

A COMPARISON OF "TAKE-HOME TESTS" WITH "IN-CLASS TESTS"
ON FINAL EXAMINATION ACHIEVEMENT

by

BARRY PHILIP DAVIS

A professional paper submitted to the School of Education
in partial fulfillment of the requirements for the degree

of

MASTER OF EDUCATION

in

Secondary School Administration

MONTANA STATE UNIVERSITY
Bozeman, Montana

August, 1967

TABLE OF CONTENTS

CHAPTER	PAGE
I. THE PROBLEM	1
Background of the Problem	1
Statement of the Problem.	2
II. REVIEW OF RELATED RESEARCH AND SUMMARY.	3
Related Research	3
Summary	4
III. METHODS	5
Population.	5
Procedures.	5
Data	6
Analysis of Covariance.	6
IV. RESULTS	9
Analysis of Covariance.	9
Reliability and Validity.	9
V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS.	10
Summary	10
Conclusions	11
Recommendations	12
LITERATURE CITES	13
APPENDIX	15

LIST OF TABLES

TABLE

I.	Analysis of Covariance on Final Examination	
	Scores of Groups I and II	8
II.	Grade Point Averages and Final Examination	
	Scores of Group I	16
III.	Grade Point Averages and Final Examination	
	Scores of Group II.	17

CHAPTER I

THE PROBLEM

I. BACKGROUND OF THE PROBLEM

The background of the problem studied in this paper arose from an actual field situation at Montana State University. At this University, the instructor of Principles of Secondary Education (Education 413) classes often used frequent objective quizzes as learning devices in his classes. The problem studied in this paper deals with the use of these quizzes in the fall quarter Education 413 classes of 1964 and 1965. Prior to the beginning of the fall quarter 1964 classes the instructor prepared a set of fifteen short objective quizzes for use in the Education 413 classes. These quizzes were designed to cover certain sections of the course work as it was studied during the quarter. The quizzes were then administered as regular in-class quizzes to the fall quarter 1964 Education 413 classes as the material for which they were constructed was covered. The same set of quizzes were administered in the following year to the fall quarter 1965 Education 413 classes. The quizzes were administered then, however, as take-home quizzes rather than as regular in-class quizzes. It is apparent that the take-home type quiz is not a test in the usual sense. Therefore, in the classes using take-home quizzes the character of the daily instruction was more relaxed and did not possess the

motivational effect which is often claimed for frequent testing such as was carried on in the classes where the in-class quizzes were given.

It then became the purpose of this study to determine by comparing final examination achievement which method of instruction was most effective...the use of the in-class quizzes or the use of the take-home quizzes.

II. STATEMENT OF THE PROBLEM

The purpose of the study was to compare the final examination achievement of a group which was subjected to frequent in-class quizzes with the final examination achievement of a group using the same set of quizzes as take-home quizzes.

CHAPTER II

REVIEW OF RELATED RESEARCH AND SUMMARY

I. RELATED RESEARCH

Some researchers have attempted to determine if the use of frequent quizzes during the term actually leads to improved final examination achievement. Two studies closely related to the one at hand will be given here.

Turney (1) conducted an experiment with two classes in educational psychology, one having frequent short objective tests while the other had but two tests during the semester. The first-mentioned class, on a pre-test covering the subject matter of the course, had an average score substantially lower than that of the other group. At the end of the course the two were practically equal in achievement. Whether the greater gain of the experimental group was the result of more frequent tests is not certain. The groups were not equated.

Hertzberg, Heilman and Leuenberger (2) report an investigation of the value of objective tests as teaching devices in a course of educational psychology. Two groups matched in scholastic aptitude were the subjects. One group had a test every third day which was scored and returned to them. They were permitted to keep these until the final examination. The other group had no quizzes. On the final examination covering the entire course, the non-tested group was slightly superior.

II. SUMMARY

The results of the two studies on frequent testing are somewhat equivocal. Furthermore, an inadequacy in the two studies just cited is that the non-tested groups apparently were not allowed to study the same tests that the tested groups took. Since tests are learning devices as well as evaluation devices, the non-tested groups may have been at a disadvantage is not being allowed to study the same tests as the tested groups in preparation for the final examination.

This inadequacy was removed in this study since both groups of Education 413 classes were presented with the same quizzes during the time they took the course. Both groups were allowed to keep and study the quizzes in preparation for the final examination.

