

CORRELATIONS AMONG GRE SCORES FOR
DOCTORAL PROGRAM APPLICANTS IN EDUCATION

by

Gary Edward Ackerson

A Thesis Submitted to the Faculty of the
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY
In Partial Fulfillment of the Requirements
For the Degree of
MASTER OF ARTS
In the Graduate College
THE UNIVERSITY OF ARIZONA

1 9 7 6

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Gary E. Anderson

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Darrell Sabers
DARRELL SABERS
Associate Professor
of Educational Psychology

12/17/75
Date

TABLE OF CONTENTS

	Page
LIST OF TABLES	iv
ABSTRACT	v
CHAPTER	
I. INTRODUCTION	1
The Graduate Record Examination: Aptitude Test	2
The Graduate Record Examination: Advanced Test in Education	2
Assumptions Underlying the Usage of GRE Scores	3
Assumption 1	4
Assumption 2	5
Assumption 3	7
Assumption 4	7
II. PROCEDURES	10
III. RESULTS	12
IV. DISCUSSION	17
LIST OF REFERENCES	21

LIST OF TABLES

Table	Page
1. Pearson r , E , r^2 , and E^2 for the six possible relationships of the Verbal, Quantitative, and Advanced Test in Education	13
2. E^2 , r^2 , F ratio, and probability of the six relationships of the three GRE scores	14
3. Obtained E^2 (E^2_o), E^2 corrected for inflation (E^2_c), and the r^2 of the 6 relationships of the three GRE scores	15

ABSTRACT

The purpose of this study was to examine the inter-correlations among Graduate Record Examination scores obtained by applicants to doctoral programs in Education within the Graduate School of The University of Arizona. Of special interest was the assumption that each GRE Test provides unique information which is not provided by the other two tests.

Pearson Product Moment correlation coefficients were computed for the three possible pairs of the GRE Verbal, GRE Quantitative, and GRE Advanced Test in Education. The Verbal and Advanced Test in Education were found to be highly correlated ($r = .78$). The other two pairs, Verbal and Quantitative and Quantitative with the Advanced Test in Education were not highly correlated ($r = .34$ for both pairs).

The conclusion drawn from this study is that the GRE Verbal and the GRE Advanced Test in Education are measuring much the same thing and as a result it is recommended that the GRE Advanced Test in Education should be dropped from the requirements for admission to doctoral programs in Education.

CHAPTER I

INTRODUCTION

All applicants to doctoral programs with majors in departments of the College of Education of The University of Arizona are required to report scores earned on the Graduate Record Examination (GRE). These scores in combination with other sources of information are used to screen potential candidates for admission to the doctoral programs. Each of the Departments within the College assesses the significance of the test scores individually and thus each has established its own criteria for the relative importance of GRE scores in their respective admission criteria.

The College-wide requirement is that the applicant report three scores: Verbal (GRE-V), Quantitative (GRE-Q), and an Advanced Test. Nine Departments require the Advanced Test to be the GRE Advanced Test in Education (GRE-AE). The Educational Psychology Department allows the student to submit either the Advanced Test in Education or the Advanced Test in Psychology (GRE-AP).

Normative data has not been compiled either College-wide or by Department on the achievement of applicants on the GRE.

The Graduate Record Examination:
Aptitude Test

The Aptitude Test of the GRE was designed to provide information about the general scholastic ability of students at the senior college or graduate school level. It provides separate measures of Verbal (GRE-V) and Quantitative (GRE-Q) ability. The GRE-V measures verbal reasoning and reading comprehension using questions drawn from the humanities, sciences, and social studies as well as activities and relationships of daily life. The GRE-Q consists of questions involving algebra, graph and diagram interpretation, and general mathematics reasoning.

A description of the abilities measured by the Aptitude Test as well as normative data are provided by the service (Educational Testing Service [E.T.S.] 1975).

The Graduate Record Examination:
Advanced Test in Education

The Advanced Test in Education (GRE-AE) has four general areas of coverage: Psychological Foundations, Social Foundations, Instructional Methods and Curriculum, and Organization and Administration. The examination is not divided into discrete sections but rather the order of questions is randomized. A description of the Advanced Test in Education, sample questions, and normative data are provided by the service (E.T.S. 1973a).

