

DOCUMENT RESUME

ED 055 096

TM 000 812

AUTHOR

Aleamoni, Lawrence M.; Eitelbach, Sarah B.

TITLE

Comparison of Six Examinations Given in Rhetoric 101,  
at the University of Illinois, Fall, 1965.

INSTITUTION

Illinois Univ., Urbana. Office of Instructional  
Resources.

PUB DATE

Jun 70

NOTE

10p.

EDRS PRICE

MF-\$0.65 HC-\$3.29

DESCRIPTORS

\*College Freshmen; \*Comparative Analysis;  
\*Composition Skills (Literary); Correlation; Grades  
(Scholastic); Rhetoric; Statistical Analysis; \*Test  
Reliability; \*Test Validity; Writing Skills

ABSTRACT

Two forms of the CEEB English Composition Test and four tests constructed by the University of Illinois Rhetoric department were compared as final examinations for the freshman English composition course. Results from 2545 students indicate that while the CEEB tests discriminate more between students and are more reliable, the departmental tests are more valid and correlate more highly with final course grades. (AG)

# research report

Comparison of Six Examinations Given  
in Rhetoric 101, at the  
University of Illinois, Fall, 1965

Lawrence M. Aleamoni  
Sarah B. Eitelbach

Measurement and Research Division  
Office of Instructional Resources  
University of Illinois  
507 East Daniel Street, Champaign

June, 1970

measurement and research division  
office of instructional resources  
university of illinois  
urbana, illinois

U.S. DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
OFFICE OF EDUCATION  
THIS DOCUMENT HAS BEEN REPRO-  
DUCED EXACTLY AS RECEIVED FROM  
THE PERSON OR ORGANIZATION ORIG-  
INATING IT. POINTS OF VIEW OR OPIN-  
IONS STATED DO NOT NECESSARILY  
REPRESENT OFFICIAL OFFICE OF EDU-  
CATION POSITION OR POLICY.

Comparison of Six Examinations Given in Rhetoric 101,  
at the University of Illinois, Fall, 1965

At the end of Fall semester, 1965, six separate final examinations were administered to 2,545 students enrolled in Rhetoric 101, the basic English composition course required of all freshmen. Two of these examinations were the College Entrance Examination Board (CEEB) English Composition Tests. The other four were final examinations constructed by the Rhetoric department. The 2,545 students were a randomly selected sample out of the 4,100 students enrolled in the course. The two forms of the CEEB tests were administered to randomly selected groups, each of which was to contain approximately 1,000 students. The remaining students were administered the various forms of the departmental examinations according to the regular final examination schedule. Each student, therefore, took only one of the objective tests.

The CEEB English Composition Test is available in several forms, two of which were used in the present study: NPL and KPL 1. It is a one hour objective test designed to assess indirectly a student's ability to write. The test has three parts: Part A measures correctness and effectiveness of expression, Part B measures ability to organize ideas and materials, and Part C measures sensitivity to language.

The four Rhetoric final examinations (M49, W59A, IR54, and IW54) are also one-hour objective tests designed to assess indirectly a student's ability to write. These tests have four sections: Section A - vocabulary, Section B - spelling, and Sections C and D - knowledge of what constitutes good usage and effectiveness in sentence construction.

The students' answers were coded on DIGITEN answer sheets and then processed, yielding cards with all item information. The students' responses (now on cards) were then processed by the Measurement and Research Division's item analysis program which provided the statistics necessary for making a comparison among the six tests.

In Table 1 all the relevant test statistics obtained from the item analysis are presented for each of the tests.

Table 1

Test Statistics for the Six Rhetoric 101 Examinations

Test	Mean (Highs Only)	$\sigma$	$\sigma^2$	Raw Score Range	N	R	Number of Alternatives	Skewness	Kurtosis	K-R 21 Measure	Standard Error of Difficulty	Mean Difficulty Range	Difficult Point Range	Mean Point Biserial Range	Point Biserial Range
CEEB-NPL	59.02	12.86	165.38	16-95	846	100	4-5	-.13	-.05	.862	4.330	.590	.165 to .937	.260	-.093 to .460
CEEB-KPL1	56.03	12.83	164.61	17-92	855	100	4-5	-.19	.07	.859	4.407	.563	.092 to .937	.282	.121 to .416
M49	78.39	7.07	49.98	60-93	132	70	4	.38	-.13	.668	3.567	.784	.159 to 1.000	.190	-.063 to .507
W59A	78.01	7.31	53.44	60-93	212	70	4	-.26	-.50	.685	3.677	.780	.340 to 1.000	.190	-.071 to .397
IR54	72.12	8.07	65.12	48-87	119	100	4	-.39	-.24	.698	3.884	.721	.151 to 1.000	.204	-.464 to .404
IV54	73.51	7.74	59.91	45-95	381	100	4	-.48	.80	.681	3.767	.735	.003 to .984	.201	-.085 to .380



The standard deviations ( $\sigma$ ), variances ( $\sigma^2$ ), and raw score range indicated that the scores varied over a wider area for the CEEB tests than for the Rhetoric tests. This could be interpreted as meaning that the CEEB tests are discriminating between more students.

The number of students (N), and the number of items (k), and the number of alternatives are presented for each test.

The skewness measure indicates how well the sample distribution compares to a normal one. If the high raw scores were more numerous than the low scores, then the distribution would be negatively skewed. On the other hand, if the low raw scores were more numerous than the high, then the distribution would be positively skewed. The NPL and KPL 1 distributions were the most nearly normal of the six.

