DOCUMENT RESUME

ED 035 941 24 CG 004 959

AUTHOR GROHSMEYER, FREDERICK; JOHNSON, RICHARD R.
TITLE A COMPARISON OF TWO TECHNIQUES OF TEACHING SCIENTIFIC METHOD IN INTRODUCTORY PSYCHOLOGY LABORATORIES: STAGE 1, THE DEVELOPMENT OF AN EVALUATIVE INSTRUMENT. FINAL REPORT.
INSTITUTION EARLHAM COLL., RICHMOND, IND.
SPONS AGENCY OFFICE OF EDUCATION (DHEW), WASHINGTON, D.C. BUREAU OF RESEARCH.
BUREAU NO BR-7-E-194
PUB DATE 14 JUN 68
GRANT OEG-0-8-000194-1809-(010)
NOTE 20P.

EDRS PRICE EDRS PRICE MF-$0.25 HC-$1.10
DESCRIPTORS *DISCRIMINANT ANALYSIS, FACTOR ANALYSIS, *GROUP TESTS, *MEASUREMENT INSTRUMENTS, *SCIENTIFIC LITERACY, STUDENTS, *TEST CONSTRUCTION, TESTS, TRAINING
IDENTIFIERS EXPERIMENTAL METHOD TEST (EMT)

ABSTRACT

DURING THE FALL TERM, PRELIMINARY FORMS OF THE EXPERIMENTAL METHOD TEST (EMT) WERE GIVEN TO INTRODUCTORY LEVEL PSYCHOLOGY STUDENTS AS WELL AS A NUMBER OF ADVANCED STUDENTS. THE RESULTS OF THIS PRELIMINARY TESTING LED TO THE DEVELOPMENT OF A SECOND FORM OF THE TEST WHICH WAS EVALUATED OVER A NUMBER OF GROUPS OF STUDENTS. THE RESULTS OF THE FINAL FORM OF THE TEST SEEM TO INDICATE: (1) THAT THE TEST MAY BE A GOOD DISCRIMINATOR AMONG GROUPS OF STUDENTS WHO HAVE DIFFERING BACKGROUNDS OF KNOWLEDGE OF SCIENTIFIC METHOD, AND (2) THAT THE TEST APPEARS TO DISCRIMINATE CHANGES IN STUDENTS AS A FUNCTION OF HAVING HAD TRAINING ORIENTED TOWARD SCIENTIFIC METHOD. THE TEST, HOWEVER, SEEMS TO BE A POOR DISCRIMINATOR AMONG INDIVIDUALS IN A GROUP. FURTHER CONCLUSIONS ARE: (1) THE TEST WILL SERVE QUITE ADEQUATELY TO DISCRIMINATE BETWEEN GROUPS OF STUDENTS WHO ARE TAUGHT SCIENTIFIC METHOD BY TWO DIFFERENT TECHNIQUES, (2) FURTHER DEVELOPMENT OF THE TEST OUGHT TO INVOLVE FACTOR ANALYSIS TO DISCRIMINATE AMONG SUBSCALES, AND THE PRODUCTION OF MORE ITEMS SO THAT THE TEST COULD BE LENGTHENED, AND PARALLEL FORMS COULD BE OFFERED. (KJ)

FINAL REPORT
Project No. 7-E-194
Grant No. OEG-0-8-000194-1809-(010)

A Comparison of Two Techniques of Teaching Scientific Method in Introductory Psychology Laboratories: Stage 1, The Development of an Evaluative Instrument

June 1968

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

FINAL REPORT

Project No. 7-E-194
Grant No. OEG-0-8-000194-1809(010)

A Comparison of Two Techniques of Teaching Scientific Method in Introductory Psychology Laboratories: Stage 1, The Development of an Evaluative Instrument

Investigators: Frederick Grohsmeyer, Professor of Psychology & Richard R. Johnson, Assoc. Professor of Psychology

Earlham College

Richmond, Indiana

June 14, 1968

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

CONTENTS

ACKNOWLEDGEMENT

We wish to thank Lary Jones in the Psychology Department at Syracuse University for cooperating with us in giving a form of the test developed at Earlham to a group of undergraduate students at Syracuse. We appreciated very much, his willingness not only to give the test, but the computer analysis of data which he then sent along to us to be included as part of this report.

