DOCUMENT RESUME

ED 048 844 JC 710 080

AUTHOR Cohen, Arthur M.
TITLE A Procedure for Assessing Students' Ability to Write Compositions.
PUB DATE 71
NOTE 18p.; Paper presented at the Annual Meeting of the California Educational Research Association, April 29-30, 1971, San Diego, California

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Composition (Literary), *Composition Skills (Literary), Evaluation Methods, *Evaluation Techniques, *Junior Colleges, Observation, *Performance Criteria, Performance Specifications, *Writing Skills

ABSTRACT

This investigation developed a procedure for scoring English compositions that would be simple enough for use by junior college instructors with minimal statistical assistance, and still yield data that would allow sound inferences regarding student placement procedures and assessment of instructional effects. Twenty-one instructors from 14 junior colleges developed a scoring key that included 19 dichotomous criteria and learned to use it reliably. They collected pre- and post-compositions from students in their classes during the first and last week of the fall semester, and scored the compositions without their knowing the student's name, course level, institution, or whether the composition was a pre- or post-sample. Comparing class means, significant differences were found between remedial and transfer groups and between pre- and post-test performance on item clusters relating to "content" and "organization," but not on "mechanics." The procedure was found to be feasible for use in departmental settings. (Author/CA)

Paper Presented at The Annual Meeting
California Educational Research Association
April 29, 1971

A PROCEDURE FOR ASSESSING STUDENTS' ABILITY TO WRITE COMPOSITIONS*

Arthur M. Cohen

ERIC Clearinghouse for Junior Colleges
University of California, Los Angeles

The problem was to develop a procedure for scoring English compositions that would be simple enough for use by junior college instructors with minimal statistical assistance, yet yield data that would allow sound inferences regarding student placement procedures and assessment of instructional effects.

Twenty-one instructors from 14 junior colleges developed a scoring key that included 19 dichotomous criteria and learned to use it reliably. They collected pre- and post-compositions from students in their classes (total N = 878) during the first and last week of the Fall Semester, and scored the compositions without their knowing the students' names, course level, institution, or whether the composition was a pre- or post-sample.

Comparing class means, significant differences were found between remedial and transfer groups and between pre- and post-test performance on item clusters relating to "content" and "organization" but not on "mechanics." The procedure was found to be feasible for use in departmental settings.

*The full study will be reported in detail in a monograph by Arthur M. Cohen, M. Stephen Sheldon, and James Chadbourne, to be published in fall, 1971 in the ERIC Clearinghouse for Junior Colleges/ American Association of Junior Colleges Monograph series.

Many arguments may be raised regarding the need for valid studies of instruction in written composition in the junior college. Most of these arguments revolve around the contention that although English composition is a required course in nearly all colleges, the instructors assigned to the task find it impossible to agree on what constitutes good writing, how it should be taught, or even if it should be taught (Bossone, 1966, 1969). Several studies have called for continuing inquiry into the nature and effects of teaching composition--as for example those reported by Shugrue (1970), Archer (1965), and Weingarten and Kroeger (1965). However, the question of what students learn in the courses--if anything--is still hotly debated.

Student learning in English composition classes is typically assessed in various ways. Scores on alternate forms of various normative tests may be compared or grade marks issued by the instructors examined. However, these types of measures lack several key elements. The grade marks tell little about student learning: Were they issued in response to students' participation in class? Did they depend on written assignments, on performance on quick score tests, on the preparation of research papers? Were they based on the common practice of combining various types of measures into a single score? When one instructor grades his students according to how well they learned to write, is he applying the same standard as his colleague? The standardized tests have different types of problems. First, they offer, at best, analogous assessments of students' ability to write; second, and most important, the instructors doubt their veracity. As Stake points out, "...indirect measurement of achievement is irrelevant, even offensive, to many curriculum developers and supervisors of instruction. They want to know <u>what</u> has been learned. They want to know what deficiencies remain in student understanding. The standardized test does not tell them" (1967, p 6).

