

**AN EXAMINATION OF THE CONSTRUCT AND CONTEXT
VALIDITY OF THE AREA CONCENTRATION ACHIEVEMENT
TEST (ACAT) IN PSYCHOLOGY**

RHONDA MICHELLE CRAIN

AN EXAMINATION OF THE CONSTRUCT AND CONTEXT
VALIDITY OF THE AREA CONCENTRATION ACHIEVEMENT
TEST (ACAT) IN PSYCHOLOGY

An Abstract
Presented to the
Graduate and Research Council of
Austin Peay State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Rhonda Michelle Crain
July 1989

ABSTRACT

The importance of assessing major field curricula as part of outcomes assessment is becoming evident to faculty and administration in higher education. Newly developed instruments should demonstrate both construct and contextual validity, and must be an accurate reflection of the structure of the major and its objectives.

The Area Concentration Achievement Test (ACAT) in Psychology, consisting of 120 multiple-choice questions in 10 content areas, was administered to 84 students enrolled in General Psychology courses taught at Austin Peay State University during the 1988-89 academic year. Tests were administered either at the beginning or end of the course. Their scores were compared with those of 21 graduating senior psychology majors to determine whether the ACAT accurately assesses curricular impact.

The results indicated that students tested at the beginning of Psychology 1020 (Group I) and Psychology 1030 (Group III) were approximately equivalent in their knowledge of psychology. Taking Psychology 1020 (Group II) seemed to have little effect on test outcome, while taking Psychology 1030 (Group IV) appeared to increase scores on several ACAT subtests. Seniors scored significantly higher than students in Groups I and II on all subtests except physiological, and also produced higher total scores.

Seniors also scored significantly higher than students in Group III, both overall and on all subtests except for social and physiological. Seniors scored higher than students in Group IV, both overall and on all subtests except abnormal, developmental and physiological.

Self-reported overall grade point averages of Groups I through IV correlated significantly with performance on several ACAT subtests. Seniors' self-reported overall GPAs did not correlate significantly with any aspect of ACAT performance. GPA in psychology correlated significantly with overall test score. For Groups I through IV, scores on the experimental design and history and systems subtests were the best predictors of overall scores. For seniors, a combination of scores on the human learning and cognition, abnormal and social subtests were the best predictors of total scores. The evidence indicates that the ACAT possesses both construct and contextual validity.

AN EXAMINATION OF THE CONSTRUCT AND CONTEXT
VALIDITY OF THE AREA CONCENTRATION ACHIEVEMENT
TEST (ACAT) IN PSYCHOLOGY

A Thesis
Presented to the
Graduate and Research Council of
Austin Peay State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Rhonda Michelle Crain

July 1989

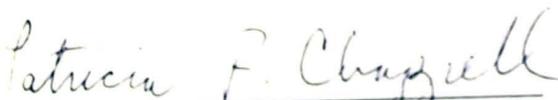
To the Graduate and Research Council:

I am submitting herewith a Thesis written by Rhonda Michelle Crain entitled "An Examination of the Construct and Context Validity of the Area Concentration Achievement Test (ACAT) in Psychology." I have examined the final copy of this paper for form and content, and I recommend that it be accepted in partial fulfillment of the requirements for the degree Master of Arts with a major in Psychology.


Major Professor

We have read this thesis and recommend its acceptance:


Second Committee Member


Third Committee Member

Accepted for the Graduate and Research Council:


Dean of the Graduate School

ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Anthony Golden for his time, patience, and invaluable assistance during the course of this study.

My appreciation is also extended to my other committee members, Dr. Patricia Chappell and Dr. Susan Kupisch, for their support and guidance.

Most of all, I would like to thank my parents for instilling in me the determination to further my education as far as I am able.

