

THE GRADUATE RECORD EXAMINATIONS

GRE

THE PREDICTION OF DOCTORATE ATTAINMENT IN
PSYCHOLOGY, MATHEMATICS AND CHEMISTRY

GRE Board Research Report GREB No. 69-6aR

GRE Board Preliminary Report
(August 1972)

June 1974

This report presents the findings of a
research project funded by and carried
out under the auspices of the Graduate
Record Examinations Board.



EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY □ BERKELEY, CALIFORNIA □ EVANSTON, ILLINOIS

THE PREDICTION OF DOCTORATE ATTAINMENT IN
PSYCHOLOGY, MATHEMATICS AND CHEMISTRY

GRE Board Research Report GREB No. 69-6aR

GRE Board Preliminary Report
(August 1972)

June 1974

The Prediction of Doctorate Attainment in
Psychology, Mathematics and Chemistry

Abstract

The GRE Board-sponsored project utilized data from the National Science Foundation Fellowship applicant records and the NRC Office of Scientific Personnel Doctorate Records File to evaluate the potential of GRE Aptitude and Advanced Tests as predictors of whether or not the candidate attained the doctorate within a period of from seven to ten years. In addition, the study sought to determine whether there were particular subgroups within each field as described by variables such as age, "quality" of the institution or graduate department, for which the GRE tests have varying degrees of predictive accuracy.

Sample sizes ranging from 643 to 779 were obtained for the three fields and divided into two samples so that cross-validation could be performed.

The results of the study by field are summarized as follows:

Psychology - Of the predictor tests the GRE Advanced Psychology Test had the most consistent relationship with the criterion. Undergraduate GPA had a surprisingly low predictive validity with Ph.D. attainment. Sex had a strong relationship to the criterion; women were less likely to attain the doctorate in psychology than men. Age level provided a basis for defining more and less predictable groups. A "U" shaped relationship existed with younger and older groups being more predictable than the middle or 25- and 26-year old applicants. There is a slight tendency for students attending "lower quality" psychology departments, as defined by Cartter (1964) to be more predictable.

Mathematics - The criterion was generally more predictable for mathematics than for psychology. The GRE Advanced Mathematics Test was the single best predictor with correlations of .38 and .44 for the two samples. GRE Verbal and Quantitative followed in order of magnitude. It may be that the successful completion of the Ph.D. Program in mathematics depends upon the assimilation of a relatively structured body of knowledge which in turn leads to more accurate assessment of any one individual. There was little or no consistently different prediction for groups defined by age or by departmental quality indices. When age and departmental quality were combined, however, the young who attend "lower quality" departments appear to be more predictable than the remaining groups.

Chemistry - The validity coefficients for chemistry were similar in level and pattern to those of mathematics. Again, the GRE Advanced Test was the single best predictor in both samples. Correlations of GRE Verbal and Quantitative, while lower than for mathematics, were still reasonably strong. Age, when included as a predictor, added significantly to the prediction, contrary to the case in psychology and mathematics. There was little or no consistently different prediction for subgroups based on age, "quality" indices or both together.

THE PREDICTION OF DOCTORATE ATTAINMENT IN
PSYCHOLOGY, MATHEMATICS AND CHEMISTRY

Introduction

Researchers seeking to demonstrate the validity of test scores for predicting graduate school performance have encountered a number of operational as well as logical difficulties. Reilly (1971) lists three major difficulties. These are: (1) the small samples usually available which in turn lead to unstable estimates of the parameters, (2) homogeneity of the sample itself due to previous selection with respect to ability and achievement variables, and (3) the establishment of an adequate criterion. Graduate grade point average (GPA), while being the more accessible criterion, has also been the most severely criticized. Lannholm et al, (1968) probably levels the most valid and serious criticism of GPA when he concludes that it represents only a limited aspect of graduate school performance. It is also subject to an understandable unwillingness on the part of faculty to discriminate among individuals all of whom are members of a highly selected population.

