DOCUMENT RESUME

ED 079 365 TM 002 964

AUTHOR Willingham, Warren W.
TITLE Predicting Success in Graduate Education.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE [73]
NOTE 23p.; Revised version of paper presented at a Council of Graduate Schools meeting, November, 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Academic Achievement; *Admission Criteria; *Competitive Selection; *Graduate Study; Higher Education; Literature Reviews; *Prediction; Predictive Ability (Testing); Predictive Measurement; Predictor Variables

ABSTRACT

The available objective evidence suggests that the accuracy of predicting which students will succeed in a particular graduate school is often no better than modest, especially if such predictions are based only upon a test or a grade record. Taken together these two types of predictors do a reasonably good job, considering the restricted range of ability involved. The best way to improve selection of graduate students will be to develop improved criteria of success. This is no small job for graduate faculties, but it carries the promise of more effective utilization of talent and greater assurance of equity in admitting students to advanced levels of training and the privilege associated with such programs. (Author)

FILMED FROM BEST AVAILABLE COPY

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 079365

TM 002 964

Predicting Success in Graduate Education

Warren W. Willingham*

Educational Testing Service

In recent decades graduate schools have assumed a major responsibility for the advanced training of a talented segment of American society. Compared with lower forms of schooling, most graduate programs are costly as well as intellectually demanding. Students who complete these programs feed the professions and academic disciplines and constitute a critical national resource. Traditionally most graduate students have been selected with great care but until the past decade or so there were relatively few formal statistical studies of that selection process. Such investigations are now common.

There are several possible explanations for recent interest in prediction studies of success in graduate education. In earlier times space in graduate schools and the number of applicants were in a rough equilibrium but burgeoning applicant groups in the fifties and sixties focused attention on how some were selected and others turned away. These larger numbers of students permitted statistical studies in many departments which previously had too few students to make this type of systematic evaluation worthwhile. Finally, increasing use of selection tests (Graduate Record candidates increased from 100,000 to 280,000 during the 1960's) suggested the prediction studies with which similar tests are closely associated at the undergraduate level. The purpose of this report is to summarize the results of the substantial number of such studies that have accumulated, to suggest some practical implications for selecting graduate students, and to indicate further research needed on the topic.

Correlational analysis is the principal research design for evaluating the selection process. One or more predictors (measures of student potential) are evaluated with respect to the extent to which they accurately forecast one or more criteria (measures of student success typically taken after a year or more in graduate school). The value of a predictor for selecting students varies directly with the size of its correlation with the criterion (Cronbach, 1971). This correlation, called a validity coefficient, ranges from a chance relationship of .00 to a perfect

*An earlier version of this paper was presented at a presession of the Council of Graduate Schools meeting in November, 1972. The author expresses appreciation to Jane Porter for assistance in compiling the data on which this report is based.

relationship of 1.00, though negative coefficients can occur and perfect validity is not closely approached in practice. Usually more than one predictor is involved (e.g., a test and a grade average) and in such cases a statistically weighted composite of the predictors is typically more useful for selection purposes than either predictor alone.

There are a variety of measures that can be used as predictors of success; there are also various measures that can serve as criteria after admission to graduate school. None are entirely satisfactory. Any predictor or criterion should have reasonable construct validity, reliability, and acceptability. By construct validity we mean that the predictor or criterion should be relevant to what we intend to measure. More specifically, it should represent what we want to measure, all we want to measure, and nothing but that which we want to measure (Thorndike and Hagen, 1969). By reliability we mean that a measure provides a stable estimate from one measuring occasion to another and is free from distortion. By acceptability, we mean that a measure is economically feasible, administratively practical and socially ethical. It is in this context of construct validity, reliability, and acceptability that we can review briefly some of the strengths and weaknesses of predictors and criteria before addressing the empirical and utilitarian question of the relationship between the two.

Predictors

Undergraduate Grade Point Average. The student's undergraduate average has obvious relevance as a predictor because it represents the same sort of behavior one is trying to forecast. It is a matter of psychometric as well as everyday experience that past behavior is usually the best predictor of future behavior. Undergraduate GPA is readily available, widely assumed to be fair and equitable, and almost universally used by graduate departments in selecting students. The measure has two important weaknesses. It often has a narrow range - from 3.0 to 4.0 in many departmental candidate groups and thus doesn't differentiate applicants very well. Also the meaning of a B average varies considerably from one undergraduate college to another.

Recommendations. References from undergraduate professors are widely used despite the fact that they are time consuming to prepare and sometimes difficult to quantify or even interpret. Recommendations can be highly relevant, particularly in the sense that an informed person can judge a student's suitability for a particular graduate program. In many situations the Achilles heel of personal references is the unreliability or lack of comparability among judges.

Test Scores. A principal advantage of standardized tests is their seventy year history of research and theoretical development. On the one hand this work has produced reliable, standard measures highly suitable for national administration under security conditions. More importantly, this work has developed detailed conceptual frameworks of human abilities and established relationships between underlying abilities and socially valued observable behavior such as scholastic competence. Due to this psychometric development it is possible to construct tests to measure any of an extremely wide range of human abilities. A corresponding weakness of most individual tests, however, is their tendency to focus attention on fairly limited aspects of competency. Another weakness is the lingering suspicion (despite substantial evidence to the contrary; see Linn, 1973, Stanley, 1971) that standardized tests are intrinsically biased against individuals from cultural minorities.

Biographical Information. Various characteristics of an applicant's background are used implicitly if not formally when graduate departments select students. Both the strengths and weaknesses of background information lie in its construct validity or relevance to success in the graduate program. Special accomplishments or experience of students can be highly relevant. Characteristics such as age, sex, or race may be quite irrelevant but nonetheless used for legitimate social or administrative reasons. On the other hand some particular characteristic of applicants may be easy to collect but treacherous to use in selection decisions if there is no logical and defensible explanation for an observed relationship between that characteristic and success in graduate school.

