Review of the Collegiate Assessment of Academic Proficiency (CAAP)

Institutional Research, Planning, and Assessment

Changing Lives,
Creating Futures

About the Authors

Molly J. Vaughn, is Coordinator of Analytical Service in the Office of Institutional Research, Planning and Assessment. She has been employed at UNA since September of 2006. Mrs. Vaughn received her M.A. degree in Quantitative Psychology from Middle Tennessee State University in Murfreesboro, TN and completed her B.A., with a dual major in Psychology and Spanish, from the University of North Alabama.

Andrew L. Luna, is Director of Institutional Research, Planning, and Assessment. He has served over 20 years in higher education, with 12 of those years in institutional research. He has published research studies on many topics including salary studies, assessment, market research, and quality improvement. Dr. Luna received his Ph.D and M.A. degrees in higher education administration and his M.A. and B.A. degrees in journalism, all from the University of Alabama.

EXECUTIVE SUMMARY

For the past 14 years, the University of North Alabama has participated in the Collegiate Assessment of Academic Proficiency (CAAP) family of testing instruments with varying degrees of success. This report outlines the following areas of concern:

- There is a significant relationship between actual CAAP scores and student motivation while taking the test. Because over half of UNA's respondents were not motivated, test scores were lower than expected and potentially unreliable
- Most CAAP scores fall below the national average; Math scores are below the national average 100 percent of the time while Science Reasoning scores are below the national average 85 percent of the time
- Recommendations from a 1992 UNA committee report, which would have placed greater responsibility on students to do well on the CAAP exam, were not initiated
- The CAAP exam is not being effectively used at UNA, and trend data indicate that the already expensive exam will continue to become more expensive

Recommendations

- If UNA decides to continue with the CAAP exam, administration cycles should be re-reviewed
- Investigate other General Education assessment instruments used by other colleges and universities and determine cost effectiveness
- Address the 1992 recommendations, especially as they relate to improving student motivation
- Ensure that outcomes from any General Education assessment instrument should be effectively used by UNA to improve academic programs

INTRODUCTION

In 1988, American College Testing (ACT) developed the Collegiate Assessment of Academic Proficiency (CAAP) family of testing instruments. Used as an assessment of General Education attainment, CAAP offers six individual test components: Reading, Writing Skills, Writing Essay (not administered at UNA), Mathematics, Science Reasoning, and Critical Thinking.

The purpose of the CAAP exam is to measure whether or not a student has effectively mastered the basic fundamentals of General Education. Within a college or university, the teaching of General Education courses is designed to improve a student's skills in critical thinking and problem solving, oral and written communication, and the ability to analyze issues from multiple perspectives. Specifically, General Education provides a broad foundation in the social and natural sciences, humanities, and fine arts.

The importance of General Education was first realized by the medieval university and based on the foundations of the trivium (grammar, logic, and rhetoric) and the quadrivium (arithmetic, geometry, music, and astronomy). It was believed that the trivium encompassed the basic subjects, while the quadrivium followed immediately afterward as preparatory work for the serious study of philosophy and theology.

Today, most students have only rudimentary skills (at best) in General Education when they enter college as freshmen. It is believed by the institution that when students take certain courses, they will be subjected to and will retain the skills necessary for General Education mastery. The CAAP exam is given to students during their junior year in college after most students have taken these courses. Therefore, the CAAP is designed to test the retention of General Education skills that students learned during their first two years of college.

In 1992, the University of North Alabama's Board of Trustees agreed to start using CAAP as UNA's primary assessment tool for General Education. During the fall of 1993, UNA administered all of the test modules except Critical Thinking to its students and continued to do so for a few years. Afterwards, test administrations became more sporadic. In fall 2003, UNA started using the Critical Thinking module and also started a cyclic pattern for administering each module.

One of the five components is issued once a semester (summer is excluded). The testing time is approximately 40 minutes, with a total administration time of 50 minutes.

The five components administered at UNA are multiple-choice format with the range of scoring between 40 (low) – 80 (high), including a mean of 60 and a standard deviation of approximately five. For more information on the CAAP instrument and scoring, please refer to the CAAP link on ACT's website: (http:// www.act.org/caap/tests/index.html).

