DOCUMENT RESUME

ED 373 820   JC 940 469

AUTHOR   Chen, HongYu; And Others
TITLE   Assessing General Education: Selection and
              Implementation of an Instrument To Satisfy Internal
              and External Constituencies.
PUB DATE   2 Aug 94
NOTE   33p.; Paper presented at the Annual Conference of the
              Southeastern Association for Community College
              Research (23rd, Savannah, GA, August 1-3, 1994).
PUB TYPE   Reports - Resea _ᵇ/Technical (143) --
              Speeches/Conference Papers (150)

EDRS PRICE   MF01/PC02 Plus Postage.
DESCRIPTORS   Academic Achievement; *College Outcomes Assessment;
              Community Colleges; *Comparative Testing;
              *Educational Assessment; Educational Testing;
              *General Education; Scores; Test Norms; *Test
              Selection; *Test Validity; Two Year Colleges

ABSTRACT
              As part of an effort to develop strategies for
measuring institutional effectiveness, a study was conducted at West
Virginia University at Parkersburg (WVU-P), a two-year institution
with an enrollment of nearly 4,000 students, to determine the most
appropriate instrument for measuring gains in general education
between incoming freshman (i.e., with 15 or fewer credit hours) and
near completers (i.e., with more than 45 but fewer than 75 credit
hours). WVU-P reviewed five assessment instruments available for
measuring general education attainment, electing to pilot test short
forms of the Academic Profile (AP) and the College BASE (CB). The AP
was administered to 69 students and the CB was administered to 48,
while the total sample population had a mean age of 24.6 and was
57.4% female. The mean score for the AP was 444, with a standard
deviation of 15.7, compared to a mean AP score of 443 nationwide. The
CB mean score for English was 268, with a standard deviation of 64,
and 196 for mathematics, with a standard deviation of 46. Test scores
indicate that entering WVU-P students are below the national average
in general education, but after completion of 45 credit hours,
students are generally above the national average. Moreover, the AP
was found to be a more consistent and statistically meaningful
measure than the CB with respect to differences between incoming
freshmen and near-completers in general education. Report of Student
Outcome Assessment Tests Analysis is appended. Contains eight
references. (KP)

Assessing General Education: Selection and Implementation

of an Instrument to Satisfy

Internal and External Constituencies

HongYu Chen

Director, Institutional Research

Phil O. McClung

Chair, Outcomes Assessment Committee

Associate Professor in Psychology

Eldon Miller

Campus President

WVU at Parkersburg

Parkersburg, WV 26101

Presented at the

Southeastern Association for Community College Research

Twenty-third Annual Conference

Savannah, Georgia

August 2, 1994

Table of Content

/

List of Tables

Table   Page

List of Figures

Assessing General Education: Selection and Implementation
of an Instrument to
Satisfy Internal and External Constituencies

West Virginia University at Parkersburg (WVU-P) is committed
to the support of efforts that study, develop, evaluate, and
implement strategies related to increasing institutional
effectiveness. More recently, the focus has been on developing
strategies to measure outcomes in the area of general education.
Although there are a number of professionally developed
instruments available to measure gains in general education, the
instruments: the College Outcome Measures Program (COMP), the
Collegiate Assessment of Academic Proficiency (CAAP) both from
American College Testing (ACT), the Academic Profile from
Educational Testing Service (ETS), and the College Basic Academic
Subjects Examination (College BASE) from the Riverside Publishing
Company (RPC) were considered for implementation at the WVU-P.

Past studies indicated that both COMP and the Academic
Profile are not appropriate for evaluating the impact of general
education outcomes (Pike, 1988). As for the instruments from
ACT, both COMP and CAAP are not adequately reliable and are not
sensitive to the institution's general education course work
(Pike, 1998). Welch (1989) stated that black students did not
perform as well as did their white peers on CAAP. Doolittle
(1989) found no overall performance differences on CAAP between
males and females. Regardless of the drawbacks pointed out by

2

the researchers, many higher education institutions still adopted ACT-COMP and/or CAAP as their assessment tools. As for the instrument from ETS, Thorndike (1992) found that the College BASE has a good test-retest reliability but also shows growth in subject areas. However, as pointed out by Thorndike, College BASE adds little information about entering students that is available from other sources. Stiggins (1985) stated that self-developed tests are far better in aiding teachers in their classroom-based assessment than standardized tests. However, due the emphasis of public awareness and accountability, WVU-P was to use a nationally recognized standardized instrument.