TABLE OF CONTENTS

CHAPTER	PAGE
1. REVIEW OF THE LITERATURE	1
Introduction	1
Assessment in the Major	3
Area Concentration Achievement Test	5
Issues of Validity	6
Summary and Hypotheses	9
2. METHOD	10
Subjects	10
Instrument	12
Procedure	13
3. RESULTS	14
4. DISCUSSION	19
APPENDIX	28
REFERENCES	32

LIST OF TABLES

TABLE	PAGE
1. ACAT Subtest Scores and t -Tests Among Groups	29
2. Correlations of ACAT Subtest Scores with Overall Grade Point Average	30
3. Correlations of Senior ACAT Subtest Scores with Overall and Major Grade Point Average	31

CHAPTER 1

Review of the Literature

Introduction

Outcomes assessment as a measure of educational effectiveness, although not a new idea, is currently a focal issue in American higher education. For example, over 1,100 conferees were registered for the recent Fourth National Conference on Assessment in Higher Education, sponsored by the American Association for Higher Education (Hutchings, 1989). Although institutions are using this concept more frequently to provide evidence of accountability to the public and to governing boards, there is a lack of agreement about purposes, definitions and methods of assessment. Disagreements between those outside institutions such as legislatures and state governance boards, and inside, represented by administration, faculty and students, have led to confusion in recent years (Banta & Fisher, 1986).

Data from a recent 50-state survey revealed that two-thirds of the states now have formal initiatives labeled "assessment", compared to only a few states in previous years (Boyer, Ewell, Finney, & Mingle, 1987). Although there is tremendous variety among approaches to outcomes assessment, a strong trend among state governing boards is to consider the development of specific

assessment programs an institutional prerogative (Boyer et al.). Indeed, a first principle of outcomes assessment is that no single approach will be applicable to every institution since each differs in its missions, goals and desired instructional outcomes for different programs (Banta & Fisher, 1986).

Tennessee was one of the first states to implement such a program and was the first to provide financial incentives for public colleges and universities to define student objectives more clearly and to establish assessment procedures to demonstrate institutional effectiveness in meeting these objectives (Boyer & McGuinness, 1986). In 1979, the Tennessee Higher Education Commission (THEC) introduced a performance funding policy that provided a financial supplement of two percent of each institution's instructional budget for engaging in the following evaluation activities: 1) testing undergraduate students in general education; 2) conducting peer reviews and/or testing undergraduates and master's degree candidates in their major field; 3) surveying current students, dropouts, alumni and/or employers to obtain their assessments of academic program quality and related student services; and 4) using assessment data results to make needed improvements in programs and services (Bogue & Brown, 1982). In 1984, the policy's financial supplement was increased to five percent for institutions meeting the

point criteria. The attainment of accreditation for eligible programs was added as an area of instructional evaluation.

The flexibility of Tennessee's policy may well be its greatest strength. Although institutions receive additional rewards for high test scores, they also receive credit for undertaking assessment initiatives and demonstrating that the results have been useful in improving academic programs. Moreover, the policy does not require institutions to use standardized tests in major fields; faculty in each department are encouraged to select or develop a test that measures the achievement of its student objectives (Banta & Fisher, 1986).

Assessment in the Major

The practice of assessing the major is now considered a very important part of the outcomes assessment process, since it is critical to the development of disciplinary curricula (Marchese, 1989). In addition, major field assessment directly involves faculty, who are in the position to give intensive attention to objectives and goals of their curriculum. Matching assessment to departmental goals is critical, and should lead naturally to the development of assessment instruments (Erwin, 1989).

Many institutions have found, however, that the available national field examinations do not meet their

needs, as these tests are designed to measure individual student performance rather than curricula. These standardized tests with fixed content, such as the Educational Testing Service's Major Field Achievement Tests (MFAT) and Graduate Record Examination (GRE) field tests, have proven unsatisfactory to many faculty because of departments' differing content emphases and program goals. Departments often encounter difficulty comparing their students' performance on the GRE to that of similar departments since the norm groups on which the GRE is based are comprised to a large extent of students bound for graduate study. Thus, these tests do not appear to be appropriate for assessment of the major. Additional information, such as a percent correct score, is needed in order to more accurately interpret departmental performance. Departments also need to be able to determine the level of difficulty at which material is learned and whether transfer students' scores on a test reflect their courses or courses taken elsewhere (Golden & Wood, 1988).

Another approach to major field assessment that many departments have chosen is the development of local tests. Although this approach is advantageous in that it allows departments to tailor tests to individual needs, it is almost impossible to demonstrate the validity of these institutionally based tests against external criteria. However, some state governing boards now strongly recommend

comparisons to state and national norms. This emphasis on curricular evaluation as part of outcomes assessment has created a need for specialized measurement instruments (Golden & Wood, 1988).

