audio and taped responses. In addition, it cannot be easily integrated into classroom administration both because it is too long--2 hours--and because its administration requires specific equipment and singular student taping abilities. This adds to a major drawback, namely getting the faculty involved in the testing so there is strong "faculty ownership" and therefore an increased chance for faculty use once the scores are returned.

If students are not sufficiently "prepped" for the COMP ARC, there is often great student confusion regarding the relevancy of the exam to their curricular experience since both the administration of the COMP ARC and the actual items are not usually experiences in the test taking lives of students. Therefore, to be most successful, institutions have reported substantial front-loading and faculty have reported substantial back-loading (if the exam is locally scored) which is most often the choice because of price.

Compared to multiple-choice, machine scored tests, the cost is often prohibitive with the COMP ARC costing $15.75 per student up to 500 students, with graduated breaks up to 2501 or more students costing $9.00. This scoring does NOT include ACT rating (scoring) of the written or speaking portions of the exam. Rather, it only includes the test use and analysis of scores (rendered by trained on-campus or centralized faculty).

APPENDIX D2
REVIEW OF COLLEGIATE ASSESSMENT OF ACADEMIC PROFICIENCY (CAAP)

Published by: American College Testing

Developed by: American College Testing. Items are developed by external item writers who teach at the college level. Deans of curriculum are often contacted to ask faculty who teach problem solving or critical thinking on their campuses at the college level to write items. No items are generated at ACT; however, items are "cleaned up" at ACT, e.g., biased language, incorrect punctuation, etc. are corrected, then the item is returned to the item writer to see if the changes have altered anything essential. The item writer gives the final O.K. If a consensus cannot be reached, the item is not used. One item writer generates only a few items and item writers are selected from all over the nation.

Original Publication Date: 1989

Forms Available: Form 88B is available and was reviewed. Concurrent forms are available for pre and post testing.

Format: The CAAP is a multiple choice, 32 item test with an additional 9 items which can be developed by local faculty and which ACT will score and on which they will run some comparative data. Some campuses use these 9 items to formulate their own critical thinking questions, and some use the nine items to get more data on the individual student (for example, how many hours of math a student has taken excluding remedial courses).

Administration Time: 40 minutes. CAAP can easily be integrated into a classroom hour.

Scoring: Individual single number scores are produced from an objective test scoring process. The Writing Essay (which is an exam in which students write two independent essays in response to two situational prompts) is scored on a "modified holistic basis on a 1-6 scale in increments of .5 with set "criterion" used in the scoring of the essay." From: *Assuring Academic Excellence in General Education Foundational Skills*, ACT.

For all CAAP Tests, there is an individual score report which tells the individual how he/she has done in comparison with all other students who have taken the CAAP test that the student took, e.g., reading, math, or critical thinking for that norming period. The score report gives some suggestions regarding what the student can do to work on improving scores. There is an institutional score report which is an extensive package that gives lots of comparison information, grouping students into different groups (gender, age, type of institution norms. etc.).

The most recent norms are for Fall, 1995. Almost 20,000 students from 157 institutions from all over the nation took the Critical Thinking Test. The mean was 61.5 with a S.D. of 5.3. Reference groups from two year public, two year private, 4 year public and 4 year private are provided as well as information regarding demographics (self-reported race, full/part time status, gender and geographic origin).

Types of Scores Produced: Individual Student
Institutional Scores (norms, percentiles)

General Description of the Test: The CAAP Critical Thinking Test is a 32-item, 40 minute test that measures students' skills in clarifying, analyzing, evaluating, and extending arguments. An argument is defined as a sequence of statements which includes a claim that one of the statements, the conclusion, follows from the other statements. The test consists of four passages that are representative of the kinds of issues commonly encountered in a postsecondary curriculum. A passage typically presents a series of subarguments in support of a more general conclusion or conclusions. Each passage presents one or more arguments and uses a variety of formats, including case studies, debates, dialogues, overlapping positions, statistical arguments, experimental results, or editorials. Each passage is accompanied by a set of multiple-choice test items.

The approximate proportions of test devoted to each content area are:

CONTENT CATEGORY	*P. OF TEST*	*# OF*
ITEMS		
Analysis of the elements of an argument	*.62*	*20*
Evaluation of an argument	*.19*	*6*
Extension of an argument	*.19*	*6*

(The above description taken from CAAP *Technical Handbook.*)

The CAAP Critical Thinking Test is a fairly standardized format; that is, there are two reading passages for each set of items. The passages express opposing viewpoints. The student is asked to draw conclusions, infer material, analyze information, etc. There are three passages with about 11-12 questions for each passage. The reader must be familiar with comparing and contrasting viewpoints and must understand logical steps to conclusions. The stems (answer choices) to the questions are often difficult (there is often no obviously wrong answer--that is, the differences between 2, 3 or 4 possible answers are often fine differences.

Because there is an opportunity for faculty to develop their own 9 questions, the CAAP may be a stronger choice than some other multiple choice exams since about a quarter of the exam can be faculty driven (although multiple choice answers must still be generated).

Only right answers are counted. There is no penalty for guessing. This leaves the exam open to the "standard multiple choice criticism."

Relevance:

1. Directly measures CT knowledge: Given our criterion that an exam can't measure critical thinking if it is a multiple choice exam, this exam, as with many other multiple choice objective exams, does not rank highly. However, if a faculty defines the most basic element of critical thinking to be problem solving from a point of view, the exam would do better on this criterion.

2. Requires demonstration of CT skills: Again, the exam measures what it purports to if the definition of critical thinking is similar to the definition defined by the test.

3. Is at an appropriate level and range: This test may be too difficult in reading level for many college students. Therefore the thing being measured would be reading level and not critical thinking. A sample of students might have to be tested with the CAAP Reading test to see if reading levels are sufficiently high to do well on the Critical Thinking Test.

4. Reflects a continuum of knowledge and skills: Items are more or less difficult on a continuum of difficulty. The difficult items may be more difficult for people who recognize and can defend multiple points of view since there is only one correct answer.

5. Includes appropriate results' comparisons: The CAAP reports can be broken into a variety of subgroups. Institutions can choose peer institutions with which to get score comparisons if they so choose.

6. Technical Characteristics e.g., norming groups, efficacy studies, predictive to real-world tasks. etc. CAAP was devised to measure college level outcome skills. It can be directly hooked to the ACT Assessment, ASSET or COMPASS, to show "educational change."

Accurate:

7. Provides analytical as well as global scores: CAAP Critical Thinking does not have any subscores. The best testing model would be to "spiral" the CAAP examinations first. That is, the Reading, Writing, Mathematics, Science Reasoning and Critical Thinking exams would be taken at the same hour across campus, one test each by a different student. Then the scores would have more program information than just critical thinking, and one could see the strengths and weaknesses across the board. If this were the testing methodology chosen, then there would be a total score rendered for each exam and subscores rendered for Writing (usage/mechanics; rhetorical skills); Reading (arts/literature; social/natural sciences); and Mathematics (through College Algebra. Institutions that require achievement through college algebra should use only the Math subscore and not the total score).

8. Affords formative as well as summative score(s): CAAP Critical Thinking as a stand-alone module does not have any subscores.

9. Has sufficient specificity of score reporting: Single numerical score render for each student.

Diagnostically useful:

10. Shows student strengths and weaknesses: It seems as if the reading level of the test is very high; therefore, one would have to review the test with the reading level of students in mind. Because the test does not give subscores, diagnostic information is weaker than one would want.

11. Relates directly to educational program: The major way the critical thinking skills are defined in the CAAP Critical Thinking is through a comparison and contrasting of two different opinions. Since the CAAP is linked very directly to the curriculum (or at least the item writers devised the questions from their own classroom examples and experiences), one could presume that the CAAP is very curricular based. However, this means the items "look" more academic in nature than those of ACT's COMP for example.

12. Affords maximum generalizability of results: There is good external validity.

D. RECOMMENDED SECONDARY CRITERIA

1. EASILY INCORPORATED INTO HIGHER EDUCATION PROGRAMS: The CAAP Critical Thinking Test scores very highly here since it can be administered in a 50 minute time period..

2. ECONOMICAL IN RESOURCE REQUIREMENTS: It is relatively low cost and requires "low maintenance" administration skills.

3. QUICKLY AND/OR LOCALLY SCOREABLE: The CAAP is not locally scoreable except for the CAAP Essay Exam. The score reports take about 3-4 weeks after all tests have been returned to ACT.

4. CLEAR AND EASY TO ADMINISTER, SCORE AND INTERPRET: This is very much in the CAAP's favor. Score interpretation is very simple.

5. INTERESTING TO STUDENTS: It is about the same as any standardized academic, curricular based test such as the GRE, SAT, or ACT.

Major Strengths of the CAAP:

Overall evaluation of the CAAP Critical Thinking Exam is that it is a solid academic multiple choice examination. Its major strength is that it gives good score reports (comparisons by demographics, programs and institutional types) that can be used to make broad changes in a curriculum. For example, if psychology majors were doing much better then philosophy majors, one could study the course taking patterns of both and perhaps ascertain which courses were doing a better job on teaching/learning in critical thinking as measured by the exam. Certain teaching/learning strategies could be implemented in those courses or group(s) of courses that consistently score below a desired benchmark.