The University requires each student to register and complete the CAAP exam before graduation. The student must have completed 44 semester hours of coursework, but must not have completed more than 70 hours. Transfer students with more than 70 hours of coursework must register for the CAAP within their first semester at the University of North Alabama. All students must have completed the following General Studies Curriculum:

- English 111 and English 112
- At least one semester of the literature requirement
- Math 110 or Math 112 or higher level math course as appropriate
- At least one semester of the laboratory science requirement

In the past, students who met these requirements would register for the preparatory course (CAP 299). The non-credit course was required of all students during their junior year. The course consisted of two one-hour test review sessions and completion of the examination. CAP 299 was a pass/fail course, with a score recorded on the student's transcript. However, there is no longer a preparatory class required of the students. Currently, the students are prepped 15 minutes before the exam begins.

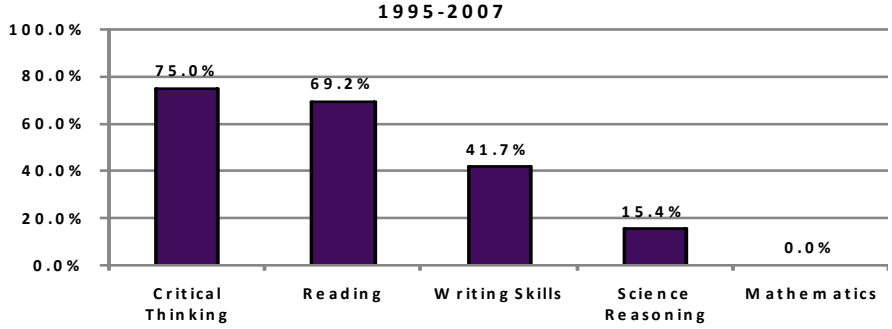
In addition to measuring General Education outcomes, each module of the CAAP exam contains a Likert-type, self-reported performance effort question used to measure the level of motivation while taking the test. For this question, respondents are given the following options: tried my best, gave moderate effort, gave little effort, and no response. As will be noted later in this report, the motivational responses become a significant predictor of test performance.

CAAP SCORES: UNA vs. National

Since the spring of 1995 when UNA started receiving national norm data, there have been 55 modules administered. Of these 55 modules, UNA scored below the national average on 35 of them, indicating that only 36 percent of the time UNA students scored above the national standard.

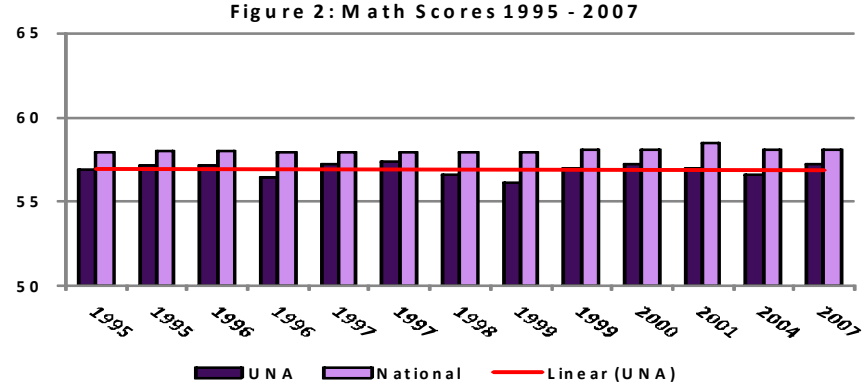
Disaggregating by module, (**Figure 1**) UNA has never scored above the national average on its 13 administrations of the Mathematics exam, and on two of these administrations the University scored in the 37th percentiles. Science Reasoning has been administered 13 times with similar results. UNA scored below the national average eleven times, indicating that only 15.4 percent of the administrations were above the national average. Writing Skills has been administered 12 times with five of those administrations (41.7 percent) scoring above average.