Area Concentration Achievement Test

The Area Concentration Achievement Test (ACAT) represents an alternative in that it is designed to reflect the content area emphases of individual departments and can provide feedback about program strengths and weaknesses. (Golden & Wood, 1988). Developed in 1983 by Dr. Anthony Golden, the ACAT model currently includes tests in Psychology, Social Work and Political Science. Parallel forms of the test are used to increase both the overall amount of material sampled within a department and test security. Tests are revised periodically using previously field-tested items. To date, over 1,500 ACATs have been administered in Tennessee, Kentucky, North Dakota, Missouri, North Carolina, Ohio and Georgia (Golden & Hardy, 1989). Although the ACAT model serves as a middle ground between locally and nationally developed major field tests, there are several issues which must be investigated. One of these issues is validation, with which the current study is concerned.

Issues of Validity

The first step in conducting a validation study is to identify clearly the desired inference to be derived from test scores (Crocker & Algina, 1986). For this study, the inference was that studying in the psychology major would affect ACAT performance. The second step is to design an empirical study to collect evidence of the usefulness of test scores for supporting such inferences (Crocker & Algina). The three major types of validation studies are as follows.

Content validation is used in situations where the test user wishes to know whether examinees' responses to test items will be representative of responses that would be made to a larger domain of similar items. Criterion-related validation can be used for situations where the test user wants to relate examinees' test scores to performance on a behavioral variable of practical importance (Crocker & Algina, 1986). And finally, construct validation is used in situations where neither of the above definitions is sufficient to define what is to be measured, but the test user wishes to relate examinees' test performance to behaviors or performances that can be grouped under the label of a particular construct or trait (Crocker & Algina).

Messick (1975) states that any discussion of the meaning of a measure should center on the concept of

validity, and specifically on the concept of construct validity. One view of validity is that it is an evaluative judgment of the extent to which inferences and decisions derived from test scores are supported by evidence and rationale (Messick, 1988b). Thus, it is not the test itself that is to be validated but rather the inferences about meaning and interpretation derived from test scores and about action taken as a result of these inferences (Messick, 1988a).

Anastasi (1988) also defines construct validation as determining the extent to which a test measures a theoretical construct or trait. Constructs derive from established interrelationships among measures and are developed in order to explain observed response consistencies. Thus, each construct measured by a particular test can only be defined adequately by data gathered in the validation process. The definition of the construct should take into account the variables with which the test correlates significantly, the conditions found to affect its scores, and the groups that differ significantly in such scores (Anastasi).

The concept of construct validity is critical in determining the method by which content domains are defined and external criteria conceptualized and measured. Therefore, according to Messick (1988a), it is unnecessary to consider separately content and criterion validity when

in actuality the construct validity of score interpretation underlies all score based inferences, including those derived from content and criterion methods of validation. It also is important to consider the context in which the assessment instrument is used. The concept of construct validity is especially applicable to the ACAT model since analysis of student performance throughout the psychology major can be used as the basis for curricular modifications.

Summary and Hypotheses

This study compared ACAT scores of students enrolled in general psychology courses with each other and with scores of graduating senior psychology majors. The specific hypotheses to be tested were as follows.

1. Those students enrolled only in the first semester of General Psychology (PSYC 1020) should obtain higher scores on the ACAT subtests which are equivalent to that course's content areas, i.e., human learning and cognition, physiological, animal learning and motivation, and history and systems, than on the other subtests.

2. Students enrolled only in the second semester of General Psychology (PSYC 1030) should obtain higher scores on the ACAT subtests which correspond to that course's content areas, i.e., development, personality, abnormal and social, than on the other subtests.

3. Since the test is designed to assess curricular impact, graduating senior majors should score higher than General Psychology students on all subtests of the ACAT, including total score.

4. Graduating senior majors with higher overall and psychology grade point averages should produce higher total scores on the ACAT.

CHAPTER 2

Method

Subjects

Participants in this study were 106 students enrolled in General Psychology courses taught at Austin Peay State University during the academic year 1988-89. Data for 16 students were omitted because the students did not consent to its use and data were omitted for six additional students because these students fell into more than one experimental group. Of the remaining 84 volunteers, 60 were females, 23 were males, and one gave no gender response. The mean age of the students was 22 years. Eleven were transfer students and 51 reported that they planned to attend graduate school, although not necessarily in psychology.