Scores on Graduate Record Examinations are always three digit numbers with the last number being 0. Scores on the GRE-V and GRE-Q range from 200 to 900. Scores on GRE Advanced Tests range from 200 to 990. Percentile ranks are provided which show the relative position of an individual's score in relationship to a normative sample. The normative sample is made up of all students taking the GRE Test during the most recent three year period. While the Educational Testing Service provides percentile ranks for all examinees, it advocates the usage of local institutional norms. It does not provide guidance for setting minimal performance levels.

The reliability of the GRE Tests is measured by the Kuder-Richardson formula 20. The reported reliabilities of the GRE-V, GRE-Q, and GRE-AE are .93, .92, and .92 respectively. They were computed from an unspecified period of time using equivalent forms of the respective tests. The figures reported were adjusted to a hypothetical sample with a standard deviation of 100 (E.T.S. 1973b).

Assumptions Underlying the Usage of GRE Scores

There appear to be four underlying assumptions for the use of the GRE scores in this College. The first is that the GRE Tests provide information which will be different from other available sources of information. The second

is that each of the scores is unique and measures different abilities. The third is that each of these abilities is required to some degree for successful completion of a doctoral program. The fourth is that each Department's doctoral program is unique and thus the departments should establish for themselves the relative importance of each score in accordance with their own criteria for acceptance.

Assumption 1

The first assumption has been tested somewhat indirectly. Numerous research studies have been done to evaluate the predictive ability of the GRE in conjunction with other sources of information.

Willingham (1973) reviewed 43 studies involving 21,214 subjects in 138 independent sets of data which resulted in 618 separate validity coefficients. These studies used several predictors besides GRE scores and had several different criterion variables. The criterion variables usually were some measure of success in graduate school. The results of this review suggest that GRE scores are more accurate predictors of success than undergraduate GPA. They also suggest that the GRE-Q is a more valid predictor in fields such as science or mathematics than more verbally oriented tests. Conversely the GRE-V has more predictive ability in more verbally oriented disciplines. Williams, Harlow, and Gab (1970) used the GRE-Q, GRE-V, and GRE-AE

with 14 other predictor variables using Graduated--Not Graduated and doctoral GPA as criterion variables and found the GRE scores to be the best predictors for the criterion of Graduated--Not Graduated. The above two studies suggest that since the GRE scores yield different predictive validity coefficients than other predictor variables they are probably measuring different sources of variance. If they are in fact accounting for variance unexplained by other predictors then the first assumption is somewhat validated.

It should be noted that while the assumption that GRE scores provide different information than other sources is supported by the above research, this research does not support the assertion that GRE scores are accurate predictors of success in graduate studies. Thacher and Williams (1974) in an extensive review of research involving GRE scores as predictors of success in graduate studies concluded that the widespread usage of GRE scores as predictors should be questioned as the majority of predictive correlations obtained in the studies they reviewed were low and several studies yielded non-significant correlations.

Assumption 2

The second assumption, that each of the three required GRE scores is unique and measures different abilities, has not been directly studied. Again, most studies involving the GRE scores are predictive ones, using GRE scores as

predictors and some measure of success in graduate school as a criterion variable. Most studies have found low correlations and have explained the low correlations as being due to the restricted range of the variables. Some studies report the correlations between the predictor variables but do not discuss these correlations in relation to the obtained predictive coefficients. Rupiper (1959) found the correlation between the GRE-V and the GRE-AE to be .90, the correlation between GRE-Q and GRE-AE to be .60, and the correlation between GRE-V and GRE-Q to be .48.

Other studies reporting correlations between GRE scores are: Sleeper (1961) who found a correlation of .55 between GRE-V and GRE-Q; Madaus and Walsh (1965) who found a correlation of .45 between GRE-V and GRE-Q; and Tully (1962) who found a correlation of .47 between GRE-V and GRE-Q.

The above reported correlations suggest that the GRE-V and GRE-Q are not highly correlated and thus are probably measuring different abilities. The obtained correlations between GRE-V and GRE-AE (.90) and between GRE-Q and GRE-AE (.60) as reported by Rupiper raise serious doubt as to whether the GRE-AE provides information that is unique from the information provided by the GRE-V and GRE-Q.

Assumption 3

The third assumption, that the abilities measured by the GRE tests are abilities which are necessary for successful completion of doctoral programs is not addressed by the previously cited literature.

The consistently low predictive correlations yielded by these studies suggest that there are other abilities or conditions which have a greater bearing on success in doctoral programs. A possible alternative suggestion is that the abilities purportedly measured by the GRE scores are in fact necessary but probably present in sufficient amounts in all applicants. Applicants to graduate schools have already completed at least undergraduate training and at the College of Education of The University of Arizona all doctoral candidates have completed some graduate work. This prior education may have worked as a selective agent eliminating those students who do not have the minimal amounts of the abilities measured by the GRE Tests. Thus those students accepted into the graduate program may have higher amounts of the measured abilities but these amounts may exceed the minimal amounts required.

