5-1955

Comparison of the Utah State Agricultural College Mathematics Entrance Examination Computational Form Versus Multiple-Choice

Wilson Walker
Utah State University

COMPARISON OF THE UTAH STATE AGRICULTURAL COLLEGE MATHEMATICS ENTRANCE EXAMINATION COMPUTATIONAL FORM VERSUS MULTIPLE-CHOICE

by

Wilson Walker

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Psychology

UTAH STATE AGRICULTURAL COLLEGE
Logan, Utah

1955

ACKNOWLEDGEMENT

I wish to express appreciation to Dr. Arden N. Frandsen, my thesis director, for suggesting the problem for this thesis and for the assistance he has given me in bringing about its completion.

Wilson Walker

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

It is the standard practice in colleges and universities to administer to all entering students a battery of tests which are designed to measure the aptitude and ability of students. Each institution uses in its test battery those tests which are best suited to its particular needs, but in the main they consist of measures of general mental ability, the ability to understand and use the English language and the ability to understand and to use mathematical symbols. These tests are available to the institution through companies which construct and standardize tests of all kinds. In addition to these tests, others may be constructed by the departments of an institution and validated for use in that department or school. The test battery of the Utah State Agricultural College consists of a combination of these two types. The English examination is the Educational Testing Service Cooperative English test Battery, the Mathematics test is a department constructed test, originally designed and constructed by Dr. Arden N. Frandsen.

The original form of the test is constructed as an expendable booklet in which the student computes the answer and enters it in a space provided. The original form was used as part of the Utah State Agricultural College entrance examination until 1947. At that time David Stone, William Dobson and Glenn Hawkes revised the test into a multiple-choice form. In revising the test, the authors used the same problems as those in the original form. To this they added as five possible answers the correct answer and the four most frequently

given wrong answers for each item. The theory behind the revision was to provide a test which would be more economical, easier to score and in harmony with the new trend in test construction.

It is the scope of this paper to present some of the history of the Utah State Agricultural College Mathematics test, its construction, validation and revision, more specifically, to compare the original and revised editions as to predictive value.

REVIEW OF LITERATURE

In 1942, Arden Frandsen (1) constructed the original form of the Utah State Agricultural College Mathematics Test to be used as a predictor of mathematical ability. The author describes the test as follows:

> The test consists of 50 problems sampling the basic and most frequently used phases of arithmetic and elementary algebra. The selection of problems was guided by the hypothesis that performance on the common mathematical problems within the experience of every person who has completed high school would measure mathematical abilities independent to some degree from the specific training. To free further the functions measured from specific memory, rules for many of the processes were supplied. The problems included addition, subtraction, multiplication and division of integers, percentage problems; ratio and proportion; solving for an unknown in simple literal equations; substitution in formulae; computing areas of simple geometric figures; determining position of decimal place; conversion of fraction to percentages; and dealing with signed numbers.

The test has a Spearman-Brown chance half reliability of .93. It is scored for the number of items correctly answered in the 45 minute period designated for the test. The test has been used on numerous occasions as a predictor of scholastic achievement. It has been used mainly as a predictor of mathematical ability but on occasions it has been used to predict general achievement and achievement in specific courses.

The test was administered to 512 of the entering freshmen at the Utah State Agricultural College in 1941-42. The means for this group on forms A and B were 22.3 and 23.1 respectively. The mean for the entire group was 22.5, with a standard deviation of 7.10. (2, p. 20)

Table 1, published by Frandsen, (2, p. 21) shows correlations between the Utah State Agricultural College Mathematics Test and achievement.

TABLE 1. CORRELATION OF U.S.A.C. MATHEMATICS TEST WITH ACHIEVEMENT

Course	Correlation
Average grade for Freshmen Students	.44
Beginning Algebra	.69
Physical Science	.42
Chemistry	.75
English Usage	.41
Vocabulary	.34
Reading	.36

The above correlations indicate that there is a moderate to high relationship between the U.S.A.C. Mathematics test and courses which require mathematical ability, a low relationship to the command and use of the English language, and a moderate relationship to average college success.