An additional strength of the CAAP is that the entire exam is devoted to Critical Thinking so there are enough items to cover the topic and render valid data (if critical thinking is defined on the basis of arguments or opposing viewpoints). By using all CAAP tests (Writing, Math, Critical Thinking, Science Reasoning) an institution could get more information on a variety of General Education processes of which critical thinking is only one; however, even by using the Critical Thinking Exam as a stand-alone much useful information can be gleaned.

Major Weaknesses of the CAAP:

Extensive diagnostic materials cannot be provided by the CAAP since it does not have subscores in the three areas in which it tests content (analysis of elements of an argument, evaluation of an argument and extension of an argument) probably because there are not enough items in the last two to provide subscores. This reviewer suggests approaching ACT to see if the development of subscores could be explored if a state (or several states) wished to partner with ACT to do so.

A student must be a good reader to do well on the critical thinking test as it is "written word" based.

APPENDIX D3

REVIEW OF COLLEGE OUTCOME MEASURES PROGRAM (COMP), OBJECTIVE TEST.

This review will focus primarily on the aspects of the test relevant to assessing critical thinking/reasoning as defined in the Revised Criteria (2/16/96)

Published by: American College Testing

Developed by: American College Testing through a FIPSE grant which involved college faculty in the field who developed the format, criteria, and question types. Original development of the COMP in 1976 and the OBJECTIVE TEST shortly thereafter makes the COMP the "oldest" standardized outcomes exam available. It has been used by hundreds of institutions, and over 100,000 students in the nation have been tested using the COMP OBJECTIVE TEST.

Original Publication Date: 1976 (In existence for 16 years); Updated forms available

Forms Available: Except where otherwise noted this review will concentrate on the COMP OBJECTIVE TEST. This reviewer reviewed the specimen copy of FORM 9 which is a non-secure form of the test. Secure forms are in current use.

Format: The COMP OBJECTIVE TEST was derived from the COMP Composite Exam (a 4.5 hour administration time instrument). It consists of fifteen simulation activities based on real-life stimulus materials drawn from television, radio, magazines, and other media. The activities are NOT "content or academic discipline based," but rather they are taken from real-life situations. For example, a group of questions on the COMP OBJECTIVE may center on a blueprint drawing of an "environmentally sound" house with a short article from Newsweek regarding the house.

The COMP OBJECTIVE TEST outcome areas are Communicating, Solving Problems, Clarifying Values, Functioning within Social Institutions, Using Science and Technology, and Using the Arts. The COMP measures these skills on a 'MATRIX' basis; that is, one item will measure two things, a process and a content area.

Administration Time: 2.5 hours

Scoring: Scored by ACT, usually a 4-5 week turnaround time

Types of Scores Produced: There are student scores as well as an institutional composite package. The following is taken from the **COMP GUIDE:** " The OBJECTIVE TEST yields a **total score and six subtest scores in the following areas:**

PROCESS AREAS:

Communicating: measures the ability to send and receive information (including mathematical calculations) in a variety of modes (oral, written, graphic) for a variety of purposes.

Solving Problems: reflects the ability to define a variety of problems, select approaches to solve them, generate solutions, collect information, check logical consistency, select a good solution, and evaluate the process by which a problem was solved.

Clarifying Values: indicates the ability to identify one's own values and the values of others, understand how values develop, and analyze the implications of decisions made by oneself or others based on those values.

CONTENT AREAS:

Functioning Within Social Institutions: reflects the ability to identify those activities and institutions which constitute the social aspects of a culture, understand their impact on individuals, and analyze the functioning of oneself and others within social institutions.

Using Science and Technology: indicates the ability to identify scientific/technological aspects of a culture, understand their impact on individuals, and analyze the consequences of using technological products for oneself and the culture.

Using the Arts: reflects the ability to identify those activities and products which constitute the artistic aspects of a culture, understand the impact that various forms have on individuals, and analyze the use of works of art by oneself and others.

THE COMP OBJECTIVE TEST IS A PROXY OF THE COMP COMPOSITE EXAM. THE MAJOR DIFFERENCE IS THAT THE OBJECTIVE TEST IS MULTIPLE CHOICE AND THE COMP COMPOSITE (AND THE COMP ARC EXAM) ARE PERFORMANCE BASED, NON-MULTIPLE CHOICE TESTS.

There are extensive data files that may be used for comparison to similar institutions in the nation on such things as score differences in gender, race and age.

The scoring is normative based on comparison groups. That is, the total score and the six subscores are plotted on a graph comparing those scores with other institutions or other student groups. Each student has a score report returned, and the institution has an institutional score report returned. The institutional score report compares the institution in such areas as gender, grade level (freshmen, sophomore and senior), by institutional type, etc. This is a strength of the COMP score reports. There are special reports available on a fee basis.

Multiple norms and comparison groups are available including data on gains from longitudinal studies. The ACT Assessment can be used as the "front-end" measurement and the COMP OBJECTIVE EXAM can be used as an outcome measure.

General Description of the Test: The student listens to a tape and answers objective questions about situations which are drawn from real-life, e.g., magazines, radio, etc. There are fifteen simulation activities. They test problem solving and clarifying values as they occur in social institutions, science and technology and the arts. There are two correct answers and students are penalized for guessing.

Relevance:

1. Directly measures CT knowledge: Although it may measure Critical Thinking in its Solving Problems area, this is a multiple choice form of the exam, and therefore is highly criticized by groups. Because of our definition of this criterion, this would be a weakness for the objective form of the COMP.

The COMP OBJECTIVE EXAM is especially appropriate "where emphasis across the curriculum is on all skills that the COMP measures and not just on Critical Thinking" (Problem Solving).

In constructing the COMP, fundamental assumptions were made: it should assess higher order cognitive skills using realistic problems rather than discipline-or content-based outcomes; it should go beyond the assessment of academic information (recall) and assess the ability to apply skills and knowledge useful in civic, volunteer, and on-the-job roles outside of the classroom; problem settings should focus on fundamental social issues such as energy, pollution, waste disposal, housing, transportation, etc.

2. Requires demonstration of CT skills: Again, according to our definition of this criterion any multiple choice exam would not score well on "demonstration."

3. Is at an appropriate level and range: The COMP OBJECTIVE EXAM has been used to measure growth from freshmen to senior exit in the areas which the COMP measures (including Problem Solving). The 1995 norms are based on student data collected during the 1996 year and the three years preceding it. This is a very strong point in favor of the COMP OBJECTIVE TEST--that is, the normative data is very recent and does not include students prior to 1991; therefore, it includes "current" students as opposed to some exams that have lumped all students together since the first testing date to get normative tables.

4. Reflects a continuum of knowledge and skills: Although there may be a continuum of knowledge and skills measured by the COMP ARC, there is little evidence that the application of those critical thinking skills is taught across the curriculum or by "real-life" examples in the South Carolina higher institutions. Critical thinking may be taught from more of a "content or discipline base" in S.C. institutions; unfortunately, there is little research that tells us if this is the case.

5. Includes appropriate results comparisons: The original criteria for the COMP were set by a group of over 100 higher education practitioners including Trudy Banta working through a FIPSE grant awarded to ACT to develop " a criterion referenced, application oriented, general education exit exam." There are many studies attesting to both national norming data and institutionally based comparison data.

There are many nation-wide norms for the COMP ARC including those for freshmen, seniors, age, gender and institutional type. This reviewer did not find any comparative norms based on race.

6. Technical Characteristics e.g., norming groups, efficacy studies, predictive to real-world tasks, etc. COMP OBJECTIVE TEST fairs well on technical characteristics. However, some institutions have reported a maturation function on the COMP scores. Although COMP national studies do not bear this out, it is extremely difficult to ignore that for adult students their experience with real-life problems before entering the classroom often might affect COMP scores particularly in solving problems (reasoning). *Also see below: Diagnostically useful*

Accurate:

7. Provides analytical as well as global scores: In the 16 years or so of the COMP's use, there is extensive reliability and validity data on the instrument. The correlation coefficients between the scales of the instruments show that distinctly different skills are measured by six subtests of the instrument. Clearly the reasoning skills measure something different from the writing skills measure and from the speaking skills measure.

8. Affords formative as well as summative score(s): This may be a weak area for the COMP OBJECTIVE TEST in S.C. institutions as it is often difficult to increase COMP scores without broad based changes in the curriculum as opposed to singular course changes.

However, ACT reports that some institutions which have used COMP to actually assist them in changing the curriculum have generally made broad based changes to the curriculum after much faculty investment.

9. Has sufficient specificity of score reporting: Although the score reporting is more difficult to understand perhaps than non-matrix scoring, the COMP OBJECTIVE TEST renders both individual and group information in the standard scoring service.

Diagnostically useful:

10. Shows student strengths and weaknesses: Renders student and group level scores. Shows student strengths and weaknesses relative to other students in a chosen peer group of institutions. The scoring of the COMP OBJECTIVE EXAM is on the number of items the student answered correctly and there is a penalty for guessing. Each item is cross scored; that is, it can be given a +2, +1, -2, or -1. The score reports are on a normative, comparison basis and a rather comprehensive packet comes from ACT to the institution breaking the scores into groups and norms.