A study conducted at WVU-P in early Spring, 1992 (Appendix A) indicated that the CAAP was the least preferred instrument among "general education core-curriculum" faculty. Due to financial reasons, WVU-P elected to use the Academic Profile and the College BASE in this pilot study.

The pilot study was conducted during the next to last week of April, 1992. The test instruments utilized were the short form of the Academic Profile (AP) from ETS and the short form of the College BASE from the RPC. The purpose of the pilot study was to find the most appropriate instrument that would accurately measure the gain between students who are incoming freshmen (zero credit hours earned) and near completers (earned 45 credit hours or more) in the general education curriculum.

Sample Selection

WVU-P is a non-residential, two year institution with an enrollment of nearly 4,000 students, consisting of 63% females and 2% minorities. The majority of the students are employed while they take classes and the mean age of the student body is 26.6 years. Due to the nature of the commuting/working students, it is almost impossible to select a true random sample and expect the sampled students to take a test without monetary or tangible incentives, especially at the end of the semester. Therefore, WVU-P's Outcomes Assessment Committee (OAC), with the consent of course instructors, selected four classes* for the pilot testing. Among the selected classes, there was a morning class, an evening class, and two afternoon classes. The total number of students tested was 117, with a gender distribution of 42.6% male, 57.4% female, and a mean age of 24.6 years.

Pilot Testing

The pilot tests were implemented by Phil McClung and HongYu Chen in the presence of the class instructor. The tests were administered using guidelines designed by the testing companies. Based on the observations, the majority of the students took the tests seriously and answered questions thoughtfully. However, because the results of these tests would not influence the students' grades in any way, it was felt that some students marked the answers without careful thinking, especially in the afternoon sessions.

WVU-P, SACCR 23rd Annual Conference, August, 1994

4

Scoring

The tests were scored by the testing companies. It took approximately four weeks to score AP and six weeks to score College BASE (CB).

Results

The short form of the AP test includes English, math, natural and social sciences, with an administration time of 40 minutes. The ETS reports a single score for each student. The mean score for AP (total of 69 students) was 444, with a standard deviation of 15.7. Compared to the mean score (443) of two-year colleges that had used AP nationwide. WVU-P's score is slightly higher the national norm.

The short form of CB includes only English and math, with a 60 minutes completion time. The mean CB score for English (total of 48 students) was 268, with a standard deviation of 64. The mean score for mathematics was 196, with a standard deviation of 46. There were no national norms available for comparison. However, the RPC reports both students' scores as well as their proficiency level of high, medium, and low. The majority of WVU-P students scored within the levels of medium and low as shown in Table 1 and Figure 1.

Data Analysis

All the analyses were done in the Office of Institutional Research at WVU-P.

Table 1

Percentage of Student Who Reached in Each level for the College

BASE

Subject	High	Medium	Low
English	13.3%	51.3%	35.3%
Math	2.1%	43.1%	54.7%

Figure 1: WVU-P Students Proficiency on College Base

Data from a sample of 48 students.