Figure 1: CAAP Results at or Above the National Average 1995-2007



The University demonstrated stronger outcomes in both Reading and Critical Thinking, with nine of the 13 Reading administrations (69.2 percent) scoring above the national average, and three out of four Critical Thinking administrations (75.0 percent) scoring above the national average. Furthermore, as shown in **Figures 2-6**, Math, Reading and Critical Thinking scores had a flat or decreasing trend since administration began on each module.

Figure 2: Math Scores 1995 - 2007



"Of these 55 modules, UNA scored below the national average on 35 of them, indicating that only 36 percent of the time UNA students scored above the national standard."

Figure 3: Writing Skills Scores 1995 - 2005

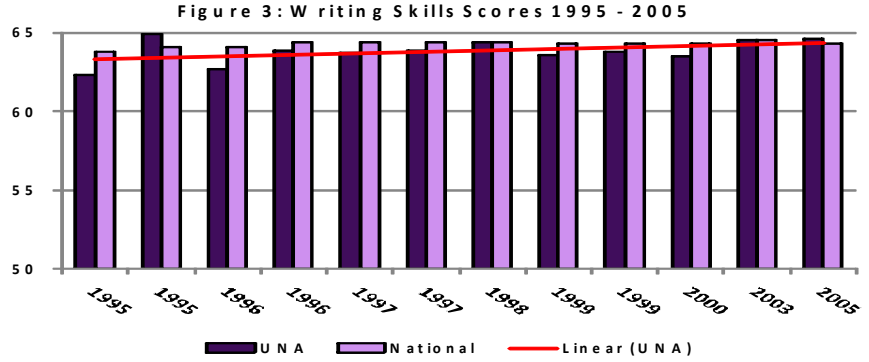


Figure 4: Reading Scores 1995 - 2007

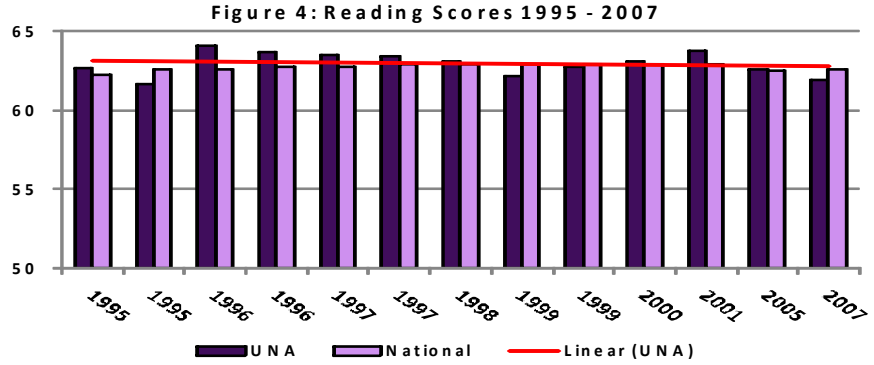


Figure 5: Critical Thinking Scores from 2000-2006

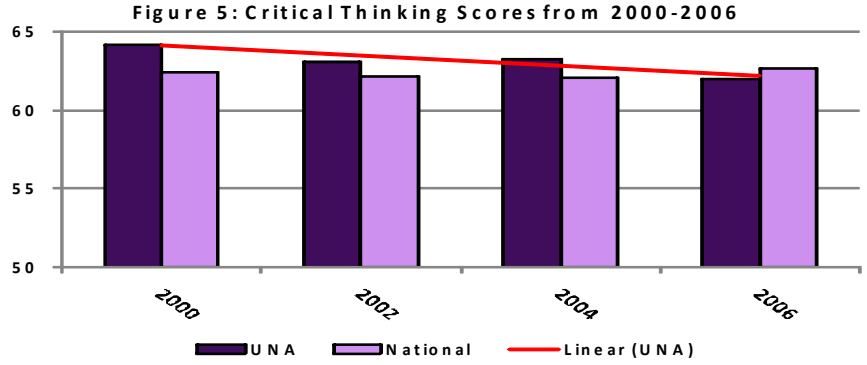
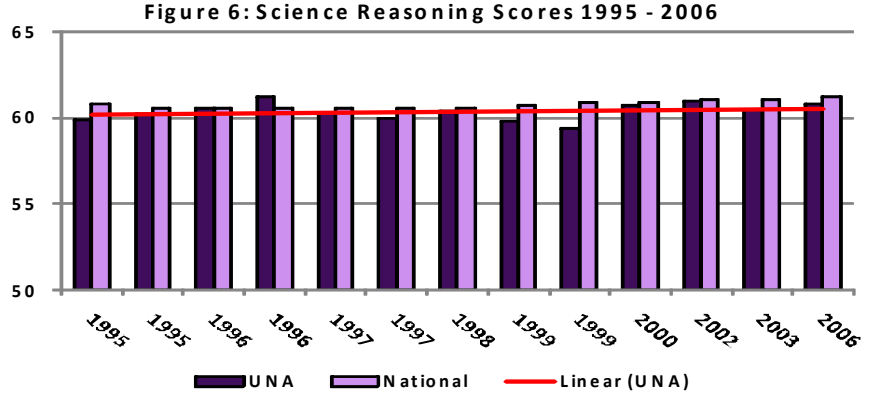