Criteria

Graduate Grade Point Average. The grades a student makes in graduate school are a readily available and certainly relevant indication of success. But traditionally, grades in many graduate schools have consisted largely of A's and B's. Thus the range is so narrow that differences among GPA's do not usually represent reliable differences in student accomplishment. Furthermore, many faculty members doubt even that reliable grades represent the most important outcomes of graduate education.

Comprehensive Examination. Many departments require students to pass a qualifying or comprehensive examination at some point during their graduate program. In theory, a properly constructed examination could provide the most reliable and valid criterion of subject competence. In practice such examinations likely vary widely in quality. In any event an objective examination should not serve as a sole criterion since it measures a limited aspect of success. Furthermore it suggests a logical circularity when test scores are used mainly to predict test scores.

Faculty Ratings. A principal advantage of collective faculty judgment is versatility in measuring important aspects of graduate success other than knowledge of the subject field. Faculty who know a student well are in the best position to say whether he is able to execute independent research or motivate a class of undergraduates. A weakness is the fact that many faculty ratings are unreliable and not carefully designed to represent observable outcomes of graduate training.

Attain Ph. D. Regardless of what other judgments a faculty may make of a doctoral student the carefully considered acid test is whether he or she is granted the degree. Consequently, this is probably the single most defensible and relevant criterion of success at the Ph. D. level. On the negative side, one must wait a long time for this criterion. In fact years lapsed between BA and Ph. D. attainment is a corollary criterion used in some studies. Another difficulty is the fact that whether a student graduates may frequently depend upon extraneous influences rather than demonstrated competence. In any event this criterion places a premium on academic persistence and probably does not differentiate very well the most promising scholars and professionals.

Nature of the Data

By far the most common predictors used in studies of success in graduate school are undergraduate average and Graduate Record Examinations scores. This review is based upon correlation studies using these variables that were cited by Lannholm (1968, 1972) or located through searches of appropriate journals and abstracts. Forty-three studies were found for the period 1952-1972 though about half were dated during the last five years of that span (see list appended). Half of these studies were published; the remainder were institutional reports or theses.

The 43 studies included 138 independent sets of data, usually corresponding to departments though occasionally representing some broader group such as first year students across several departments. Individual sets of data were based upon 20 to 1479 students (Median N 80). The total number of students included in all studies was 21,214. The total number of validity coefficients was 616. These coefficients are summarized in Table 1.

The first two predictors in Table 1 refer to the Verbal and Quantitative sections of the GRE Aptitude Test. The GRE Advanced Test evaluates achievement in the student's chosen field; thus the content varies, depending upon the department involved. The fourth predictor varied somewhat from study to study. It was usually the average of two or three GRE scores though this composite was occasionally weighted statistically.

The undergraduate GPA was undoubtedly computed in various ways in different studies but seldom specified very carefully. The data concerning recommendations came almost exclusively from three extensive studies of National Science Foundation fellowship applicants (Creager, 1965; Rock and Harmon, 1972) and represent the average rating of several letters of reference.

With respect to criteria of success, the exact nature of faculty ratings varied from study to study but typically represented the composite judgments of several faculty members concerning professional promise or overall success as a graduate student. Very few studies reported validity data with departmental exams as the criterion. "Attain Ph. D." typically means attaining the degree within a certain number of years, so a time element is also involved. That factor is formalized in the "time to Ph. D." criterion by assigning criterion scores to students according to years elapsed between BA and Ph. D. All of the data concerning this last criterion comes from two very large studies by Creager (1965).

Predictability of Graduate Success

The studies represented in Table 1 vary widely in quality and scope. Some are based on small samples, making individual correlations unreliable. But those medians based on more than just a few coefficients should give a dependable idea of how valid these predictors are and how predictable are the various criteria of graduate success. Insofar as possible the same data have been sorted by major field and presented in Table 2 to illustrate differential validity of the predictors for different disciplines. Several observations can be made from these tables.

Validity coefficients for the various predictors and composites (against the GPA criterion) tend to be about .15 lower than corresponding median coefficients at the undergraduate level (Fishman and Pasanella, 1960).

The undergraduate GPA is a moderately good predictor of graduate GPA and faculty ratings; it is a poor predictor of whether a student will attain the Ph. D. Depending upon the success criterion used, the GRE composite is either slightly more valid or substantially more valid than the undergraduate GPA.

The GRE Quantitative is typically a better predictor in those scientific fields where quantitative ability counts most. The reversal in the field of mathematics may be due to restriction in the range of quantitative scores because of heavy emphasis on this variable in selection.

Correspondingly, the GRE Verbal tends to be more valid in verbally oriented disciplines such as English and education. Otherwise the pattern of validity coefficients is fairly similar from one discipline to the next.

The GRE Advanced is evidently the most generally valid predictor among those included. In seven of the nine disciplines in Table 2 it has the highest validity among the three GRE scores. In eight of the nine fields it has higher validity than undergraduate GPA.

Recommendations appear to be a fairly poor predictor of whether a student will successfully complete a doctoral program.

The comprehensive departmental examination seems a somewhat more predictable criterion than the others examined here. This is an uncertain conclusion because the available data are sparse but the conclusion is consistent with the reasonable assumption that such a criterion should be more reliable than the others represented.

A weighted composite including undergraduate GPA and one or more GRE scores typically provides a validity coefficient in the .40-.45 range for various criteria of success and for different academic fields. This is somewhat higher than the validity of GRE alone. The composite of undergraduate GPA and GRE provides substantially more accurate prediction than does undergraduate GPA alone. This is the case for each success criterion and practically every academic discipline.

The Utility of Current Predictions

What overall evaluation can be made of the extent to which success in graduate school is predictable? Cronbach (1971) describes the following considerations in determining the utility of a predictor for selection purposes. First, the utility of the predictor is directly related to the size of the correlation coefficient. Thus Table 2 indicates that in most fields the value of the GRE-GPA composite prediction amounts to about 40% of the benefit that could accrue if prediction were perfect. How useful that level of validity is in practical terms depends upon the cost of gathering the predictor information and two other considerations. A small correlation can produce a large benefit if the proportion of students selected is low. Finally, a given validity coefficient will have more practical value if the selection decision is important, and the selection decision is more important if it is irreversible.