If the indirect measure of achievement is "irrelevant" to the curriculum developer, it is <u>anathema</u> to the instructor who maintains that no one knows his students as he does! The instructor frequently insists on relying on his <u>own</u> judgment even if that judgment is reported out in the form of grade marks which subsume a variety of student skills. Accordingly, in the usual junior college situation, an unbridgeable gulf exists between the institutional researcher who is charged with validating placement procedures and assessing the efficacy of instructional techniques and the English instructor who wants no part of any outsider's studies. If the researcher expects the instructors to attend to his findings, to change their procedures accordingly--or even to acquiesce to his collecting data directly from their students--the instructors must be convinced that his research design is valid <u>in their terms</u>. They must understand the design, believe it will aid their own instructional operations, yield data of use to them. One way of gaining their support--or at least mitigating their dissatisfaction--is to involve them in the design and conduct of the study itself. This means implicating them at every step of the way, not merely reporting the results to them or soliciting their aid as data tabulators. In many investigations, even when the instructors have been involved in decisions regarding the measurable variables, the collection of data, and the analysis of results, the designs employed have lacked one or more crucial elements; hence, the

findings have been equivocal. Sound experimental design and the instructors both must be satisfied.

Since many instructors seem to believe--perhaps rightfully so--that it is not desirable to measure analogous behavior, any experimental procedure must incorporate equivalents--that is, compositions written under classroom conditions as samples of student performance. And, because the instructors do not trust the judgment of anyone other than themselves in reading these compositions, they must be employed in the composition scoring. In addition, instructor bias--intended or otherwise--must be mitigated through multiple blind scoring; the instructors must not know whose paper they are reading and whether it was written prior to, or subsequent to, instruction. Nor should they know in whose class it was written or the "level" of the students. The design must also insure reliability of reading or at least mitigate the effects of unreliability. In other words, extreme care must be taken to keep the instructors involved at every step of the way without allowing them to prejudge the results.

This is a report of an investigation which directly involved junior college English instructors in designing and conducting a study of student learning. The experimental procedures maximally involved the instructors, while yielding reliable and valid measures of their students' abilities to write compositions. The study was based on the assumptions that one of the major purposes of composition courses is to enhance students' ability to write compositions; that this change in ability can be measured by assessing compositions written prior to, and again after, instruction; and that compositions can be validly assessed using a multiple blind technique. The design was developed at a workshop sponsored by the League for Innovation in the Community College at UCLA. Twenty-one instructors from fourteen junior colleges met for two weeks. They selected topics on which their students would write, developed a scoring key, familiarized themselves with the categories in the key, and committed themselves to conducting the investigation. The investigation was subsequently coordinated by the ERIC Clearinghouse for Junior Colleges with statistical analyses made by M. Stephen Sheldon.

THE DESIGN

The instructors selected a pair of topics because one "before" and one "after" composition had to be collected from each of their students. Certain topics were avoided--for example, those that might tend to invite triteness and those that would be biased against students who might prefer not to reveal personal matters or who might not believe the presenting statements. The instructors also decided that rhetorical devices should not be suggested. The topics they chose were "What makes a good advertisement?" and "What makes a good entertainer?" The instructors also developed a scoring key (Figure 1) and practiced using it on sample compositions.

At the beginning of the fall term, the investigator prepared bluebooks with instructions to the students noted on the cover (Figure 2) and sent them to the instructors. During the first week of the semester each

Figure 1

SCORE SHEET

	YES	NO		
Content I.	____	____	1.	Ideas themselves are insightful.
	____	____	2.	Ideas are creative or original
	____	____	3.	Ideas are rational or logical.
	____	____	4.	Ideas are expressed with clarity.
Organization II.	____	____	5.	There is a thesis.
	____	____	6.	Order of thesis idea is followed throughout the essay.
	____	____	7.	Thesis is adequately developed.
	____	____	8.	Every paragraph is relevant to the thesis.
	____	____	9.	Each paragraph has a controlling idea.
	____	____	10.	Each paragraph is developed with relevant and concrete details.
	____	____	11.	The details that are included are well ordered.
Mechanics III.	____	____	12.	There are many misspellings.
	____	____	13.	There are serious punctuation errors.
	____	____	14.	Punctuation errors are excessive.
	____	____	15.	There are errors in use of verbs.
	____	____	16.	There are errors in use of pronouns.
	____	____	17.	There are errors in use of modifiers.
	____	____	18.	There are distracting errors in word usage.
	____	____	19.	The sentences are awkward.