Figure 6: Science Reasoning Scores 1995 - 2006



"These flat or decreasing trends indicate that UNA has either shown no improvement in General Education since 1995, or more likely that students are showing decreasing motivation in completing the exam."

While Mathematics has a somewhat flat trend line, the lines for both Reading and Critical Thinking scores show a decreasing trend. These flat or decreasing trends indicate that UNA has either shown no improvement in General Education since 1995, or more likely that students are showing decreasing motivation in completing the exam. Again, the trendlines for Writing and Science Reasoning indicate some improvement over time.

MOTIVATION OF UNA STUDENTS

As mentioned previously, each CAAP exam module contains a question that attempts to address the respondent's level of motivation at the time they were taking the exam. Since 1995 the average reported percentage, of students who "tried their best" on the CAAP exam was 46.3 percent indicating that over half of the students may not be taking the test seriously. UNA students seem to be aware that the CAAP exam is an administration-based test for the benefit of the University and that the score has no impact on them academically.

"...when student motivation decreases, the CAAP test scores follow suit."

This assumption can be supported statistically. A Pearson Product Moment correlation-coefficient was computed by comparing the percentage of students who stated they "tried their best" on the exam in relation to actual CAAP scores over the 63 administrations **(Figure 7)**. At the .01 level of significance, the Pearson correlation of .413** indicated a significant relationship between motivation and CAAP test scores.

Figure 7: Correlating Scores and Motivation

		UNA mean score	% of students who tried their best
UNA mean score	Pearson Correlation	1	.413(**)
	Sig. (2-tailed)		0.001
	N	63	63
% of students who tried their best	Pearson Correlation	.413(**)	1
	Sig. (2-tailed)	0.001	
	N	63	63

**Correlation is significant at the 0.01 level (2-tailed).

Specifically, the test shows that as the percentage of students who "tried their best" increases, UNA's average score increases as well. Likewise, when student motivation decreases, the CAAP test scores follow suit. Furthermore, because respondents may see no direct benefit in taking the CAAP exam, reliability over the 46.3 percent who reported they "tried their best" could be questioned.

Many Universities (Arkansas State, Iowa State, Central Missouri, and Dakota State to name a few) have created a policy incorporating a minimum score requirement for the CAAP exam to ensure more reliable results.

It is beneficial to include that most of these institutions administer all modules of the CAAP exam to a random sample of students, and these students must meet a minimum score. Most of these universities require enhancement courses if the student does not meet the minimum requirement or the student must remediate and retest in that area within one year. A similar policy at the University of North Alabama could greatly increase motivation, which in turn will increase reliability of the instrument and increase the average score.

The significance of respondents' motivation while taking the test may be discussed statistically. If it is assumed that the CAAP exam effectively measures student mastery of General Education, it can also be assumed that student exposure to General Education increases as they take more college-level courses.