The most desirable criterion, of course, would be some measure of achievement as a scholar or scientist. Aside from the logical difficulties in arriving at any sort of agreement as to what is a relevant measure of scientific achievement, we are faced with the operational problem of time lapse which must occur before such data can be collected. An alternative criterion of a more intermediate nature is whether or not one has attained his or her doctorate within a reasonable period of time. Attainment versus non-attainment of the doctorate is appealing on logical grounds since: (1) it is one test of the effectiveness of the overall selection process, i.e., the decision to admit a student to graduate education or to admit him to candidacy for a higher degree implies an expectation that his formal graduate education will be completed. The attainment of the doctorate degree is the primary indicator that such an expectation has been fulfilled; and (2) more often than not attainment of the doctorate is a necessary pre-requisite to gaining entry into the scientific-academic arena.

From an operational viewpoint doctorate attainment is readily quantifiable, and, of course requires less time to mature than "on-the-job" measures of effectiveness. One criticism, however, is that it lacks sensitivity in the sense that it cannot take into account the various qualitative levels of performance among individuals attaining or not attaining the Ph.D. Although the latter criticism may well be valid, it was felt that the ease of quantification, and availability were sufficiently compelling reasons for its use in this study. It was also felt that if it was sufficiently lacking in sensitivity, this in turn would be reflected in the relative level of its predictability.

The immediate focus of this research project was to evaluate the potential of GRE aptitude and advanced tests as predictors of a dichotomous criterion of whether or not the candidate attained the doctorate within a specified length of time. More specifically, the project attempted to: (1) define subgroups for which the GRE tests have varying degrees of predictive accuracy, and (2) provide biographical profiles of each of these subgroups.

Methodology

Approximately 1,000 National Science Foundation (NSF) applicant records were collected within each of the areas of Psychology, Chemistry and Mathematics from the merging of the National Research Council Office of Scientific Personnel (OSP), Doctorate Records File, and the National Science Foundation Fellowship applicant file. These file records indicated time to Ph.D. for all those who attain the Ph.D. Additional biographical information available in the Doctorate Records File and on the OSP tape included sex, age, marital status, number of dependents, number of NSF applications made and awards received. The OSP records also provided Office of Education Codes for the institution each applicant had chosen for graduate study.

Predictor information available from the OSP records included the GRE test scores, verbal, quantitative and advanced as well as undergraduate grade point average and reference report average. The reference report average (Harmon, 1966) was a quantification of an overall rating of the reference letters submitted on behalf of an NSF applicant.

The criterion of doctorate attainment required a judgment to be made concerning the time lapse to be allowed before assigning an individual to the attainment versus non-attainment category. It is, of course, rare that one completes a doctorate within three years after

the baccalaureate. In the science fields the mean time lapse is approximately eight years (Creager, 1965), with greater deviations above the mean than below. If time were allowed for almost everyone to complete a doctorate, the study might well suffer on both operational and rational grounds. That is, not only would more of the people attaining doctorates have more time out and extensive study time (thus complicating the interpretation), but more persons of low measured ability would have achieved a doctorate under possibly lower standards of dissertation and course quality. From the viewpoint of efficient use of resources as well as cost of graduate education, it would seem to be desirable to select those individuals capable of successfully finishing the program in a reasonable amount of time. Conversely, too short a time lapse would eliminate many high quality people, possibly those very able persons who take on more ambitious dissertation projects and/or more difficult course offerings.

These considerations lead, for criterion definition purposes, to setting limited cutoff times for doctorate completion. The doctorate completion cutoff was June, 1968. Since most of the subjects included applied for first-year fellowships in 1958-1961, they had 7 to 10 years from fellowship application time to attain the doctorate.

Cartter's (1964) report on the quality of graduate departments furnished the quality indices which were then assigned to each candidate according to the ranking of the department which he attended. Additional institutional quality information was also collected from an Office of Education tape which included such information as: (1) proportion of faculty with doctorate, (2) per student expenditure, (3) number of books in the library, (4) income per student, and (5) student/faculty ratios. These particular "quality indices" suffer from the fact that they apply to the total institution and thus are not necessarily an accurate picture of the graduate school or more specifically the graduate department itself.