CODE NO. _____

Figure 2

Code Number_____
(LEAVE BLANK)

INSTRUCTIONS TO STUDENT

1. Fill in the form below:

Name_____
LAST NAME FIRST NAME MIDDLE INITIAL

School_____

Course_____

Date_____

Sex: Male____ Female____

Have you a high school diploma? _____ _____
Yes No

Have you attended any college prior to this term? _____ _____
Yes No

Your age: (Check one) _____ _____ _____
under 17 17 18

_____ _____ _____ _____ _____ _____ _____ _____
19 20 21 22 23-26 27-30 31-35 36

2. Write a composition in this bluebook.

Write in ink on one side of the page only.

Write on alternate lines.

You are to write on the topic: (to be selected by participating instructors)

instructor randomly distributed the bluebooks to his students. Half had the notation, "Write a composition on the topic, 'What makes a good advertisement?'"; an equivalent number called for the student to write on the "entertainer" topic. Thus, half of each instructor's class wrote on one topic at the beginning of the term and half on the other. These compositions were collected and sent to the investigator.

At the end of the semester, each student received a bluebook with his name and the directive, "Write on the topic, 'What makes a good entertainer (if he wrote previously on 'advertisement')'?" and vice versa. Each student, then, wrote on both topics, preparing one composition before instruction began, the other at the end of the course. These compositions also were sent to the investigator who removed all identifying marks from each, entered its author's name on a list, and assigned a code number to it. The code numbers did not reveal the time when, or by whom--or the college at which--the composition was written.

Because each instructor had used the students in one or more of his own classes as subjects, there were a total of twenty-four classes; five of these were considered pre-college English and the remaining nineteen, normal college classes for which the students received credit that could be transferred to four year institutions. In order to lessen the number of compositions that had to be scored, a random half of each of the pre- and post-test essays were selected. Student absence, drop-out, and other factors reduced the number of students who wrote compositions at the end of instruction as compared with the number who had written at the beginning. For the pre- test, 535 essays were scored--105 from remedial classes and 430 from the college English classes. The post-test sample consisted of 343 essays--47 from the remedial English classes and 296 from the transfer courses. Within the total group, 184 students had both pre- and post-test essays scored.

So that each participating instructor would score an approximately equivalent number of compositions from each class, the bluebooks were mixed together prior to distribution. Then, using the key he had helped develop during the workshop, each instructor scored approximately 50 compositions. He did not know whether the composition was a pre- or post-test essay, whether it was from a remedial or transfer class, or, indeed, what student or college was represented. The scoring sheets were then returned to the investigator for tabulation.

The Criterion Variable

Though much effort had gone into the development of the scoring key, until the study was conducted, there was no way of determining the key's reliability, validity, or internal consistency. In order to get some indication of reliability, four freshmen essays were duplicated and read independently by 15 instructors. Each instructor marked each of the 19 items on the scale for all four essays. The proportion of concurrence for each item was computed independently for each of the essays. This index of concurrence was simply the proportion of instructors who agreed that the item was either a zero or a one. If, for instance, 12 of the 15 marked an item either "yes" or "no," the index for that item would be .80. Table 1 shows the results of this reliability study.

Inspection of the table indicates that the index of concurrence on the items of the four essays ranged from .50 to 1.00. In interpreting this table as an indication of reliability, the reader should keep in mind that on a dichotomous variable, a chance score is .50. In other words, an index of .50 would indicate zero reliability.

It is interesting to examine the variability that occurs in the reliability, both across essays and across items. Examining the index for each of the items, it seems evident that some--e.g., item 4 ("clarity of ideas") and item 6 ("order of thesis idea is followed throughout the essay")--have relatively low reliabilities, while others--e.g., item 2("creative or original") and item 17 ("errors in use of modifiers")--appear quite reliable. Continuing to examine the individual items, the variability in the index of concurrence is also striking for some items. As most English teachers know by insight, an essay that is clearly good or bad would receive a much higher concurrence than one which is in between.

The validity of a criterion instrument that is purported to measure achievement is difficult to establish empirically; one must resort to construct validity. By the very nature of the development of this instrument, validity was established. If 21 college English instructors agree that 19 items reflect the quality of an essay, one can assume construct validity.

