File

Report on Graduate Admissions

J. Baron

What follows is a brief summary of the literature I could find on the validity of various criteria for graduate admissions, and of a study of our own admissions procedures. The issue arose because of some doubts about our reliance on GRE scores, with reinforcement for these doubts provided by M. A. Wallach's article (American Scientist, 1976, 64, 57-63), "Tests tell us little about talent". Wallach argues that test scores and grades are evidently good predictors of future test scores and grades but poor predictors of real accomplishment, no matter how this is measured. Accomplishment can, however, be predicted from past accomplishment. The implication is that if we want to select graduate students who will accomplish things in and after graduate school, we should weigh more heavily the evidence we have of their accomplishments as undergraduates, such as the papers they send us. (Test of creativity, by the way, do as poorly at predicting creative accomplishments as do aptitude tests.)

The most relevant studies cited by Wallach are those in which the quality of scientific accomplishments of scientists (including one study of psychologists by Marston, American Psychologist, 1971) is rated by peers. These ratings have low or sometimes negative correlations with the most widely used predictive criteria, such as grades and test scores. Wallach cites a number of studies showing prediction of accomplishment from past accomplishment, but none of these studies concerns itself with graduate school admissions, and none showed that accomplishment can be predicted from past accomplishment when it cannot be predicted from scores or grades in the same study. The possibility remains that the failures to predict accomplishment from grades and scores were due to the selection of samples with restricted ranges both in accomplishments and scores; studies showing that accomplishment can

be predicted from past accomplishment were based on less restricted samples, such as high-school students. It is possible that nothing can predict scientific accomplishment within the small range of ability examined. (Note: Accomplishment measures are no lower for minority high-school students, while grades and scores are.)

An interesting study is that of R. M. Dawes (Science, 1975, 187, 721-3) who found a negative correlation between GRE scores and undergraduate GPA among graduate students admitted to the University of Oregon in psychology. The correlation was positive for the applicants, however. This study testifies to the magnitude of the problem of restricted range. Since grades and GREs were used to admit students, the use of a high criterion for a weighted sum of these indices will insure such a negative correlation (if the usual assumption about bivariate normal distributions, etc., are approximately true).

The most informative paper, I think, is that of Willingham (Science, 1974, 183, 273-8), who reviewed a number of studies using various predictors and various criteria of success in various fields of science. The following table reports median validity coefficients from available studies:

	Criterion of Success:			
Predictor:	Graduate GPA	year 1 faculty ratings	attains Ph.D.	Psych. (GPA or Ph.D.)
GRE verbal	.24	.31	.18	.19
GRE quant	.23	.27	.26	.23
GRE advanced	.30	.30	.35	.24
undergrad GPA	.31	.37	.14	.16
recommendations			.18	
GRE composite	.33	.41	.31	

Another apparent finding was that the quantitative GRE was a better predictor than the verbal GRE for quantitative fields such as economics and engineering, and the verbal GRE was better than the quantitative for verbal

fields. Finally, letters or recommendation seemed to have both low reliability and low validity. (See also Hackman, et al., Ed. & Psych. Meas., 1970, 30, 365-374, for a smaller-scale study with some surprising , but ungeneralizable, findings.)

I should point out one problem in extrapolating studies of this sort to a particular case. All of these studies examined students who had been admitted to graduate school. To the extent to which a measure was used as a criterion for admission, the range on that measure will be restricted, and the measure will not predict anything very well. When a measure is not used as a criterion, it may predict better as a result of its range being greater. Thus, these studies tell us only about current practices of the graduate schools examined. High validity coefficients mean that a measure is not being weighed heavily enough by the schools in the sample considered. Low validity coefficients may mean not that a measure is useless, but that it is overused relative to other measures. An ideal admissions procedure (by one definition) would create a situation in which all predictors had equally low validity coefficients. The implication of this argument for us is that we need to examine our own admissions procedure. It is likely that our present criteria are different from those of others, and what we really need to know is what criteria we are overusing and what we are underusing. The second part of this report will deal with that problem.

It can be seen from Willingham's data that different predictors are good for different criteria of success, just as Wallach argued. This raises the question of what criterion is best, which, in turn, raises the question of the purpose of graduate education. If one were to decide that the purpose of graduate education is to produce contributors to the literature, one should rely heavily on evidence of undergraduate research productivity. This issue is not a simple one; we all know, for example, that number of publications does not measure real scientific advancement, and one paper by a smart

(and lucky) person might be worth the life output of a dummy. Inevitably, we all have our own views on such matters. However, I do think that we can agree on one criterion, namely, we want students who will complete the Ph.D. Given this criterion, Willingham's study suggests that grades are being somewhat overused and GRE scores are being somewhat underused by the schools studied.