WVU-P, SACCR 23rd Annual Conference, August, 1994

11   12
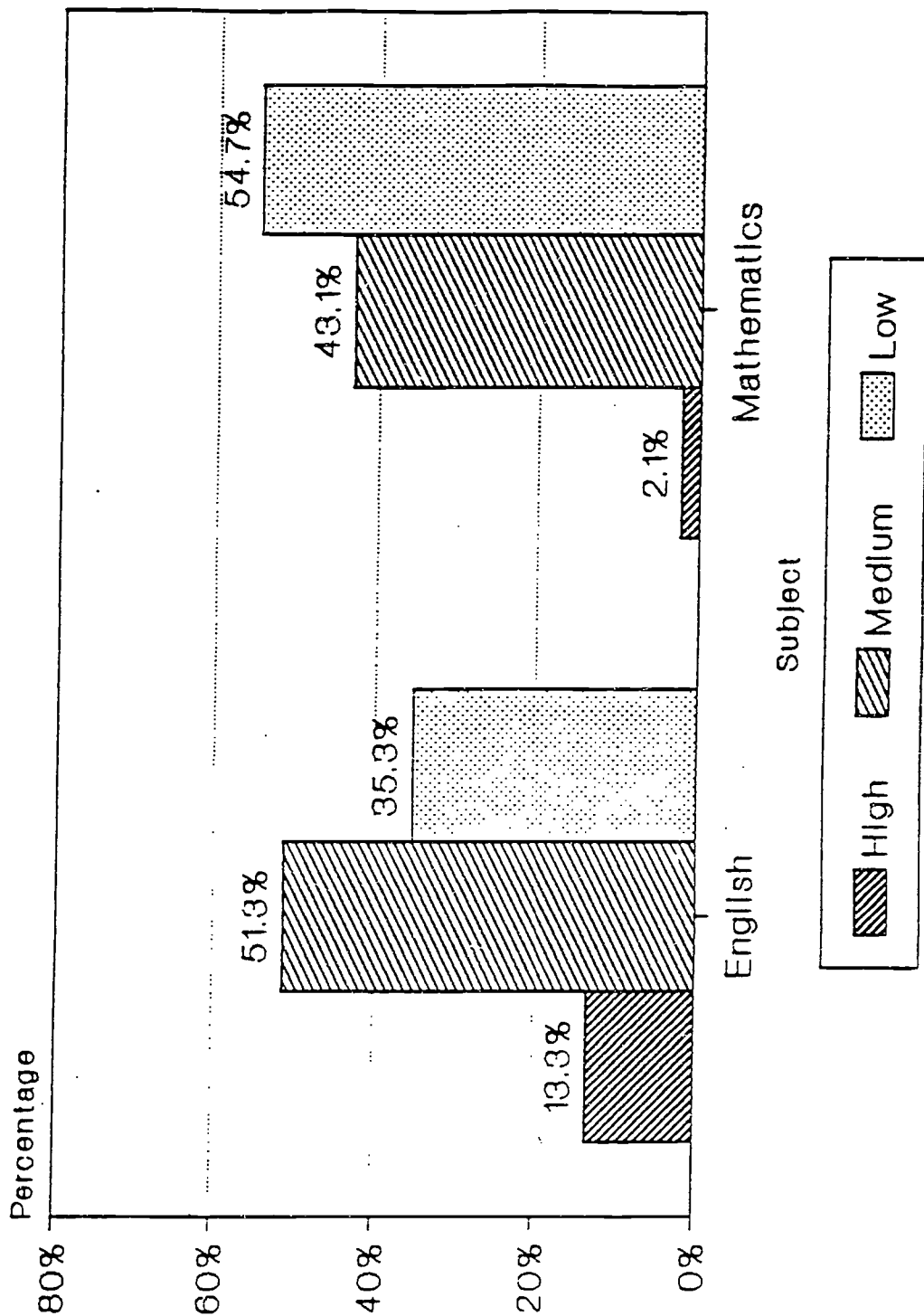
Analysis on the Academic Profile

In order to see whether AP showed a significant gain between incoming freshmen and near completers, the analysis of variance (ANOVA) was used to detect the significance of the mean difference. Because there were few students who had earned exactly zero and 45 hours, the grouping was completed based on: a) incoming freshmen - students who had earned 15 or fewer cumulative hours at the time of administration; b) near completers - students who had earned more than 45 but fewer than 75 hours by April, 1992. There were 24 students in the incoming freshmen group and 18 students in the near completers group. Taking into account the consideration of unequal group size (the larger variance is associated with the larger group) as shown in Table 2, the F statistic is robust (Stevens, 1990).

Compared to students in other comprehensive colleges and universities, WVU-P freshmen were more academically disadvantaged (ETS, 1991). However, compared with the sophomore student group, WVU-P students' mean score was higher than the scores of those colleges and universities (Figure 2). This indicates that entering WVU-P students are below the national average in general education, but after completion of 45 credit hours, these students are above the national average.

