

which the cell is a part, so that the two kinds of organisms then live together in a state of mutualism or symbiosis (26).

#### References and Notes

1. L. A. Stauber, in *Immunity to Parasitic Animals*, G. J. Jackson, R. Herman, I. Singer, Eds. (Appleton-Century-Crofts, New York, 1970), vol. 2, pp. 739-765.
2. M. A. Rudzinska, P. A. D'Alesandro, W. Trager, *J. Protozool.* **11**, 166 (1964).
3. R. Lainson, P. C. C. Garnham, R. Killick-Kendrick, R. G. Bird, *Br. Med. J.* **2**, 470 (1964).
4. E. Weidner, *Z. Zellforsch.* **105**, 33 (1970).
5. K. Ohshima, *Parasitology* **29**, 220 (1937).
6. E. Weidner, *Z. Parasitenkd.* **40**, 227 (1973).
7. R. Ishihara, *J. Invertebr. Pathol.* **12**, 245 (1968).
8. J. K. Frenkel, J. P. Dubey, N. L. Miller, *Science* **167**, 893 (1970).
9. W. Brommer, H. H. Heunert, B. Milthaler, *Z. Tropenmed. Parasitol.* **20**, 450 (1969).
10. H. G. Sheffield and M. L. Melton, *J. Parasitol.* **54**, 209 (1968).
11. R. Norby, *Infection and Immunity* **2**, 250 (1970).
12. T. Jones and J. Hirsch, *J. Exp. Med.* **136**, 1173 (1972).
13. R. Ladda, M. Aikawa, H. Sprinz, *J. Parasitol.* **55**, 633 (1969).
14. R. Herman, *ibid.*, p. 626.
15. D. M. Hammond, E. Scholtyseck, B. Chobotor, *J. Protozool.* **14**, 678 (1967).
16. M. A. Rudzinska, *Int. Rev. Cytol.* **25**, 161 (1969).
17. W. Trager, S. G. Langreth, E. G. Platzer, *Proc. Helminthol. Soc. Wash.* **39** (special issue), 220 (1972).
18. L. A. Stauber, H. A. Walker, A. P. Richardson, *J. Infect. Dis.* **89**, 31 (1951).
19. S. G. Langreth and W. Trager, unpublished data.
20. M. A. Rudzinska and W. Trager, *J. Protozool.* **9**, 279 (1962).
21. W. A. Siddiqui, J. V. Schnell, Q. A. Geiman, *Mil. Med.* **134**, 929 (1969).
22. T. P. Bennett and W. Trager, *J. Protozool.* **14**, 214 (1967).
23. P. J. F. Henderson and H. A. Lardy, *J. Biol. Chem.* **245**, 1319 (1970).
24. W. Trager, in *Comparative Biochemistry of Parasites*, H. Van den Bossche, Ed. (Academic Press, New York, 1972), pp. 343-350.
25. E. Weidner and W. Trager, *J. Cell Biol.* **57**, 586 (1973).
26. W. Trager, *Symbiosis* (Van Nostrand Reinhold, New York, 1970), pp. 1-100.
27. This article is based on a lecture given on 17 March 1972 at a symposium at Rutgers University honoring Prof. Leslie A. Stauber on his retirement and is submitted in memory of Prof. Stauber who died on 27 March 1973. The illustrations were prepared with the skilled assistance of Erminio Gubert. The work was supported in part by grants from the National Institute of Allergy and Infectious Diseases, U.S. Public Health Service (training grant AI 00192 and research grant AI 08989).

## Predicting Success in Graduate Education

Improved selection procedures are likely to come from  
better definitions of "success."

Warren W. Willingham

In recent decades, graduate schools have assumed a major responsibility for the advanced training of a talented segment of American society. Compared with lower levels 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 the recent interest in predicting success in graduate education. In earlier times, space in graduate schools and the number of applicants were in a rough equilibrium, but burgeoning numbers of applicants in the 1950's and 1960's focused attention on how some

were selected and others turned away. These larger numbers of students made it possible to undertake statistical studies in many departments that had previously had too few students to make this type of systematic evaluation worthwhile. Finally, increasing use of selection tests [Graduate Record Examination (GRE) candidates increased from 100,000 to 280,000 per year during the 1960's] suggested the prediction studies with which similar tests are closely associated at the undergraduate level. The purpose of this article 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 where further research is needed.