Within each major field the sample was split into two random halves for validation and cross-validation purposes; that is, any relationships found in the validation sample could then be examined on the second sample (cross-validation sample) to see if the findings were indeed replicable. The data were then analyzed using the moderated regression technique (Rock, et al, 1972). This technique not only furnishes the researcher with the usual multiple prediction validity information, but also provides a systematic means of searching for consistent biographical patterns associated with "types" of individuals who, in turn, may be characterized by varying levels of

predictability. For example, this type of analysis enables one to determine if any one subgroup such as older NSF applicants may be more or less predictable than the younger ones. If the moderated regression technique were used with two possible classification variables such as age and department quality index, it might, for example, identify a group of older individuals attending lower quality graduate departments who are unpredictable with respect to Ph.D. attainment when GRE test scores were used as predictors. Since this technique requires complete information, the sample sizes were reduced to 779^a, 845, and 643 for Psychology, Mathematics and Chemistry, respectively.

Potential moderators which were analyzed with respect to their impact on accuracy of prediction were age, sex, marital status, university quality indices, and graduate department quality indices. Students attending the same institution were assigned that particular institution's quality ratings as well as department rating.

Results and Discussion

Psychology

Among the predictor tests, the GRE Advanced Psychology Test appears to have the most consistent relationship with the criterion when considered across both samples. Undergraduate grade point average had a surprisingly low predictive validity with respect to Ph.D. attainment. Sex proved to be a good predictor of Ph.D. attainment. As indicated in Table 1, sex has the highest single variable correlation with the criterion (-.45 in Sample 1, and -.34 in Sample 2) among all the potential predictors or moderators. The negative sign indicated that women are less likely than men to attain the doctorate in Psychology. Further inspection of Table 1 indicates that the GRE Verbal and Quantitative, one college quality index (department rating), reference average, and number of NSF applications have consistent (non-zero) relationships in both samples with the criterion. The department rating's relationship with the criterion carries a negative sign, since the quality code indices range from 1-4 with one signifying the highest quality and

^aAnalyses in Psychology were based on a total of 930 observations when department quality indices were not part of the analysis. This was due to the fact that a substantial number of cases had to be dropped when the quality indices were included.

TABLE 1

**O-ORDER VALIDITY COEFFICIENTS, MEANS, AND STANDARD
DEVIATIONS OF THE PREDICTORS FOR SAMPLES 1 AND 2
(PSYCHOLOGY)**

	Sample 1 N=380			Sample 2 N=398		
	r	\bar{X}	σ	r	\bar{X}	σ
Criterion	-	1.40	.49	-	1.39	.49
Sex	-.45*	1.41*	.49*	-.34	1.34	.48
Age	-.05	23.37	3.70	-.03	23.68	4.15
No. of Books	.06	9.87	.57	.02	9.85	.67
Income/Student	.15	8.75	1.77	.05	8.71	1.97
Faculty/Student	.13	6.36	3.04	.07	6.45	2.98
Percent with Ph.D.	-.02	7.19	3.06	.01	7.13	3.11
Dept. Rating	-.14	1.67	1.14	-.16	1.66	1.29
GRE-V	.12	63.52	8.25	.19	63.47	9.29
GRE-Q	.33	59.89	11.34	.14	60.96	10.82
GRE-Adv.	.19	60.98	8.90	.24	60.87	9.05
Ref. Average	.16	43.86	8.36	.14	43.93	8.49
UGGPA	.02*	241.70*	40.10*	.02	236.78	42.96
App. Made	.31*	1.63*	1.11*	.33	1.64	1.10

* N=462

four the poorest quality. The remaining institution quality indices appear to be too general and thus do not necessarily reflect the quality of the Psychology departments. The correlations between the department rating and the college quality indices range from a low of $-.13$ for percentage of faculty with the Ph.D., to a high of $-.51$ with income per student, indicating a large proportion of the variation in the department rating is not accounted for by the more general college indices.

The relatively high correlation between the number of NSF applications made and Ph.D. attainment is somewhat artifactual, since a large percentage of the NSF applicants in this study were required to reapply for their grant every year. Many of those students who did not reapply may have either dropped out of the program or possibly felt that their past performance record would not be supportive of a grant extension. Thus, applications made may be considered an intermediate progress report on the way to the Ph.D. in Psychology. See Appendix A for list of variable definitions.