The ANOVA table presented in Table 3 shows no significant statistical difference between the mean score of incoming freshmen and near completers at .05 alpha level (although it is

8

Table 2

Descriptive Information of Grouping for Academic Profile

Group	Mean	Std Dev	Min.	Max.
Incoming Freshmen (24)	439.9	16.8	413	467
Near Completers (18)	449.6	13.0	428	476

14

Figure 2: WVU-P Student Performance on Academic Profile

Mean Test Score

450
440
430
420

443
439.9
447.9
449.6

Incoming Freshmen   Near Completers

Student Group

☐ Other College/Univ.   ▨ WVU-P
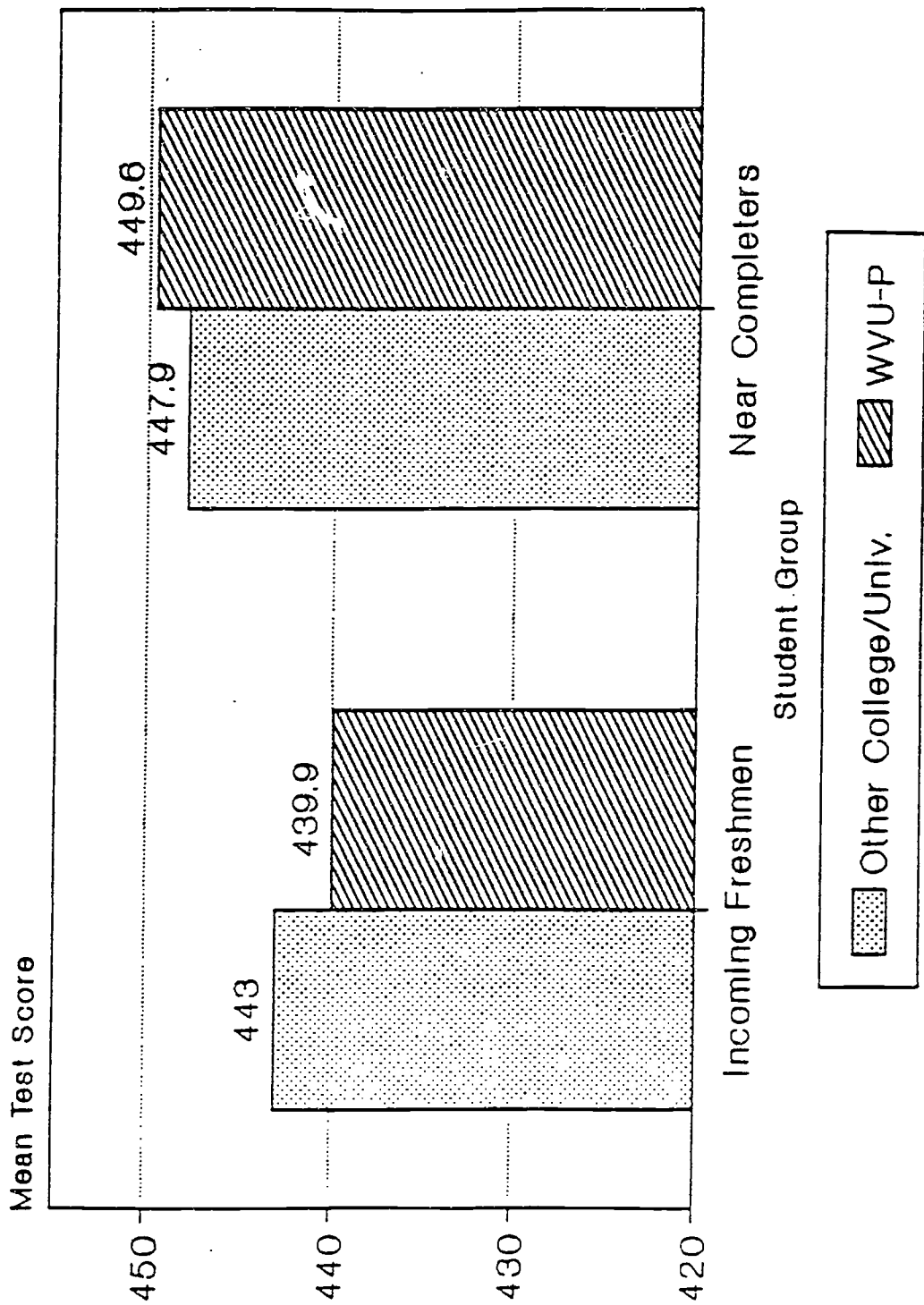
Data is from a sample of 42 students.

WVU-P, SACCR 23rd Annual Conference, August, 1994

16

15

very close to be significant). However, a near 10 point gain of
the near completers group may have practical meaning although it
is not statistically significant. A post hoc estimation of power
based on the available sample presented in Table 4 indicates that
if one were to look for a large gain the power was sufficient for
detecting existing differences (power $=$ 0.93). However, if one
were interested in detecting small differences, the power would
be insufficient (power $=$ 0.10) based on the sample size selected.