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 (1). This correlation, called a validity coefficient, ranges from a chance relationship of .00 to a perfect relationship of 1.00, although negative coefficients can occur and perfect validity is not closely approached in practice. Usually more than one predictor is involved (for example, 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 is 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. Specifically, it should represent what we want to measure, all we want to measure, and nothing but that which we want to measure (2). By reliability we mean that a measure provides a stable estimate from one measuring occasion to another. By acceptability, we mean that a measure is economically feasible, administratively practical, and socially ethical. It is in this context of construct

The author is executive director for program research, Educational Testing Service, Princeton, New Jersey 08540.

Table 1. Median validity coefficients for various predictors and criteria of success in graduate school. (The number of coefficients upon which each median is based is given in parentheses. Coefficients involving dichotomized criteria were sometimes reported as biserials and sometimes as point-biserials.)

| Predictors        | Criteria of success |                        |                          |              |               |
|-------------------|---------------------|------------------------|--------------------------|--------------|---------------|
|                   | Graduate GPA        | Overall faculty rating | Departmental examination | Attain Ph.D. | Time to Ph.D. |
| GRE-verbal        | .24 (46)            | .31 (27)               | .42 (5)                  | .18 (47)     | .16 (18)      |
| GRE-quantitative  | .23 (43)            | .27 (25)               | .27 (5)                  | .26 (47)     | .25 (18)      |
| GRE-advanced      | .30 (25)            | .30 (8)                | .48 (2)                  | .35 (40)     | .34 (18)      |
| GRE-composite     | .33 (30)            | .41 (8)                | *                        | .31 (33)     | .35 (18)      |
| Undergraduate GPA | .31 (26)            | .37 (15)               | *                        | .14 (30)     | .23 (9)       |
| Recommendations   | *                   | *                      | *                        | .18 (15)     | .23 (9)       |
| GRE-GPA composite | .45 (24)            | *                      | *                        | .40 (16)     | .40 (9)       |

\* No data available.

validity, reliability, and acceptability that I review briefly some of the strengths and weaknesses of predictors and criteria before addressing the empirical and utilitarian question of the relation between the two.

### Predictors

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 grade point average (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, however. 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.

References from undergraduate professors are widely used, despite the fact that they are time-consuming to pre-

pare 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 of, or lack of comparability among, judges.

A principal advantage of standardized tests is the 70-year history of research and theoretical development behind them. On the one hand, this work has produced reliable, standard measures that are highly suitable for national administration under security conditions. More important, 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. Because of this psychometric development, it is possible to construct tests to measure any of an extremely wide range of human abilities. On the other hand, a corresponding weakness of most individual tests is their tendency to focus attention on fairly limited aspects of competency. Another weakness is the lingering suspicion [despite substantial evidence to the contrary (3)] that standardized tests are intrinsically

biased against individuals from cultural minorities.

Various characteristics of an applicant's background are considered implicitly, if not formally, when graduate departments select students. Both the strengths and the 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 irrelevant, but they are used nonetheless 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. For example, one might find that students from a particular kind of neighborhood typically do poorer in a given university than would be expected from other information, but it is difficult to justify using such a finding in selecting students for admission.

### Criteria

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 appropriate way to describe the important outcomes of graduate education.

Many departments require students to pass a qualifying or comprehensive examination at some point during their graduate program. In theory, a prop-

Table 2. Median validity coefficients for five predictors of success in graduate school in nine fields. (The number of coefficients upon which each median is based is given in parentheses. Coefficients involving dichotomized criteria were sometimes reported as biserials and sometimes as point-biserials. In those sets of data where two criteria were included, one was selected in the following order of priority: GPA, attain Ph.D., departmental examination, and faculty rating.)

| Predictors                   | Biological science | Chemistry | Education | Engineering and applied science | English | Mathematics | 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)         |

erly constructed examination could provide the most reliable and valid criterion of subject competence; in practice, such examinations 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, using test scores mainly to predict test scores suggests a logical circularity. The point is less often made, but the same criticism applies to predicting GPA from GPA.