SUMMARY

The Psychology Department at Earlham College, as a part of its continuing efforts to evaluate and improve the teaching of scientific methods in psychology, has begun developing a paper and pencil objective instrument by which it can evaluate different methods of teaching in the laboratory. As a part of the first stage of evaluating two different methods of teaching, further studies were run on developing this criterion instrument during the academic year in 1967-68 under an Office of Education grant.

During the fall term preliminary forms of the Experimental Method Test (EMT) were given to Introductory level psychology students as well as a number of advanced students. The results of this preliminary testing led to the development of a second form of the test which was evaluated over a number of groups of students at Earlham and Syracuse Universities during the second half of the year.

The results of the final form of the test seem to indicate: (1) that the test may be a good discriminator among groups of students who have differing backgrounds of knowledge of scientific method and (2) that the test, in the only pre and post test comparisons available, appears to discriminate very well, changes in students as a function of having had training orientated toward scientific method. However, the present evidence indicates that the test functions relatively poorly as a discriminator among individuals in a group, especially if that group is relatively homogeneous and is at an advanced level of knowledge in this field.

It is suggested in conclusion that: (1) the test in its present form will serve quite adequately to discriminate between groups of students who are taught scientific method by two different techniques and (2) that further development of the test ought to involve factor analysis to discriminate among sub-scales and the production of more items so that the test could be lengthened, parallel forms could be offered, and items of greater difficulty could extend the range of the test in discriminating among individuals.

INTRODUCTION

Over a period of two years the Earlham College Psychology Department has worked on the development of a diagnostic-evaluative paper and pencil instrument to be used in conjunction with its Introductory Psychology laboratory. The need for this instrument was twofold: (1) Because the students entering Introductory Psychology at Earlham College range from freshmen with no college experience to seniors majoring in other natural sciences, we have needed some way to evaluate understanding of scientific procedures so that the laboratory experience could be adjusted to the student's level of competence and, (2) because we have been very interested in studying different methods of teaching the laboratory aspect of the Introductory course, we have needed an evaluative instrument separate from the actual grading in the course or students' comments to provide some objective standard against which various techniques could be measured. In line with this second objective, a proposal was submitted to the Office of Education in the spring of 1967 to study two techniques of teaching scientific method in the Introductory Psychology laboratory. As a part of this proposal, an early form of the Experimental Method Test (EMT) developed here at Earlham was proposed as an evaluative instrument. In the negotiations concerning how this experiment should be carried out, it was finally decided that the investigation should proceed in two phases. The first phase would involve further development of our laboratory test so that it could be shown to be an effective measuring instrument. After this we would resubmit our proposal to study the actual processes and end results of the two methods of teaching in the laboratory.

While our original proposal for the development of this test suggested a grant period to cover one calendar year and a large enough budget to evaluate the test over a number of different student populations, the final agreed upon grant period was eight months and the final budget figure was too small to involve a significant number of students beyond the Introductory Psychology laboratories at Earlham.

Working within these limitations we decided that the major focus of the development of our instrument would have to be limited to the students in the Introductory Psychology laboratory at Earlham (about two hundred students in a year). We planned also to evaluate the instrument using whatever other outside groups we could, to insure that the instrument was not too closely related to the particular way we teach Introductory Psychology here at Earlham.

METHODS

The form of test which we set out to develop during the period of this grant was specifically designed to evaluate different methods of teaching scientific method in the laboratory at Earlham College. This evaluation of the instrument should be confined to these purposes since the test criteria we set imposed several constraints on our work in developing the test. First, we needed an instrument which would be short enough to be administered in one hour and yet would cover a variety of topics which might be included under the heading of scientific method. Second, the purpose of the test was not so much to discriminate among students as to provide an instrument for discriminating between groups and changes of scores in a group from pre to post laboratory testing. In line with these goals, we began rewriting the existing form of our Criterion Test during the summer of 1967. The various forms of the test have contained about 50 items which we have found to be easily administered in the one hour laboratory session.

During Term 1 of 1967 (September through December), we administered the first form of the Experimental Method Test (EMT1) on the first day of class. The results from this administration of the test were item analysed and a new form of the test was developed utilizing items which showed good difficulty and discrimination levels, rewriting other items and developing new items.