Analysis on the College BASE

Three separated ANOVAs on English, math, and total score
were implemented for testing the anticipated difference.
However, due to a lack of "qualified" subjects, the grouping for
these analyses was slightly different: a) incoming freshmen -
students who had earned 15 or fewer cumulative hours when the
test was implemented; b) near completers - students who had
earned more than 45 hours by April, 1992. There were equal
numbers of students (18) in both the incoming freshmen and the
near completers groups. The descriptive information of the CB
respondents is presented in Table 5 and Figure 3.

Compared to the AP scores, CB shows a larger gain. However,
the standard deviation for CB is much greater than AP. This
could be the result of the difference in the seriousness of
students who completed the test as well as the difference in
backgrounds (general education) of our sampled students.

The ANOVA results presented in Table 6 indicate that there

WVU-P, SACCR 23rd Annual Conference, August, 1994

Table 3

ANOVA Table for Academic Profile

Source	DF	Sum of Squares	Mean Squares	F Value	P
Model	1	955.6	955.6	4.08	.0503
Error	40	9380.3	234.5		
Total	41	10335.9			

12

Table 4

<u>Post Hoc Power Estimation</u>

Effect Size	Sample Size	Harmonic Mean	Power
Small (0.1)	42	20.6	.10
Medium (0.3)	42	20.6	.52
Large (0.5)	42	20.6	.93

19

was no statistical significant difference between the mean score
of incoming freshmen and near completers at .05 alpha level for
all three scores (English, mathematics, and total).

Conclusion

The Academic Profile, in comparison with the College BASE,
was found to be a more consistent and statistically meaningful
measure of differences between the incoming freshmen and the
near-completers at WVU-P. This study adds credibility to a
previous study conducted in which a majority of the WVU-P faculty
rated the Academic Profile as preferred instrument to measure
gains in general education studies.

* The selected classes were Dr. Robert McCloy's business 381 and
410, and two of Mr. Phil McClung's psychology 111 classes.

14

Table 5

Descriptive Information of Grouping for College BASE

Group	Mean	Std Dev	Min.	Max
English				
Incoming Freshmen	247.8	68.0	126	352
Near Completers	279.1	52.7	185	405
Mathematics				
Incoming Freshmen	185.2	59.5	142	274
Near Completers	207.6	50.8	117	304
Total (English + Math)*				
Incoming Freshmen	443.1	100.1	268	626
Near Completers	486.7	82.2	350	629

* The addition for minimum and maximum may not be for the same person.

2±

WTI-P, SACCR 23rd Annual Conference, August, 1994

Figure 3: WVU-P Students Performance on College Base

Mean Test Score

500 — 400 — 300 — 200 — 100 — 0

English: 247.8, 279.1
Mathematics: 185.2, 207.6
Total (Eng + Math): 443, 486.7

Subject

☐ Incoming Freshmen   ▨ Near Completers

Data is from a sample of 48 students.