A principal advantage of faculty judgment as a criterion is versatility in measuring important aspects of graduate success other than knowledge of the subject. Faculty who know a student well are in the best position to say whether he is able to execute independent research or to 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.

Regardless of what other judgments a faculty may make about a doctoral student, the acid test is whether he or she is granted the degree. Consequently, this is probably the single most defensible criterion of success. On the negative side, one must wait a long time for this criterion. Another difficulty is the fact that whether or not 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.

Long-range criteria also have special significance for selecting students. Predicting career success is beyond the scope of this article, but better use of intermediate criteria is strongly recommended.

### Validity Studies

By far the most common predictors used in studies of success in graduate school are undergraduate average and GRE scores. This article is based upon correlation studies that used these variables and that were either cited by Lannholm (4) or located through searches of appropriate journals and abstracts. Forty-three studies were found for the period of 1952 to 1972. Half of these studies were published; the remainder were institutional reports or theses (5).

The 43 studies included 138 independent sets of data, usually corresponding to departments, although 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, although this composite was occasionally weighted statistically. The undergraduate GPA was undoubtedly computed in various ways in different studies, but it was seldom specified very carefully. The data concerning recommendations came almost exclusively from three extensive studies of National Science Foundation fellowship applicants (6, 7) 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 examinations 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 B.A. and Ph.D. All of the data concerning this last criterion come from two very large studies by Creager (6).

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 an 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. Sev-

eral observations can be made from these tables:

1) 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 (8).

2) 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.

3) The GRE quantitative test is typically a better predictor than the GRE verbal test 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 than the GRE quantitative in verbally oriented disciplines such as English and education. Otherwise, the pattern of validity coefficients is fairly similar from one discipline to the next.

4) The GRE advanced test 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.

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

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

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

What overall evaluation can be made of the extent to which success in graduate school is predictable? Cronbach (1)

describes alternate interpretations of the utility of a validity coefficient ( $\rho_{xy}$ ) for selection purposes. With a fixed training program, "the coefficient expresses the benefit from testing as a percentage of the benefit one could get from perfect prediction of outcomes," but when the nature of the program is responsive to the quality of entering students, "the value of a test is more nearly proportional to  $\rho^2_{xy}$  than to  $\rho_{xy}$ ." When there are two predictors (for example, GPA and GRE), the value of one depends upon how much it adds to the other and upon the cost of gathering the additional information. In general, a small validity coefficient can produce a large benefit if the proportion of students selected is small. Finally, a given validity coefficient has more value if the selection decision is important and irreversible.

We might say that a validity coefficient of .40 is moderate and one of .20 is, at best, modest and of marginal utility in many applications. The conditions of graduate student selection are often 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 combined GRE and undergraduate GPA are providing quite useful information in most situations.

Figure 1 shows the benefit likely to accrue from using predictors that are valid to the extent indicated. The three coefficients illustrated are biserial correlations between GRE advanced test scores and whether or not the student attained the Ph.D. (6, 7). As indicated, students at high levels of ability are 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. In the samples reported by Creager (6), there were substantial differences in attainment rates among fields (chemistry, 51 percent; physics, 36 percent; and psychology, 26 percent).

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 predictions 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, such as 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 aptitude tests can enhance prediction to any significant degree.

The undergraduate GPA suffers shortcomings similar to those of 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 (9) 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 a different conclusion [for example, see Mehra-bian (10)]. There are indications, however, that success in some graduate business schools is enhanced somewhat by considering the quality of the undergraduate school (11).

Anyone with long experience in selecting or training students in higher education is 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 (12) documents a number of instances where such student self-report devices have made small but significant contributions to predicting grades. But Kendrick (13) describes 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 measures of ability. 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 (7).