CHAPTER III

METHODS

I. POPULATION

The population for this study consisted of two groups of fall quarter Education 413 students. The first group consisted of the 108 students enrolled in Education 413 during the fall quarter, 1964. The second group consisted of the 122 students enrolled in Education 413 during the fall quarter of 1965. These two groups subsequently had to be reduced to 100 and 111 students, respectively, because of the unavailability of data on eight of the students in the first group and eleven of the students in the second group.

The first group will be designated as Group I and the second group as Group II throughout the remainder of the paper.

II. PROCEDURES

Group I was subjected to fifteen short objective quizzes given as regular in-class quizzes during the fall quarter of 1964. Group II was given the same set of quizzes as take-home quizzes during the fall quarter of 1965. The instructor attempted to keep all other assignments and instruction for both groups as similar as possible in order to give a maximum degree of validity to the results in terms of final examination achievement.

The same final examination was administered to both groups. It was made up of one hundred items of the multiple choice type to insure objectivity in scoring. The items were constructed to sample points studied in all areas of the course material, but not already tested in the quizzes.

The reliability of the final examination was measured by the split-half method (3). There was no acceptable independent criterion by which the validity of the examination could be measured, but the examination was studied for "face validity" (4).

III. DATA

The data used in this study consisted of the final examination scores and the grade point averages of all the subjects computed just prior to taking the Education 413 course.

The data are compiled in Table II and Table III of the Appendix.

IV. ANALYSIS OF COVARIANCE

Due to the difference in size of Group I and Group II it would have been difficult to equate the two groups at the start. To remedy this the method of analysis of covariance was used. For, as stated by Garrett (5), "through covariance analysis one is able to effect adjustments in final or terminal scores which will allow for differences in some initial variable".

The initial variable in this study was taken to be the grade point average.

The method of analysis of covariance as explained by Garrett (6) was followed. Because of the length of the method only the essential data and results will be presented. The important data derived from the original data are listed in Table I.

In Table I, the grade point averages are identified as the initial independent variable (x). The final examination scores are identified as the dependent variable (y). As can be seen in Table I, the ratio of the among means variation to the within groups variation known as the "F" ratio is 0.26. According to Table F of Garrett (7) this was not a significant "F" ratio, consequently there was no significant difference between the mean scores achieved by Group I and Group II.

TABLE I

ANALYSIS OF COVARIANCE ON FINAL EXAMINATION SCORES OF
GROUPS I AND II

Source of Variation	df	SS _x	SS _y	SS _{xy}	SS _{y.x}	V _{y.x}	SD _{y.x}
Among Means	1	3450	3	110	12	12	- - -
Within Groups	208	340,186	12,877	32,478	9777	47	6.86

$$F_{y.x} = \frac{\text{Among means variance}}{\text{Within groups variance}} = \frac{12}{47} = 0.26 \text{ (not significant)}$$

Abbreviations:

x represents grade point averages

y represents final examination scores

df - degrees of freedom

SS_x - sum of squares for x

SS_y - sum of squares for y

SS_{xy} - sum of squares for the cross product of x and y

SS_{y.x} - sum of squares for y, x held constant

V_{y.x} - variance of y, x held constant

SD_{y.x} - standard deviation of y, x held constant

CHAPTER IV

RESULTS

I. ANALYSIS OF COVARIANCE

The method of analysis of covariance when applied to the grade point averages and final examination scores of the two groups yielded an "F" ratio of 0.26. This ratio was not significant at the .05 level and indicated that there was no significant difference in mean final examination achievement of Group I and Group II (8).

II. RELIABILITY AND VALIDITY

The reliability of the final examination was computed from a sample of twenty-one answer sheets of a group of Education 413 students who took the course in the spring quarter of 1967. The answer sheets of Group I and Group II could not be used as they had been discarded prior to the writing of this paper.

The reliability of the test as calculated from the small group sample using the split-half method was 0.50. This coefficient was less than what would have been found for the large Groups I and II as larger groups produce greater test variability and higher reliability coefficients (9). The standard deviation of the small sample ($SD = 5.93$) was much

less than the combined standard deviation of Groups I and II. (SD = 6.86) With the aid of a formula by Garrett (10), the reliability of the final examination was approximated at 0.63 for Groups I and II.