At the time of test administration, students were asked to provide information on non-survey content area courses in psychology taken at Austin Peay State University. Sixty-three students reported that they had taken no other psychology courses at Austin Peay State University other than the General Psychology course in which they were enrolled at the time of testing. Of the 21 students who reported taking other courses, nine students had taken one other course, five students had taken two other courses and seven students had taken three or more

psychology courses offered at Austin Peay State University. Credit for participation in the study was awarded at the discretion of the course instructors.

Thirty-three graduating senior psychology majors also were tested as mandated by the Tennessee Higher Education Commission. Data for 12 of the students were omitted because consent was not given for its use. Of the remaining 21 students, 14 were females and seven were males. The mean age of the students was 28 years, 10 were transfer students, and 17 reported plans to attend graduate school. These students reported taking all categories of psychology courses offered at Austin Peay State University with the exception of sensation and perception.

The students enrolled in the General Psychology courses were assigned to groups as follows.

Group I. Seventeen females and nine males enrolled in the first semester of General Psychology (PSYC 1020) who were tested near the beginning of the course in the Spring semester of 1989.

Group II. Thirty females, 12 males, and one student who did not give a gender response enrolled in Psychology 1020 who were tested near the end of the course in the Fall semester of 1988.

Group III. Seven females enrolled in the second semester of General Psychology (PSYC 1030) and tested near the beginning of the course in the Spring semester of 1989.

Group IV. Two males and six females enrolled in Psychology 1030 who were tested near the end of the course in the Fall semester of 1988.

Group V. Twenty-one graduating senior psychology majors (14 females, seven males) who were tested near the end of the Spring semester of 1989.

Instrument

The Area Concentration Achievement Test (ACAT) in Psychology consists of 120 multiple-choice items divided into two sections. The ten content areas which are contained in the ACAT version used by Austin Peay State University and in the present study are as follows: Abnormal; Animal Learning and Motivation; Development; Experimental Design; Personality; Physiological; History and Systems; Social; Human Learning and Cognition; and Statistics. The ACAT currently is normed on 408 cases from nine institutions. Testing materials, scoring and access to the data were provided by the Project for Area Concentration Achievement Testing (PACAT) which is supported through a grant from the Fund for the Improvement of Post-Secondary Education (#P116B81711) to Dr. Anthony Golden and Austin Peay State University.

Procedure

After participants entered the classroom in which the test was to be administered, the examiner described the ACAT and explained the purpose of the experiment. Answer sheets were provided and the accompanying instructions for their completion given. Each answer sheet included a consent statement which, when signed by the student, allowed PACAT access to test data and other educational information deemed necessary for research purposes. The test booklets were then distributed in numerical order and instructions given in accordance with the administration manual for responding to test items. The test was divided into two 60-minute sections with a 10-minute rest period allowed between sections.

CHAPTER 3

Results

Students' standard scores (\underline{M} = 500, \underline{SD} = 100) on the test were used for the analyses. Grade point averages, overall and in the major, were reported by the students in ranges rather than as exact values. The SPSS-X statistical package (Release 3.0) running on a VAX 8250 mainframe computer was used to perform the analyses.

Differences between the means of each group's scores were obtained by using t -tests for independent samples, as shown in Table 1. Groups I and II were compared in order to determine what differences in content area knowledge might be present in students who were completing Psychology 1020. No indications of gains in any area were found and Group II scores were significantly lower on the subtests in abnormal, $t(67) = 2.21$, $p < .05$, history and systems, $t(67) = 2.61$, $p < .05$, and statistics, $t(67) = 3.50$, $p < .001$.

Scores for Groups I and III were compared in order to determine whether students began each course with an approximately equivalent knowledge base in psychology. Students tested near the beginning of their enrollment in Psychology 1020 (Group I) scored significantly higher on the statistics subtest than did students tested near the

beginning of their enrollment in Psychology 1030 (Group III), $t(31) = 2.55$, $p < .05$. No other differences between subtests were found.