WVU-P, SACCR 23rd Annual Conference, August, 1994

23

22

16
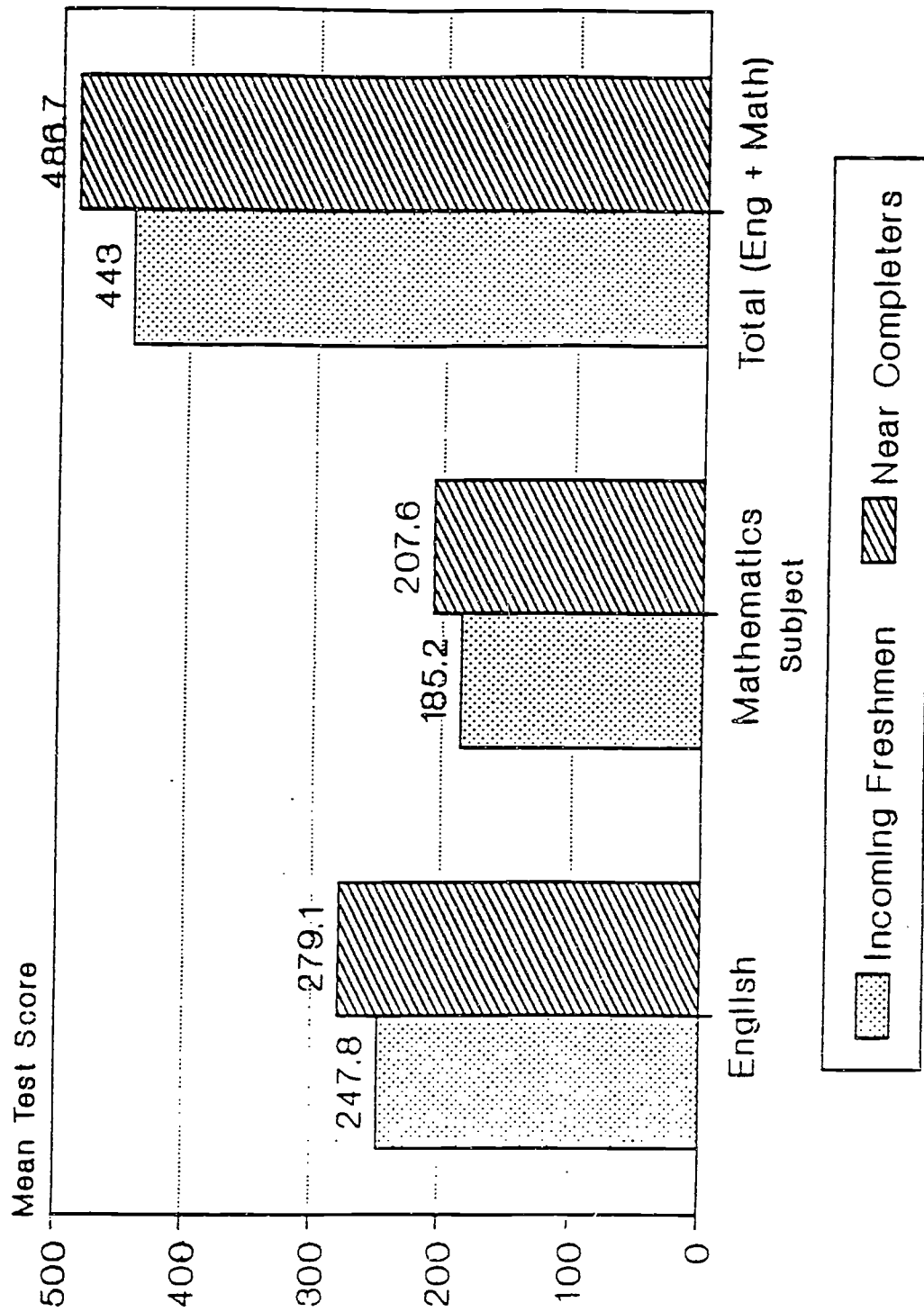
Table 6

<u>ANOVA Results for College BASE</u>

Group	Degree of Freedom	F Value	P Value
English	1, 34	2.37	.13
Mathematics	1, 34	1.47	.23
Total	1, 33	2.32	.14

References

Doolittle, A. & Welch, C. (1989). Gender differences in performance on a college-level achievement test. San Francisco, CA: American Educational Research Association Annual Meeting. (ERIC Document Reproduction Service No. ED 306 237)

ETS. (1991). The Academic Profile: Comparative data guide. NJ: Princeton.

Pike, G. R. (1988). A comparison of the College Outcome Measures Program (COMP) Exam and the ETS Academic Profile. Knoxville, TN: Tennessee University, Learning Research Center. (ERIC Document Reproduction Service No. ED 348 369)

Pike, G. R. (1989). A comparison of the College Outcome Measures Program (COMP) and the Collegiate Assessment of Academic Proficiency (CAAP) Exams. Knoxville, TN: Tennessee University, Learning Research Center. (ERIC Document Reproduction Service No. ED 320 911)

Stevens, J. (1992). Applied multivariate statistics for the social sciences. (2nd Ed.). NJ: Hillsdale.

Stiggins, R. J. (1985). Improving assessment where it means the most: In the classroom. Educational Leadership, 43(2), 69-74.