The estimated reliability coefficient of 0.63 for the final examination taken by Groups I and II was high enough to differentiate between the means of the two groups had any true difference existed (11).

The final examination exhibited "face validity" as it sampled a wide range of course material and appeared to be relevant to the course objectives.

CHAPTER V

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

I. SUMMARY

The final examination achievement of a group of Education 413 students subjected to frequent in-class testing was compared to the achievement on the same examination of a group of Education 413 students who had not been subjected to the testing, but were allowed to study the same tests as take-home quizzes.

The method of analysis of covariance was used to adjust the mean final examination scores for any differences that may have existed between the groups on the basis of the initial variable which was taken to be the grade point averages of the students at the time the course was taken.

No significant difference was found between the mean final examination scores of the two groups.

The reliability of the final examination was calculated and found to be approximately 0.63. This reliability coefficient was considered high enough that any true difference in the means of the final examination scores would have been revealed in the examination.

II. CONCLUSIONS

The main conclusion to be drawn from this study is that in terms of final examination achievement neither frequent in-class or take-home testing is superior to the other as long as both groups are presented with the same materials including the frequent tests to study in preparation for the final examination. Apparently, the motivational effects of frequent testing, if they do indeed exist, are not sufficient to significantly effect final examination achievement.

Some of the major weaknesses of this study as seen by the author should be mentioned now.

First, the two groups studied were equated only on the variable of overall grade point averages. The use of other possible initial variables such as I.Q., college entrance scores, grades in education courses, and so on may have revealed differences in the population groups not detected by grade point average only.

Second, the groups studied did not take the course at the same time. The background information possessed by these two groups may have been different enough to give one group an advantage over the other.

Third, the population for this study was two groups of college students, mostly seniors, who undoubtedly would not be affected as much by any particular teaching method as grade

school or secondary students. If there was superiority of one method over the other it may not have been revealed using this population group.

III. RECOMMENDATIONS

The results of this study are certainly conclusive enough so that further study is warranted. Some of the specific recommendations that the author would like to make in regard to any further experimentation dealing with frequent testing are:

(1) The population could be groups of grade or high school students who are presumably more effected by differences in teaching methods than college students.

(2) The test scores should be analized on a basis of several different initial variables; one may not be sufficient.

(3) The experimentation should take place within one span of time to control the effects of history on the population.

(4) Greater consideration should be given to the use of take-home tests as learning devices thus allowing more class time for lecture and discussion.

LITERATURE CITED

LITERATURE CITED

1. Turney, A. H. "Effect of Frequent Short Objective Tests Upon the Achievement of College Students in Educational Psychology," School and Society, Vol. 33 (June 6, 1931), pp. 760-762.
2. Hertzberg, O. E., Heilman, J. D., and Leuenberger, H. W. "The Value of Objective Tests as Teaching Devices in Educational Psychology Classes," Journal of Educational Psychology, Vol. 23 (May, 1932), pp. 371-380.
3. Garrett, Henry E. Statistics in Psychology and Education (sixth edition; David McKay Company, 1966), pp. 339-340.
4. Ibid., p. 355.
5. Ibid., p. 295.
6. Ibid., pp. 296-303.
7. Ibid., p. 466.
8. Ibid., p. 301.
9. Ibid., p. 352.
10. Ibid., p. 352.
11. Ibid., p. 351.