The population on which the first form of the EMT was developed consisted of 112 introductory level psychology students, a group of senior psychology majors (N=28), and an experimental section of an introductory biology laboratory (N=22). At the end of Term 1 in December, the second form (EMT2), was given to all Introductory Psychology students as a part of the final examination. Because the test had been changed radically from its original form, no pre and post comparison scores were possible on this group of students.

The second form of the test (EMT2) was given as a pre test on the first day of class (January 1968) in the second term to 81 Introductory Psychology students. This group was composed of 60% freshmen, 22% sophomores, 10% juniors and 6% seniors at Earlham College. The average verbal SAT score for the group was 570.2 and the average mathematic SAT score for this group was 588.9.

The instructions given the group taking the test were that this was a diagnostic test to be used by the staff to find out what aspects of the laboratory work the student already knew. It was emphasized that there would be no grade given on this test. No indication was given that we contemplated later regiving a version of this test at the conclusion of the laboratory experience.

The third form of the test (the EMT3) was given as a part of the final examination at the conclusion of the course in March. The EMT3 differed from the EMT2 only in the addition of three more items. All pre and post comparisons between the two tests were made between the EMT2 and the sub-scale of items on the EMT3 that corresponds. All internal estimates of reliability on the two forms of the test, however, are made using all items on each of the two forms.

The students learned that we would give the EMT3 as a part of the final examination only during the last week of class and no test or pre-information concerning the content of this aspect of the final examination was given to them. Eighty students took this form of the test and of the material completed we were able to develop comparable data for a group of 74 students on both pre and post tests.

Our previous experience in giving the EMT2 at the end of Term 1 indicated that merely requiring the students to take this test during the final examination period led to such low levels of motivation that many students completed the test without actually spending much time looking at the items. To counteract this tendency to rush through the examination we indicated to the students that this part of the examination would count five extra points towards their total grade if they achieved a score of 35 or more out of the 50 items. No points were given and no penalty was assessed if they scored below that level. While this device certainly may have slightly heightened the motivation of the students in taking the post test, the relatively low pay-off probably did not have a significant influence. The five points, if the students received them, would have counted less than 6% of the final examination grade and less than 2% of the total grade of the course. Thus, while we cannot argue that the motivation of the students in taking the pre and post tests was exactly identical, it seems unlikely, under the circumstances, that the students spent any significant extra time in studying for the post test and the grade pay-off probably only ensured that the students took the test seriously rather than rushing through to fill out their answer sheets.

While the major focus in developing the EMT was on the Earlham Introductory Psychology student population, we tried in a number of ways to sample other populations to ensure that the test was not totally linked to the particular content which we teach in our Introductory Psychology Laboratory.

EASTERN INDIANA CENTER (EVENING) STUDENTS

Pre and post test (EMT2 and EMT3) were administered to an evening class of students in Laboratory Psychology. This course, which is the second level course taught at Indiana University, is the first acquaintance with laboratory aspects of psychology these students would have had. These students (N=20) in this evening class tend to be more variable in age, academic ability, and background than Earlham day students.

EXPERIMENTAL PSYCHOLOGY

A single form of the test (EMT3) was given to the second level class in psychology at Earlham. These students range from sophomores to seniors and have typically had one or two psychology courses before taking Experimental Psychology. While all the students in the class took the test, the results were considered only on students who had had no previous contact with this test (N=13) in Introductory Psychology.

EDUCATIONAL THEORY AND CURRICULUM

A single administration of the EMT2 was given to a senior level class of education students (N=20). These students in general would have had little contact with laboratory aspects of psychology, although most of them would have had at least one psychology course.

EARLHAM COLLEGE SENIORS

At the end of the final examination period in June we attempted to sample the entire population of seniors graduating from Earlham College to determine if students in various majors having different numbers of science courses achieved different scores on the EMT. We obtained a relatively small sample (N=51) which we could not reliably break down by majors. This sample is probably also biased since taking the test was voluntary and the students knew that this was a test of scientific knowledge.

SYRACUSE UNIVERSITY, INTRODUCTORY EXPERIMENTAL PSYCHOLOGY

Through the cooperation of a graduate student at Syracuse University, we were able to give a different form of the test (Test of Experimental Knowledge-TEK, developed from our form EMT2 at Syracuse) to a beginning level class in Experimental Psychology. The test was given midway in the course and did not count toward the course grade. The students in the course (N=79) probably had a greater range of student ability than the population sampled in the Introductory level class at Earlham.