If these assumptions are correct, then we can posit that students who have taken these courses have a higher degree of exposure and knowledge of General Education than those students who have not taken these courses (i.e. new freshmen). Therefore, a "proxy" null hypothesis of the CAAP exam may be that there is no difference between students who have taken General Education courses and those students who have not. If this were an actual null, one would have to give the CAAP exam during students' freshman year and then during the junior year in a pre-post test scenario. The pre-post test method is actually one of ACT's options in evaluating the CAAP exam.

If scores from the CAAP exam indicate that students within a particular university are below or significantly below national averages, one may be willing to accept the null hypothesis stated above and believe that students who took the General Education courses are not much better off than the new freshmen who have not yet taken them. If, however, CAAP exam respondents indicated that they were not motivated to take the test, the results of the exam may not reflect the true nature of the students' mastery of General Education. Therefore, a low motivation score on the CAAP exam could force the outcome of the research into a Type II error.

Type II error is also known as an "error of the second kind", an error, or a "false negative." It is the error of failing to reject a null hypothesis when the alternative hypothesis is the true state of nature. In other words, this is the error of failing to observe a difference when in truth

"Many Universities have created a policy incorporating a minimum score requirement for the CAAP exam to ensure more reliable results."

there is one. This type of error can only occur when the statistician fails to reject the null hypothesis. This theory is exemplified in **Figure 8**.

Figure 8. Type I and Type II Error

		Actual Condition	
		Present	**Absent**
Test Result	**Positive**	Condition Present + Positive Result = True Positive	Condition absent + Positive Result = False Positive **Type I Error**
	Negative	Condition present + Negative result = False (invalid) Negative **Type II Error**	Condition absent + Negative Result = True (accurate) Negative

When students are not motivated to do well on the CAAP exam, the probability of them not doing their best on the exam significantly increases. Therefore, the results of the exam taken by unmotivated respondents may not accurately measure their mastery of General Education and may not be generalized to all juniors within the college or university.

In addition to Type II error causing interpretation problems for the CAAP exam, accepting a false null hypothesis can also be problematic when multiple assessments are used. For example, General Eduction is normally assessed at a college or university using multiple methods. If the results of one assessment method have been falsely interpreted due to Type II error, the conclusions derived from this method may contradict the results of other assessment methods.

IMPLICATIONS

The CAAP exam is the standardized, nationally normed assessment program from ACT that enables postsecondary institutions to assess, evaluate, and enhance the outcomes of their General Education programs.

In 1988 the University of North Alabama adopted the CAAP exam as the instrument that would be used to assess General Education. The administration of the CAAP exam began in 1993 and continues to be a consistent instrument in the evaluation of General Education at UNA. However, many of the objectives the University intended to achieve through this instrument have not been accomplished.
On March 2, 1992 the University prepared a list of future recommendations for the CAAP:

- Establishment of minimum passing scores for the CAAP modules. If students fail to achieve the minimum scores, they might be required to take the examination a second time
- "Use" of CAAP data: A pattern in the data over several semesters or years would be more likely to convince the academic departments involved and/or the University Curriculum Committee to alter the program
- Comparison of CAAP norms of students at UNA to norms from similar institutions in the country
- Linkage reports to establish if the student has improved since taking the ACT
- Subgroup comparisons (for example: first-time freshman vs. transfer students)

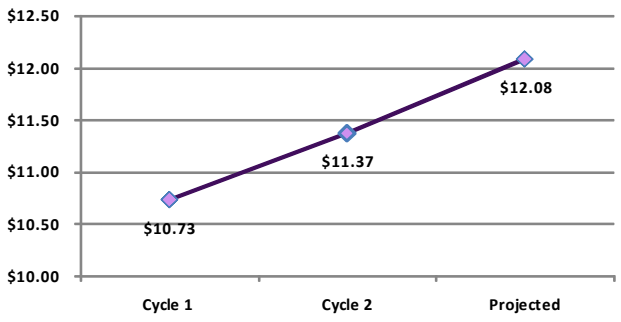
Over the years, ACT representatives, CAAP workshops, and Planning and Institutional Effectiveness meetings have been addressing the uses of CAAP, motivation of students and how to apply the CAAP results to SACS requirements. However, none of these five recommendations were implemented during the 14 years of administration at UNA.