Groups III and IV were compared in order to determine what content area performance differences might exist as a result of completing Psychology 1030. When the scores of students tested near the beginning of 1030 were compared with those tested near the end of the semester, it was found that the latter group scored significantly higher on the abnormal psychology subtest, $t(13) = -2.21$, $p < .05$.

To determine what material relevant to ACAT content areas was learned in each semester of General Psychology, t -tests were performed on the scores of Group II students (tested near the end of 1020) and Group IV students (tested near the end of 1030). Group IV students scored significantly higher on the subtests in abnormal, $t(49) = -4.65$, $p < .001$, developmental, $t(49) = -3.15$, $p < .005$, history and systems, $t(49) = -2.71$, $p < .01$, personality, $t(49) = -2.82$, $p < .01$, and total score, $t(49) = -2.37$, $p < .05$. No areas were found in which Group II scored higher than Group IV.

Data were subjected to stepwise linear regression to determine the best predictors of ACAT total score. For Group I, a combination of history and systems, social, and experimental design scores were the best predictors of the

total score, accounting for 82.6% of the variance, $F(3,22) = 27.26$, $p < .001$. For Group II, experimental design and history and systems scores were the best predictors, accounting for 65.8% of the variance, $F(2,40) = 42.53$, $p < .001$. Being a transfer student to Austin Peay was the best predictor of total score for Group III, accounting for 68.7% of the variance, $F(1,5) = 14.19$, $p < .05$. For Group IV, a combination of abnormal, experimental design, and statistics subtest scores were the best predictors, accounting for 96.6% of the variance, $F(3,4) = 67.68$, $p < .001$.

Pearson's product-moment correlation coefficient was used to examine the relationship between subtest scores and students' self-reported overall grade point averages (see Table 2). Participants in Group I reported modal overall GPAs between 2.1 and 2.5. A significant positive relationship was found between overall GPA and the statistics subtest, $r = .46$, $p < .05$. Group II students reported GPAs between 2.6 and 3.0. Significant positive relationships were found between their GPAs and the subtests in animal learning and motivation, $r = .33$, $p < .05$, experimental design, $r = .32$, $p < .05$, history and systems, $r = .37$, $p < .05$, physiological, $r = .34$, $p < .05$, social, $r = .33$, $p < .05$, and total score, $r = .51$, $p < .001$. Group III students reported GPAs between 2.1 and 2.5, but no

significant relationships were found between subtest scores and GPA for this group. Group IV reported GPAs between 2.1 and 2.5. Significant correlations were found between their GPAs and the subtests in abnormal, $r = .92$, $p < .001$, developmental, $r = .85$, $p < .01$, and total score, $r = .87$, $p < .005$.

For comparisons between graduating senior majors' test scores and introductory students' test scores, t -tests for independent samples were used (see Table 1). Seniors scored significantly higher than did students in Group I on all ACAT subtests except physiological, and also produced higher total scores. Similar results were found in comparing seniors' scores with Group II students' scores. Graduating psychology majors also scored significantly higher than did students in Group III on all of the subtests with the exceptions of physiological and social, and also produced higher total scores. In comparing scores of Group IV students with scores of the senior majors, no significant differences were found between these groups on the abnormal, developmental and physiological subtests. On all of the other subtests as well as total score, seniors scored significantly higher than did students in Group IV.

Since 10 of the 21 senior majors reported that they had transferred to Austin Peay State University, additional t -tests were performed to determine whether differences in subtest and total scores existed for these two subgroups of Group V. No significant differences were found.

In order to determine the best predictors of total score for the graduating psychology majors, stepwise linear regression was used. Analysis indicated that a combination of human learning and cognition, abnormal, and social scores were the best predictors, accounting for 83% of the variance in total scores, $F(3,17) = 33.9$, $p < .001$.

Pearson's product-moment correlation coefficient was used to determine the relationship between seniors' self-reported overall and major GPAs and ACAT subtest scores (see Table 3). The modal overall GPA reported for seniors was in the 2.6 to 3.0 range. These grade point averages did not correlate significantly with subtest scores or with total scores. Seniors' modal psychology GPAs fell in the 2.6 to 3.0 range. Psychology GPA correlated significantly only with total score, $r = .47$, $p < .05$.