RESULTS

Any evaluation of the results of the administration of the final forms of the EMT need to be made within the context of the purposes of that test. Because we intended to produce a relatively short instrument, the final form of the EMT (see appendix), comprised only 46 four alternative multiple choice items and one four choice item scored as four true-false decisions. This means that the scores fell within a restricted range (a total of only 50 points) and that this relatively small group of items had to cover what probably amounted to two and perhaps as many as four sub-scales. In the development of the test, we attempted to produce sets of items which were related to: (1) identification of problems and recognition of appropriate hypotheses, (2) recognition of assumptions and weakness of experimental design, (3) identification of experimental variables in terms of how they function, (4) measurement of outcomes and evaluations of various measuring techniques.

The probable lack of homogeneity of items, the restricted range

of test scores and the relative homogeneity of the Earlham College student population led us not to be surprised by the relatively low reliability scores obtained on our instrument. A Kuder-Richardson #20 estimate of reliability for the EMT2 was .484. This same estimate of reliability gave a figure of .360 for the EMT3 and .778 for the TEK. Rulon estimates of reliability on split halves of the EMT3 test gave estimates of .537 for odd versus even items and .533 for a split between items by face validity. (See Table 1 for means, standard deviations and reliability estimates across the test forms and over the various sample populations). The higher Kuder-Richardson #20 estimate of reliability on the TEK form of the test probably represents the fact that the test was slightly longer (60 items) and that the population sampled varied more (note the standard deviation for this group is the very largest, 6.78).

While these estimates of reliability would certainly be very disappointing if we were attempting to develop a standized test, or if we needed to make discriminations among individual students, this level of reliability seems quite adequate for a test which is only used for group discriminations. Kelly (1927), by assuming that a test should make discriminations of differences as small as .26 times the standard deviation of a grade-group with a chance of 5 to 1 of being correct, suggested that reliability levels would need to be only about .50 to evaluate levels of <u>group</u> accomplishment.

An examination of the individual items in the last forms of the EMT (see Table 2) shows that across the pre test population (EMT2) and among a more heterogeneous sample (the Syracuse group) the difficulties of the various items centered around .5 (which should give a very good level of discrimination) and the differential discrimination between high and low groups is relatively good with few reversals. The same criteria when applied to the EMT3 test which was taken after the laboratory experience show the test to be a much poorer discriminator. This occurred because the general increase in test scores led to a bunching of the scores on the post test.

Attempts to estimate the validity of the EMT by correlation with criterion variables are difficult. One external variable which would seem to be fairly well correlated to the students acquisition of knowledge of the scientific method would seem to be his scores on the laboratory aspect of the Introductory Psychology course. However, a correlation between the pre test (EMT2) and the total laboratory grade achieved was .039 and the correlation between the post test (EMT3) and the laboratory grade was .008. These failures to achieve correlations between the laboratory grades and the EMT probably represent differences in what is being measured by the two instruments. The laboratory grades are heavily weighted with a verbal ability and writing skill component and also represent more clearly an achievement measure of students ability to handle one specific problem. In

TABLE 1
PREFORMANCE ACROSS GROUPS AND FORMS
OF THE EXPERIMENTAL METHOD TEST

Group	Test Form	N	Mean	Standard Deviation	Reliability Rulon	Reliability K-R #20
EC Introductory Psychology	EMT2	81	28.37	5.9	-----	.484
EC Introductory	EMT3	80	34.79	4.98	.537 (odd-even) .532 (face-validity)	.360
EIC Evening Class	EMT2	20	24.00	4.80	------	----
EIC Evening Class	EMT3	20	27.55	5.25	-----	----
EC Seniors	EMT3	51	32.72	6.38	-----	----
EC Experimental Psychology	EMT3	13*	32.38	6.71	----	----
EC Education Seniors	EMT2	20	24.15	4.71	----	----
Syracuse Introductory Psychology	TEM	79	30.99**	6.71	----	.778

* The number of students in this class who made up the sample of students not familiar with this test.