Another way of perceiving validity is to use the concept of criterion groups. One would, for instance, expect remedial English student essays to be considerably poorer than transfer English essays. One would also expect post-test essays to be considerably better than pre-test essays. To the extent that the criterion instrument reflects these differences, it can be considered valid for measuring the quality of the freshmen English essays. This concept of validity will be discussed when looking at the results of the experiment.

The internal consistency of the instrument is reflected by how well each item is measuring that which the scale is purported to measure. For the instrument in question, there are two ways of looking at this internal consistency. One would be the correlation between each item and the subtotal for each of the three areas, i.e., Content, Organization, and Mechanics. A second would be reflected by the correlation between the subtotals and the grand total. The matrices of these correlations appear in Tables 2 and 3.

Table 1

Percentage of Concurrence
on Four Essays Graded Independently by Fifteen Instructors

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
Essay 1	.80	.87	.60	.67	.80	.64	.69	.72	.60	.73	.73	.93	.73	.73	.67	.93	.93	.53	.93	15
Essay 2	.73	.93	.87	.67	.80	.53	.67	.73	.73	.93	.57	.87	.93	.93	.50	.53	.87	.53	.53	13
Essay 3	.67	.93	.93	.60	.73	.50	.71	.79	.80	.93	.69	.93	.50	.87	.87	.67	1.0	.80	.73	16
Essay 4	.53	.87	.73	.50	.87	.57	.67	.93	.93	.80	.57	.60	.80	.73	.77	.53	.73	.80	.60	12
																				14
Mean	.68	.90	.78	.61	.80	.56	.69	.79	.77	.85	.64	.83	.74	.82	.70	.67	.88	.67	.70	.74

Table 2

Correlations Between Items and Subscales

Item	1	2	3	4	5	6	7	8
Content	.73	.59	.73	.66				
Organization	.61	.78	.66	.78	.66	.60	.63	
Mechanics	.54	.58	.55	.59	.53	.44	.57	.65

Table 3

Correlations Between Subscales and Total Score

	Organization	Mechanics	Total
Content	.58	.35	.78
Organization		.20	.81
Mechanics			.69

Examination of Table 2 indicates there is indeed acceptable correlation between each item and the sub-scale. Normal psychometric procedure would indicate a .30 correlation as satisfactory and a .50, very good. The correlations in Table 2 appear exceedingly high and suggest a great deal of internal consistency for the sub-scales. The reader must keep in mind, however, that there were relatively few items comprising each of the sub-scales--four for Content, seven for Organization, and eight for Mechanics. As a consequence, the correlations contain a significant ipsative factor. Said another way, there is a large element of correlating numbers which contain a self-sameness.

The correlations of the sub-tests with the total and with each other appear in Table 3. Very high correlations of the sub-tests with the total are again influenced by the ipsative nature of the numbers. The relatively low relationships between the scales can be perceived as a favorable characteristic suggesting that each scale is measuring an independent variable in the quality of the English essays.

RESULTS AND DISCUSSION

This study sought answers to certain general questions:

1. Was there any empirical validity to the scale that was developed?

2. Were freshmen students learning to write better as measured by this scale?

3. Were there differences in the writing ability between students assigned to remedial English and those assigned to transfer English?

The answer to the first question, of necessity, hinges on the answers to the second and third. If one considers the pre- and post-essays as one set of criterion samples and the remedial and transfer essays as another, the validity of the scale can be determined by the mean differences between these criterion groups.

Broadly stated, question number two asks, "Is anyone learning to write?" To answer this question, a number of data analysis techniques were employed. First, the scores assigned to the pre- and post-essays for the remedial and transfer English groups separately were scrutinized carefully. The means and standard deviations for these groups appear in Table 4. These means and standard deviations are broken out by sub-scale as well as total. Inspection of the table indicates that the post-test means are higher in every case than the pre-test means, with the exception of Scale 3, Mechanics, for the remedial classes.