We might say that a validity coefficient of .20 is modest and one of .40 moderate. The conditions of graduate student selection are generally favorable to using predictors of even modest validity. In many departments only a small proportion of students are accepted; the decisions are quite important to the student and to broader interests; and the decisions are typically irreversible. There seems little doubt that the GRE and the undergraduate GPA are providing quite useful information in most situations.

Figure 1 illustrates graphically the level of benefit likely to accrue from using predictors that are valid to the extent indicated. Students at high ability levels were far more likely to attain the Ph.D. than those at low levels. The figure also illustrates that many students fail to attain the degree, even among talented NSF fellowship applicants. And in these samples reported by Creager (1965) there were substantial differences in attainment rates among fields (Chemistry 51%, Physics 36%, Psychology 26%).

It should be emphasized also that validity studies at particular schools and departments give varying results. Such variability is exacerbated by the small samples often used, but real variations do occur. It is important to undertake local studies in order to justify selection procedures and utilize available information to maximum benefit.

Can Predictions be Improved?

What are the prospects of improving prediction of graduate success? One cause for pessimism is the very restricted range of talent involved. Many of the studies summarized here are based upon highly selective departments or groups like NSF fellowship applicants. For this reason alone one would expect substantially lower validity coefficients than are typical at the undergraduate level. Consequently, it does not follow that the predictors are inherently any less valid. The GRE aptitude test, for example, is basically similar to the less difficult and "more valid" Scholastic Aptitude Test. Judging from considerable research at the undergraduate level it seems unlikely that other types of aptitudes can enhance prediction to any significant degree.

The undergraduate GPA suffers similar shortcomings as does the high school average in predicting success at the next educational level. The range of the grade average is greatly restricted by selection and the grade scale varies considerably depending upon the origin of the student. There have been many efforts to develop both simple and highly sophisticated methods of adjusting grade averages to correct for grading

variations from school to school. Linn's (1966) review of the extensive work on this problem at the undergraduate level indicates that such adjustments result in little if any improvement in prediction beyond that offered by joint consideration of an aptitude test and the grade record. There has been only spotty work on this problem at the graduate level and none of it suggests any different conclusion (e.g., see Mehrabian, 1969). There is some indication, however, that success in some graduate business schools is enhanced somewhat by considering the quality of the undergraduate school (Fitcher and Schrader, 1972)

Anyone with long experience in selecting or training students in higher education is very inclined to plead for some way to measure student motivation--through personality scales, interest inventories, background information or whatever. There have been many pertinent studies at the undergraduate level and Freeberg (1965) documents a number of instances where such student self-report devices have made small but significant contributions to predicting grades. But Kendrick (1964) describes well the host of practical problems and ethical objections that have inhibited formal use of such information in selection.

A slightly different and perhaps more acceptable use of a motivation measure would be for the purpose of identifying groups of students who differ considerably in the extent to which success is predictable from traditional ability measures. Don Rock, with support from the Graduate Record Examinations Board, is presently studying that possibility as an outgrowth of earlier efforts to locate such moderating effects on the basis of age or quality of undergraduate school (Rock and Harmon, 1972).

One might suppose that motivation to undertake graduate work would be one important quality reflected in letters of recommendation, but the validity of such references is disappointingly low. In extensive studies of NSF fellowship applicants, the reliability of single references was reported to be in the low .30's (Harmon, 1966). This may be the main reason why recommendations are poor predictors, but careful efforts to improve that reliability with multiple ratings did not result in good validity for the NSF fellowship recommendations. Such results do not suggest that improved letters of reference are a promising possibility for increased accuracy of prediction.

So much for predictors. What are the prospects for improving the criteria? Graduate point average is traditionally subject to varying interpretation and practices which tend to make it unreliable. In recent years graduate faculty seem even more dubious regarding the value of the GPA as a criterion. With different grading procedures the GPA could theoretically be a quite good criterion but there is little reason to

expect that to happen in the foreseeable future. Systematic faculty ratings of different aspects of graduate success seems to be a more feasible development but there has been limited theoretical rationale to guide such extension. The comprehensive departmental examination, if properly developed, could very likely serve as a highly reliable criterion but it would place primary emphasis upon that aspect of success that is associated with content knowledge reproducible in a written test.

In some respects "Ph. D. attainment" (and its corollary "time to Ph. D.") is the most defensible criterion of those represented here. Not only does it represent the final reality of success; it also includes all those personal characteristics like ability, organization, and persistence that are normally considered necessary in the successful doctoral candidate. Unfortunately the researcher must wait a long time for this criterion. It is seldom applied to the MA degree and may be much less appropriate at that level, particularly in graduate departments with heavy emphasis on the Ph. D. A more serious shortcoming of this criterion is the fact that it is not easy to predict. There are similar types of behavior (e.g., employee turnover or withdrawal from military flight training) which are also dependent upon voluntary persistence. Such even , are notoriously difficult to predict accurately--probably because lack of persistence may be due to a wide variety of independent contingencies, any one of which may be unimportant for most people but critical for a few.

We might sum up the preceding discussion as follows. There is no doubt that present predictors, taken together, are providing a useful means of reducing some of the guesswork in selecting graduate students. Nonetheless, attrition of able students is disturbingly high. To the extent that attrition represents a mismatch of students and programs it is important to improve the validity of selection procedures.

Unfortunately the foregoing paragraphs do not present an optimistic picture of the possibilities for improving prediction of success in graduate school. There is no obvious way to improve the validity of existing measures of student potential. From past experience there is little reason to expect that new measures will do a substantially better job of predicting conventional criteria. Improving the focus and reliability of present criteria might well improve validity coefficients somewhat; it would not likely have much effect on which students are selected (i.e., one would still simply choose the students with the highest scores on the same predictors). The main problem is that we operate almost exclusively with one prediction strategy dominated by the notion of scholastic aptitude. There are, however, alternate prediction strategies that suggest additional predictors and additional criteria.