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 (14). 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 would increase accuracy of prediction.

So much for predictors. What are the prospects for improving the criteria? The GPA is traditionally subject to varying interpretations 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 very good criterion;

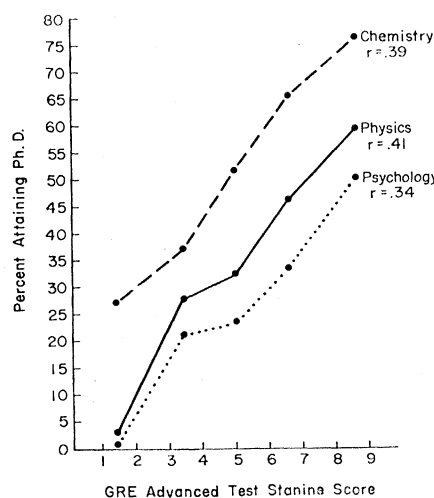


Fig. 1. Proportion of students at various levels of GRE advanced test scores in chemistry, physics, and psychology who attained the Ph.D. within 10 years (6).

but there is little reason to expect that to happen in the foreseeable future. Systematic faculty ratings of different aspects of graduate success seem 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 which 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 such as 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 M.A. degree and may be much less appropriate at that level, particularly in graduate departments with heavy emphasis on the Ph.D. A more serious difficulty is the fact that Ph.D. attainment is not easy to predict. There are similar types of behavior (for example, employee turnover or withdrawal from military flight training) that are also dependent upon voluntary persistence. Such behavior is notoriously difficult to predict accurately—possibly because lack of persistence may be due to a 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. Judging 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

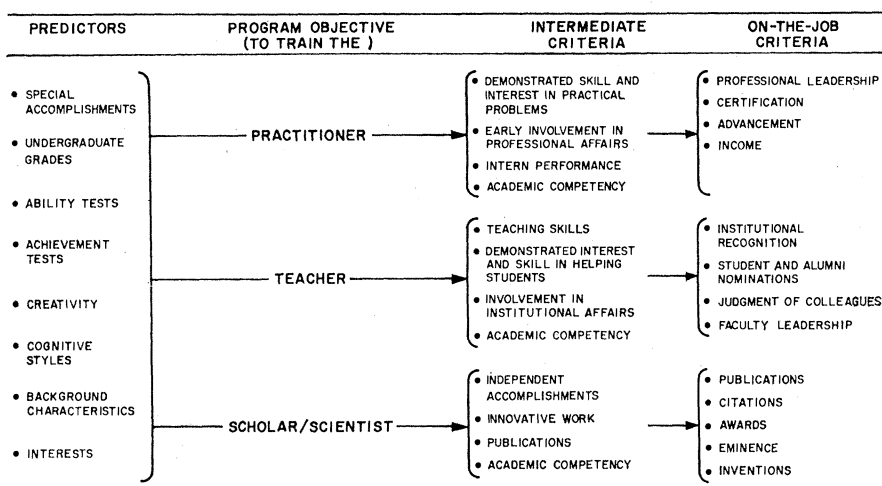


Fig. 2. Alternate prediction strategies in graduate education.

and reliability of present criteria might well improve validity coefficients somewhat; it would not likely have much effect on which students are selected (that is, 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 a 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 (for example, independent scholarship) not explicitly represented in conventional criteria and that there are important student abilities not represented among traditional selection measures (for example, creative potential). In general, training programs are characterized by multiple criteria of success that may not be highly related to one another and that 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 another.

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 success, but an elaborate study by Creager and Harmon (15) 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, and .29; income: .11, -.05, and .03; citations of the individual's publications: .28, .07, and .12. Since the study involved 1600 students in seven fields, these median correlations 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 their 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, and so forth. It is preferable to develop such criteria in the actual job situation, but for most practical purposes this would require prohibitive time and expense. It is possible and desirable, however, to develop better intermediate criteria that can be measured during graduate training.