CHAPTER 4

Discussion

Students tested at the outset of General Psychology 1020 (Group I) and 1030 (Group III) appeared to have an equivalent knowledge base in psychology, with the exception of the statistics content area. Group I students scored higher on this subtest than did students in Group III, a finding from which two interpretations may be derived. First, Group I students may have had a more extensive background in mathematics and/or statistics than had Group III students. Second, the small size of Group III ($n = 7$) may have produced a sample that was not representative of the overall group of students taking Psychology 1030. In the latter case, however, more than just the one difference should have been obtained.

To determine what contributions Psychology 1020 made to ACAT subtest performance, the scores of students tested at the beginning of the semester (Group I) were compared with the scores of students tested close to completion of the course (Group II). The hypothesis was that Group II students would score higher than Group I students on those ACAT subtests which reflected the course's objectives. Data analysis revealed, however, that Group I students scored significantly higher on the abnormal, history and

systems, and statistics subtests than did Group II students. Contrary to the hypothesis, no areas were found in which Group II students scored significantly higher than Group I students.

Caution should be taken in interpreting this aspect of the results, however. Absence of the differences that were anticipated in Group II scores could be a result of the failure of Psychology 1020 to contribute to knowledge in these areas, of Group I students being unusually knowledgeable in these areas, or of a lack of sensitivity of the ACAT to introductory level skills. However, the two groups tested at the outset of their respective courses (Groups I and III) differed only in the area of statistics, suggesting that the score discrepancy on the two remaining subtests may have derived from some source other than knowledge prior to taking the course. Furthermore, the significantly higher scores achieved by students completing Psychology 1030 (Group IV) on the abnormal psychology subtest when compared to Group III further suggests that the lowered performance of Group II may actually have resulted from some aspect of completing Psychology 1020. However, since only a small number of subtests are involved, a lack of sensitivity of the ACAT to introductory level skills still cannot be ruled out.

The scores of Groups II and IV were compared in order to determine the contributions of each course to ACAT subtest performance. Group IV students performed better on overall score and on the abnormal, developmental, history and systems, and personality subtests. No areas were found in which Group II students scored higher than Group IV students, although it was expected that Group II would have scored higher on the subtests corresponding to Psychology 1020 course objectives. The subtests on which Group IV students scored higher did in fact reflect Psychology 1030 course objectives. This finding further supports the observations that the ACAT does have some limited sensitivity to introductory level skills and that Psychology 1020 does not contribute to ACAT performance. It should be noted, however, that within the Austin Peay State University general education curriculum, Psychology 1020 and 1030 serve somewhat different "audiences" and therefore may be expected to produce different learning outcomes.

As was expected, graduating senior majors scored higher than did students in all introductory level groups on most of the ACAT subtests. The only area in which no differences existed between any groups was in physiological psychology. The failure to find discrepancies between group means on this subtest probably reflects introductory

students' exposure both to high school and to college level biology courses. Furthermore, only 6 of the 21 senior majors reported taking a physiological psychology course, which probably contributed to the lack of score differences.

Although taking Psychology 1020 does not seem to affect ACAT scores, taking Psychology 1030 has a positive impact in certain areas of test performance. While seniors scored higher both on overall score and on most of the subtests than did Group III students, scores on the social and physiological subtests did not differ for the two groups. Also, comparisons between Group IV students' scores and seniors' scores showed no statistically significant differences on the abnormal, developmental, and physiological subtests. The subtests in social, abnormal, and developmental correspond to common objectives for Psychology 1030 among departmental faculty at Austin Peay State University.

It was not the case that seniors performed poorly in these areas, but rather that Psychology 1030 students performed better than was expected on these subtests. This finding raises some interesting questions. The 15 students comprising Groups III and IV may well have been among the most advanced introductory students in terms of the number of psychology courses taken, which would put them at a

somewhat higher level of knowledge. The lack of differences between Groups I and III, however, argues against this interpretation. Another possible explanation is that these subtests may have a greater degree of sensitivity to introductory level skills than do the other subtests.

Students' self-reported overall grade point averages were found to be related to test performance for Groups I, II and IV. For Group I, the correlation was significant for the statistics portion of the test. Higher GPAs among students in Group II was associated with better performance on the animal learning and motivation, experimental design, history and systems, physiological, and social subtests, as well as on the overall scores. For Group IV, significant positive correlations were found between overall GPAs and overall scores, as well as on the abnormal and developmental subtests. Thus, these findings indicate that higher overall grades tend to predict better performance on several portions of this test.