Jacobsen (3, p. 38 & 44) used the Utah State Agricultural College Mathematics Test in a research study designed to predict achievement in the Mechanic Learner program of the college. The results of the study indicate that the test predicts best, achievement in Machine Shop, having a correlation of .44. A multiple correlation using the U.S.A.C. Mathematics Test and the Bennett's Test of Mechanical Comprehension yielded a correlation of .49.

Brite (2, p. 35) used the Utah State Agricultural College Mathematics Test in a research study in which it was used as a predictor

of success in Elementary Electricity and Radio Material at the Utah State Naval Training School. The results of the study show that of all the tests used as predictors, the U.S.A.C. Mathematics Test was the best single predictor of success. The correlation with mathematics is .73; with laboratory achievement, .73; with Radio-Physics, .66; and with average grade achievement, .73. The average correlation of the U.S.A.C. Mathematics test with success in the criterion of the study is .67.

Egbert (4) in a thesis study designed for evaluation of the guidance tests given at the Utah State Agricultural College used the mathematics test as a predictor of success. Table 2 shows correlations between the test and college courses.

TABLE 2. CORRELATIONS BETWEEN U.S.A.C. MATHEMATICS TEST AND COLLEGE COURSES

Course	No.	Correlation	Significance
Typewriting	23	0.265	Not Sig.
Shorthand	33	0.146	Not Sig.
Aeronautics 5	22	0.246	Not Sig.
Physiology 4	177	0.55	Sig. at 1% level
Speech 1	102	0.219	Sig. at 1% level
Forestry 1	97	0.452	Sig. at 1% level
Mathematics 34	121	0.633	Sig. at 1% level
Mathematics 35	54	0.490	Sig. at 1% level
Psychology 3	105	0.47	Sig. at 1% level

Bromley and Carter (5) in a study at the University of Illinois, of the incoming freshmen and sophomores, indicate a correlation between grades in college mathematics and scores on two standardized mathematic achievement tests, the Cooperative General Achievement and the American Council on Education Psychological Examination Q Test, of .32 and .28 respectively.

Keller and Jonah (6) in a similar study conducted at Purdue concluded that the score on the mathematic placement test was the best single predictor of success in the first course in college mathematics.

The results of these studies and the information supplied by Frandsen, the author of the original form of the U.S.A.C. Mathematics test, indicate that the original form of the test is a valid predictor of success in various educational areas. It also indicates that the test compares favorably with other standardized mathematics tests used for predictive purposes.

In 1947, Stone, Dobson and Hawkes (7) revised the original form of the test into a multiple-choice form. The authors in revising the test were directed by the hypothesis that the original form of the test could be changed into a multiple-choice form which would be more economical, easier to score and yet retain the same predictive value.

The authors, in choosing to revise the test, were following a trend in test construction which was gaining popularity and wide spread use among test constructors. Cronbach (8, p. 51) indicates that the reason for the increased use of the multiple-choice items in test of knowledge, aptitude and interest is that for pencil and paper tests

the multiple-choice test seems to be the most free from response sets. Again, Cronbach (8, p. 73) indicates the acceptance of the multiple-choice test.

> Current practices in group testing favors the multiple-choice form. Although this measures recognition of correct answers rather than recall, it is satisfactory for many purposes. The College Entrance Examination Board, for example, reports that in mathematics tests at the college level multiple-choice questions had reliability coefficients and correlation with grades in later mathematics essentially the same as those for the free answer questions.