Alternate Prediction Strategies

It is well known that there are training objectives in graduate education (e.g., independent scholarship) not explicitly represented in conventional criteria, and there are important student abilities not represented among traditional selection measures (e.g., creative potential). In general, many training programs are characterized by multiple criteria of success which may not be highly related to one another and may depend upon different abilities. Or as may be more likely in graduate education, one department may emphasize one set of objectives while another department in the same discipline may stress other outcomes.

It may be easier to appreciate multiple criteria of success when examining actual job performance. There has been relatively little research on the relationship of performance in graduate school to later professional success but one elaborate study by Creager and Harmon (1966) includes the same predictors examined in this review.

The median validity coefficients for GRE Advanced, Recommendations, and Undergraduate GPA were as follows for three on-the-job criteria--rating of scientific knowledge: .27, .23, .29; income: .11, -.05, .03; citations of the individual's publications: .28, .07, .12. The study involved sixteen hundred students in seven fields so these median corelations can be assumed to be fairly stable. It is evident that each predictor has a modest correlation with a later rating of scientific knowledge, no predictor is related to income, and only the GRE Advanced predicts scholarly citations. (The authors report the latter to be a very promising measure since accumulated yearly data should provide a more reliable index than that available for this study.)

These limited data suggest that different predictors (or, in the case of income, no predictors) are relevant to different criteria. There are many quite defensible criteria of professional success: eminence as a scholar, teaching skill, professional leadership, etc. It is preferable to develop such criteria in the actual job situation, but for most practical purposes this would require prohibitive time and expense. But it is possible and highly desirable to use a stepping stone procedure by developing better intermediate criteria that can be measured during graduate training.

Students exhibit many forms of incipient professional behavior in graduate school, though we typically make little effort to evaluate such behavior in relation to selection procedures and training objectives. It can be useful to develop alternate prediction strategies that reflect the reality of varied training objectives. Figure 2 illustrates the intended connections in the case of three possible program objectives: to train the

practitioner, teacher, or scholar/scientist. Of course the criteria of success for a practitioner will vary from field to field, particularly if professional schools are considered.

Figure 2 speaks mostly for itself. It implies that different departments or programs within departments may emphasize different training objectives which, in turn, should be related to the way students are selected and the way their performance is evaluated. It is assumed that academic competency in the subject field is always an important success criterion, but beyond that, different training objectives imply multiple and often different criteria.

The first requirement in opening the possibility of alternate models of selection-training-evaluation is development of the necessary criteria. More than likely these would need to be specially constructed and then combined into a composite to be predicted by an appropriate combination of predictors. Developing the criterion components might involve faculty ratings of different student competencies, a common examination of subject matter competence, systematic identification of accomplishments, special means of collecting outside judgments, or whatever procedures may be required to obtain information that is relevant to the specific training objectives considered important. Some recent work by Reilly (1971) is a good example of progress along these lines. The notion of multiple criteria related to different training objectives has several advantages.

First, it encourages the use of additional predictor variables which may not enhance prediction of conventional criteria but are nonetheless relevant to important aspects of success in some programs. In this way it becomes feasible to demonstrate the validity of creativity tests, cognitive styles, or special accomplishments (see Frederiksen and Ward, 1972, Witkin, 1972, and Wallach, 1972, for discussions of recent developments in these areas). The simple reason is that specialized criteria can give such predictors something to shoot at. Using measures of this sort for selecting graduate students has the very desirable effect of broadening the conception of talent.

Second, the model depicted in Figure 2 is more likely to result in appropriate matching between student characteristics and program characteristics than one would expect to occur under a single, aptitude-dominated mode of prediction. Improved matching should result in more student satisfaction and overall competency in a class of graduate students.

Third, the proposed view assumes that prediction and selection are inseparable from program design and evaluation. Consequently, the

process of defining an appropriate prediction strategy forces desirable attention to the intended outcomes of the program and the relationship of the curriculum to those outcomes.

In summary, the available objective evidence suggests that the accuracy of predicting which students will succeed in a particular graduate school is often no better than modest, especially if such predictions are based only upon a test or a grade record. Taken together the two types of predictors do a reasonably good job, considering the restricted range of ability involved. The best way to improve selection of graduate students will be to develop improved criteria of success. This is no small job for graduate faculties, but it carries the promise of more effective utilization of talent and greater assurance of equity in admitting students to advanced levels of training and the privilege associated with such programs.

Table 1. Median Validity Coefficients[#] for Various Predictors and Criteria of Success in Graduate School

Predictors	Criteria of Success				
	Graduate GPA	Overall Faculty Rating	Dept. Exam.	Attain Ph.D.	Time to Ph.D.
1. GRE--Verbal	.24 46	.31 27	.42 5	.18 47	.16 18
2. GRE-Quantitative	.23 43	.27 25	.27 5	.26 47	.25 18
3. GRE-Advanced	.30 25	.30 8	.48 2	.35 40	.34 18
4. GRE-Composite	.33 30	.41 6	*	.31 33	.35 16
5. Undergraduate GPA	.31 26	.37 15	*	.14 30	.23 9
6. Recommendations	*	*	*	.18 15	.23 9
7. GRE-GPA Composite (weighted)	.45 24	*	*	.40 16	.40 9

[#] The lower number in each pair (set in smaller type) represents the number of coefficients upon which each median is based