Using the subtest scores themselves to predict overall score on the ACAT, the most consistent predictors for the introductory student groups were scores on the history and systems and experimental design subtests. Therefore, if the ACAT reflects the major at Austin Peay, then these two

content areas would appear to be a foundation for entry into the major.

An unexpected finding was that seniors' self-reported overall GPAs did not correlate significantly with subtest or total scores on the ACAT. GPA in psychology was found to correlate significantly only with total scores. Thus, higher GPAs in the major tend to predict higher total test scores but not higher scores on subtests which correspond to courses taken in the major. The test appears to be less sensitive to the general knowledge or ability of psychology majors which would be reflected by overall GPA.

Since senior performance was superior to that of the introductory students on almost all subtests as well as overall scores, it seems that the Psychology ACAT can be useful for measuring what it was designed to measure, namely, the impact of the psychology curriculum on student majors. The evidence gathered from this study indicates that the ACAT possesses a high degree of validity when it is used in the context of a senior exit exam and can play an important role in the growing practice of major field assessment.

How well test data can support inferences and decisions made depends a great deal on the validity of the instrument. According to Messick (1988b), this is the crux of construct validity. For this study, the construct was

perceived as curricular impact on students majoring in psychology. To measure this, the ACAT was designed with a unique system for curriculum matching, which enables it to accommodate different departments' program emphases. The ten content areas listed earlier in this paper form the emphasis of Austin Peay State University's psychology curriculum.

The second step in defining the construct was to find relationships between the test and other variables. It seems obvious that GPA is an intervening variable, since analysis of student scores indicated that ACAT performance correlated significantly with the self-reported overall GPAs of three of the four introductory student groups and with self-reported major GPAs of the graduating senior psychology majors. The latter finding suggests that if a student does well in his or her psychology courses, he or she will score well on the ACAT.

The next step was to identify conditions that affect test scores. It is apparent that a student who has not taken many psychology courses will not perform as well on the test. This conclusion can be seen clearly when comparing scores of the introductory students with those of the graduating senior majors. Thus, the most practical use of this test would be to administer it as closely as possible to completion of the major. This finding also

lends support to the contextual validity of the test. For departments wishing to measure the progress of their majors, however, the test could be given as students declare the major and again at its completion.

Faculty and administration in various disciplines have a need to explore the meaningfulness of their curricula and the amount of benefit that study in the major provides to their students. The development of new instruments to measure achievement in many different fields can provide a large part of this information. Of course, student performance on tests will not accurately reflect the broader objectives of the departmental major. Additional sources of information, collected over long periods of time, are imperative for proper assessment of the major. Student and faculty surveys, evaluation of student advising programs, peer reviews, review of accreditation standards, rates of graduate school placement, interviews with graduates' employers and periodic faculty evaluations of curricular objectives are all examples of other assessment procedures for the major. Departmental use of instruments such as the ACAT can, however, provide suggestions for needed curricular modifications based on student performance.

Several issues raised in this study have not been fully answered by the data collected. Further research is

necessary to determine, for example, whether the findings in this study are typical of introductory level student performance on the ACAT. Perhaps different trends would be found if the same students were tested both at the beginning and end of each General Psychology course, using parallel forms of the test to prevent overfamiliarity. Longitudinal studies should also be conducted to examine test performance across the major. Information deriving from investigations such as these could greatly enhance faculty understanding of the structure and objectives of the psychology major at Austin Peay State University.

APPENDIX

TABLE 1
ACAT Subtest Scores and t -Tests Among Groups

Content Area	Group				
	I ($n=26$)	II ($n=43$)	III ($n=7$)	IV ($n=8$)	V* ($n=21$)
Abnormal ^{acde}	393	347	394	511	534
Animal Learning and Motivation ^{ef}	377	419	376	373	505
Developmental ^{ce}	367	367	404	472	534
Experimental Design ^{ef}	415	385	368	395	530
History and Systems ^{acef}	418	360	386	441	519
Human Learning and Cognition ^{ef}	396	415	403	436	554
Personality ^{cef}	373	343	429	425	547
Physiological	472	491	521	466	513
Social ^f	450	414	437	410	522
Statistics ^{abef}	426	368	344	373	548
Total Score ^{cef}	354	325	349	388	553

Significant differences with $p < .05$

a Groups I vs II

b I vs III

c II vs IV

d III vs IV

e III vs V

f IV vs V

* All comparisons of Groups I and II with Group V were significant with the exception of the physiological subtest.