Assumption 4

The fourth assumption, which is that each department's doctoral program is unique, and thus should establish acceptance criteria based on its own individual needs is

supported in the literature. Madaus and Walsh (1965) found that by computing individual prediction equations for each department they could increase the predictive ability for the GRE Tests if the Departments had a large number of students. They first computed a Multiple R for the entire College between GPA and the two predictors GRE-V and GRE-Q. The resultant Multiple R was .22, significant at the .01 level. They then computed separate Multiple R's for each department and found higher Multiple R's for those departments which had a larger number of students. This procedure suggests that the variance accounted for is different among departments.

The purpose of this study was to examine the second assumption which assumes that each GRE test is unique and provides different information from the other GRE Tests. The previously cited research suggests that in fact the GRE-V and GRE-Q are somewhat unique yielding intercorrelations between .40 and .50. The GRE-AE on the other hand appears to be more highly correlated with the GRE-V and GRE-Q.

Previous research concerning the GRE Tests has dealt with predictive validity and as such has not centered on the intercorrelation between predictor variables. Most of these studies have yielded low predictive coefficients. The most frequent explanation for these low correlations has been that the criterion and predictor variables have restricted ranges and thus yield spuriously low prediction coefficients.

An alternative explanation is that the three GRE Tests may be highly intercorrelated and taken together have little discriminatory power. Discriminatory power is a measure of how much unique variance in the criterion variable is accounted for by the predictor variable. The higher the correlation between two predictor variables the lower the probability that one will account for variance in a criterion variable not accounted for by the other predictor variable. This being the case, if any one of the GRE Tests is highly correlated with either of the other two GRE Tests then it probably does not add discriminatory power to the correlation coefficient obtained using just two GRE Tests.

The research hypothesis was that the two correlations involving the GRE-AE (GRE-AE and GRE-V; GRE-AE and GRE-Q) are higher than the correlation between GRE-V and GRE-Q. The sample chosen for the study consisted of doctoral applicants to the College of Education of The University of Arizona.

CHAPTER II

PROCEDURES

Data for this study were provided by The Office of Graduate Studies in Education of the College of Education from files of 296 doctoral applicants. All applications had been made between 1971 and 1974 and GRE-V, GRE-Q, and GRE-AE scores had been reported by all applicants.

While the files have scores on all applicants who applied since 1960 the sample was restricted to those applicants whose scores were dated from 1971 to 1974 because the Educational Testing Service revises the norms for the test each three years. The scores for foreign born applicants were excluded. The rationale for this exclusion was based on recommendations included in a pamphlet (E.T.S. 1973b).

As was stated previously the Educational Psychology Department allows students to substitute GRE-AP for GRE-AE. Students who substituted the GRE-AP were not included in the study.

Since this study is descriptive in nature there were no independent or dependent variables. Three Pearson product-moment correlation coefficients were computed. The three correlations computed were GRE-V with GRE-Q, GRE-V with GRE-AE, and GRE-Q with GRE-AE. Additional measures of

the relationships (Eta, Eta squared) were computed for all possible pairs of the three tests. Tests of significance, linearity, and homoscedasticity were utilized to examine the relationships.

CHAPTER III

RESULTS

Pearson Product Moment correlation coefficients (r), Eta correlation ratios (E), and their squared values (r^2 , E^2) were computed for each of the six possible combinations of the three GRE scores. There are six possible combinations if one treats each pair as being reversible. For example in the relationship of Verbal and Quantitative scores two correlations exist; Verbal to Quantitative, and Quantitative to Verbal. The r , (and r^2) is the same for each pair, but the respective E , (and E^2) changes according to which score is used as the predictor variable and which is used as the criterion variable.

Table 1 lists the Pearson r , r^2 , E , and E^2 for each possible pair of GRE scores.

Two tests of significance were conducted on the three obtained coefficients. The first tested whether the obtained r was significantly different from 0. All three coefficients were found to be significantly different from 0 at the .01 level. The second significance test examined whether the obtained coefficients were significantly different from each other. The Verbal and Education r was found to be significantly different from both the Verbal and Quantitative r and

Table 1. Pearson r , E , r^2 , and E^2 for the six possible relationships of the Verbal, Quantitative, and Advanced Test in Education.