APPENDIX

TABLE II

GRADE POINT AVERAGES AND FINAL EXAMINATION SCORES OF GROUP I

Student	GPA	Exam.	Student	GPA	Exam.	Student	GPA	Exam.
1.	3.88	93	35.	2.83	83	68.	2.52	83
2.	3.68	77	36.	2.83	75	69.	2.49	81
3.	3.65	86	37.	2.83	70	70.	2.48	63
4.	3.65	83	38.	2.82	71	71.	2.47	76
5.	3.51	83	39.	2.82	68	72.	2.46	55
6.	3.47	78	40.	2.80	79	73.	2.45	61
7.	3.43	72	41.	2.77	69	74.	2.44	68
8.	3.38	90	42.	2.76	78	75.	2.38	73
9.	3.36	91	43.	2.75	85	76.	2.36	64
10.	3.34	79	44.	2.73	78	77.	2.34	62
11.	3.32	75	45.	2.73	73	78.	2.32	66
12.	3.29	76	46.	2.73	62	79.	2.31	63
13.	3.22	69	47.	2.72	73	80.	2.30	74
14.	3.20	70	48.	2.72	65	81.	2.29	77
15.	3.16	77	49.	2.66	80	82.	2.28	65
16.	3.12	73	50.	2.66	79	83.	2.27	61
17.	3.11	76	51.	2.66	72	84.	2.26	59
18.	3.08	67	52.	2.65	80	85.	2.23	69
19.	3.07	87	53.	2.65	76	86.	2.23	68
20.	3.06	69	54.	2.65	76	87.	2.21	69
21.	3.05	85	55.	2.65	64	88.	2.19	66
22.	3.02	79	56.	2.64	80	89.	2.18	65
23.	3.02	77	57.	2.64	78	90.	2.16	66
24.	3.00	77	58.	2.64	69	91.	2.16	63
25.	2.99	78	59.	2.63	61	92.	2.15	71
26.	2.97	72	60.	2.62	76	93.	2.12	74
27.	2.96	90	61.	2.60	74	94.	2.11	69
28.	2.91	79	62.	2.60	74	95.	2.11	65
29.	2.89	76	63.	2.59	74	96.	2.09	66
30.	2.89	67	64.	2.57	79	97.	2.09	59
31.	2.88	62	65.	2.54	71	98.	2.05	72
32.	2.87	69	66.	2.53	73	99.	2.04	69
33.	2.84	87	67.	2.53	69	100.	1.80	72
34.	2.83	85						

TABLE III

GRADE POINT AVERAGES AND FINAL EXAMINATION SCORES OF GROUP II

Student	GPA	Exam.	Student	GPA	Exam.	Student	GPA	Exam.
1.	3.72	79	38.	2.71	75	75.	2.43	74
2.	3.59	86	39.	2.71	70	76.	2.41	77
3.	3.55	86	40.	2.69	71	77.	2.40	76
4.	3.48	80	41.	2.68	76	78.	2.40	73
5.	3.41	85	42.	2.67	80	79.	2.40	69
6.	3.34	83	43.	2.67	77	80.	2.38	72
7.	3.27	81	44.	2.67	56	81.	2.38	70
8.	3.26	70	45.	2.66	74	82.	2.37	71
9.	3.24	80	46.	2.66	66	83.	2.36	68
10.	3.21	78	47.	2.64	81	84.	2.36	56
11.	3.19	83	48.	2.61	73	85.	2.34	68
12.	3.12	83	49.	2.61	70	86.	2.34	64
13.	3.12	78	50.	2.60	74	87.	2.33	79
14.	3.11	81	51.	2.60	57	88.	2.33	72
15.	3.09	82	52.	2.59	83	89.	2.32	72
16.	3.08	76	53.	2.59	70	90.	2.31	64
17.	3.05	84	54.	2.59	63	91.	2.27	57
18.	3.04	82	55.	2.58	80	92.	2.25	65
19.	3.02	78	56.	2.58	75	93.	2.24	80
20.	3.02	74	57.	2.58	69	94.	2.22	82
21.	3.00	74	58.	2.58	60	95.	2.20	76
22.	2.98	76	59.	2.57	70	96.	2.20	71
23.	2.96	84	60.	2.57	66	97.	2.19	64
24.	2.95	82	61.	2.56	70	98.	2.17	82
25.	2.95	59	62.	2.54	85	99.	2.17	59
26.	2.90	72	63.	2.54	70	100.	2.16	85
27.	2.89	73	64.	2.53	69	101.	2.15	69
28.	2.87	86	65.	2.52	67	102.	2.12	71
29.	2.87	74	66.	2.52	66	103.	2.12	60
30.	2.87	72	67.	2.51	79	104.	2.11	69
31.	2.83	80	68.	2.50	59	105.	2.11	63
32.	2.81	68	69.	2.48	76	106.	2.09	80
33.	2.78	73	70.	2.47	78	107.	2.09	67
34.	2.77	79	71.	2.47	64	108.	2.09	65
35.	2.75	73	72.	2.44	79	109.	2.04	65
36.	2.74	75	73.	2.44	74	110.	2.04	54
37.	2.73	79	74.	2.43	76	111.	2.00	60