**Mean of items correct out of 60 in-tead of 50 (adjusted would be 25.83)

TABLE 2
ITEM ANALYSES BY GROUPS

	EARLHAM COLLEGE		SYRACUSE (TEK)		
Item No.	(EMT2) Pre-test Bi-Serial	(EMT3) Post-test Bi-Serial	Dif.	Discrim.	Point Bi-Serial
1	.293	.226	55.7	.61	.54
2	.233	.446	78.5	.38	.40
3	.370	.297	34.2	.33	.27
4	.442	.431	55.7	.43	.31
5	.351	.539	31.6	.43	.33
6	.438	.569	51.9	.33	.26
7	.529	.513	74.7	.19	.21
8	.688	.455	60.8	.57	.50
9	.297	.296	38.0	.38	.29
10	.488	.401	51.9	.62	.42
11	.017-	.327	No Item		
12	.396	.475	72.2	.14	.14
13	.073	.094	36.7	.14	.12
14	.545	.230	60.8	.19	.17
15	.158	.273	25.3	-.05	-.08
16	.358	.132	30.4	.10	.05
17	.062	.426	No Item		
18	.431	.279	48.1	.24	.23
19	.034	.063	55.7	.09	.02
20	.028-	.355	No Item		
21A	.447	.602			
21B	.605	.397			
21C	.221	.350	53.2	.14	.17*
21D	.203	.121			
22	.412	.703	68.3	.38	.36
23	.501	.411	57.0	.38	.32
24	.245	.438	77.2	.10	.11
25	.618	.426	62.0	.33	.33
26	.377	.376	38.0	.38	.29
27	.359	.343	78.5	.52	.44
28	.158	.148	89.9	.33	.43
29	.188	.248	60.8	.43	.34
30	.249	.452	35.4	.29	.23
31	.412	.271	53.2	.48	.31
32	.363	.499	51.9	.76	.57
33	.497	.108	34.2	.24	.25
34	.349	.364	60.8	.48	.44
35	.574	.321	57.0	.38	.23
36	.646	.134-	72.2	.42	.40
37	.105	.212	45.6	.29	.32
38	.186	.455	78.5	.00	.04

*This was computer analysed as only a single item, although it was scored as four items at Earlham

TABLE 2 (Cont.)
ITEM ANALYSIS BY GROUPS

	EARLHAM COLLEGE		SYRACUSE (TEK)		
	(EMT2)	(EMT3)			
Item No.	Pre-test Bi-Serial	Post-test Bi-Serial	Dif.	Discrim.	Point Bi-Serial
39	.194	.324	78.5	.33	.23
40	.313	.446	63.3	.57	.42
41	.298	.186	29.1	.38	.33
42	.631	.297	67.1	.00	.16
43	.420	.285	63.3	.09	.21
44	.230	.129	48.1	.09	.03
45	.232	.445	86.1	.29	.29
46	.367	.586	50.6	.24	.24
47	.482	.242	30.4	.19	.17
48			53.2	.29	.92
49			65.8	.29	.27
50			70.9	.24	.21
51			13.9	.29	.30
52			65.8	.48	.42
53			65.9	.38	.37
54			73.4	.33	.37
55			35.4	.26	.24
56			15.2	.33	.40
57			53.2	.38	.29
58			32.9	.22	.25
59			35.4	.29	.24
60			34.2	.17	.12
61			22.8	.38	.29

contrast, the EMT asks the student to respond appropriately in a variety of problem situations related to methodological decisions and the logic of scientific method.

Correlations of EMT2 (pre test) scores with the SAT scores of the introductory students at Earlham during Term 2 were .344 with verbal SAT scores and .428 with mathematics SAT scores. These correlations are roughly the same level that SAT measures seem to correlate with other achievement measures at the college level. It is, however, interesting to note that the correlation is higher with the mathematics SAT scores, which might be expected if this test covers scientific content. The post laboratory experience administration of the EMT, however, correlates .195 with verbal SAT and .109 with mathematics SAT. This seems to suggest that whatever is measured by the EMT seems to change as a function of having had the laboratory experience in a way which is not well predicted by the SAT scores. This would certainly meet our expectations if the EMT measures knowledge of scientific method.

Undoubtedly, the most important means of validating this instrument for our purposes has to do with how well it can discriminate between groups which have had different backgrounds in training and how well it can evaluate the achievement of a group as measured by the shifting of scores from pre to post laboratory experience. A comparison of the mean scores achieved by various groups taking the test seems to show differences which are all in the right direction (see Table 1). The pre test mean scores range from 24 to 28.37 while the post test scores are in all cases higher and the scores are also higher for groups which would be expected to have some of this knowledge (seniors and students in Experimental Psychology). The only advanced group tested which showed a low mean score in relation to introductory students were the senior majors in education.