Predictor	Criterion	r	E	r^2	E^2
Educ.	Verb.	.7879	.8180	.6207	.6692
Verb.	Educ.	.7879	.8547	.6207	.7306
Educ.	Quant.	.5454	.6145	.2974	.3777
Quant.	Educ.	.5454	.6843	.2974	.4684
Verb.	Quant.	.5473	.6481	.2995	.4201
Quant.	Verb.	.5473	.6373	.2995	.4062

the Quantitative and Education r at the .01 level of significance. No statistically significant difference was found between the Verbal and Quantitative r and the Quantitative and Education r .

Each of the six possible relationships of the three GRE scores was tested for linearity. The test for linearity entailed computing an F ratio based on the E^2 and r^2 of each relationship as suggested by Walker and Lev (1953, p. 278). An obtained F which is statistically significant is considered to be an indication that the relationship is not linear.

Table 2 lists the 6 possible relationships, their E^2 , r^2 , F ratio, and the resultant probability.

Table 2. E^2 , r^2 , F ratio, and probability of the six relationships of the three GRE scores.

Predictor	Criterion	E^2	r^2	F
Educ.	Verb.	.6692	.6207	.8797
Verb.	Educ.	.7306	.6207	1.813*
Educ.	Quant.	.3777	.2974	.7742
Quant.	Educ.	.4684	.2974	1.497*
Verb.	Quant.	.4201	.2995	.9243
Quant.	Verb.	.4062	.2995	.8364

* $p < .05$

If a relationship is found to be nonlinear E is considered to be more representative of the true relationship than the Pearson r . A drawback to the usage of an E obtained from a sample is that it is usually higher than the E one would obtain if one could test the entire population.

This inflated E is due to the fact that it is affected by the number of categories of the independent variable as well as the inherent extent of the correlation. In order to get an E independent of the number of categories Kelley (1947, p. 452) developed the formula:

$$E^2_c = 1 - \frac{N-1}{N-K} (1 - E^2_o)$$

In this formula E^2_c is the corrected E^2 , N is the number of subjects in the sample, K is the number of categories in the independent variable, and E^2_o is the E^2 obtained from the sample.

Table 3 lists the 6 relationships with obtained E^2 , E^2 corrected for inflation, and r^2 .

Table 3. Obtained E^2 (E^2_o), E^2 corrected for inflation (E^2_c), and the r^2 of the 6 relationships of the three GRE scores

Predictor	Criterion	E^2_o	E^2_c	r^2
Educ.	Verb.	.6692	.6130	.6207
Verb.	Educ.	.7306	.6686	.6207
Educ.	Quant.	.3777	.2873	.2974
Quant.	Educ.	.4684	.2515	.2974
Verb.	Quant.	.4201	.2873	.2995
Quant.	Verb.	.4062	.2756	.2995

The six possible relationships were tested for homoscedasticity using the Bartlett-Box F Test. A relationship whose probability exceeded .50 was considered to be homoscedastic. The relationship which used the Verbal scores as the predictor variable and the Quantitative score as the criterion variable was the only relationship which

did not meet this criterion. The probability level obtained for this relationship was .167. As a result this relationship was considered to be heteroscedastic.

CHAPTER IV

DISCUSSION

One assumption underlying the requirement of GRE Tests for applicants to doctoral programs is that each of the three tests is unique and measures some different ability. This assumption is not confirmed by the Pearson Product Moment correlation coefficients obtained from this sample. The Pearson r was used for these comparisons in spite of the fact that both relationship involving the GRE-AE were linear in one direction and nonlinear in the other. Pearson r was chosen as the statistic for comparing these relationships as it is the more conservative measure in a nonlinear relationship and is more frequently used in correlational research.

The correlation between the GRE-V and the GRE-AE is .78, suggesting that these tests are not particularly unique and are probably measuring much the same thing. The GRE-Q appears to be unique, as the relationship between it and either of the other two tests was an identical .54. This suggests that either the GRE-V or the GRE-AE could be eliminated from the requirements for applicants. If the two tests are measuring much the same thing then it is redundant

to require both tests. The GRE-AE appears to be the logical choice for elimination for four reasons.