"...this is an average increase of $369.02 per semester."

INFLATING COSTS OF CAAP

The yearly cost of the CAAP exam is increasing based on two factors - escalating cost per test module and increasing enrollment at UNA. The CAAP exam is priced per student. For example, in fall 2007, 520 students participated at a cost of $12.00 per student, totaling $6,240.00. On average the ACT has increased the price of the CAAP exam 71 cents per module. Using the 520 students listed above, this is an average increase of $369.02 per semester.

Figure 9: Rising CAAP Cost



If this increase continues, the University will be paying nearly $10,000 per semester to administer the CAAP exam by 2012 during UNA's re-accreditation visit from SACS **(Figure 9)**.

While the national norm comparisons provided by CAAP are informative, UNA has not consistently made improvements towards meeting those norms based on the data. Again, because lack of student motivation has affected the CAAP scores, the trend data may be unreliable and, therefore, adequate conclusions or generalizations cannot be made.

Furthermore, UNA is the only four-year public college in Alabama that administers the CAAP; it has no customers in Tennessee and only four in Mississippi. Many institutions may be looking elsewhere for their General Education Assessment, perhaps due to the inflating price or reliability issues. Therefore, more research needs to be done to establish the assessment tools being used by other universities.

"UNA does not currently participate in any of the recommended measurement options."

ADDITIONAL MEASUREMENT OPTIONS FOR CAAP

The CAAP results have been analyzed over the past several years and posted on the Office of Research's website (www2.una.edu/research), but this simplistic report did not produce longitudinal results. ACT lists many ways in which universities can benefit from administering the CAAP exam; however, UNA does not currently participate in any of these options.

- **CAAP as an Outcomes Measurement**
  - ☐ Establish benchmarks
  - ☐ Implement improvements to the specific program
  - ☐ Monitor change in student performance over time

- **Longitudinal study (CAAP to CAAP):** This study requires the administration of CAAP to incoming students and then again to the same students, usually at the end of the sophomore, junior, or senior year. By looking at the variation between the CAAP mean scores of both groups, institutions can infer value-added performance gain.

- **Longitudinal (CAAP to ACT):** CAAP is administered to rising juniors, students who have just completed the general education courses. ACT can link their CAAP

results back to their results on the ACT to indicate the level of change on a group basis.

- **Cross-Sectional:** This enables an institution to obtain an initial reading on program outcomes. Incoming freshmen are tested at the beginning of the fall term, and a similar group of sophomores/juniors/seniors is tested at the end of the spring term in the same academic year. The effectiveness of a program may then be inferred from the differences between the two mean group scores. The challenge of this design is matching student characteristics across groups.

- **Content Analysis Reports:** The report provides more detailed data regarding your students' relative strengths and weaknesses in the specific content areas of each test module - these can be compared to national norms as well. Institutions can use this information to evaluate what specific areas of the General Education programs may need enhancement.

  - **Writing Skills**
    - Punctuation
    - Grammar
    - Sentence Structure
    - Strategy
    - Organization
    - Style

  - **Reading**
    - Reasoning Skills
    - Referring Skills

  - **Mathematics**
    - Pre-Algebra
    - Elementary Algebra
    - Intermediate Algebra
    - Coordinate Geometry
    - College Algebra
    - Trigonometry

  - **Science Reasoning**
    - Analysis
    - Generalization

- Understanding

☐ **Critical Thinking**

- Analysis of Elements of Arguments
- Evaluation of Arguments
- Extension of Arguments

CONCLUSION

It is clear that the University has not taken full advantage of the CAAP instrument and, based on past trends, costs for administering the exam will significantly increase. While General Education assessment is vital for UNA, instruments used to obtain these assessments must be used consistently and results should be used to make improvements. The following are recommendations from this report:

- If UNA decides to continue with the CAAP exam, administration cycles should be re-reviewed
- Investigate other General Education assessment instruments used by other colleges and universities and determine cost effectiveness
- Address the 1992 recommendations, especially as they relate to improving student motivation
- Ensure that outcomes from any General Education assessment instrument should be effectively used by UNA to improve academic programs