A closer examination of the pre and post test scores in the Introductory Psychology class shows that 68 out of 74 students for whom there is comparable data, showed an increase in scores between the first and second administration of the test. The average shift in scores was an increase of 6.65. (See Figure 1 for graph of the change in scores). A Wilcoxen test for differences between paired scores gave a z score of 6.98 which is significant well beyond the P=.001 level. A similar shift was found between the pre and post test scores of the students in the evening class. While the change in scores is not as dramatic (an increase of 3.55), a Wilcoxen test of differences between paired scores indicated the differences were beyond the P=.01 level. Thus, in both cases where pre laboratory and post laboratory tests are available there were significant shifts in the groups as measured by the paired scores. As noted above, inspection of the various group means sampled also seems to suggest that this particular form of the

FIGURE 1

DISTRIBUTION OF INDIVIDUAL SCORE CHANGES FROM PRE TO POST TESTS FOR INTRODUCTORY PSYCHOLOGY

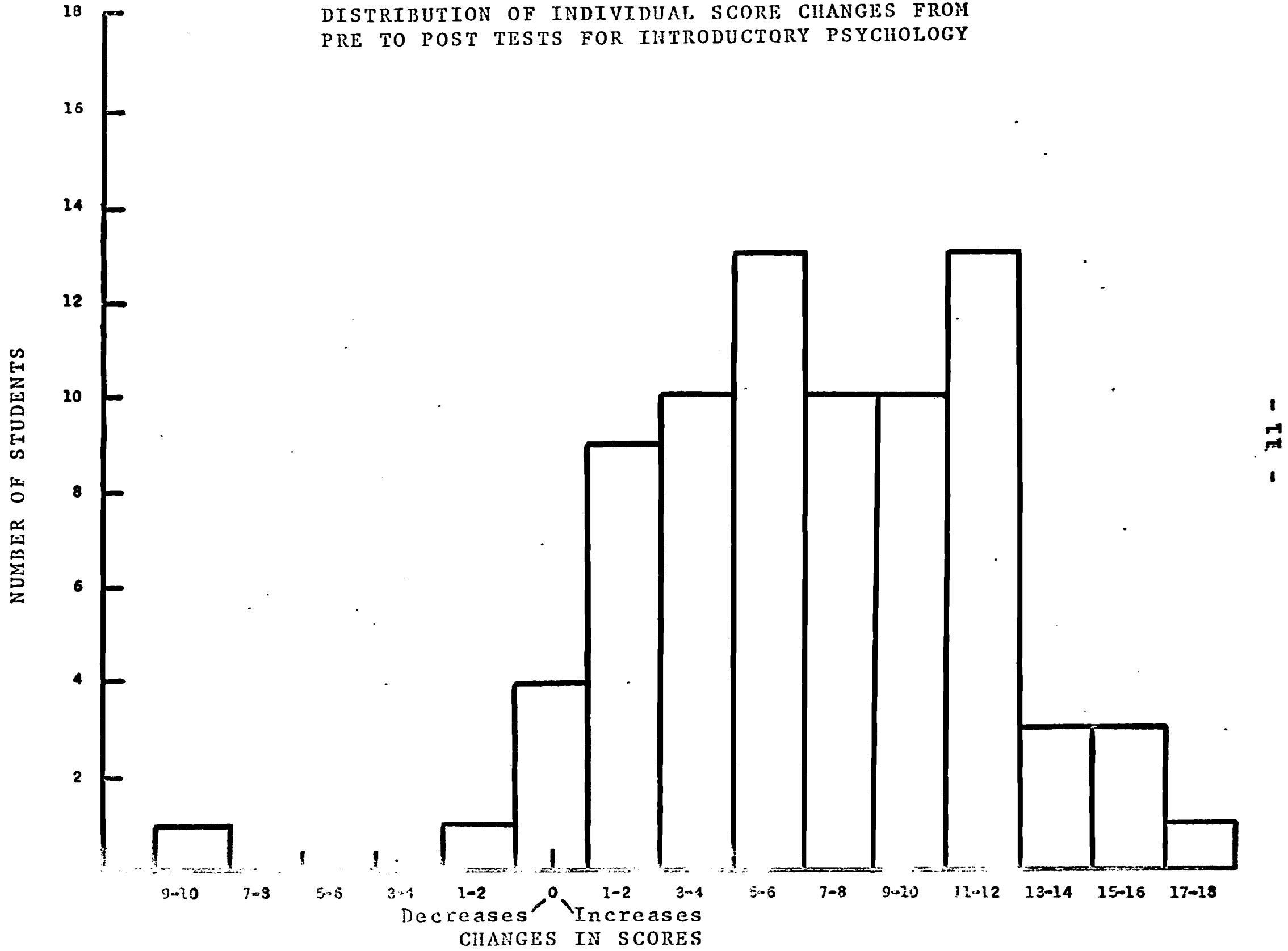

FIGURE 2

Score Distributions for Pre (EMT2) and Post (EMT3) Tests for Introductory Psychology at Earlham

EMT2 (January 1968)		EMT3 (March 1968)	
Score	Intro. Psych.	Score	Intro. Psych.
45		45	1
44	1	44	
43	1	43	11
42	11	42	11111
41		41	1
40		40	1111
39	1	39	1111
38	1	38	11111111
37	1	37	1111111
36		36	111111
35	111111	35	111111111
34	1	34	11111
33	1	33	1111
32	1111111	32	1111
31	1111111	31	111
30	1111	30	111111
29	1111	29	11
28	11111111	28	111
27	111	27	
26	11111111111	26	1111
25	11	25	
24	1111	24	1
23	111	23	
22	11	22	
21	11	21	
20	111	20	
19	11	19	
18	11	18	1
17	1	17	
16	1	16	
15		15	

test does a good job of differentiating among various groups.

CONCLUSIONS AND RECOMMENDATIONS

While the data available on the present form of the EMT suggest that the test is not particularly sensitive to individual differences especially in a relatively homogeneous population, there is evidence to indicate that the test will do an adequate job of differentiating among groups and also is sensitive enough to show shifts in scores as a function of learning in a laboratory. We plan to continue refining the test by: (1) factor analyzing the items to see if there are actually the sub tests which we attempted to construct, (2) developing new items in line with the various sub tests so that the test length could be extended by similar items and some of greater difficulty to differentiate among a more homogeneous population and (3) by running further tests on other outside, control populations to determine the range of generality and application of this particular test.