11. Relates directly to educational program: There is much discussion as to whether COMP OBJECTIVE EXAM relates directly to the traditional academic, content oriented curriculum since its items are so distinctly application oriented using real-life examples. Therefore, much discussion is centered around what happens when a specific COMP score is low (for example Solving Problems). The question usually asked by faculty is which course(s) should we change to improve a low score on this subtest? COMP does not directly answer that question; rather, low COMP scores in Solving Problems would indicate that an entire weak area (problem solving) should be addressed across the curriculum as a whole. That is, all course(s) would systematically increase their use of application problem solving; for example, on in-class tests and in-class exercises, thereby increasing the teaching and learning of the weak area (e.g., problem solving). As a by-product of this type of curricular change, higher COMP scores should result.

This possible solution is often far more difficult for faculty to do in a large, "open" general education curriculum than one in which there is a set-progression of courses. If there is a set progression of courses which is required, then of course any single course or required group of courses could be specifically altered to include more problem solving types of activities.

12. Affords maximum generalizability of results: Although ACT has information on how institutions have used COMP OBJECTIVE EXAM scores to strengthen their curricular offerings, the examples tend to be from ACT (as compared to SAT) based states (not the East Coast or South to South Carolina). The COMP COMPOSITE can show both longitudinal growth information (either by test/retesting a small number of students) or by using ACT scores as the front end score and COMP scores as the exit scores. The COMP OBJECTIVE EXAM can also show point-in-time evaluation information.

D. RECOMMENDED SECONDARY CRITERIA

1. EASILY INCORPORATED INTO HIGHER EDUCATION PROGRAMS: The results of the COMP are not easily used in many higher education curricula because most curricula have not identified which courses are teaching their critical thinking skills and which are not. Therefore, to increase scores on the COMP, one is left with the two solutions: 1) increase the teaching of all critical thinking in all general education courses making it an "across the curriculum" strategy, or 2) identify specifically by score differentiation of students or course grouping exactly which courses are **best teaching** critical thinking skills and a) requiring those courses of all students or b) increasing the teaching of critical thinking in other courses in which the majority of the students enroll, with the subsequent result of teaching across a "specified, selected " curriculum. In either case, low COMP ARC scores require a genuine willingness on the part of the faculty to make broad based changes in the curriculum.

2. ECONOMICAL IN RESOURCE REQUIREMENTS: The COMP's greatest weaknesses for broad-base application in S.C. institutions of higher education are (as evaluated by this rater) in this

area. COMP OBJECTIVE TEST is not a simple test to administer (since it involves audio and taped responses). In addition, it cannot be easily integrated into classroom administration both because it is too long - 2 hours- and because its administration requires specific equipment. This adds to a major drawback--namely, getting the faculty involved in the testing so there is strong "faculty ownership," and therefore an increased chance for faculty score use once the scores are returned.

If students are not sufficiently "prepped" for the COMP, there is often great student confusion regarding the relevancy of the exam to their curricular experience since both the administration of the COMP and the actual items are not usually experiences in the test taking lives of students who are used to content oriented multiple choice tests or content oriented essay exams. Therefore, to be most successful, institutions have reported substantial front-loading preparation of both students and faculty.

Cost is often prohibitive for the COMP OBJECTIVE TEST which costs $15.75 per student up to 500 students, with graduated breaks up to 2501 or more students costing $9.00. This scoring does include ACT scoring the multiple-choice scan sheets and a "standard extensive" institutional score report which groups the scores by different groups and norms. Specialized score reporting which more extensively connects scores to the curriculum is available depending on how the test was administered and what the testing number was (population size).

APPENDIX D4

STRENGTHS & WEAKNESSES
THE CALIFORNIA CRITICAL THINKING SKILLS TEST (CCTST)

Strengths:

1. Definitions used by test developers closely follow our definition.
2. Items included in test range in level of difficulty. For example, some items require an analysis of the meaning of a given sentence in relation to those requiring more complex integration of thinking skills.
3. Questions are set in a context that address topics familiar to college-aged persons.
4. Knowledge needed for the test should be achieved through typical maturing of students through elementary and secondary education.
5. Available in 2 forms, A and B.
6. Does not require a large block of time for completion (45 minutes).
7. Test has 5 sub-scales--analysis, evaluation, inference, deductive reasoning and inductive reasoning. The test developers suggest that use of sub-scale scores is inappropriate since these mental activities are not independent. They suggest that these individual scores be used only as "gross indicators" of possible critical thinking strengths and weaknesses.
8. Discussion on item analysis, construct validity, and reliability are good.
9. Research found that the CCTST correlates with college level grade point average, verbal SAT, & math SAT. Data revealed no significant correlation between critical thinking ability and age and number of semester units of college work earned.
10. Test does not differentiate unfairly between women and men, nor between people based on their ethnic or racial heritage, nor among students based on their academic majors.
11. Test can be scored using computer assistance.

Weaknesses:

1. Test developers recommend "local norming." This takes time.
2. The test questions are wordy and can be seen as difficult for students. These types of comments were made by with the students in the validation study.
3. Information needed on this test's "track record." How has use of data from this test helped other schools?
4. There is no way to understand how students arrived at their answers. Scored as right or wrong.
5. Since the sub-scales should not be viewed as single scores, it would be difficult to identify individual student strengths & weaknesses.

Summary of the California Critical Thinking Skills Test (CCTST)

The California Critical Thinking Skills Test was developed to measure critical thinking as defined by the APA Delphi Report of 1990. The CCTST was developed by Peter A. Facione in 1990, and it is published by The California Academic Press. This standardized 34 item test is available in 2 forms, Form A and Form B. The format of the test is multiple choice with 4-5 answer options. Items on the test range from analyzing the meaning of a given sentence to more complex integration skills. The test should take approximately 45 minutes to complete.

Scoring can be done by hand or computer, and students are not required to use any special scoring sheet. Scoring by computer is done with use of the computer program SAS. The student's score is the total of correct answers. The CCTST provides 6 scores--analysis, evaluation, inference, deductive reasoning, inductive reasoning, and an overall score. Norms are provided for the overall score and each of the 5 sub-scales. Data is provided on the norming procedure completed in 1989-1990.

APPENDIX D5

REVIEW OF COLLEGE BASE
(College Basic Academic Subjects Examination)

Note: This review will focus primarily on the aspects of the test relevant to assessing critical thinking/reasoning.

Published by: The Riverside Publishing Company

Developed by: University of Missouri, Columbia, Assessment Resource Center

Original Publication Date: 1989

Forms Available: Regular Edition and Institutional Matrix form (Note: Except where otherwise indicated, this review will focus on the regular edition of the test.) Institutions may also administer any combination of one to four subject sections from the regular edition.

Format: Multiple choice and writing exercise

Number of items: 180 plus optional writing exercise (Institutional Matrix form contains either one section of about 36 items or the writing exercise)

Administration time: 3 1/2 hours (any one subject - 50 minutes; Institutional Matrix form - 50 minutes)

Scoring: Scores for the Long Form and Institutional Matrix Form are processed at the Center for Educational Assessment.

Types of Scores Produced: (from *Presenting . . . College BASE* published by the Riverside Publishing Company): "*College BASE* yields a variety of scores, each referenced to specific criteria. Individual scores are reported on the *Student Score Report*, and institutional averages are reported on the *Institutional Summary Report*. Scores for both reports are organized into the following categories:

- A Composite Score, representing the overall performance of a student (or group of students) on the exam;
- Subject Scores, representing the performance of a student (or group of students) on each subject tested (i.e., English, mathematics, science, and social studies);
- Cluster Scores, representing performance on closely related skills within a subject (e.g., laboratory and field work in science);

- Skill Scores, representing a degree of mastery (High, Medium, or Low) on each of the 23 individual skills tested; and

- Competency Scores, representing performance on the three cross-disciplinary processing skills (i.e., interpretive reasoning, strategic reasoning, adaptive reasoning).

Although *College BASE* is criterion referenced, it can nevertheless yield information about student or institutional performance relative to the performance of other students and institutions. For the purpose of determining relative ranking on specific criteria, *College BASE* offers tables of percentile ranks for individuals as well as for institutions."

General description of test (from "Guide to Test Content" published by the Riverside Publishing Company, 1989): The *College BASE* is a "criterion-referenced achievement test that assesses student proficiency in English, mathematics, science, social studies, and three cross-disciplinary, cognitive competencies--interpretive reasoning, strategic reasoning, and adaptive reasoning. All of these subjects and competencies represent levels of cognitive functioning developed as a result of education and practice.

Each of the four academic subjects is organized into levels of increasing specificity: from subjects, to clusters, to skills, to enabling subskills. As *College BASE* is a criterion-referenced test, the specific skills and factual knowledge defined by the 23 skills form its heart. Each skill is defined by two to six enabling subskills deemed by content experts to be instrumental in a student's ability to master that skill.

College BASE is intended to assess content knowledge and skill development at a level commensurate with students completing the general education component of their college experience. At most institutions this will be near the end of the sophomore year. While the test is not course specific, students who have completed a typical lower-division course of study should be well prepared for *College BASE."*

A. Relevance

1. The primary stated purpose of the *College BASE* is to provide a "diagnostic assessment of . . . academic attainments in subjects and competencies usually covered through a general education program during the first two years of undergraduate study" (from *College BASE Technical Manual*). It assesses clearly defined objectives stemming from the summary report of the Educational Equality.