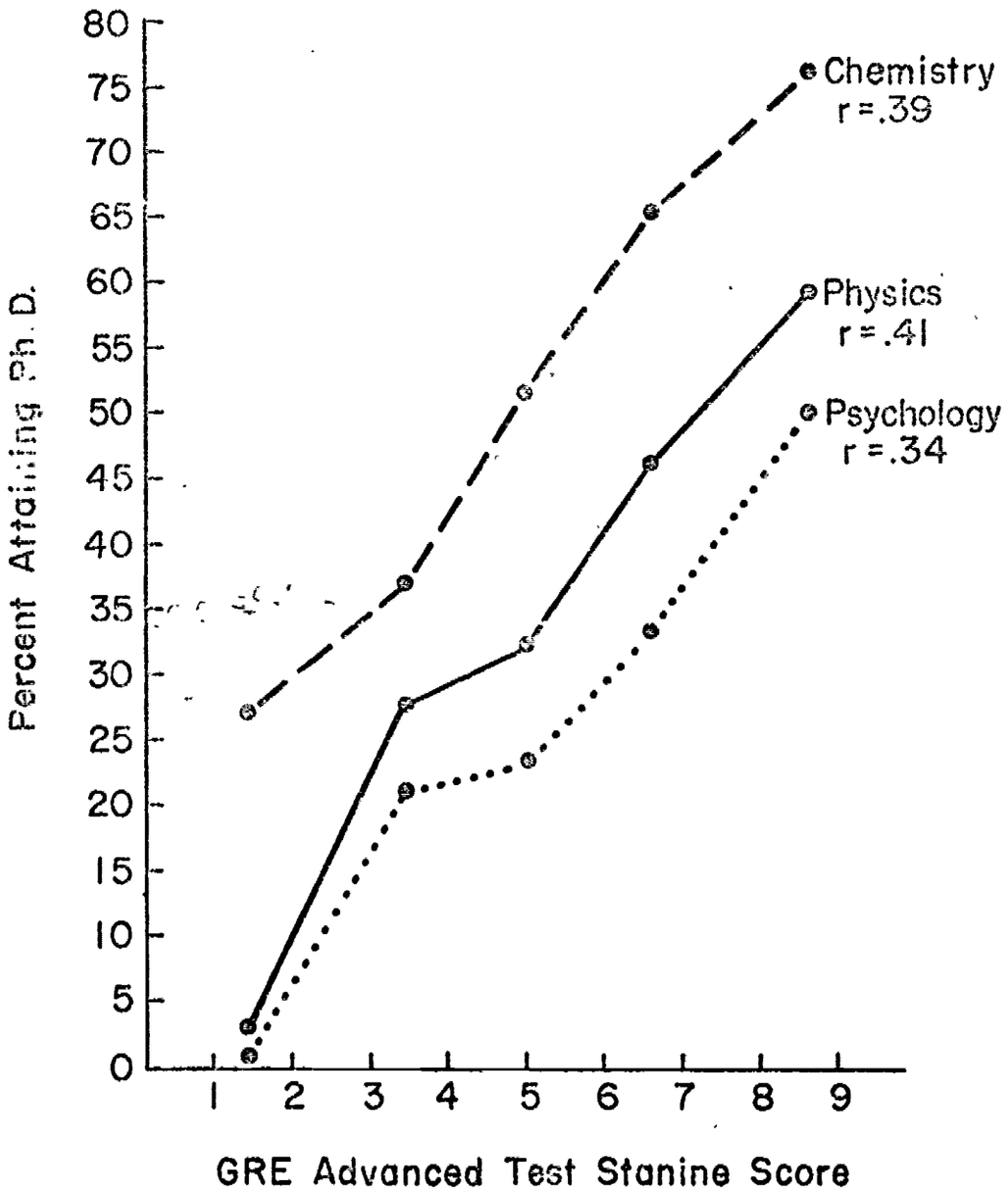
* No data available

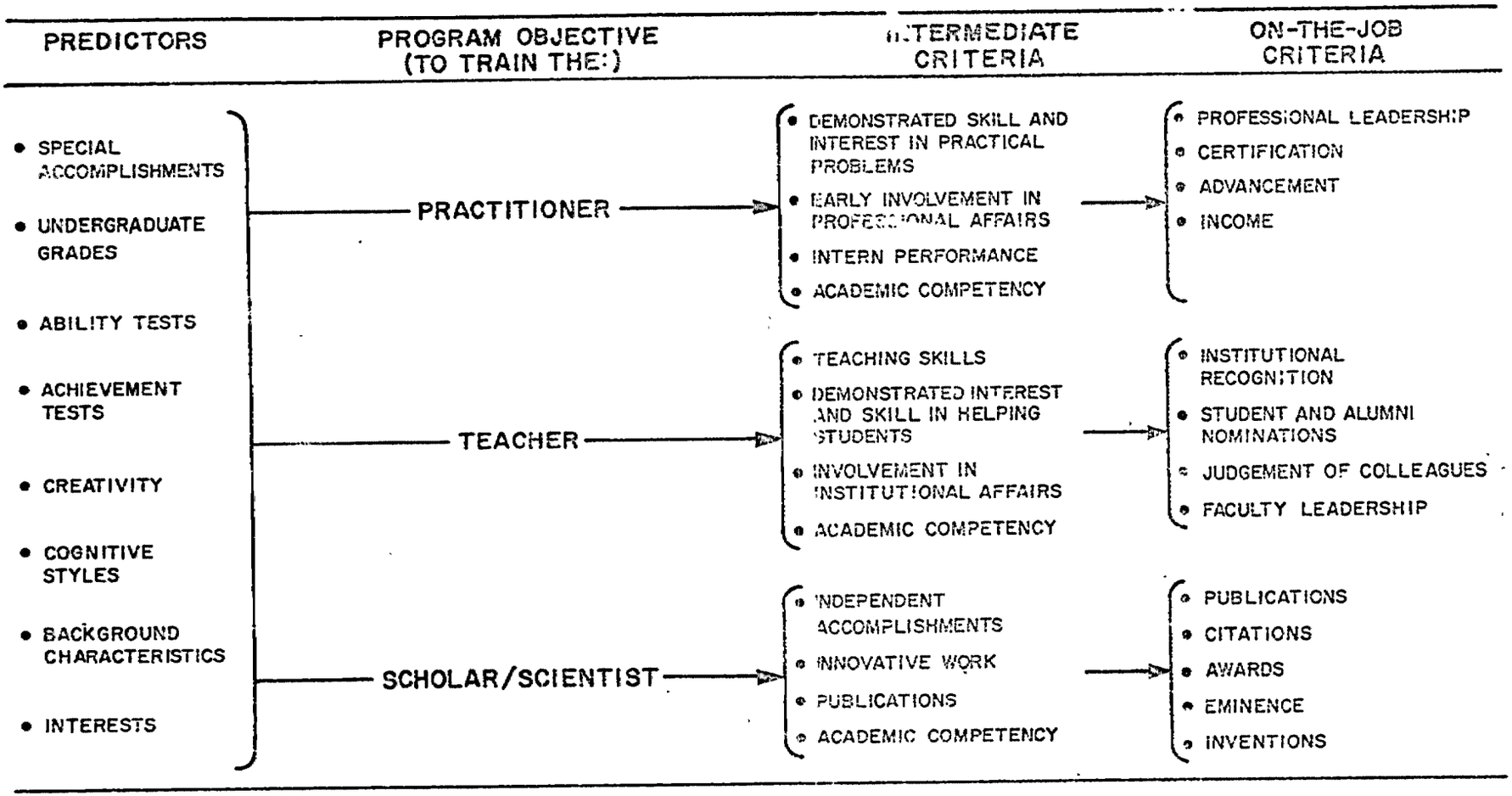
Table 2. Median Validity Coefficients[#] for Five Predictors of Graduate Success (variously defined*) in Nine Fields