To test the significance of the differences between these means, a two-way analysis of variance was computed for each of the sub-scales and the total. The main effects were pre-post essays and remedial-transfer essays. The results of these analyses appear in Tables 5, 6, 7, and 8. Table 5 indicates that for the sub-scale Content the post-test is significantly higher than the pre-, and, further, that the transfer English essays were significantly

Table 4

Means and Standard Deviations on Subscales and Total
Separated by Pre and Post Test and by Remedial College Classes

	Pre-test			Post-test		
	$\bar{X}$	SD	N	$\bar{X}$	SD	N
Remedial Classes						
Content	1.21	1.21	105	1.77	1.17	47
Organization	2.64	2.14	105	3.81	2.18	47
Mechanics	5.58	2.18	105	5.49	2.15	47
Total	9.43	4.13	105	11.06	3.76	47
Transfer Classes						
Content	1.78	1.22	430	2.11	1.26	296
Organization	3.34	2.21	430	3.97	2.28	296
Mechanics	5.85	1.87	430	5.99	1.78	296
Total	10.96	4.01	430	12.07	4.14	296
All Classes						
Content	1.66	1.24	536	2.07	1.25	345
Organization	3.20	2.21	536	3.93	2.26	345
Mechanics	5.79	1.94	536	5.92	1.83	345
Total	10.64	4.08	536	11.92	4.09	345

Table 5

Nova on Subtotal for Content
Main Effects Pre and Post Essay and Remedial College English

Source	Sum of Square	D.F.	Mean Square	F
Pre/Post	20.94	1	20.94	13.89
Remedial College	19.37	1	19.37	12.84
Pre/Post Remedial College	1.48	1	1.48	0.98
Error	1309.08	868	1.51	

Table 6

Nova on Subtotal for Organization
Main Effects Pre and Post Essay and Remedial College English

Source	Sum of Square	D.F.	Mean Square	F
Pre/Post	85.23	1	85.23	17.40
Remedial College	13.88	1	13.88	2.83
Pre/Post Remedial College	8.70	1	8.70	1.78
Error	4251.46	868	4.90	

Table 7

Nova on Subtotal for Mechanics
Main Effects Pre and Post Essay and Remedial College English

Source	Sum of Square	D.F.	Mean Square	F
Pre/Post	0.01	1	0.01	0.00
Remedial College	7.20	1	7.20	2.13
Pre/Post Remedial College	1.69	1	1.69	0.50
Error	2936.00	868	3.38	

Table 8

Nova on Total for Content, Organization, Mechanics
Main Effects Pre and Post Essay and Remedial College English

Source	Sum of Squares	D.F.	Mean Square	F
Pre/Post	187.59	1	187.59	11.82
Remedial College	116.84	1	116.84	7.36
Pre/Post Remedial College	8.21	1	8.21	0.52
Error	13773.14	868	15.87	

better than those written by the remedial students. Inspection of Table 6, which tests the means for sub-scale 1, Organization, indicates post-test scores significantly higher than pre-tests, but no significant difference between transfer and remedial English students.

On the sub-test for Mechanics, none of the F-tests are significant, indicating that, at least as far as these data are concerned, little had been learned in the Mechanics of English. For the total scores, the F-ratios again indicate that there is significant growth between pre- and post-test essays and a significant difference in the total mean scores of the remedial and transfer classes.

In summarizing these results, one can look at the magnitude of the differences in means. Though in all instances except Mechanics it would appear that significant growth had taken place, it would also seem that the magnitude of the difference in means between the pre- and post-test is relatively small. On the other hand, if one examines the pre-test means, it is evident that more students knew more about the Mechanics of English than the other two areas measured by these scales and reflected in the essays. In Content and Organization, where the best possible scores are 4 and 7 respectively, the pre-test means are between 1 and 2 and between 3 and 4 respectively. In Mechanics, where the maximum score is 8, the overall mean is 5.79. This would give the students less "room at the top" to demonstrate growth. Said more appropriately, the ceiling of the test was too low.

Another way of determining the answers to questions two and three is to "stare at the data." To do this, frequency distributions were plotted and line graphs drawn for the pre-test and post-test totals. These appear in Figure 3. It is evident from inspection that the post-test totals have a greater negative skew than the pre-tests. Indeed, considering the general low ceiling for a significant proportion of the subjects, one could assume that the scale does not reflect even greater growth in a significant portion of the subjects.