   The project was sponsored by the College Board. Although not specifically designed to measure critical thinking, the test does address intellectual processes involved in three types of reasoning as defined below.

   - interpretive reasoning - cognitive process by which one begins to understand information that has been remembered or observed. This includes translating, paraphrasing, summarizing, and explaining.

- strategic reasoning - cognitive process by which we expand the translation provided by interpretive reasoning. This includes defining, comparing, classifying, and analyzing.

- adaptive reasoning - cognitive process by which we extend our knowledge beyond the boundaries established by strategic reasoning. This includes synthesizing, hypothesizing, predicting, and expressing judgments of value, merit, or worth.

2. The test and associated items were developed according to clear specifications and included a variety of appropriate if not exemplary analyses designed to assure validity while reducing the likelihood of bias. Nonetheless, a weakness of the test which is especially relevant to assessing critical thinking/reasoning is that it is limited to multiple choice items except for the writing exercise that contributes to the grade in the English cluster only. Thus, reasoning skills must be inferred from answers to subject-related questions. No demonstration of critical thinking skills is required.

3. A strength of the *College BASE* is the range of subject clusters and skills measured. The test appears to be at a level appropriate for assessing general education in most South Carolina college curricula. Extensive validity research including content-related evidence, criterion-related evidence, and construct-related evidence including factor analytic studies is presented in the technical manual for the test. Moderate correlations between *College BASE* results and those on aptitude measures such as the SAT and ACT are strong enough to indicate that the test measures some of the same intellectual processes but not so high as to indicate that it is measuring virtually the same general abilities (a criticism of other national college achievement tests). *College BASE* subject area scores have been found to have statistically significant moderate correlations (ranging from .34 to .43) with GPA. The magnitude of correlations for English, science, and social studies and GPA is higher than that for ACT, SAT-Verbal and SAT-Quantitative and GPA. Unfortunately, except for content-related evidence and a few intercorrelations among competencies, there is a lack of evidence presented regarding the validity of the three reasoning competencies.

4. The *College BASE* is described as a "criterion-referenced," "diagnostic" "achievement" test. The technical manual provides criteria for each of these terms and an in-depth discussion of how, according to the test developers, the *College BASE* meets the criteria. Actually, the test is cited by the developers as being both criterion-referenced and norm referenced. While the test has the strength of being designed from a criterion-referenced perspective, in fact the scores that it yields reflect relative standing as opposed to degree of mastery or attainment of skill or knowledge based on established criteria. Scores on reasoning competencies are reported as "low," "medium," or "high" based on ±1 standard deviation from the mean score of the standardization population.

5. Despite the fact that *College BASE* scores are based on performance relative to the standardization population, little information is provided in the technical manual as to the demographic characteristics of that population. No information is provided in the technical manual regarding geographic regions or institutional types from which the standardization population was drawn, and demographic data that are provided suggest reason for concern

about the ethnic and gender representativeness of the population (e.g., only 3.7% Black and 15.8% male). This makes the cut-off criteria for reasoning competency levels, which were selected by a "group of knowledgeable persons who compared the frequency distributions of the standardization population on each skill," difficult to translate in a meaningful way. Recent samples for the test, which are used to recalibrate items but not to redefine scoring criteria, have been much larger and more representative according to data provided by the test developer.

B. Accuracy

1. Substantial information is provided on the reliability of *College BASE* scores.

2. *College BASE* yields 40 scores: one score for each of four subjects, nine clusters, 23 skills, and three competencies in addition to a composite exam score. However, only three scores (competency scores for three reasoning areas) are directly relevant to critical thinking.

3. Scores are strictly summative in nature. Formative information as to how students attained their answers must be inferred from multiple choice selections.

4. Scores are reported based on a standard scale but more global scales ("low," "medium," "high") are used for reporting reasoning competency scores.

C. Diagnostic Usefulness

1. *College BASE* was designed to be a "diagnostic" test and does yield a wide range and variety of scores useful for that purpose (although clearly limited by the fact that the scores are based strictly on answers to objective multiple choice items except for the writing exercise). Both individual and institutional score reports are provided (except for the Institutional Matrix form, which yields an institutional report only since each student completes only one section of the overall test). However, results pertinent to critical thinking assessment are only global in nature and are very limited.

2. Although *College BASE* results may be useful generally to determining educational program strengths and weaknesses, they are of very limited usefulness in determining strengths and weaknesses relevant to critical thinking.

3. Results are based on objectively scored items except for the writing exercise, which is scored by Riverside based on clear criteria. Comparisons to the standardization population enhance generalizability. Lack of information about the population of examinees, however, is a serious problem in interpreting results.

Other Evaluative Information

1. The 3 1/2 hours administration time for the long form makes it difficult to incorporate the *College BASE* into existing program structures/schedules.

2. The test is fairly reasonable in cost considering the range of materials and services provided. Test booklets are reusable (which further reduces costs).

3. Scoring by the publisher is conducted in an efficient and effective manner. In addition to a Student Score Report and Institutional Summary Report, interpretive guides are provided. Reports are well-organized and clearly written. A centralized data base of all examinees' scores is permanently maintained at the University of Missouri-Columbia. This may be helpful in the case of examinees who transfer. Special score reports designed for disaggregating group performances (e.g., for transfer status, ethnicity, or gender) and comparisons to similar institutions are available at extra charge.

4. The test requires minimal "front-loading" and "back-loading" except for the time that might be needed to plan for its administration (e.g., arranging for examinees, scheduling administration dates, etc.). Results are obtained quickly and easily.

5. A variety of ancillary materials is available, including well-organized manuals for test administration. An extensive technical manual is available.

6. The partnership between the University of Missouri Assessment Center (ARC) and Riverside Publishing which produced the test provides an added degree of academic credibility, but sometimes makes it less clear as to where and how to receive particular test services or information. In such cases, users are encouraged to contact ARC for staff assistance.

SUMMARY OF STRENGTHS AND WEAKNESSES

The *College BASE* is a well-developed, standardized instrument designed to measure college achievement. The examination includes objectively scored multiple choice items and one writing exercise scored according to clearly established criteria. Four subjects (English, mathematics, science, and social studies), nine clusters, and 23 skills are assessed, thus providing a variety of results potentially useful for determining individual and institutional strengths and weaknesses. However, only three competency areas on the test--adaptive reasoning, strategic reasoning, and interpretive reasoning--relate in any direct way to critical thinking.

The *College BASE* may be more relevant to assessing actual achievement than other national standardized exams. However, as with other objective tests, examinee possession of knowledge and skills must be inferred from responses to multiple choice items. This is a particular limitation in assessing a skill as complex as critical thinking.

College BASE was designed from the onset to be criterion-referenced (an advantage in comparison to instruments which are strictly normative or which have criterion-referenced components that were "afterthoughts"). Scores, however, are based on relative standing in comparison to a standardization population that is not adequately described by the developers or publisher. Scores on various reasoning competencies relevant to critical thinking consist solely of "low," "medium," or "high" based on criteria of ±1 standard deviation from the mean.

College BASE is available in Regular Edition and Institutional Matrix form. The Regular Edition administration time of 3 ½ hours may be impractical for inclusion in many ongoing college and university programs/schedules. The availability of a 50-minute Institutional Matrix form is an advantage that comes at the price of a greatly reduced sample (because it takes 12 examinees to collectively comprise two tests) and the loss of *individual* analyses and reports. Institutional analyses and reports, a national database of examinee scores, and a variety of helpful ancillary materials and support are available for the *College BASE*.
relate

APPENDIX D6

REVIEW OF CORNELL CRITICAL THINKING TEST

Published by: Midwest Publications

Developed by: Robert Ennis, Jason Millman, Thomas Tomko

Original Publication Date: 1985

Forms Available: X (appropriate for grades 4-10); ·Z-G (appropriate after high school is completed); X version 3 (appropriate for college students)

Format: Multiple-choice

Administration Time: more or less 50 minutes, but 10% will not finish in 50 minutes.

General Description: Test taker will read research findings, arguments, etc. and answer multiple choice questions. The questions relate to such things as the facts given in the written material support the hypothesis; whether the observation made is credible; whether a logical prediction can be supported by the material supplied. This test has low reliability for an objective test.

Strengths: Short testing time
See chart page 14 of this document

Weaknesses: Low reliability
Does not seem relevant
See page 14 of this document

APPENDIX D 7

REVIEW OF ENNIS-WEIR CRITICAL THINKING TEST

Definition: Reasonable and reflective thinking that is focused upon deciding what to believe or do. Critical Thinking in context of argumentation.

Published by: Midwest Publications

Developed by: Robert Ennis & Eric Weir

Original Publication Date: 1985

Forms Available: Single form

Format: Write essay in response to 8-9 paragraph argument in the form of a letter

Administration Time: 40 minutes

Scoring: Subjective criteria supplied, no examples of good or bad (range finders). Only content validity. Reliability is inter-rater reliability between 2 highly trained judges.