Of the biographical data for the Psychology students, only age level led to a consistent pattern of differential predictability, that is, the pattern from Sample 1 was replicated in Sample 2. It is interesting to note that there is a "U" shaped relationship between age and predictability. That is, the relatively young and the relatively older groups were considerably more predictable than the 25- and 26-year old applicants. Although the oldest group was the most predictable, they had the smallest probability of getting the doctorate. That is, while almost 50% of each of the remaining age groups did obtain the doctorate within the specified time, only 28% of the older group did likewise. The mean on the predictor scores for the various age groups indicated that both the older and the "middle" age groups had similar means, both of which are consistently lower than the youngest group. Thus the "middle" age group (25- and 26-year olds) consistently produce a greater proportion of graduates than either the younger or the older groups. Since the "middle" age group tends to have lower predictor scores on the average, yet possesses the highest level of Ph.D. attainment, they are generally under-predicted if the usual prediction equation were used. Thus they are what is commonly referred to as over-achievers in the psychometric literature.

It may well be that the 25- and 26-year olds have overcome their somewhat mediocre ability-achievement credentials by a higher level of motivation and consequently have a higher rate of Ph.D. attainment. Unfortunately, we do not have the data to determine what, if

any, other age-related characteristics are operating here. These findings of differential predictive accuracy, as well as possible motivational differences point out the need for more biographical information about graduate applicants, if we are to understand and/or infer the causal pattern underlying their differential performance.

As in the case of the biographical variables, only one quality index led to a replicable pattern of differential predictability. That is, there is a slight but seemingly consistent tendency for students attending "lower quality" Psychology departments to be more predictable. This result certainly comes as no surprise, since the so-called "higher quality" schools are more selective of applicants with respect to the GRE test scores and thus attainment of the Ph.D. is likely to depend on some unmeasured quality. It is, however, interesting to note that at the "lower quality" Psychology departments, the probability of obtaining the Ph.D. is consistently less than at the "higher quality" departments.

When grouping of students was done on both age and department quality index, the pattern of predictability is less clear-cut. There remained a tendency for the older students attending "lower quality" institutions to be more predictable. In this four-way break-out, the sample sizes are rather small and the resulting instability of the parameter estimates makes any further interpretations of these results rather tenuous.

In order to determine the utility of age and departmental quality as potential predictors, they were incorporated into prediction equations along with the usual predictors. In no form did they consistently lead to an increment in prediction over the original five predictors (GRE-verbal, quantitative, advanced, UGGPA, and reference report average).

It would appear that for NSF applicants in Psychology, the utility of age information lies primarily in separating out those individuals for which: (1) we have varying degrees of confidence in their predicted or expected achievement, in this case Ph.D. attainment, and (2) motivational levels may differ.

The results also suggest that where there was differential prediction, the overall equation used within the groups was not noticeably inferior to the unique group equation with respect to predictive

accuracy. This suggests that different weightings of the same predictive variables for different types of people (older versus younger, etc.) does not appear to be the answer. That is, some individuals seemed to be more or less predictable regardless of whether you use overall weights or their own unique weights. It is possible that entirely different predictor measures must be developed for the "unpredictable" people. This, of course, is beyond the scope of this study.

Mathematics

Table 2 presents the single variable validity coefficients for the predictors and potential moderators of grouping variables. In general it appears that the criterion of Ph.D. attainment in Mathematics is considerably more predictable from achievement-aptitude measures than was found to be the case in Psychology. Of particular interest in Table 2 are the correlations of .38 and .44 for the Advanced Mathematics Test against the criterion for Sample 1 and 2, respectively. The GRE verbal and quantitative as well as undergraduate grade point average have respectable although lower relationships with the criterion. Institutional quality indices such as income/student, student/faculty, and departmental quality index, also demonstrate stronger relationships with Ph.D. attainment in Mathematics. It may well be that the successful completion of the Ph.D. program in Mathematics depends upon the assimilation of a relatively structured body of knowledge which in turn leads to more accurate assessments of any one individual's standing with respect to this body of knowledge.

The multiple correlation between the five predictors (GRE verbal, quantitative, advanced, reference reports and undergraduate grade point average) and Ph.D. attainment is a quite respectable .40 in Sample 1 and cross-validates to a surprising .44 in Sample 2 indicating relatively accurate prediction. Unlike Psychology, there was little or no consistent differential prediction by age group. It was also found that the older the NSF applicant, the less likely he is to attain his doctorate within the cutoff time of this study. As in Psychology, the "middle" and "older" NSF applicants had similar aptitude-achievement test scores, and when considered as a whole, had consistently lower test scores than the younger candidates. The one exception to the above findings was the advanced test where the "older" NSF candidates were not only lower than the younger candidates, but were also one-half a standard deviation below the "middle" age candidates.