The authors of the new test began the revision by using the same problems which Frandsen had included in the original form, to this was added five possible answers from which the student could choose one. The criteria by which the authors selected the five alternative answers were to, 1. select the correct answer to the problem, 2. select four alternate answers which had a high probability of being computed by the student taking the test. The method used in selecting the four wrong answers was to choose them on the basis of their order of frequent appearance on the original form of the test. Cronbach (8, p. 278) concludes that this is the best method of selecting the incorrect answers. He states that;

> Considerable research has been done to determine if the objective tests are adequate. The findings have been consistently favorable to the use of well constructed objective tests. When a multiple-choice test is made up of correct answers, together with incorrect alternatives chosen from among the wrong answers given by students who have answered the same questions with free response, the multiple-choice test has a high correlation with the free response test.

It is generally accepted among those who construct tests that where selection of the position for the correct answer among the alternate choices is concerned, one criterion which needs consider-

ation is that each choice have an equal number of correct answers. In so far as the sequence of the correct answer is concerned, it may appear as many times in sequence as there are number of choices. It is unlikely that the sequence will normally run past three or four.

This author was unable to determine the exact method used by Stone, Dobson and Hawkes for selecting the position of the correct answer among the five choices. In the absence of such data it is felt that a description of the position of the correct answer is needed. The following table indicates the number of times the correct answer appears for each choice.

TABLE 3. POSITION OF CORRECT ANSWER IN U.S.A.C. MATHEMATICS TEST MULTIPLE-CHOICE FORM

Choice	1	2	3	4	5
Number of times correct answer appears	6	16	10	17	6

In the multiple-choice form of the U.S.A.C. Mathematics test there are 55 problems. According to the literature cited above there should be 11 correct answers in each of the 5 choices. Observation of table 3 shows that none of the five choices has the correct number of right answers.

The problem of chance guessing on multiple-choice tests has some significance to this study. The test instructions for the revised edition of the Utah State Agricultural College Mathematics Test does not

include any information relative to guessing, neither is there any correction formula applied in the scoring of the test. Present literature on the subject indicates, that on tests of this nature, guessing has an effect on the score. Cronbach (8, p. 90) states that, "if there is no penalty for wrong answers, the guesser receives a better score than the non-guesser of equal ability." Cronbach also indicates that if people have a knowledge that they will be penalized for guessing, some will be more hesitant, while others will take more chances and guess freely. One point seems evident, the more severe the penalty the less guessing on the part of the student. In general, when the instructions indicate "do not guess" the test will give a better measure of the ability of the group.

Stanley (9), in a study directed toward the problem of correction for chance states;

> Correcting test scores for "guessing" is tantamount to correcting them for differing numbers of unanswered items among the testees. If O, the standard deviation of the number of items omitted, is zero, then each person's z-score is unaffected by the usual correction for chance success. Therefore, if each testee leaves blank the same number of items as every other testee, the correlation rRS between "rights" scores (R) and corrected scores
>
> $$(S=R - \frac{W}{c-1})$$
>
> will be $\neq 1.00$. Here W is the individual's "wrong" score and c the number of choice (options) each item has.

In the article cited above Stanley quotes Gulliksen as saying:

> ....there is no reason for considering any of these correction for chance formulas if, for most of the people, $R \neq W$ is essentially equal to the total number of items in the test. Such formulas are to be used if, and only if, the number of unmarked items is fairly large for some persons and fairly small for others.

Stanley also indicates that "even when a correction for chance

is useless from a statistical standpoint, it may nevertheless be warranted by a certain attitude-effect produced on college students, once the logic of the correction formula is understood by them."

Analysis of the answer sheets of the multiple-choice test for this study shows a range of 0 to 40 for the unanswered items. There was a high concentration of unanswered items in the range of 10 to 20.

One other test was used in connection with this study. The School of Engineering gives to all prospective engineers a test which is designed to predict achievement in the first year of engineering study. The test which was given to the subjects of the present study was the Educational Testing Service Pre-Engineering Ability Test. This test was developed primarily for guidance purposes in engineering schools. The content of the test includes: (10)

> ....reading passages, tables and graphs, covering various types of scientific information, with one or more questions pertaining to each. Problems which are concerned mostly with arithmetic, algebra, and plane geometry, although a small number require a knowledge of solid geometry and trigonometry. The items included sample not only the entire range of difficulty, but the types of questions and subject material shown to be most effective by research and item-analysis at one leading engineering institution.