Types of Scores: Not mentioned but subscores could be possible based on scores of 8 paragraphs.

General Description: Read letter (10 minutes), write 1 paragraph in response to each of 8 paragraphs, write conclusion (30 minutes). Too reliant on writing skills, too easy for college level (mean is more or less 1 standard deviation from the maximum score)

Strengths and Weaknesses: See page 14 of this document

APPENDIX D8

REVIEW OF ETS TASKS IN CRITICAL THINKING

Published by: Educational Testing Service

Developed by: The Tasks in Critical Thinking were developed as the General Intellectual Skills Test for the New Jersey Department of Higher Education by Educational Testing Service and the College Board.

Original Publication Date: 1988

Forms Available: There appear to be nine (9) Tasks available

Format: A "task" asks a student to do several things, e.g., fill in graphs and make charts from numerical information provided, find relationships and draw conclusions from such charts and graphs, and write a coherent essay using this and other provided information to present a hypothesis and defend it. The information is presented in written form (including numerical) and the response is asked for in written form. There are usually 8-10 questions per Task, and they appear to grow increasingly more difficult in nature until the student must write a short essay using all he/she has learned from the Task.

Administration Time: Total Administration Time is 90 minutes

Scoring: The scoring is done by faculty at the institution or is centralized in some manner. ETS will score the Tasks but only on a contractual basis. The scoring rubric is fairly well defined. The score range for most questions is 9-12 with 8 as the "core score" or that score which means a "satisfactory answer to the question; the student was proficient in using the skills being assessed in this question." Higher or lower scores are always relative to the core score. Although the "8" core score was presented in the *Student's Guide to Tasks in Critical Thinking*, in all the actual tasks (per the scoring guide) a score of "4" was the "core score." A student is provided a score only when the student fills out a self-mailer. The student is told " As your Task is read and the scores for each questions are recorded on a special score sheet, those scores will come through on the inside of a self-mailer, which is the Student Score Sheet. When scoring is complete Score Sheets are folded and sealed with the scores inside and mailed to the address you have provided."

Results for the institution are summarized across all the 9 Tasks and reported as Group Data in percentages only. Scores are not reported for the Tasks; they are reported for the skills that were used in working through the different Tasks. **The skills assessed by the Tasks are inquiry, analysis and communication (written not verbal). Percentages are provided for the student body who scored above or below a "4=Fully Proficient" in each of the three skills (inquiry, analysis and communication). The score report can include subgroups (minimum 100 students per subgroup). There are generally 6 levels stated for the institutional report, i.e., 6 = superior, 5=exceeds, 4= fully proficient, 3=some proficiency, 2=limited proficiency and 1=not proficient. Subscores render only three levels: superior performance, fully proficient, and limited proficiency.**

Clarification is needed as to what the "core (acceptable) score" really is -- 8 or 4.

Types of Scores Produced: There are single digit scores produced for inquiry, analysis and communication for the institution as a whole in a formal score report. There are no formal student score reports. The confidential self-mailer includes a reminder of which Task the student took and what it was about, along with a brief description of each question and what the scores means in the context of what was expected. The scoring of the ETS Tasks in Critical Thinking is labor intensive, dependant on training and maintaining a faculty willing and able to continually use the scoring that ETS has devised.

General Description of the Test:

Relevance:

1. *Directly measures CT knowledge:* The assessment models the production of ideas rather than selecting right answers. However, sometimes the right answer is required to produce the correct responses; that is, a chart must be graphed correctly by a student or all the conclusions drawn from the chart will be flawed. Therefore, it is not unusual that "Information A" must be correct in order to produce "Information B;" if a student incorrectly does A, he/she will sometimes (not always) incorrectly produces B since A and B are not independent answers or questions within a singular Task.

2. *Requires demonstration of CT skills:* Each Task can be a demonstration of one type of critical thinking skill, e.g., one task centers on comparison and contrast; one task centers on the use of maps and calculations to draw conclusions, one task centers on analyzing art from drawings and reproductions, etc.

3. *Is at an appropriate level and range:* There is little or no data on whether the ETS Critical Thinking Tasks would be appropriate for any other students than New Jersey students, since the original development was for a singular state and all faculty writers were originally drawn from that state and set the types and difficulty ranges for the questions.

4. *Reflects a continuum of knowledge and skills:* Although there may be a continuum of knowledge and skills measured by the Tasks, there is little evidence that the application of those critical thinking skills is taught across the curriculum or by "real-life" examples in the South Carolina higher institutions. Critical thinking may or may not be taught from more of a "content or discipline base" in S.C. institutions, we simply do not know. This is a problem with several other exams which were reviewed.

5. *Includes appropriate results comparisons:* There are no large scale norms (or perhaps any norms at all) for the ETS Critical Thinking Tasks. This reviewer did not see any norming information for the exams. In the 1989 publication "Validity and Reliability of the COEP GIS Assessments" there is very little data given. Originally THIS TEST WAS CALLED THE "GENERAL INTELLECTUAL SKILLS TEST" AND PURPORTED TO MEASURE "GIS;" THE SAME TEST WAS RENAMED BY ETS AS TASKS IN CRITICAL THINKING.

6. *Technical Characteristics, e.g., norming groups, efficacy studies, predictive to real-world tasks, etc.* See number 5 above. There are no large norming groups, subscores, or predictive tables.

However, one concern in the technical area is that there would seem to be a very high correlation between these Critical Thinking Tasks and the skills of reading and writing (as correlated skills). Since all tasks are presented in written format and the most heavily weighted responses are called for in written format (e.g., short answers, essays), it would seem there may be a high likelihood that a high score on critical thinking would require a high skill in both reading and writing. One would need to see intercorrelations between reading, writing and critical thinking to be comfortable that one is not essentially dependant upon the other.

Accurate:
7. *Provides analytical as well as global scores:* This depends on whether the "core components of Critical Thinking" are defined as inquiry, analysis and communication.

8. *Affords formative as well as summative score(s):* There is no overall indication that formative information could be used in the curriculum from these assessments. That is, if a student has never been exposed to a specific way in which the information might be presented, i.e., a style of painting, a type of graph, a specific kind of argument, it is difficult to know whether the student is lacking in the skill itself (i.e., being able to read a graph) or is lacking a skill in critical reasoning. Since only one Task per student is administered in 90 minutes--rather than 2 or 3, or a choice of 1 of 3, for example--then a student could be weak in X skill (numerical manipulation in graphing) but could be strong in critical thinking.

9. *Has sufficient specificity of score* reporting: The scoring is very simplistic. Grammar, sentence structure, etc. are not scored since this would be more of a writing score than a critical thinking score. Scores above the "core score" are given only if the basic solution appears and is expanded upon. Lower scores are given for such things as partial responses, minimal development of an idea, or not providing support for the position taken. Essentially scores are reported for superior performance (above core), fully proficient (core) and limited proficiency (below core).

Diagnostically useful:
10. *Shows student strengths and weaknesses*: This test does not render individual student information that can be used by the student in any meaningful way.

11. *Relates directly to educational program:* There is little evidence that the types of Critical Thinking Tasks relate directly to South Carolina's higher education curriculum. That is not to say that South Carolina institutions are not teaching critical thinking, it is to say that South Carolina higher education institutions may be teaching more or fewer critical thinking components than inquiry, analysis, and communication in different ways than this assessment tests them. For example, the communication skills tested in ETS's Critical Thinking Tasks is only written communication; there is no verbal communication tested or in fact mentioned, although verbal communication is a skill which often requires critical thinking.

12. *Affords maximum generalizability of results:* It is highly debatable whether faculty in South Carolina would define Critical Thinking in the same way as faculty in New Jersey defined it, and even less possibility that faculty in one state would test it in the same way.

D. RECOMMENDED SECONDARY CRITERIA

1. EASILY INCORPORATED INTO HIGHER EDUCATION PROGRAMS: The results of the ETS Critical Thinking Tasks are not easily used in many higher education curricula because most curricula have not identified which courses are teaching critical thinking skills using what types of measures (graphs, charts, essays, comparisons, analysis, etc.) and which are not. Therefore, it is difficult to ascertain what exactly to do in a curriculum to increase scores on the Critical Thinking Tasks..

2. ECONOMICAL IN RESOURCE REQUIREMENTS: The greatest weaknesses for broad-base application in S.C. institutions of higher education are (as evaluated by this rater) in this area. First, when an assessment doesn't give individual student information regarding what the students should strengthen in critical thinking, how to do it, and why they should do it, is very difficult to sell to a student body. If an institution administered the Tasks in one discipline, it would very much skew the scores; therefore, a random sample or whole population sample is required. It is very difficult to get students to buy into an exam that doesn't give them individual feedback (or enough to benefit them specifically). Student motivation would be a major problem.

Secondly, the grading of the Tasks by faculty (trained by a faculty member from New Jersey) is thought to be the best scoring method and allows faculty to really understand the process. ETS does say that " normally, after a day and a half to two days of training (which presumably is not included in the cost of the test), five to ten faculty members can score 200 Tasks in two or three days or even in several half day sessions over a period of a few weeks."