Thorndike R. M. & Andrieu-Parker, J. M. (1992). Growth in knowledge: A two-year longitudinal study of changes in scores on the College Basic Academic Subjects Examination. (ERIC Document Reproduction Service No. ED 346 134)

18

Welch, C. & Others (1989). <u>Differential performance on a direct measure of writing skills for black and white college freshmen</u>. Iowa City, IA: American College Testing Program. (ERIC Document Reproduction Service No. ED 320 955)

Appendix A - REPORT OF STUDENT OUTCOME ASSESSMENT TESTS ANALYSIS

Given the time and expertise of department chairpersons and faculty members, we have completed an analysis of content quality/ suitability and difficulty for WVU-P students on three well known test sets (All three tests claim to be suitable for measuring student learning outcomes in a two-year college program). The test sets examined were:

Test set A - "Academic Profile" from Educational Testing Services,
Test set B - "College BASE" from the Riverside Publishing Company,
Test set C - "College Assessment of Academic Proficiency" (CAAP) from American College Testing Program (ACT). The four subjects analyzed were: English, Mathematics, Natural Sciences, and Social Sciences.

The ratings from faculty analyzing the test (total of 26, 10 from English, 3 from Mathematics, 5 from Natural Sciences, and 8 from Social Sciences Department), indicates a preference for Academic Profile (Test set A). The overall faculty ratings on the three test sets are:

Test Sets	Quality/Suitability New Student	45 Hrs Plus	Test Difficulty New Student	45 Hrs Plus	Faculty's Preference
A	3.12	2.87	2.28	2.64	First
B	3.08	3.04	2.52	3.13	Second
C	3.58	3.34	1.91	2.57	Third

Note: The numbers reported above are weighted mean values based on the following criterion:
a) Test Quality/Suitability
   1 = Excellent 2 = Good 3 = Neutral 4 = Poor 5 = Very Poor,
b) Test Difficulty
   1 = Too Difficult 2 = Somewhat difficult 3 = About Right
   4 = Somewhat Easy 5 = Too Easy.

The participating faculty indicated that test A (Academic Profile) is suitable for both our incoming freshmen (zero credit hour earned) and near completers (students near completion of the AA, AS, or AAS program - have earned more than 45 credit hours). The faculty ratings on Academic Profile shifted from "somewhat difficult" for our incoming freshmen toward "about right" for our near completers.

Some faculty members from the English, Mathematics, and Natural Sciences Departments commented that none of the three test sets are a perfect fit for WVU-P. Some of the test questions fell outside the content of our curriculum or courses. Endeavors in searching for a more suitable testing instrument or developing our own test is recommended. However, the Board of Trustees has mandated student testing for outcome purposes. It is anticipated that one of the "nationally recognized" test instruments will be mandated in the near future.

WVU-P, SACCR 23rd Annual Conference, August, 1994

The following is the test content information about each test (long form).

Test A - Academic Profile (test designed with three proficiency levels)

Subject	Percent	No. of Items	Test Time
Humanities		36	
Literature	50%		
Philosophy	25%		
Music	10%		
Art	10%		
Film	5%		
Social Sciences		36	
History, Political Science, International Relations	47%		
Behavioral Sciences (Psychology, Sociology, Urban Studies)	31%		
Economics	20%		
Anthropology	2%		
Natural Sciences		36	
Biology*			
Chemistry			
Physics			
Experimental Finding			
College-Level Reading		36	
College-Level Writing		36	
Critical Thinking		36	
Mathematics		36	
Total		144**	2.5 hr.

  * In Natural Sciences, Biology weights more than Chemistry and Physics.

  ** Reported from the user's manual, somehow it is not the summation of the above.

Note: One can ask up to 50 locally written questions.

Test B - College BASE (without 40 minutes writing exercise)

Subject	No. of Items*	Test Time

English
  Reading & Literature
  Writing
	41	40 min.

Social Studies
  History
  Social Sciences
	42	40 min.

Science
  Laboratory & Field Work
  Fundamental Concepts
	41	40 min.

Mathematics
  General Mathematics
  Algebra
  Geometry
	56	40 min.

Total	180	160 min.

* Approximation of the number of questions.

Note: No locally written question can be added.