Predictors	Biological Science	Chemistry	Education	Engineering and Applied Science	English	Math	Physics	Psychology	Social Science
GRE-Verbal	.18 (7)	.22 (14)	.36 (15)	.29 (11)	.21 (6)	.30 (6)	.02 (6)	.19 (23)	.32 (11)
GRE-Quantitative	.27 (8)	.28 (13)	.28 (14)	.31 (10)	.06 (6)	.27 (6)	.21 (6)	.23 (22)	.32 (10)
GRE-Advanced	.26 (5)	.39 (9)	.24 (6)	.44 (7)	.43 (3)	.44 (5)	.38 (5)	.24 (17)	.46 (5)
Undergraduate GPA	.13 (2)	.27 (7)	.30 (5)	.18 (4)	.22 (4)	.19 (4)	.31 (4)	.16 (15)	.37 (6)
GRE-GPA Composite (weighted)	.35 (3)	.42 (6)	.42 (7)	.47 (4)	.56 (2)	.41 (3)	.45 (2)	.32 (4)	.40 (5)

[#] The lower number in each pair (set in smaller type) represents the number of coefficients upon which each median is based

* In those sets of data where two criteria were included, one was selected for the purposes of this table in the following order of priority: GPA, Attain Ph.D., Dept. Test, and Faculty Rating


80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Percent Attaining Ph. D.
1
2
3
4
5
6
7
8
9
GRE Advanced Test Stanine Score
Chemistry
r = .39
Physics
r = .41
Psychology
r = .34


PREDICTORS
PROGRAM OBJECTIVE (TO TRAIN THE:)
INTERMEDIATE CRITERIA
ON-THE-JOB CRITERIA
• SPECIAL ACCOMPLISHMENTS
• UNDERGRADUATE GRADES
• ABILITY TESTS
• ACHIEVEMENT TESTS
• CREATIVITY
• COGNITIVE STYLES
• BACKGROUND CHARACTERISTICS
• INTERESTS
PRACTITIONER
• DEMONSTRATED SKILL AND INTEREST IN PRACTICAL PROBLEMS
• EARLY INVOLVEMENT IN PROFESSIONAL AFFAIRS
• INTERN PERFORMANCE
• ACADEMIC COMPETENCY
• PROFESSIONAL LEADERSHIP
• CERTIFICATION
• ADVANCEMENT
• INCOME
TEACHER
• TEACHING SKILLS
• DEMONSTRATED INTEREST AND SKILL IN HELPING STUDENTS
• INVOLVEMENT IN INSTITUTIONAL AFFAIRS
• ACADEMIC COMPETENCY
• INSTITUTIONAL RECOGNITION
• STUDENT AND ALUMNI NOMINATIONS
• JUDGEMENT OF COLLEAGUES
• FACULTY LEADERSHIP
SCHOLAR/SCIENTIST
• INDEPENDENT ACCOMPLISHMENTS
• INNOVATIVE WORK
• PUBLICATIONS
• ACADEMIC COMPETENCY
• PUBLICATIONS
• CITATIONS
• AWARDS
• EMINENCE
• INVENTIONS

Validity Studies summarized in
"Predicting Success in Graduate Education,"

Warren W. Willingham
Educational Testing Service
March 1973

Alexakos, Constantine E. The Graduate Record Examinations: Aptitude Tests as screening devices for students in the College of Human Resources and Education. Unpublished report, West Virginia University, 1967. Reported by G. V. Lannholm in Graduate Record Examinations Special Report 68-1. Educational Testing Service, Princeton, New Jersey, 1968.

Besco, Robert O. The measurement and prediction of success in graduate school. Ph. D. dissertation, Purdue University, 1960. Reported by G. V. Lannholm in Graduate Record Examinations Special Report 60-3, Educational Testing Service, Princeton, New Jersey, 1960.

Borg, Walter R. GRE aptitude scores as predictors of GPA for graduate students in education. Educational and Psychological Measurement, 1963, vol. 23, no. 2, pp. 379-389.

Capps, Marian P., and Decosta, Frank A. Contributions of the Graduate Record Examinations and the National Teacher Examinations to the prediction of graduate school success. Journal of Educational Research, 1957, vol. 50, no. 5, pp. 383-389.

Clark, Henry. Graduate Record Examination correlations with grade-point averages in the Department of Education at Northern Illinois University, 1962-1966. Unpublished Master's thesis, Northern Illinois University, June 1968.