Sustaining faculty year after year and training and retraining different members is a monumental job. Faculty state-wide would have to not only agree that these Tasks are valid (that is, that similar kinds of things are being taught and tested) but they must agree each time they come together that the scoring rubics are valid also. Cost of ETS scoring is NOT included in the $15.00 per Task fee; neither is there an interrater reliability check done by ETS regarding faculty scoring. Therefore, there is little outside validity to this instrument, other than if a different faculty in a different state devised a particular set of "tasks" to classify what they believed should measure critical thinking and then devised a common scoring rubic for the tasks.

Additional ETS Tasks Strengths and Weaknesses

Strengths
-The tasks require student performance rather than selecting the right answer.
-The tasks are interesting and might motivate students to perform at their best.
-The tasks may assess critical thinking, depending upon institutional definitions.

Weaknesses
-Administration time is long - 90 minutes.
-Scoring is expensive in terms of faculty time for training and actual scoring.
-There are no norms provided.
-Student performance of tasks may be highly dependent on reading and writing skills.
-There are no individual student scores.
-Scores will not provide information sufficient for instructional improvement.
-The tasks may or may not relate to critical thinking as taught in South Carolina's higher education institutions.

APPENDIX D9

Watson-Glaser Critical Thinking Appraisal (CTA) Strengths & Weaknesses

Strengths:

1. Follows author's definition of critical thinking.
2. Easy to administer, score, interpret, and is inexpensive.
3. Available in 2 forms.
4. Sufficient information included on how to use the results.
5. Statistical information included on reliability and validity.
6. Information included on how the CTA correlates with other tests.
7. Scoring can be done by OpScan scoring machine.
8. Test divided into 5 sub-scales; however, authors do not recommend use of sub-scales as individual scores.

Weaknesses:

1. "Wordy" test items which stress the need for strong reading skills. Students may label the test as hard and may not attempt to complete.
2. Items too specific for definition of critical thinking.
3. Tasks in tool are not real world tasks. Test was last published in 1980 which makes test out-of-date.
4. No formative information provided the student. Students cannot get specific feedback on how well they performed.
5. Authors encourage users to avoid using individual subscales.

Summary of the *Watson-Glaser Critical Thinking Appraisal* (CTA)

The *Watson-Glaser Thinking Appraisal* was developed by Goodwin Watson and Edward M. Glaser. The tool evaluated by this task force was published in 1980. The CTA was developed to measure critical thinking, which was identified by the authors as a composite of attitudes, knowledge, and skills. This tool is available in 2 forms with each consisting of 5 subtests--namely inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. There are 80 items on the test with questions requiring various answers. Some items require the student to decide if the question is true, probably true, insufficient data, probably false, and false. Other items require a decision as to whether a person is making an assumption or not. The last items require the student to decide whether a conclusion follows from the information given. This test should take approximately 50 minutes to complete.

Students use an OpScan answer sheet. A scoring sheet is available for each form of the test. Sheets are available for OpScan machine scoring. One overall score is given. The raw score is the total number of correct answers. Raw scores are converted into norm scores. Norm scores are available for high school students and college students.

APPENDIX E

GUIDELINES FOR CHOOSING COMMERCIAL TESTS

APPENDIX E

GUIDELINES FOR CHOOSING COMMERCIAL TESTS

The overall validity ratings presented in this report are based on the following three assumptions: 1) the instruments reviewed may be used to make decisions about students, programs, or institutions; 2) performance-based tests are potentially more accurate measures of critical thinking than are multiple choice tests; and, 3) critical thinking test items/tasks should be grounded in critical thinking theory, but must also have direct relevance to problem solving skills necessary for success in today's complex, information-rich world. Since the needs and philosophies of practitioners will vary greatly, no one instrument can be recommended for the assessment of critical thinking for all institutions. The following guidelines are presented to assist practitioners in choosing the best instrument for the assessment needs of their institutions.

1. Definition - Appendices D and G contain a brief description of each test, including the definition of the construct being measured. Reduce the list of prospective instruments by eliminating those instruments which differ significantly from the definition of critical thinking adopted by your institution. If your institution has no formal definition of critical thinking, the definitions presented in Appendix A (Literature Review) may be a good starting point. A faculty committee could review the definitions and hopefully adopt a definition used by existing instruments for measuring critical thinking.

2. Use of results - Determine the use of the test results. If decisions are to be made at the student level, accuracy of the measure should be of utmost importance. If the results are to be used to make decisions at the program or institutional level, perhaps a proxy in the form of a multiple choice test, a one task performance test, or a test which allows matrix sampling (each student takes only one section of the test) may be desirable. Reasonable accuracy and ease of administration may be the goals for measurement at the program or institutional level.

3. Relevance of each criterion - Page 14 contains the ratings on the 12 criteria of validity for all instruments reviewed by the committee. Based on your philosophy regarding the measurement of critical thinking and the use of the test results, weight each criterion to derive your own overall validity score for the instruments still under consideration. For instance, if you believe multiple choice items are adequate for your needs, criteria one and eight may be eliminated from consideration entirely.

4. Overall institutional assessment plans - Once the instruments are rated, determine how the top ranked instruments would fit into your overall institutional assessment plans. For example, if critical thinking is a high priority, then the highest ranked instrument may be the best test for your needs. However, if general education assessment is also a high priority, you may want to consider the possibility of using a general education test which also measures critical thinking. Or, if measuring student learning at the content level is a

high priority, perhaps an achievement test which also measures critical thinking may be most appropriate for your institution.

5. Feasibility - Finally, using the feasibility ratings of each instrument contained on page 14, determine which of the top ranked instruments are feasible as measured by cost, test length, and motivation of students to perform well on the instrument.

6. Finally, having taken into consideration the above information, contact the publishers of the two or three tests that you are most interested in and ask for a form of the test. Only by examining the test/task items can you be sure that the test is measuring critical thinking in a manner appropriate for your institution. Where possible, actually take the test and review your test results. Some instruments which on the surface appeared to have high validity for our needs were fairly disappointing in practice.

APPENDIX F

IDEAS TO BE DISCUSSED REGARDING DIFFERENT TESTS AND FUTURE TEST DEVELOPMENT

APPENDIX F
IDEAS TO BE DISCUSSED REGARDING DIFFERENT TESTS AND FUTURE TEST DEVELOPMENT

What is needed is an instrument with acceptable levels of both validity (i.e., assurance that students' critical thinking skills are actually being measured in ways that can be directly related to curricular interventions) and feasibility (in terms of the time, effort, and expense required to administer, serve, interpret, and utilize the data). While no currently available instruments fit our needs, we are aware of at least four modifications to current models and approaches, using currently available technology, which we think offer the potential for significantly improving the quality of critical thinking measures:

1. Increasing the validity of current multiple choice approaches (like the California, Cornell, CAAP, and BASE) by using enhanced multiple choice answers and scoring (e.g., assuming a five-choice multiple choice format, have the number of "correct" answers vary from 0-5; require marking of "fully correct," "partially correct," "partially incorrect," and "fully incorrect" answers; supplement multiple choice responses with short open-ended narratives explaining why correct answers are correct, etc.).

2. Increasing the feasibility of authentic, performance-based approaches (like the COMP, ARC, and ETS Tasks) by using objectively scorable short-answer responses (i.e., scannable fill-in-the-blank or restricted responses--that are nevertheless student-produced answers).

3. Trying to improve both the validity and feasibility of current critical thinking approaches immediately by producing combos (e.g., ARC-CAAP, California or Cornell and ETS).

4. Increasing both the validity and feasibility by using the potential of computer-assisted assessment (i.e., administering, scoring, reporting, etc.) .

5. Limit uses of current measures to functions with demonstrated efficacy (i.e., with a track record of successful use for that purpose in higher education).

In addition to concerns for reliability and validity of test instruments designed to assess critical thinking skills, feasibility is a major concern. Feasibility issues include instrument cost; time and staff required for test administration; time, staff, and cost required for scoring and analyzing results; facilities and equipment required for test administration; and the usefulness of results.

One means of overcoming the feasibility weaknesses of the existing test instruments is to develop a computerized adaptive instrument to assess critical thinking skills. Such an instrument should meet at least the criteria described in the following section.

In addition to the criteria developed by the SCHEA Critical Thinking Task Force, a computerized adaptive test for assessing critical thinking skills should consist of objective/outcome based items linked to a learning taxonomy such as Bloom's Taxonomy or Gagne's Hierarchy; use multi-media delivered tasks; provide immediate scores; provide score definitions; be untimed; accept multiple forms of response entry such as voice entry, keyboard entry and/or touch screens; store individual student scores in an institutional data base; and, be available to all South Carolina higher education institutions via network.

The following section describes the advantages offered by each of the attributes listed above.

Objective/outcome based items linked to a learning taxonomy will provide information regarding the level of reasoning performed by each student. A computerized adaptive instrument could consist of banks of items designed to meet curricula needs at associate, baccalaureate, and graduate degree levels. Computerized adaptive testing also allows instruments to be used for pre- and post-testing.

Multi-media delivered tasks can be realistic and high interest, motivating student performance. In addition they eliminate testing of reading skills rather than thinking skills.

Immediate scores meet student needs and provide institutions with information necessary to make course and program improvements.

Score definitions describing the thinking skills linked to each numerical score or range of scores provide information necessary to make improvements in instruction.