Test C - ACT/CAAP

Subject	No. of Items			Test Time
Writing Skills	72			50 min.
Usage/Mechanics		32		
Punctuation			6	
Grammar			8	
Sentence structure			18	
Rhetorical Skills		40		
Strategy			15	
Organization			10	
Style			15	
Mathematics	35			50 min.
Algebra		27		
Pre & elementary algebra			7	
Intermediate algebra				
and coordinate geometry			10	
Advanced algebra			10	
Trigonometry		4		
Calculus		4		
Reading	36			50 min.
Referring		7-9		
Reasoning		27-29		
Inferring			22-26	
Applying			3-5	
Critical Thinking	32			50 min.
Analysis of an argument		20		
Evaluation of an argument		6		
Extension of an argument		6		
Science Reasoning	32			50 min.
Data Representation		15		
Research Summaries		24		
Conflicting Viewpoints		6		
Writing (Essay)				50 min.
Total*	?			?

* Each student can take any combination of test subject(s),
  therefore, the total number of questions and testing time vary
  accordingly.

Note: One can ask up to 9 locally written questions.

Stimuli: 1 = Academic Profile, 2 = College BASE, and 3 = ACT/CAAP.

RUN TITLE: All Faculty (number of responses = 26)

MATRIX OF INPUT PROPORTIONS:

	1	2	3
1	.500	.615	.654
2	.385	.500	.731
3	.346	.269	.500

MATRIX OF ORDERED PROPORTIONS:

	1	2	3
1	.500	.615	.654
2	.385	.500	.731
3	.346	.269	.500

MATRIX OF ORDERED Z-SCORES:

	1	2	3
1	.0000	.2924	.3934
2	-.2924	.0000	.6128
3	-.3934	-.6128	.0000

STANDARD DEVIATION OF STIMULI:

1	2	3
1.857	.263	.880

SCALE DISTANCE FROM STIMULUS AT RIGHT

1	2	3
.068	.328	

STIMULI NUMBER	SCALE VALUE	(Faculty Preference)
1	.396	(The Academic Profile)
2	.328	(The College BASE)
3	.000	(The ACT/CAAP)

The sample interpretation (of English faculty) is on the next page.

Stimuli: 1 = Academic Profile, 2 = College BASE, and 3 = ACT/CAAP.

RUN TITLE: English Faculty (No. of response = 10)

STIMULI NUMBER   SCALE VALUE

1	1.072	INTERPRETATION: Relative to stimuli #3, the English faculty's preference for stimuli #1 is about 5 times stronger than that of the stimuli #2.
2	.184	
3	.000	

RUN TITLE: Mathematics Faculty (No. of response = 3)

STIMULI NUMBER   SCALE VALUE

2	1.166
1	.583
3	.000

RUN TITLE: Natural Science Faculty (No. of response = 5)

STIMULI NUMBER   SCALE VALUE

2	.842
1	.129
3	.000

RUN TITLE: Social Science Faculty (No. of response = 8)

STIMULI NUMBER   SCALE VALUE

1	-.451
2	-.225
3	.000

NOTE: IF ANY SCALE VALUE IS NEGATIVE, YOU MUST RECALCULATE ON THE BASIS OF A STIMULUS DIFFERENCE TABLE. THE TROUBLE IS DUE TO THE FACT THAT THE DATA ACTUALLY HAVE MORE THAN ONE DIMENSION. SOME PROPORTIONS WERE INCONSISTENT.

Because of the multi-dimensional ratings from the Social Sciences faculty, the ratings are re-calculated without the responses from the Social Sciences Department.

RUN TITLE: All Faculty (without Social Sciences)
  PROGRAM INPUT PARAMETERS:
              NCOD =   4
                NS =   3
               NPS =   0
               NSP =   0
                NP =   18 (number of responses)
             NITAP =   0

  MATRIX OF INPUT PROPORTIONS:

          1   2   3
   1   .500   .611   .778
   2   .389   .500   .778
   3   .222   .222   .500

  MATRIX OF ORDERED PROPORTIONS:

          1   2   3
   1   .500   .611   .778
   2   .389   .500   .778
   3   .222   .222   .500

  MATRIX OF ORDERED Z-SCORES:

           1   2   3
   1   .0000   .2819   .7621
   2   -.2819   .0000   .7621
   3   -.7621   -.7621   .0000

  STANDARD DEVIATION OF STIMULI:

     1   2   3
   1.317   .653   1.029

  SCALE DISTANCE FROM STIMULUS AT RIGHT

     1   2   3
   .207   .757

STIMULI NUMBER   SCALE VALUE   (Faculty preference)

       1   .965   (The Academic Profile)
       2   .757   (The College BASE)
       3   .000   (The ACT/CAAP)

  WVU-P, SACCR 23rd Annual Conference, August, 1994