Conway, Sister Madona Therese. The relationship of the Graduate Record Examination results to achievement in the Graduate School at the University of Detroit. Unpublished Master's thesis, University of Detroit, 1955. Reported by G. V. Lannholm in Graduate Record Examinations Special Report 60-3, Educational Testing Service, Princeton, New Jersey, 1960.

Creager, John A. Predicting doctorate attainment with GRE and other variables. Technical Report No. 25, Office of Scientific Personnel, National Academy of Sciences--National Research Council, Washington, D. C., 1965.

Creager, John A. A study of graduate fellowship applicants in terms of Ph. D. attainment. Technical Report No. 18, Office of Scientific Personnel. National Academy of Sciences--National Research Council, Washington, D.C., 1961.

Dawes, Robyn M. A case study of graduate admissions: Application of three principles of human decision making. American Psychologist, 1971, vol. 26, no. 2, pp. 180-188.

Duff, Franklin L., and Aukes, Lewis E. The relationship of the Graduate Record Examination to success in the Graduate College (a supplementary comparative analysis of eight previously reported studies). Bureau of Institutional Research and Office of Instructional Research, University of Illinois, October 1966.

Eckhoff, Constance M. Predicting graduate success at Winona State College. Educational and Psychological Measurement, 1966, vol. 26, no. 2, pp. 483-485.

Ewen, Robert B. The GRE psychology test as an unobtrusive measure of motivation. Journal of Applied Psychology, 1969, vol. 53, no. 5, pp. 383-387.

Florida State University, Office of Academic Research and Planning. The prediction of grade-point average in graduate school at the Florida State University, Parts I & II. Florida State University, December 1971.

Florida State University, Office of Institutional Research and Service. Relationship between Graduate Record Examinations Aptitude Test scores and academic achievement in the Graduate School at Florida State University. Florida State University, 1958, 68 pp.

Hackman, J. Richard, Wiggins, Nancy, and Bass, Allan R. Prediction of long-term success in doctoral work in psychology. Educational and Psychological Measurement, 1970, vol. 30, pp. 365-374.

Hansen, W. Lee. Prediction of graduate performance in economics. Department of Economics, University of Wisconsin, April 1970 (Mimeographed).

Harvey, Philip R. Predicting graduate school performance in education. Unpublished ETS report. Reported by G. V. Lannholm in Graduate Record Examinations Special Report 68-1, Educational Testing Service, Princeton, New Jersey, 1968.

King, Donald C., and Besco, Robert O. The Graduate Record Examination as a selection device for graduate research fellows. Educational and Psychological Measurement, 1960, vol. 20, no. 4, pp. 853-858.

Lannholm, Gerard V., Marco, Gary L., and Schrader, William B. Cooperative studies of predicting graduate school success. Graduate Record Examinations Special Report 68-3. Educational Testing Service, Princeton, New Jersey, 1968.

Law, Alexander. The prediction of ratings of students in a doctoral training program. Educational and Psychological Measurement, 1960, vol. 20, no. 4, pp. 847-851.

Lorge, Irving. "Relationship between Graduate Record Examinations and Teachers College, Columbia University, Doctoral Verbal examinations." Letter to G. V. Lannholm, dated September 21, 1960. Reported by G. V. Lannholm in Graduate Record Examinations Special Report 68-1, Educational Testing Service, Princeton, New Jersey, 1968.

Madaus, George F., and Walsh, John J. Departmental differentials in the predictive validity of the Graduate Record Examination Aptitude Tests. Educational and Psychological Measurement, 1965, vol. 25, no. 4, pp. 1105-1110.

Mehrabian, Albert. Undergraduate ability factors in relationship to graduate performance. Educational and Psychological Measurement, 1969, vol. 29, no. 2, pp. 409-419.

Michael, William B., Jones, Robert A., and Al-Amir, Hudhail, Pullias, Calvin M., Jackson, Michel, and Goo, Valerie. Correlates of a pass-fail decision for admission to candidacy in a doctoral program. Educational and Psychological Measurement, 1971, vol. 31, no. 4, pp. 965-967.

Michael, William B., Jones, Robert A., and Gibbons, Billie D. The prediction of success in graduate work in chemistry from scores on the Graduate Record Examination. Educational and Psychological Measurement, 1960, vol. 20, no. 4, pp. 859-861.

Newman, Richard I. GRE scores as predictors of GPA for psychology graduate students. Educationl and Psychological Measurement, 1968, vol. 28, no. 2, pp. 433-436.

Office of Educational Research. Study of GRE scores of geology students matriculating in the years 1952-1961. RP--Abstract, Yale University, Special Report 68-1, Educational Testing Service, Princeton, New Jersey, 1968.

Olsen, Marjorie. The predictive effectiveness of the Aptitude Test and the Advanced Biology Test of the GRE in the Yale School of Forestry. Statistical Report 55-6, Educational Testing Service, Princeton, New Jersey, 1955.

Roberts, Pamela T. An analysis of the relationship between Graduate Record Examination scores and success in the Graduate School of Wake Forest University. Master's thesis, Wake Forest University, August 1970.

Robertson, Malcolm, and Nielsen, Winnifred. The Graduate Record Examination and selection of graduate students. American Psychologist, 1961, vol. 16, no. 10, pp. 648-650.

Robinson, Donald W. A comparison of two batteries of tests as predictors of first year achievement in the graduate school of Bradley University. Ph.D. dissertation, Bradley University, 1957.

Rock, Donald A., and Harmon, Lindsey R. The prediction of doctorate attainment in psychology, mathematics and chemistry. Graduate Record Examinations preliminary report, Educational Testing Service, Princeton, New Jersey, August 1972.

Roscoe, John T., and Houston, Samuel R. The predictive validity of GRE scores for a doctoral program in education. Educational and Psychological Measurement, 1969, vol. 29, no. 2, pp. 507-509.

Sacramento State College, Test Office. An analysis of traditional predictor variables and various criteria of success in the Master's degree program at Sacramento State College for an experimental group who received Master's degrees in the spring 1968, and a comparable control group who withdrew from their programs. Test Office Report 69-3, Sacramento State College, October 1969.