For those subjects for whom both essays were scored, a discrepancy index was computed for each, that is, the score of the pre-test essays was subtracted from that of the post-test. Figure 4 gives a frequency distribution for these data. Inspection of this distribution indicates a large number of subjects (43 per cent) had zero or less growth; 47 per cent improved 1 to 12 points. The average growth for this total score was 1.23 points.

Returning now to question one, "Is there empirical validity for the scale?", one can respond with a qualified "yes." For two of the three sub-scales and the total, the post-test essays show a significantly higher mean than the pre-test. For the total and one of the sub-tests, the transfer English students did significantly better than the remedial students. Overall, considering the large variabilities in assigning students to transfer or remedial English and in the predispositions of the reader-scorers, the scale has been shown to be valid.

For question two, "Is anyone learning to write?", we again have a qualified "yes." The data indicate statistically significant growth in the mean scores even though the magnitude of the differences might be a bit disappointing for teachers of freshman composition.

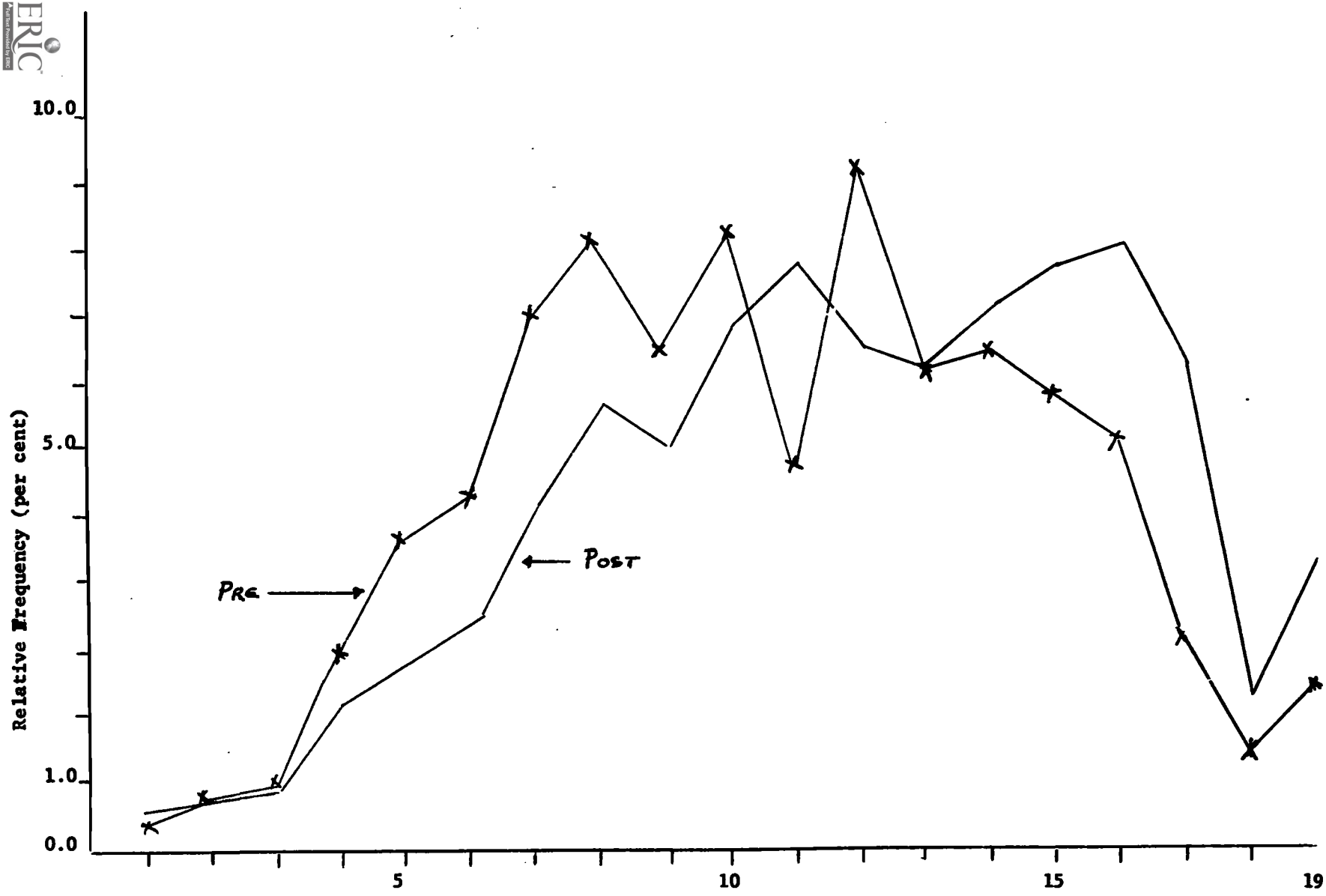

Figure 3

Frequency Distribution by Per Cent
Pre and Post Test

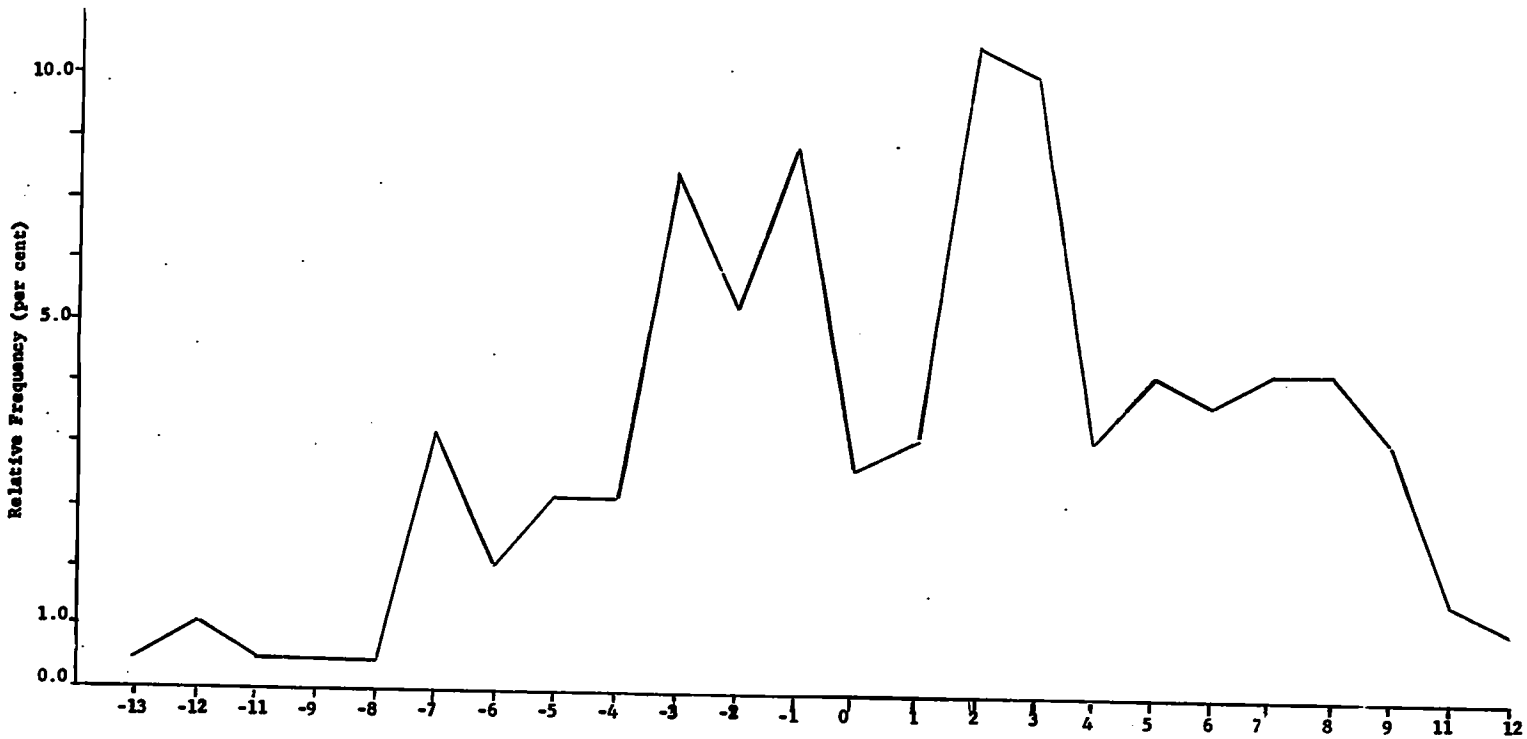

Figure 4

Frequency Distribution of Pre and Post Test Totals
By Per Cent

The third question asked, "Were there differences in the writing ability between students assigned to remedial English and those assigned to transfer English?" Here the validity of the placement procedures seemed to be established. As indicated on Table 4, the mean Content score for remedial classes was 1.21 on the pre-test and 1.77 on the post-test; for the transfer classes, it was 1.78 on the pre-test and 2.11 on the post-test. The students' Content score was almost exactly the same at the beginning of the transfer courses as it was at the end of their courses. This suggests that the screening procedures were working well. They worked less well for Organization, with remedial students beginning at 2.64 and ending at 3.81 and transfer students beginning at 3.34 and ending at 3.97. The Mechanics area showed only slight difference between the groups. Content seemed to differentiate best.