The test requires 80 minutes of working time. The test has a reliability coefficient of .90 and an average correlation coefficient between the test scores and first term grades in 12 engineering schools is .57. The test is constructed as a multiple-choice type test. The score is the number right.

HYPOTHESES

The first hypothesis to be tested in this study is, that the multiple-choice form of the Utah State Agricultural College Mathematics Test is as effective in predicting success in mathematic achievement as is the original computational form of the test.

A second hypothesis to be tested is, that the better of the two forms, the computational and the revised multiple-choice forms, of the Utah State Agricultural College Mathematics Test will predict success in engineering as well as does the Educational Testing Service Pre-Engineering Ability Test.

PROCEDURE

The subjects for the present research are those students entering the Utah State Agricultural College for the fall quarter, 1954, who have less than 96 hours of credit, or who have not taken an entrance examination in other colleges or high schools in the inter-mountain area. The total number of students included in the study is 828. This total number was divided into two major groups, those students who were registered in the school of engineering, and all the remaining students who took the test.

Each of the two major groups were again divided into two equal divisions. This division was arbitrarily set by designating two groups, one to include those students in the odd seats and the other to include those students in the even seats. No attempt was made toward arranging the seating of the students, each student was free to sit in any seat within his group. This method of seating tended to equalize the odd-even groups in so far as one being superior in ability over the other.

The two forms of the Utah State Agricultural College Mathematics Test, the original computational and the revised multiple-choice editions, were administered to the subjects of this study. Those occupying the odd row received one form, those occupying the even row received the other form.

The completed tests were scored by graduate students who assisted in the testing program. As an added check, this author and an asistant rescored each of the tests, making corrections where necessary.

The directions for the multiple-choice form of the test do not include instructions to the student relative to guessing, neither are there provisions made for applying a correction formula for chance guessing to the scores. However, on the basis of the literature cited, the correction formula (right minus $\frac{\text{wrong}}{n-1}$) was applied to all the scores of the multiple-choice form.

Through the cooperation of the registrar's office, the grade point average and the letter grade in Mathematics 34, a first course in college mathematics, were obtained for all subjects in this study. This data constituted one criterion of success in college.

Through the cooperation of the School of Engineering the scores made on the Educational Testing Service Pre-Engineering Test were obtained for the subjects of this study. These data were used for comparison with the original and revised forms of the U.S.A.C. Mathematics test.

The means and standard deviations were computed for each group taking the test and a comparison made between them. A correlation was computed between each of the tests and the criteria of achievement and a comparison made between them.

RESULTS

The results of this study are shown in two parts, one, the tables of means and standard deviations and two, the table of correlations. Each table is followed by a summary of the findings.

Part 1

TABLE 4. MEAN SCORES FOR THE ENGINEERING GROUP ON THE ORIGINAL AND REVISED FORMS OF THE U.S.A.C. MATHEMATICS TEST

Test Form	Grade Point Average			Mathematics 34 Grades		
	N	Mean	S.D.	N	Mean	S.D.
Original	62	31.0	11.6	39	27.5	10.3
Revised	65	29.3	12.5	44	26.0	11.1

For the grade point average group, the value of (t) between the means of the original and revised forms is not sufficiently high to reject the hypothesis that there is no difference between the means. It can therefore be assumed that the two samples were drawn from essentially the same population.

For the Mathematics 34 group, the value of (t) between the original and revised forms is not sufficiently high to reject the hypothesis that there is no difference between the means. It can therefore be assumed that these two samples also have been drawn from the same population.