Shaffer, Julie, and Rosenfeld, Howard. MAT-GRE prediction study--Initial results. Intra-departmental memorandum, Department of Psychology, University of Kansas, March 1969.

Sistrunk, Francis. "The GRE's as predictors of graduate school success in psychology." Letter to G. V. Lannholm, dated October 3, 1961. Reported by G. V. Lannholm in Graduate Record Examinations Special Report 68-1, Educational Testing Service, Princeton, New Jersey, 1968.

Sleeper, Mildred L. Relationship of scores on the Graduate Record Examination to grade point averages of graduate students in occupational therapy. Educational and Psychological Measurement, 1961, vol. 21, no. 4, pp. 1039-1040.

Tully, G. Emerson. Screening applicants for graduate study with the Aptitude Test of the Graduate Record Examinations. College and University, 1962, vol. 38, pp. 51-60.

University of Virginia, Office of Institutional Analysis. Correlations between admissions criteria and University of Virginia grade-point averages, Graduate School of Arts and Sciences, Fall 1964. University of Virginia, circa 1966. (Mimeographed)

Wallace, Anita D. The predictive value of the Graduate Record Examinations at Howard University. Unpublished Master's thesis, Howard University, 1952. Reported by G. V. Lannholm in Graduate Record Examinations Special Report 68-1, Educational Testing Service, Princeton, New Jersey, 1968.

White, Elizabeth L. The relationship of the Graduate Record Examination results to achievement in the Graduate School at the University of Detroit. Unpublished Master's thesis, University of Detroit, 1954. Reported by G. V. Lannholm in Graduate Record Examinations Special Report 60-3, Educational Testing Service, Princeton, New Jersey, 1960.

White, Gordon William. A predictive validity study of the Graduate Record Examinations Aptitude Test at the University of Iowa. Unpublished Master's thesis, University of Iowa, June 1967.

Williams, John D., Harlow, Steven D., and Grab, Del. A longitudinal study examining prediction of doctoral success: Grade-point average as criterion, or graduation vs. non-graduation as criterion. Journal of Educational Research, December 1970, vol. 64, no. 4, pp. 161-164.

References

Creager, John A. Predicting doctorate attainment with GRE and other variables. Technical Report No. 25. Washington, D.C.: Office of Scientific Personnel, National Academy of Sciences--National Research Council, 1965, 48 pp.

Creager, John A., and Harmon, Lindsey R. On-the-job validation of selection variables. Technical Report No. 26. Washington, D.C.: Office of Scientific Personnel, National Academy of Sciences--National Research Council, 1966, 70 pp.

Cronbach, Lee J. Test validation, in Robert L. Thorndike (Ed.), Educational measurement. (2nd ed.) Washington, D.C.: American Council on Education, 1971, pp. 443-507.

Fishman, Joshua A., and Pasanella, Ann K. College admission-selection studies. Review of Educational Research, 1960, vol. 30, pp. 298-310.

Frederiksen, Norman, and Ward, William C. Proposal for the development of measures for the study of creativity. (Submitted to the Graduate Record Examinations Board) Princeton, New Jersey: Educational Testing Service, 1972.

Freeberg, Norman E. The Biographical Information Blank as a predictor of student achievement: A review. Psychological Reports, 1967, vol. 20, pp. 911-925.

Harmon, Lindsey R. Fourteen years of research on fellowship selection. Washington, D.C.: Office of Scientific Personnel, National Academy of Sciences--National Research Council, 1966, 39 pp.

Kendrick, S.A. The personality testing tangle. College Board Review, 1964, no. 54, pp. 26-30.

Lannholm, Gerald V. Review of studies employing GRE scores in predicting success in graduate study 1952-1967. Graduate Record Examinations Special Report 68-1. Princeton, New Jersey: Educational Testing Service, 1968, 43 pp.

Lannholm, Gerald V. Summaries of GRE validity studies 1966-1970. Graduate Record Examinations Special Report 72-1. Princeton, New Jersey: Educational Testing Service, 1972, 21 pp.

Linn, Robert L. Grade adjustments for prediction of academic performance: A review. Journal of Educational Measurement, 1966, vol. 3, no. 4, pp. 313-329.

Linn, Robert L. Fair test use in selection. Review of Educational Research, Spring 1973 (in press).

Mehrabian, Albert. Undergraduate ability factors in relationship to graduate performance. Educational and Psychological Measurement, 1969, vol. 29, no. 2, pp. 409-419.

Pitcher, Barbara, and Schrader, W. B. Indicators of college quality as predictors of success in graduate schools of business. Admission Test for Graduate Study in Business, Brief Number 6. Princeton, New Jersey: Educational Testing Service, 1972, 15 pp.

Reilly, Richard R. Critical incidents of graduate student performance. Graduate Record Examinations Board Technical Memorandum No. 1. Princeton, New Jersey: Educational Testing Service, April 1971, 13 pp.

Rock, Donald A., and Harmon, Lindsey R. The prediction of doctorate attainment in psychology, mathematics and chemistry. Graduate Record Examinations Preliminary Report. Princeton, New Jersey: Educational Testing Service, August 1972. (Unpublished manuscript)

Stanley, Julian C. Predicting college success of the educationally disadvanteged. Science, February 1971, vol. 171, no. 3972, pp. 640-647.

Thorndike, Robert L., and Hagen, Elizabeth. Measurement and evaluation in psychology and education. (3rd ed.) New York: Wiley, 1969, 705 pp.

Wallach, Michael. The psychology of talent and graduate education. Paper presented at the Invitational Conference on Cognitive Styles and Creativity in Higher Education sponsored by the Graduate Record Examinations Board, Montreal, Canada, November 1972.

Witkin, Herman. The role of cognitive styles in academic performance and in teacher-student relations. Research Bulletin 73-11. Princeton, New Jersey: Educational Testing Service, February 1973.