Untimed testing insures assessment of thinking skills rather than time to think.

Multiple forms of response entry such as voice entry, keyboard entry and/or touch screens eliminate the dangers of testing writing skills rather than thinking skills, provide for disabilities, allow the testing of specific skills such as speech, and accommodate multiple learning styles.

Individual student scores in an institutional data base allow each institution to conduct research for purposes of instructional improvement, establish reliability and validity of scores, determine use of outcomes as predictors, and diagnose/prescribe for individual students.

Available to all South Carolina higher education institutions via network: the computerized adaptive test would be accessible to all institutions regardless of size.

APPENDIX G

HELPFUL CHARTS REGARDING COMMERCIAL TEST COMPARISONS

Reprinted with permission from American College Testing

COMPARISON OF COLLEGE LEVEL CRITICAL THINKING TESTS
MULTIPLE CHOICE (MACHINE SCORABLE)

INSTRUMENT	COMP OBJECTIVE TEST	CAAP CRITICAL THINKING TEST	COLLEGE BASIC ACADEMIC SUBJECTS EXAMINATION (College BASE)
Vendor Year Started Current Version	American College Testing 1976 1995	American College Testing 1988 1992	Riverside Publishing Co. 1984 1989
NO. OF FORMS	3 Forms	4 Forms	1 Form
ABILITIES ASSESSED	Communicating, Solving Problems, Clarifying Values, Functioning in Social Institutions, Using Science, & Using the Arts	Analysis of elements of an argument Evaluation of an argument Extension of an argument	English, Mathematics, Science, Social Studies
NO. OF ITEMS	120	32 Multiple choice	180 Multiple Choice
NO. OF SCALES Unit of Analysis	Total Score, Six Subtest Scores Individual & Group	Total Score Individual & Group	Composite Score, 4 Content Subscores and 3 Reasoning Subscores Individual & Group
CONTENT	Fifteen simulation activities use excerpts from television documentaries, newscasts, & magazine articles addressing significant societal issues	Passages in the form of case studies, debates, dialogues, overlapping positions, statistical arguments, experimental results and editorials.	Traditional subject matter in English, Mathematics, science and social studies. One fourth of the items measure recall of factual information.
NORMS AVAILABLE	YES (College Freshmen, sophomores, & seniors)	YES (college freshmen and sophomores)	LIMITED (college sophomores and seniors)
TESTING TIME	150 minutes	40 minutes	3-½ hours
DEFINITION OF CRITICAL THINKING	Problem solving and clarifying values as they occur in social institutions, science and technology, and the arts are viewed as decision-making components of critical thinking. Skills include the ability to analyze a variety of problems, to generate reasonable approaches to solve them, and to use a consistent value orientation recognizing perspectives that support or conflict with such solutions.	Critical thinking includes the ability to clarify, analyze, evaluate, and extend arguments. An argument is defined as a sequence of statements which includes a claim that the conclusion follows from the other statements.	Interpretive reasoning is the first level of cognitive processing beyond recalling facts by which we begin to understand information that has been remembered or observed. Strategic reasoning expands the translation provided by interpretive reasoning to recognize relationships and implications. Adaptive reasoning is the highest level of cognitive processing involving the ability to synthesize, hypothesize, make predictions, and judge worth.
TESTING COST: 100 Students	$1,575 ($15.75 per student)	$1,130 ($11.30 per student)	$1,450 ($14 50 per student)

Table 2
COMPARISON OF COLLEGE LEVEL CRITICAL THINKING TESTS

INSTRUMENT	WATSON-GLASER CRITICAL THINKING APPRAISAL	CORNELL CRITICAL THINKING TEST	CALIFORNIA CRITICAL THINKING SKILLS TEST
Vendor Year Started Current Version	Psychological Corp. 1942 1980	Midwest Publications 1979 1985	California Academic Press 1990 1992
NO. OF FORMS	2 Forms	2 Forms	2 Forms
ABILITIES ASSESSED	Inference Recognition of assumptions Deduction Interpretation Evaluation of arguments	Induction Deduction Observation Credibility Assumptions Meaning	Interpretation Analysis Evaluation Inference Explanations Dispositions
NO. OF ITEMS	80 Multiple choice	52 Multiple choice	34 Multiple choice
NO. OF SCALES	Total Score	Total Score	Total Score; 2 or 3 subscores
CONTENT	Some neutral content (weather, facts) Some controversial content (social, economic, political)	Traditional subject matter content plus descriptions of science experiments	Contexts include neutral, situational, and controversial settings
NORMS AVAILABLE	YES (high school, college, business, nursing)	YES (high school, college, adult)	YES (college)
TESTING TIME	40 minutes	50 minutes	45 minutes
DEFINITION OF CRITICAL THINKING	A composite of attitudes, knowledge, and skills including: 1) attitudes of inquiry that involve an ability to recognize the existence of problems and an acceptance of the general need for evidence in support of what is asserted to be true; 2) knowledge of the nature of valid inferences, abstractions, and generalizations in which the weight or accuracy of different kinds of evidence are logically determined; and 3) skills in employing and applying the above attitudes and knowledge.	Critical thinking is the process of reasonably deciding what to believe and do(Ennis, 1984). There are 3 types of inferences to beliefs (induction, deduction, and evaluation) and 4 types of bases for such inferences: the results of other inferences; observations; statements made by others; and assumptions. Furthermore, close attention to meaning must permeate one's dealing with the 3 types of inferences and 4 types of bases. There are considerable overlap and interdependence among aspects of critical thinking.	Critical thinking is purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual consideration upon which that judgment is based. Critical thinking includes the ability to analyze, criticize, and advocate ideas, to reason inductively and deductively, and to reach factual or judgmental conclusions based on sound inferences drawn from unambiguous statements of knowledge or belief.

Table 2 (Continued)
COMPARISON OF COLLEGE LEVEL CRITICAL THINKING TESTS

INSTRUMENT	CAAP CRITICAL THINKING TEST	COMP ASSESSMENT OF REASONING AND COMMUNICATING	ENNIS-WEIR CRITICAL THINKING ESSAY TEST
Vendor Year Started Current Version	American College Testing 1988 1992	American College Testing 1986 1991	Midwest Publications 1985 1985
NO. OF FORMS	4 Forms	5 Forms	1 Form
ABILITIES ASSESSED	Analysis of elements of an argument Evaluation of an argument Extension of an argument	Selecting approaches to solve problems Determining logical consistency Determining solution to be implemented Assessing internal consistency of values Analyzing rationales for value choices	Getting the point Seeing reasons & assumptions Stating one's point Offering good reasons Seeing other possibilities
NO. OF ITEMS	32 Multiple choice	Analytic scales to produce 15 ratings based on 3 essays and 3 speeches	Holistic scoring of written response to 9 paragraph argument
NO. OF SCALES	Total Score	Total Score; 3 subscores	Total Score
CONTENT	Passages in the form of case studies, debates, dialogues, overlapping positions, statistical arguments, experimental results and editorials.	Simulations and role-playing tasks using controversial topics related to the family, community problems, energy, health issues, music, and art reproductions.	Issues related to a city letter regarding parking between 2 and 6 a.m.
NORMS AVAILABLE	YES (college freshmen and sophomores)	YES (college freshmen, sophomores, and seniors)	YES (eighth grade and college)
TESTING TIME	40 minutes	105 minutes	40 minutes
DEFINITION OF CRITICAL THINKING	Critical thinking includes the ability to clarify, analyze, evaluate, and extend arguments. An argument is defined as a sequence of statements which includes a claim that the conclusion follows from the other statements.	Higher order thinking skills free one to act—to frame examined purposes, judge wisely, and evaluate desires by the consequences which will result from acting on them. Reasoning, in this context, involves the ability to analyze a variety of problems, generating logical and reasonable approaches to solve and implement solutions, that reflect a consistent value orientation and a recognition of conditions and perspectives that would support or conflict with such actions.	"Reasonable and reflective thinking that is focused upon deciding what to believe or do." (Ennis, 1984) The test relies on a "rea life" context in which the respondent not only applies the evaluative ability to make appropriate judgments, but also the productive ability to formulate responses and defend them logically. Points for argument analysis are awarded for recognizing and judging specific strengths and weaknesses in the original complex argument presented to respondents, and for adequately defending their judgments