The design can apparently be used to assess change in students' ability to write compositions. More important, perhaps, the investigation demonstrates that it is possible to involve English instructors in the actual conduct of a learning study and still obtain results the researcher would find respectable. In fact, with minimal coordination, the instructors themselves can conduct studies using this design. However, the procedure has certain limitations that should be noted. If it were applied to a pair of compositions written by a single student, it would be of little value, first, because the ability to write a single composition on a pre-determined topic is probably not constant and, second, because the readers' reliability is not so high that it might not prejudice a single pair of compositions. In addition, the design does not account for learning other than just in the area being measured; English instructors have goals, no less worthy, besides the teaching of Content, Organization, and Mechanics in written expression. One more limitation: because the design measures group achievement only, other assessments of individual students (for example, grade point averages, scores on other tests) cannot be correlated with the findings.

It is instructive to note a few of the criticisms that have been raised by the instructors who were involved in the study and by others to whom the design was presented. A number of instructors apparently feel that composition cannot be divorced from the writer and that judging a composition without knowledge of the writer himself is not valid. Some instructors also feel that each student should be given feedback on each composition that he writes; the design, of course, does not allow for this. Other criticisms are that a valid sample of a student's best writing cannot be collected in a one-hour exercise in which the student is asked to write on a topic previously unknown to him. And, most threatening of all, many instructors feel that the results of this type of study can be used to defend the re-sectioning of courses or even the dropping of freshman composition. That is, they feel that the results can be used against the very department which has honestly attempted to measure the learning gained by its students.