Kurtosis is used to measure the peakedness of a distribution. If the distributions were normal, kurtosis would be zero. If the distribution had a higher peak than the normal, kurtosis would be positive. If the distribution had a lower peak than the normal, kurtosis would be negative. Here, again we found that the NPL and KPL 1 distributions were the most nearly normal.

The Kuder-Richardson Formula 21 (K-R 21) provides an estimate of the reliability of a single test from a single administration. As one can see from Table 1, the two CEEB tests had the highest reliability coefficients, indicating that they were measuring ability to write more consistently than were the departmental tests.

The standard error of measurement is the degree to which test score (within one standard deviation of the mean) could vary in the total population. The standard errors were highest for the CEEB tests because the standard deviations of these tests were also higher.

The discriminating power of an item is measured by the point-biserial correlation. The point-biserial correlation is used when a dichotomous variable is to be related to a continuous variable. Here, the distribution of the dichotomous responses to an item are related to the distribution of test scores in order to see if the discriminating ability of the test score can be reflected in the item. Looking at the mean point biserials one can see that the CELB tests were doing a better job of discriminating between the students taking each test than were the Rhetoric tests. In order to see if these differences were significant an analysis of variance was run. Table 2 presents the analysis of variance summary. Because the F value was significant, the Scheffé test was used to determine the source of significance. Table 3 gives the results of this test. Here, the greatest source of difference was between the two CELB tests and the Rhetoric department tests. This reinforces the conclusion that the CELB tests are doing a better job of discriminating between students than were the Rhetoric tests.

Table 2  
Analysis of Variance Summary Table

Source of Variation	SS	df	MS	F
Treatments	1.683	5	.3365346	32.7352366
Within Treatments	6.107	594	.0102805	
Total	7.789	599		

Table 3  
Scheffé Test for Multiple Comparisons on Point Biserial for  
The Six Tests

	NPL	KPL 1	M49	W59A	IR54	IW54
NPL						
KPL 1	p < .05					
M49	p < .01	p < .01				
W59A	p < .01	p < .01	NSD			
IR54	p < .01	p < .01	NSD	FSD		
IW54	p < .01	p < .01	NSD	FSD	NSD	

At the end of the fall semester grades were reported for all rhetoric students. The grades of those students who had taken the CEEB and Rhetoric tests were correlated with their scores on the respective tests. The results are presented in Table 4.

Table 4  
Correlation of Total Score to Rhetoric 101 Grade

	Rhetoric Grade
NPL	.3810
KPL 1	.3765
M49	.5227
W59A	.4748
IR54	.4441
IW54	.5259

Table 5

Test of Significance Between Correlations of Total Score  
To Rhetoric 101 Grade

	NPL	KPL 1
M49	p = .024	p = .020
W59A	NSD	NSD
IR54	NSD	NSD
IW54	p = .001	p = .001

The Rhetoric tests correlate more highly with the grade than do the CEEB tests. This may result because the Rhetoric department tests were designed specifically to test the objectives of the rhetoric course whereas the CEEB tests were designed on a national basis with certain national rhetoric objectives in mind. Table 5 presents the results of the test of significance between the correlations in Table 4. The two CEEB test score-grade correlations were significantly different than the M49 and IW54 Rhetoric test score-grade correlations. This is understandable since the M49 and IW54 tests correlate most highly with the course grade.

The two forms of the CEEB tests were correlated by parts to the Rhetoric grade for each student. The results are presented in Table 6. This correlation was done to determine whether any one part of the CEEB tests could be used as a substitute for the whole test. Part 1 of the NPL test correlated most highly with the Rhetoric 101 grade. The addition of Parts 2 and 3 for NPL increases the correlation coefficient by .06 which normally would not justify the retention of these two parts of the test.

However, Part 1 accounts for only 10% of the variation and adding Parts 2 and 3 adds an additional 4% of the variation. For this reason one might want to keep all three parts of the test. For the KPL 1 test, Part 3 correlated most highly with the Rhetoric 101 grade. The addition of Parts 1 and 2 for KPL 1 increases the correlation by .03. In this case Part 3 of KPL 1 accounts for 14% of the variation and Parts 1 and 2 add 2% of the variation.

Table 6

Correlation of Three Parts  
of CELE Tests to  
Rhetoric Grade

Multiple of  
Three Parts to  
Rhetoric Grade

NPL

	Part 1	Part 2	Part 3	Grade
Part 1				
Part 2	.4920			
Part 3	.3849	.2830		
Grade	.3230	.2970	.2620	

Part 1	.3230
Part 2	.3598
Part 3	.3831

KPL 1

	Part 1	Part 2	Part 3	Grade
Part 1				
Part 2	.3127			
Part 3	.3103	.2823		
Grade	.2608	.1741	.3746	

Part 3	.3746
Part 1	.4043
Part 2	.4058

In summary, it appeared that there were considerable differences between the CEEB and Rhetoric tests. The CEEB tests had lower mean scores and mean difficulties with higher standard deviations and standard errors of measurement, pointing out that the students were being spread out over a larger range of scores when compared to the Rhetoric tests. The higher K-R 21's of the CEEB test indicated that they were more stable in what they were measuring and the higher mean point biserials indicated that the items were doing a better job of discriminating on the basis of the total score than the Rhetoric tests. The Rhetoric tests on the other hand were more highly related to course grade indicating that they seemed to be measuring the outcome of the course more accurately. Ideally, the Rhetoric tests should be made more reliable or the CEEB tests made more valid.