TABLE 5. MEAN SCORE FOR THE GENERAL GROUP ON THE ORIGINAL AND REVISED FORM OF THE U.S.A.C. MATHEMATICS TEST

Test	Grade Point			Mathematics 34		
	N	Mean	S.D.	N	Mean	S.D.
Original	332	20.7	11.8	105	22.0	10.4
Revised	379	19.2	11.2	135	21.7	9.7

The value of (t) between the means of the original and revised forms for the grade point average of the general group is not sufficiently high to reject the hypothesis of no difference.

The value of (t) between the means for mathematics 34 group of the general group is not high enough to reject the hypothesis of no difference.

TABLE 6. MEAN SCORES OF THE ENGINEER GROUP ON THE PRE-ENGINEERING TEST

Group	N	Mean	Standard Deviation
Original test group	39	27.5	7.5
Revised test group	44	29.9	11.4

The value of (t) between the means on the pre-engineering test for the original and revised test group is not sufficiently high to reject the hypothesis that there is no difference between the means.

In Table 7 there are presented the data for testing the hypotheses of this thesis. It presents the validity correlations between the U.S.A.C. Mathematics Test in both computational and multiple-choice forms and grades in both Mathematics 34 and average grade point ratios. And it also compares these correlations with the correlation between the Pre-Engineering Examination and both Mathematics 34 grades and average grade point ratios.

TABLE 7. CORRELATIONS OF THE U.S.A.C. MATHEMATICS ABILITY TESTS IN COMPUTATIONAL AND MULTIPLE-CHOICE FORMS AND THE PRE-ENGINEERING EXAMINATION WITH MATHEMATICS 34 GRADES AND AVERAGE GRADE POINT RATIOS.

Groups	Correlation with Achievement			
General Non-Engineering students	Math. 34 Grade	N	Avg. Grade Pt.	N
U.S.A.C. Mathematics Computational Form	.64	105	.45	332
U.S.A.C. Mathematics Multiple-Choice Form	.53	135	.51	379
"t"* for difference between these Correlations corrected to z's	.85		1.07	
Engineering students				
U.S.A.C. Mathematics Computational Form	.44	39	.46	62
U.S.A.C. Mathematics Multiple-Choice Form	.59	44	.58	65
"t"* for difference between these correlations corrected to z's	.92		.44	
Engineering Students				
U.S.A.C. Mathematics Multiple-Choice Form	.59	44	.58	65
Pre-Engineering Examination	.58	127	.64	83
"t"* for difference between these correlations corrected to z's	.11		.59	

*For significance at the .05 per cent level, these t-ratios would have to equal or exceed 1.96.

For the general non-engineering students, the difference between the correlations with Mathematics 34 grades, for the computational and the multiple-choice forms of the U.S.A.C. Mathematics Test is not significant at the .05 per cent level. The difference between the correlations with average grade point ratio for the two forms is also not significant at the .05 per cent level.

For the engineering students, the difference between the correlations with the Mathematics 34 grades, for the computational and the multiple-choice forms of the U.S.A.C. Mathematics Test is not significant at the .05 per cent level. The difference between the correlations with average grade point ratio for the two forms is also not significant at the .05 per cent level.

The difference between the correlations with Mathematics 34 grades for the Pre-Engineering Ability Test and the multiple-choice form of the U.S.A.C. Mathematics Test is not significant at the .05 per cent level. The difference between the correlations with average grade point ratio for these two tests is not significant at the .05 per cent level.

A correlation between the scores on the original computational form of the U.S.A.C. Mathematics Ability Test and scores on the Educational Testing Service Pre-Engineering Ability Test is .74 which is significant at the .01 per cent level. A correlation between the scores on the revised multiple-choice form and the scores on the Educational Testing Service Pre-Engineering Ability Test is .74 which is significant at the .01 per cent level.

It must be concluded, as far as this study is concerned, that there

is no real difference between the predictive value of the two forms, the original computational form and the revised multiple-choice form, of the U.S.A.C. Mathematics test. It must also be concluded that there is no real difference between the predictive value of the Educational Testing Service Pre-Engineering Ability Test and the two forms of the U.S.A.C. Mathematics Test.