When groups of applicants were formed based on departmental quality indices of the institutions which they attended, no consistent pattern of differential predictability was found. However, when groups were formed based on both department quality index and age, the young who attend "lower quality" departments appear to be characterized by greater predictability than the remaining groups. In general, the mean ability-achievement scores for this group was below that of both the "high

TABLE 2

0-ORDER VALIDITY COEFFICIENTS, MEANS AND STANDARD
 DEVIATIONS OF THE PREDICTORS FOR SAMPLES 1 AND 2
 (MATHEMATICS)

	Sample 1 N=423			Sample 2 N=422		
	r	\bar{X}	σ	r	\bar{X}	σ
Criterion	-	1.54	.50	-	1.54	.50
Age	-.17	22.35	2.06	-.18	22.71	2.50
No. of Books	.04	9.92	.40	.10	9.90	.48
Income/Student	.20	9.35	1.35	.17	9.24	1.53
Faculty/Student	.15	7.98	2.67	.13	7.57	2.88
Percent with Ph.D.	.06	7.55	2.88	.13	7.61	2.86
Dept. Rating	-.18	1.60	1.09	-.29	1.65	1.13
GRE-V	.27	62.95	10.96	.32	62.63	11.33
GRE-Q	.27	72.67	9.51	.26	71.54	10.14
GRE-Adv.	.38	65.93	15.39	.44	64.93	15.94
Ref. Average	.23	42.60	9.38	.27	42.59	9.69
UGGPA	.21	252.60	40.22	.24	248.77	43.13
App. Made	.52	2.32	1.52	.50	2.37	1.61

quality" young and the "high quality" old but slightly above those of the "low quality" old group. Because the "low quality" young group-size is so small, any further interpretation is probably unwarranted. As one might also expect, the findings indicated that the young applicants who attend institutions with "high quality" departments are much more likely to attain the doctorate than are the older NSF candidates who attend institutions characterized by "low quality" Mathematics departments. When age was included as a predictor, no increment was found in predictive accuracy above that which resulted from the use of the original five predictors.

Chemistry

The single variable validity coefficients for the chemistry measures are similar both in level and pattern to those of the Mathematics NSF applicants. As in Mathematics, the GRE Advanced Test is the one best predictor in both samples. However, among the Chemistry NSF applicants' undergraduate average, reference report average, and age demonstrate somewhat higher relationships with Ph.D. attainment than do their counterparts for the Mathematics applicants. In general, the level of correlations found in Chemistry yield additional support for the hypothesis that the so-called "hard sciences" may provide a more measurable domain with respect to criteria of success as well as measures of past achievements or aptitudes. It is also quite possible that it is easier to specify the necessary skills which are prerequisite to success (Ph.D. attainment in this case) in the "hard sciences."

As in the case of Mathematics, differential predictability by age groups was not found. Prediction for Sample 1 is relatively strong considering the somewhat restricted nature of the NSF applicant sample. Surprisingly enough the cross-validated multiple correlations increased from .39 in Sample 1 to .53 in Sample 2.

A considerably larger proportion of the NSF applicants in Chemistry do attain the Ph.D. than in Psychology and Mathematics NSF applicants.

As in Mathematics, when groups were formed based on the rated "quality" of their Chemistry departments, there was little or no consistent differential prediction across groups. When groups were formed on both age and quality indices, still no consistent pattern of differential predictive accuracy was evident. It appears that in the two "hard science" areas of Mathematics and Chemistry, the assimilation of knowledge in their particular area as measured by the advanced section of GRE is the one best predictor of Ph.D. attainment regardless of age group membership or quality of the institution of attendance.