To find out whether these conclusions apply to our own department, I examined (with a great deal of help from Isabelle Freedman) the records of students admitted between 1970 and 1977. Records from this period are relatively complete, and the procedures used during this period probably represent our current procedures. Based on available records, 21 students admitted during this period have since dropped out of psychology, for a variety of reasons ranging from personal to academic. Twenty-three students have obtained their Ph.D. and remain active in psychology. (Three students have obtained their degree, but their activity in psychology is in suspension; these were not considered in the analysis of the data.) The dropout group and the Ph.D. group were compared on a number of measures available from their application for graduate admission. (Numbers of students for each measure was available varied.) These were:

GRE scores;

undergraduate GPA;

presence or absence of publications (including submitted papers and convention presentations);

presence or absence of experience at research in a real laboratory;

number of psychology courses;

number of courses in mathematics and science (including logic);

quality of undergraduate institution (based on my own judgment, with Ivy League colleges and Swarthmore rated as 5, other prestigeous small colleges and U.C.S.D. as 4, major state universities as 3, and

places such as Georgetown and Syracuse as 2);

letters of recommendation (objective ratings only);

personal statements.

I read all the personal statements and rated them on four scales, trying very hard, even bending over backwards, to be objective. Each scale receives a score from 1 to 3. I do not think that my reliability at each scale was very high, although I am more confident about the total score. Sometimes I knew that there was something good about a statement, but I didn't know which category it belonged in, so I made an arbitrary decision. The scales were:

identity - evidence of a coherent and realistic plan, with origins in the students past explained and with a justification of the plan in terms of some purpose or ideology;

sophistication - knowledge of how scholarship works and what graduate school is about;

commitment to research;

display of knowledge of the complexities of some particular area, possibly by reference to papers enclosed with the application.

The results may be summarized in terms of point-biserial correlations between each index and group membership (with dropouts assigned 0, Ph.D.s 1).

Total GRE score correlated <u>negatively</u> with group membership, with r=-.39, which is significant at about the 0.01 level. Dropouts have <u>higher</u> GREs. This was slightly more true for verbal (-.33) than quantitative (-.22) scores; however, there were clear sex differences, with the negative correlation accounted for almost entirely by the quantitative score for females (-.47Q, -.08V) and the verbal score for males (-.32V, -.10Q). These results suggest that we have been relying too heavily on GREs, especially on the quantitative GREs for females and the verbal GREs for males. (Hirschberg and Itkin, <u>Am. Psychol.</u>, 1978, <u>33</u>, 1083-93, also find a negative correlation between verbal GREs and subsequent publications, p. 1089.)

Number of psychology courses showed a nearly significant negative correlation with activity (-.33) and publications a nonsignificant positive correlation (.19). Grades, ~~and~~ the ratings on the letters of recommendation and other measures showed essentially no correlation at all. Note that zero correlations suggest not that a measure should not be used but that we have been using it properly.

The negative correlation with GREs is most easily explained by the idea that we use two different kinds of criteria, GREs or something else, that the something else is the better criterion, and that the two criteria tend to be negatively correlated in the sample. The something else appears to be the personal statements and what they indicate. The correlation between the total statement rating and group membership was .41 (p less than 0.02). The best component was identity (.43) and the least useful was display of knowledge (.16). (In fact, total statement ratings and total GREs correlated -.41 over the whole sample.) These correlations do not fully capture the striking difference in the statements themselves. Many of the ~~active~~ Ph.D.-group statements and only a couple of the dropout statements read almost like summaries of fundable grant proposals; sometimes there were even specific projects proposed. Statements from the active group also tended to display a clear ideological commitment: only through physiology will behavior be understood, clinical research needs more rigor, the study of cognitive development will solve educational problems, etc. It is clearly the <u>form</u> of these commitments (as opposed to the content) that is predictive of success. Almost all of our illustrious recent graduates have changed courses in major ways.

In sum, I am convinced that we need to pay more attention to evidence of a coherent identity as a psychological researcher (including the content of recommendations, which I did not analyse), and less attention to GRE scores, if we want to get students who will not drop out. As the frequency

of dropping out increases, we should perhaps increase our efforts to prevent it, given the substantial commitment we make to our beginning students. On the other hand, we still need to worry about other criteria. For example, my own opinion is that our students' research sometimes suffers from their lack of mathematical and statistical sophistication. Unless we make more of an effort to inculcate such sophistication, I will continue to favor applicants who already seem to have it. Others will doubtless have their own pet concerns.