The first reason it should be eliminated is that when the relationships were tested for linearity the two relationships found to be nonlinear had the GRE-AE as the criterion variable. An examination of the data suggests that one possible explanation of this nonlinearity is that the GRE-AE does not spread the scores out very well. There are 80 obtainable scores on the GRE-AE test. The 296 applicants only earned 44 of these obtainable scores. The GRE-V and GRE-Q each have 71 obtainable scores, and the applicants earned 56 and 54 of these scores respectively. If one compares the percentage of obtainable scores earned on each test, one finds that 55% of the obtainable scores on the GRE-AE were obtained while 79% and 76% were obtained on the GRE-V and GRE-Q respectively. Further evidence that the GRE-AE does not spread the scores out comes from an examination of the standard deviations of the respective tests. The GRE-AE scores for this sample had a standard deviation of 90. The GRE-V had a standard deviation of 118. The GRE-Q had a standard deviation of 120. A test of the differences between standard deviations of the tests showed that the standard deviation of the GRE-AE was significantly different from the standard deviations of both the GRE-V and the GRE-Q at the .01 level of significance. There was no significant difference between the standard deviations of

the GRE-V and GRE-Q. Since the standard deviation of the GRE-AE is smaller than the other two tests, it appears that the scores for this test are clustering closer to the mean than the scores on the other two tests.

A second reason for eliminating the GRE-AE rather than the GRE-V is the fact that the GRE-V and the GRE-Q are both administered at the same time. These two tests combine to make up the GRE aptitude test. In order for an applicant to earn a GRE-AE score, he must take an additional test. If a college requires an applicant to submit both the quantitative and an advanced test, they are in effect requiring the submission of all three scores, one of which is redundant (GRE-AE).

The third reason for eliminating the requirement of the GRE-AE is related to the second reason. If applicants are required to take an advanced test, this is an added monetary expense. The GRE-AE costs the applicant \$10.50 and, as stated previously, adds little if any new information.

The final reason for eliminating the requirement of the GRE-AE is the fact that previous literature suggests that the GRE-AE is the poorest predictor of graduate school success of the three GRE Tests. Willingham (1973), in a review of six studies involving predicting graduate school success from the three GRE Tests, found the GRE-AE to be the poorest

predictor. The median validity coefficient for the GRE-AE as a predictor was .24. The median for the GRE-Q as a predictor in these studies was .28, and the GRE-V was found to be the best predictor, with a median coefficient of .36. Since the GRE-V had the highest median validity coefficient and the GRE-AE had the lowest, the GRE-V appears to be the appropriate measure to retain in conjunction with the GRE-Q.

LIST OF REFERENCES

- Educational Testing Service. Description of the Advanced Test in Education. Princeton, New Jersey, 1973a.
- Educational Testing Service. Guide to the Use of the Graduate Record Examinations. Princeton, New Jersey, 1973b.
- Educational Testing Service. Domestic 1975-76 Information Bulletin. Princeton, New Jersey, 1975.
- Kelley, T. L. Fundamentals of Statistics. Cambridge, Massachusetts: Harvard University Press, 1947.
- Madaus, G. F., and Walsh, I. J. Departmental differentials in the predictive validity of the graduate record examination aptitude tests. Educational and Psychological Measurement, 1965, 25, No. 4, pp. 1105-1110.
- Rupiper, O. I. An analysis of the graduate record examination for doctoral majors in education. Peabody Journal of Education, 1959, 36, pp. 279-285.
- Sleeper, M. L. Relationship of scores on the graduate record examination to grade point averages of graduate students in occupational therapy. Educational and Psychological Measurement, 1961, 21, No. 4, pp. 1039-1040.
- Thacher, A. J., and Williams, R. E. The Relationship of the graduate record examination to grade point average and success in graduate school. Educational and Psychological Measurement, 1974, 34, No. 4, pp. 939-943.
- Tully, G. E. Screening applicants for graduate study with the aptitude test of the graduate record examinations. College and University, 1962, 38, pp. 51-60.
- Walker, H. M. and Lev, I. Statistical Inference. New York: Holt, Rinehart and Winston, 1953.

Williams, I. D., Harlow, S. D., and Gab, D. A longitudinal study examining prediction of doctoral success: Grade-point average as criterion, or graduation vs. non-graduation as criterion. Journal of Educational Research, 1970, 64, No. 4, pp. 161-164.

Willingham, W. W. Predicting success in graduate education. Paper presented at the GRE Board Research Seminar, 12th Annual Meeting of the Council of Graduate Schools, November 29, 1972, New Orleans, Louisiana. Papers Presented at the Graduate Record Examinations Board Research Seminar at the 12th Annual Meeting of the Council of Graduate Schools. Educational Testing Service. Princeton, New Jersey, 1973.