SUMMARY AND CONCLUSIONS

It has been the purpose of this study to evaluate the two forms of the Utah State Agricultural College Mathematics Test as predictors of achievement in both general college curricula, as measured by grade point average, and success in mathematics, as measured by the letter grade in Mathematics 34, and to compare the validity of the revised multiple-choice form with the validity of the original computational form.

The history of the original form shows that it has been used effectively to predict college achievement, especially in courses which require mathematical ability. The present study yielded correlations with grade point average and grades in mathematics of equal significance to those of earlier studies. In addition to this, both forms, the original computational and the revised multiple-choice, of the U.S. A.C. Mathematics Test showed a high correlation, .74, with the Cooperative Pre-Engineering test. These data indicate that the test has a high empirical validity.

The construction of the revised form of the test has been analyzed and for the most part has complied with the accepted technique of test construction. There are two aspects of the test pointed out by this study, which need further research and evaluation. One, the use of a correction formula for guessing, and two, equal distribution of the correct answer among the number of choices.

Analysis of the means between the two forms of the test for the different groups tested, indicate that the groups were equally divided

with regards to ability. As might be expected, the means for the engineer group was higher than those for the general group.

The correlations between the two forms of the test and the criteria of achievement were all positive and significant at the 1 per cent level of confidence. The revised multiple-choice form of the test yielded slightly higher correlations with achievement than did the original computational form. The differences between the correlations with achievement for the two forms of the U.S.A.C. Mathematics tests were not significant at the .05 per cent level of confidence. The differences between the correlations with achievement for the pre-engineering ability test and the U.S.A.C. Mathematics ability test were not significant at the .05 per cent level of confidence.

On the basis of the results of this research, the following conclusions may be drawn:

1. That the first hypothesis is tenable, that is, that the multiple-choice form of the Utah State Agricultural College Mathematics Test is as effective in predicting success in mathematic achievement, general scholastic achievement, and engineering achievement, as is the original computational form of the test.
2. That the second hypothesis is tenable, that is, that the two forms, the computational and the revised multiple-choice forms, of the Utah State Agricultural College Mathematics Test will predict success in engineering as well as does the Educational Testing Service Pre-Engineering Ability Test.
3. That the revised multiple-choice form of the U.S.A.C. Mathematic Ability Test can be used with confidence as a part of the entrance examination battery, for guidance purposes and as a predictor of achievement.

BIBLIOGRAPHY

1. Frandsen, Arden N., Hadley, John M. The Prediction of Achievement in a Radio School. Journal of Applied Psychology 1943, 27; No. 4.

2. Brite, Luna R. Predicting Success at the Utah State Agricultural College Naval Training School for Elementary Electricity and Radio Material. Thesis, 1943 U.S.A.C. Library.

3. Jacobsen, Eldon E. An Evaluation of Certain Tests in Predicting Mechanic Learner Achievement. Thesis, 1943 U.S.A.C. Library.

4. Egbert, Robert L. An Evaluation of the Guidance Tests Given at Utah State Agricultural College for the Year 1946-1947. Thesis 1948 U. S. A. C. Library.

5. Bromley, Ann, Carter, Gerald C. Predictability of Success in Mathematics. Jol. of Educational Research, 1950-51, 44; 148-150

6. Keller, M. W., Jonah, H. F. S. Measures for Predicting Success in a First Course in College Mathematics. Mathematics Teacher 1949, 41;350-55

7. Stone, David R. "Unpublished Information." 1955

8. Cronbach, Lee J. Essentials of Psychological Testing. New York: Harper & Brothers, Publishers, 1949

9. Stanley, Julian C. "Psychological" Correction for Chance. Jol. Experimental Education, 1953-54, 22:297-98

10. Educational Testing Service. 1954 Catalog of Cooperative Tests New Jersey