TABLE 2

Correlations of ACAT Subtest Scores
with Overall Grade Point Average

Content Area	Group			
	I (<u>n</u> =26)	II (<u>n</u> =43)	III (<u>n</u> =7)	IV (<u>n</u> =8)
Abnormal	.047	.244	.044	.923**
Animal Learning and Motivation	.089	.336*	.288	.174
Developmental	.134	.101	.599	.850**
Experimental Design	.297	.328*	.548	.165
History and Systems	.088	.371*	-.439	.374
Human Learning and Cognition	-.002	.268	.135	.199
Personality	.193	.098	.181	.656
Physiological	-.002	.343*	.495	.700
Social	.191	.339*	.317	-.196
Statistics	.467*	-.013	-.379	.301
Total Score	.311	.515**	.531	.871**

* $p < .05$

** $p < .01$

TABLE 3

Correlations of Senior ACAT Subtest Scores
with Overall and Major Grade Point Average

Content Area	GPA	
	Overall ($n=21$)	Major ($n=21$)
Abnormal	.376	.215
Animal Learning and Motivation	.307	.296
Developmental	.384	.366
Experimental Design	-.011	.219
History and Systems	.060	.114
Human Learning and Cognition	.355	.426
Personality	.291	.197
Physiological	.077	.005
Social	-.117	.141
Statistics	.113	.340
Total Score	.372	.470*

* $p < .05$

REFERENCES

REFERENCES

- Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.
- Banta, T.W. and Fisher, H.S. (1986). Assessing outcomes: The real value-added is in the process. In K. McGuinness (Ed.), Proceedings from the Conference on Legislative Action and Assessment: Reason and Reality (pp. 81-91). Arlington, VA.
- Bogue, E.G. and Brown, W. (1982). Performance incentives for state colleges. Harvard Business Review, 60, 123-128.
- Boyer, C.M., Ewell, P.T., Finney, J.E., & Mingle, J.R. (1987, March). Assessment and outcomes measurement: A view from the states. AAHE Bulletin, 37, 8-12.
- Boyer, C.M. and McGuinness, A.C. (1986, February). State initiatives to improve undergraduate education, ECS survey highlights. AAHE Bulletin, 38, 3-7.
- Crocker, L.M. and Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.
- Erwin, T. (1989, June). Assessment in the major: Questions and answers. Paper presented at the meeting of the American Association for Higher Education, Atlanta, GA.
- Golden, A. & Hardy, R. (1989). Item banks for building instruments. Paper presented at the meeting of the American Association for Higher Education, Atlanta, GA.
- Golden, A. and Wood, M. (1988). PACAT: Program for Area Concentration Achievement Testing. (Grant #P116B81711). Fund for the Improvement of Post-Secondary Education, U.S. Department of Education.
- Hutchings, P. (1989, June). Opening remarks. In L.L. Knefelkamp, The case for parallel assessment: Getting behind outcomes. Keynote address at the meeting of

the American Association for Higher Education,
Atlanta, GA.

Messick, S. (1975). The standard problem: Meaning and values in test validation. American Psychologist, 30, 955-966.

Messick, S. (1988a). Meaning and values in test validation: The science and ethics of assessment (Research Report 88-47). Princeton, NJ: Educational Testing Service.

Messick, S. (1988b). Validity. In P. Linn (Ed.), Educational measurement (3rd ed.). New York: Macmillan.

Schneider, C. (1989, June). Remarks. In T. Marchese (Chair), First things first: Identifying and assessing outcomes of the major. Symposium conducted at the meeting of the American Association for Higher Education, Atlanta, GA.

Study Group on the Conditions of Excellence in American Higher Education (1984). Involvement in learning: Realizing the potential of American higher education. Washington, DC: National Institute of Education.