Next year a reorganization in the pattern of laboratory offerings in Introductory Psychology will provide for three different types of laboratory-field experience in psychology. The first of these will follow along the lines of the more traditional laboratory experiment training which we have done in the past. The second will involve learning observation techniques with young children in a nursery school and will focus particularly on learning objective recording techniques and observing the kinds of interactions and development in young children over time. The third laboratory experience (designed largely for education majors), will involve observations of social interactions in a public school classroom. We plan to give our latest form of the EMT to all three groups as a pre test and then give the same form of the EMT as a post test during the final examination period. One problem met this year in trying to interpret the changes in scores on the EMT between the pre and post test scores was the fact that we had no control group. The changes in scores could be attributed as easily to students learning psychology in the lecture aspect of the course instead of learning scientific method in the laboratory. For this reason, next years' design of laboratory and field experience is especially advantageous to us. Two of the three groups will be focusing on experience which is not oriented toward learning scientific method per se. This means that we can expect relatively small changes in the pre and post test scores for these two groups in relation to the changes which should occur if the laboratory is functioning to teach what is covered by the test.

While we want to continue improving the test we have developed during this grant period, we feel that the instrument is now at a stage of development where it would be appropriate for us to begin running comparisons of various techniques of teaching Introductory Psychology laboratories. For this reason, we plan

to resubmit our original proposal with some modifications. We would then begin comparison of two techniques of teaching scientific method next year. We would plan to use the form of the test we have developed over this year as one of several evaluation instruments by which we would attempt to measure the relative effectiveness of the two teaching techniques.

BIBLIOGRAPHY

Guilford, J.P.: Fundamental Statistics in Psychology and Education. McGraw-Hill Book Company, Inc., New York, New York, 1956.

Kelley, T.,: Interpretation of Educational Measurements. World Books. Utica, New York. 1927

Siegel, S.,: Non-Parametric Statistics for the Behavioral Sciences. McGraw-Hill Book Company, Inc. New York, New York. 1956.

APPENDIX

Experimental Method Test Form 3 (final form).