TABLE 3

O-ORDER VALIDITY COEFFICIENTS, MEANS, AND STANDARD
 DEVIATIONS OF THE PREDICTORS FOR SAMPLES 1 AND 2
 (CHEMISTRY)

	Sample 1 N=322			Sample 2 N=321		
	r	\bar{X}	σ	r	\bar{X}	σ
Criterion	-	1.75	.43	-	1.74	.44
Age	-.29	22.10	1.81	-.28	22.26	1.82
No. of Books	.08	9.82	.59	.01	9.84	.55
Income/Student	.12	9.44	1.12	.14	9.24	1.34
Faculty/Student	.08	7.89	2.66	.04	7.37	2.89
Percent with Ph.D.	-.05	7.11	2.99	.12	7.07	3.12
Dept. Rating	-.14	1.36	.79	-.22	1.39	.86
GRE-V	.15	59.66	10.69	.23	58.40	10.75
GRE-Q	.28	69.26	10.70	.34	67.96	10.70
GRE-Adv.	.33	67.41	11.81	.48	66.27	12.31
Ref. Average	.30	41.48	9.89	.33	42.23	9.15
UGGPA	.27	246.93	44.55	.36	247.93	42.82
App. Made	.42	2.43	1.48	.39	2.34	1.48

Age was included as a predictor and unlike Mathematics or Psychology, it did add significantly to the prediction. It was the second variable after the GRE advanced section to enter the equation. In an effort to gain some insight into this relationship, the correlation between age and whether or not the student attended on a part-time basis was examined. This correlation was effectively zero (.02). Thus, the "older" students in Chemistry are no more likely to attend on a part-time basis than the other age groups.

The significant partial regression weight associated with age indicates that after the ability-achievement variables were controlled, there remained a significant amount of variance in age which was related to Ph.D. attainment. It would appear that additional biographical information might prove helpful in untangling this relationship.

Conclusions

It was found that the GRE advanced tests were consistently the best predictors of a criterion of Ph.D. attainment. However, the predictive accuracy of the GRE advanced test varied considerably across graduate fields and in one case within a graduate field. That is, prediction on the whole was considerably more accurate in the "hard science" graduate areas of Mathematics and Chemistry than in Psychology. Within the psychology area there was a "U" shaped relationship between predictability and age. That is, the total sample prediction equation led to greater predictive accuracy for the "younger" and the "older" age groups. The "middle" age group was not only less predictable, but the errors in prediction tend to lead to underestimation of their actual rate of Ph.D. attainment. Thus, the "middle" age group was characterized by over-achievement.

Dr. Harmon participated in this project as a consultant to Educational Testing Service, not as a representative of the National Research Council.

References

- Cartter, A. M. (Ed.) American Universities and Colleges. (9th ed.) Washington, D.C.: American Council on Education, 1964.
- Creager, J. A. Predicting Doctorate Attainment with GRE and Other Variables. Technical Report No. 25, 16 November 1965. Research on Fellowship Selection Techniques Supported by the National Science Foundation. Office of Scientific Personnel, National Academy of Sciences--National Research Council, Washington, D.C.
- Harmon, Lindsey R. Fourteen Years of Research on Fellowship Selection. A summary. Progress report on research supported by the National Science Foundation. July 20, 1966. Office of Scientific Personnel, National Academy of Sciences, National Research Council. Publication 1420. Washington: National Academy of Sciences, National Research Council, 1966.
- Lannholm, G. V., Marco, G. L., & Schrader, W. B. Cooperative Studies of Predicting Graduate School Success. Graduate Record Examinations Special Report 68-3. Princeton, New Jersey: Educational Testing Service, 1968.
- Reilly, R.R. Critical Incidents of Graduate Student Performance. Graduate Record Examinations Board Technical Memorandum No. 1. Princeton, New Jersey: Educational Testing Service, 1971.
- Rock, D. A., Barone, J., & Linn, R. L. A moderated stepwise prediction system. Educational and Psychological Measurement, 1967 27, 709-713.

APPENDIX A

Variable Definition

Criterion	Coded a 2 if Ph.D. received; a 1 if not received
Sex	Coded 2 for women; a 1 for men
No. of Books	Decile rating with 10 highest
Income/Students	" " " " "
Students/Faculty	" " " " "
Percent with Ph.D.	" " " " "
Departmental Ratings	On a four-point scale with 1 highest, 4 lowest
GRE V, Q, and Advanced	Two-digit GRE score with third digit of GRE score dropped
Undergraduate Grade Point Average (UPGA)	On a four-point scale multiplied by 100
Reference Average	Zero to six multiplied by ten
Applications Made	Count of number of applicants