Students exhibit many forms of incipient professional behavior in graduate school, although typically little effort is made to evaluate such behavior in relation to selection procedures and training objectives. It would be useful to develop alternate prediction strategies that reflect the reality of varied training objectives. Figure 2 illustrates the intended connections in three possible

program objectives: to train the practitioner, to train the teacher, or to train the scholar-scientist. Including the practitioner in this model simply reflects such a training emphasis in some graduate departments; it does not imply that the data presented here are relevant to professional schools.

Figure 2 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 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 criteria would need to be specially constructed and then combined into a composite, to be predicted by an appropriate combination of predictors. Developing the components of a given criterion might involve faculty ratings of different student competencies, a common examination of competence in the subject matter, 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 (16) 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; these may not enhance prediction of conventional criteria, but they 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 (17) 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 Fig. 2 is more likely to result in appropriate matching between student characteristics and program characteristics than one would expect with a single, aptitude-dominated mode of prediction. Improved matching should result in more satisfaction and overall competency within 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, 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 training and the privilege associated with such training.

## References and Notes

1. L. J. Cronbach, in *Educational Measurement*, R. L. Thorndike, Ed. (American Council on Education, Washington, D.C., 1971), p. 496.
2. R. L. Thorndike and E. Hagen, *Measurement and Evaluation in Psychology and Education* (Wiley, New York, 1969).
3. R. L. Linn, *Rev. Educ. Res.* **43**, 139 (Spring 1973); J. C. Stanley, *Science* **171**, 640 (1971).
4. G. V. Lannholm, *Review of Studies Employing GRE Scores in Predicting Success in Graduate Study 1952-1967* (GRE special report 68-1, Educational Testing Service, Princeton, N.J., 1968); *Summaries of GRE Validity Studies 1966-1970* (GRE special report 72-1, Educational Testing Service, Princeton, N.J., 1972).
5. A list of these studies is available from the author.
6. J. A. Creager, *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).
7. D. A. Rock, "Prediction of doctorate attainment in psychology, mathematics, and chemistry" (GRE preliminary report, unpublished, Educational Testing Service, Princeton, N.J., 1972).
8. J. A. Fishman and A. K. Pasanella, *Rev. Educ. Res.* **30**, 298 (1960).
9. R. L. Linn, *J. Educ. Meas.* **3**, 313 (1966).
10. A. Mehrabian, *Educ. Psychol. Meas.* **29**, 409 (1969).
11. B. Pitcher and W. B. Schrader, *Indicators of College Quality as Predictors of Success in Graduate Schools of Business* (Admission Test for Graduate Study in Business, brief No. 6, Educational Testing Service, Princeton, N.J., 1972).
12. N. E. Freeberg, *Psychol. Rep.* **20**, 911 (1967).
13. S. A. Kendrick, *College Board Rev.* **54**, 26 (1964).
14. L. R. Harmon, *Fourteen Years of Research on Fellowship Selection* (Office of Scientific Personnel, National Academy of Sciences-National Research Council, Washington, D.C., 1966).
15. J. A. Creager and L. R. Harmon, *On-the-Job Validation of Selection Variables* (technical report No. 26, Office of Scientific Personnel, National Academy of Sciences-National Research Council, Washington, D.C., 1966).
16. R. R. Reilly, *Critical Incidents of Graduate Student Performance* (GREB technical memorandum No. 1, Educational Testing Service, Princeton, N.J., 1971).
17. N. Frederiksen, F. R. Evans, W. C. Ward, *Development of Provisional Criteria for the Study of Scientific Creativity* (research memorandum 73-3, Educational Testing Service, Princeton, N.J., 1973); H. Witkin, *The Role of Cognitive Styles in Academic Performance and in Teacher-Student Relations* (research bulletin 73-11, Educational Testing Service, Princeton, N.J., 1973); M. Wallach, "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, 1972).
18. An earlier version of this paper was presented at a pre-session of the Council of Graduate Schools meeting in November 1972. I thank Jane Porter for her assistance in compiling the data on which this article is based.