Nevertheless, Diederich (1967) insists that studies of this type not only yield convincing results but also have a beneficial effect on the professional attitudes of the instructors. The findings of this study bear out his first contention, at least. The main point is that the investigator must spell out all the premises in advance, involve the instructors at every stage of the investigation, and point out the limitations of the design. If he attends to these caveats, he may be able to enlist the participation of the instructors and even find them acting on the results.

REFERENCES

1. Archer, Jerome W. Research and the Development of English Programs in the Junior Colleges. Champaign, Ill.: National Council of Teachers of English, 1965.

2. Bossone, Richard M. Remedial English Instruction in California Public Junior Colleges. Sacramento: California State Department of Education, 1966.

3. Bossone, Richard M. The Writing Problems of Remedial English Students in Community Colleges of the City University of New York. New York: City University Research and Evaluation Unit for Special Programs, 1969.

4. Cohen, Arthur M. Is Anyone Learning to Write? Topical Paper No. 5. Los Angeles: ERIC Clearinghouse for Junior College Information, UCLA, 1969.

5. Diederich, Paul. "Cooperative Preparation and Rating of Essay Tests." English Journal, 56:4, 1967.

6. Shugrue, Michael F. "The National Study of English in the Junior College." Junior College Journal, 40:9, 1970, pp. 8-12.

7. Stake, Robert E. "Toward a Technology for the Evaluation of Educational Programs." In Ralph Tyler, Robert Gagne, and Michael Scriven (eds.), Perspectives of Curriculum Evaluation. AERA Monograph. Chicago: Rand McNally and Co., 1967.

8. Weingarten, Samuel and Kroeger, Frederick P. (eds.). English in the Two-Year College. Champaign, Ill.: National Council of Teachers of English, 1965.