COMPARISON OF COLLEGE LEVEL CRITICAL THINKING TESTS AUTHENTIC/PERFORMANCE BASED

INSTRUMENT	COMP ASSESSMENT OF REASONING AND COMMUNICATING	ETS TASKS IN CRITICAL THINKING (FORMER NEW JERSEY GIS TESTS)
Vendor	American College Testing	Educational Testing Service
Year Started	1986	1989
Current Version	1991	1994
NO. OF FORMS	5 Forms	1 Form
ABILITIES ASSESSED	Select approaches to solve problems Determine logical consistency Determine solution to be implemented Assess internal consistency of values Analyze rationales for value choices	Inquiry: plan a search, use methods of observation, extract/evaluate information Analysis: formulate hypotheses, apply techniques, evaluate evidence/reasoning Communication: organize presentation, write effectively, use quantitative info
NO. OF ITEMS	Analytic scales to rate 15 aspects of reasoning, plus separate rating of 3 essays and 3 speeches for communication skills	Holistic rating of written responses to 9 questions, plus separate reading of an essay for communication skills
NO. OF SCALES	Total Score & 3 subscores in each area (Reasoning, Writing & Speaking)	3 Scores
Unit of Analysis	Individual & Group	Group only
CONTENT	Simulations and role-playing tasks in the humanities, social sciences, and natural sciences using controversial topics related to the family, community problems, energy, health issues, media, music, and the arts.	Extended tasks in the context of the humanities, social sciences, or natural sciences that couch issues or problems requiring inquiry, analysis and communication of conclusions in written and graphic form.
NORMS AVAILABLE	YES (college freshmen, sophomores, and seniors) Longitudinal gain norms for sophomores & seniors	NO (Not interpretable for individuals)
TESTING TIME	105 minutes	90 minutes per task
DEFINITION OF CRITICAL THINKING	Higher order thinking skills free one to act—to frame examined purposes, judge wisely, and evaluate desires by the consequences which will result from acting on them. Reasoning, in this context, involves the ability to analyze a variety of problems, generating logical and reasonable approaches to solve and implement solutions, that reflect a consistent value orientation and a recognition of conditions and perspectives that would support or conflict with such actions.	Critical thinking includes the ability to comprehend and analyze data, apply techniques and provide solutions to problems, utilize information from various media to represent and project results, and develop a plan or position regarding an issue supported with detailed, reasonable evidence.
TESTING COST: 100 Students Also, centrally rated or time for local faculty	$1,575 ($15.75 per student) $2,300 ($23.00 per student) Four days for 5 faculty, including training (35-45 minutes per student)	$1,500 ($15.00 per student) ? Five days for 5 faculty, including training (45-60 minutes per student)

SHORT BIOGRAPHICAL INFORMATION ON TASK FORCE MEMBERS

Short Biographical Information on Task Force Members

Patricia R. Cook

Dr. Cook is Associate Professor and Director of the Associate Degree in Nursing Program at the University of South Carolina at Aiken. She holds a baccalaureate and masters degree in nursing from the Medical College of Georgia. Her Ph.D. is from the University of South Carolina in the area of Educational Administration. Dr. Cook has taught nursing for over 20 years and has devoted much of that time to studying critical thinking. She has made numerous presentations on critical thinking at the state, regional, and national levels. Also, Dr. Cook has authored several publications on various nursing topics in journals, newsletters, and books.

Reid Johnson

Currently the Director of Institutional Effectiveness, Planning and Assessment at Francis Marion University, Dr. Johnson is also a professor of psychology and served previously as the Director of SCHEA (South Carolina Higher Education Assessment Network). Dr. Johnson has an undergraduate degree in mathematics and psychology, and holds an MA and Ph.D. in psychology. The author of over 21 articles, book chapters, and monographs on various aspects of higher education assessment practice, he is currently under contract with Jossey Bass for *Assessing Institutional Effectiveness in Higher Education for Quality Improvement and Accountability* due out in 1997.

Phil Moore

Currently the Director of Assessment and Associate Director of Institutional Planning and Analysis at the University of South Carolina, Dr. Moore holds a B.S. in Psychology as well as an MED and Ph.D. in Educational Research. Dr. Moore has extensive experience in test form development and has served as a consultant for the Ohio Department of Education. His teaching experience includes systematic computer problem solving and educational statistics.

Phyllis Myers

Dr. Myers holds a Ph.D. in educational psychology and currently serves as the director of the Office of Institutional Research at Trident Technical College where she coordinates the college's Institutional Effectiveness process. She has conducted a variety of workshops on assessing student learning at two year colleges including workshops at the state, regional, and national levels as well as in the Bahamas.

Susan Pauly, Editor of this document

Dr. Pauly serves as the Director of Planning at the Lancaster campus of the University of South Carolina as well as serving on the South Carolina Higher Education Assessment Advisory Board. An associate professor in Humanities, she has taught a variety of courses in Women's Studies, composition, and literature and has published on the teaching of Women's Studies. Dr. Pauly holds a BSE in Education and a Masters and Ph.D. in English and has taught in Central America as well as the United States.

Faye Pendarvis, Committee Chair

Currently the Institutional Effectiveness Coordinator at Orangeburg-Calhoun Technical College, Ms. Pendarvis holds a M.Ed. in Community Occupations and Program Evaluation from the University of South Carolina and a B.S. in Business Administration. Faye brings over 13 years experience in educational settings to her current position, and has published in the ERIC Clearinghouse on Counseling and Personnel Services. In addition, she serves as a member of research associations including the South Carolina Higher Education Assessment Association, SCAIR (South Carolina Association for Institutional Research), and SACCR (Southeastern Association for Community College Research).

Joe Prus

A professor of psychology and the director of the Office of Assessment at Winthrop University, Dr. Prus has since 1988 overseen the development of Winthrop's comprehensive program to assess student learning and development. Dr. Prus holds a Ph.D. in psychology, has extensive expertise in psychometrics, and directs Winthrop's school psychology graduate program. In addition to conducting numerous presentations and workshops on assessment at state and national meetings, he has published extensively on assessment and has served as a consultant to professional organizations, colleges, and universities.

Lovely Ulmer-Sottong

Since 1993, Dr. Ulmer-Sottong has served at South Carolina's Higher Education Commission as the State Coordinator for Planning and Assessment. Dr. Ulmer-Sottong oversees South Carolina's state-mandated institutional effectiveness programs--a model which has been cited by the Education Commission of the States as one of the ten model programs in the U.S. Dr. Ulmer-Sottong received her Ph.D. in Higher Education Administration from the University of Kansas. Her previous professional experience includes her work in assessment at American College Testing, extensively consulting work in higher education assessment, and workshops on assessment in almost all fifty states. She has taught in the U.S. and abroad, has authored articles on assessment, and is listed as an outstanding educator in the international and U.S. editions of *Who's Who*.

The South Carolina Higher Education Assessment Network Advisory Board endorsed this document on May 24, 1996.

ERIC is funded by the National Library of Education / Office of Educational Research and Improvement

Clearinghouse on Higher Education

U. S. Department of Education
Educational Resources Information Center (ERIC)

Reproduction Release Form

For each document submitted, ERIC is required to obtain a signed reproduction release form indicating whether or not ERIC may reproduce document. A copy of the release form appears below or you may obtain a from ERIC/HE. Please mail two copies of your document with a completed release form to:

ERIC Clearinghouse on Higher Education
One Dupont Circle, NW
Suite 630
Washington, DC 20036-1183

If you have any questions about submitting documents to ERIC, please p
1-800-773-3742

I. Document Identification

Title: Critical Thinking Assessment

Author(s): Patricia Cook, Reid Johnson, Phil Moore, Phyllis Myers, Susan Pauly, Faye Pendarvis, Joe Prus, & Lovely Ulmer-Sottong
Date: June 1996

II. Reproduction Release

A. Timely and significant materials of interest to the educational community are announced in the monthly abstract journal of the ERIC system, "Resources in Education" (RIE). Documents are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document

Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document. If reproduction release is granted, one of the following notices is affixed to the document.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY:

Marge Tebo-Messina (signature)

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

--OR--

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN
GRANTED BY:

___________________________ (signature)

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

B. If permission is granted to reproduce the identified document, plea CHECK ONE of the options below and sign the release.

X Permitting microfiche (4" x 6" film) paper copy, electronic, and optical media reproduction (Level 1).

___ Permitting reproduction in other than paper copy (level 2).

Documents will be processed as indicated provided quality permits. If permission to reproduce is granted, but neither box is checked, docume will be processed at Level 1.

C. "I hereby grant to the Educational Resources Information Center (ER nonexclusive permission to reproduce this document as indicated. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy inform needs of educators in response to discrete inquires."

Name: Marge Tebo-Messina

Signature: Dr. Marge Tebo-Messina

Organization: South Carolina Higher Education Assessment Network

Position: Executive Director

Address: 210 Tillman Hall, Winthrop University, Rock Hill SC

Tel. No.: 803-323-2341

Zip Code: 29733

E-mail: tebomessinam@winthrop.edu

III. Document Availability Information

(Non-ERIC Source)

If permission to reproduce is not granted to ERIC, or, if you wish ERI cite the availability of the document from another source, please prov the following information regarding the availability of the document. will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware ERIC selection criteria are significantly more stringent for documents which cannot be made available through EDRS).

Publisher/Distributor:

Address:

Price Per Copy:

Quantity Price:

IV. Referral to Copyright/ Reproduction Rights Holder

If the right to grant reproduction release is held by someone other th the addressee, please provide the appropriate name and address:

About ERIC | Search ERIC | HE Library | FAQ | New and Noteworthy
HE Clearinghouse | Other HE Resources | ASHE-ERIC Report Series | Research Initiatives | HE Program Resources

New and Noteworthy | HE Clearinghouse | HE Library | ASHE-ERIC Report Series | Research Initiatives
FAQ | About ERIC | Search ERIC | Other HE Resources | HE Program Resources

GWU Home Page
ERIC-HE Home Page
U.S. Department of Education

Maintained by: *jlauver@eric-he.edu*
Last updated: *Wednesday, 12-Mar-97 12:21:40*