DOCUMENT RESUME

ED 160 647 TM 007 834

AUTHOR	Boldt, Robert F.
TITLE	Comparison of a Bayesian and a Least Squares Method of Educational Prediction.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RB-75-15
PUB DATE	Apr 75
NOTE	88p.
EDRS PRICE	MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS	*Bayesian Statistics; College Entrance Examinations; Grade Prediction; Graduate Study; Higher Education; *Multiple Regression Analysis; *Predictive Validity; Predictor Variables; Test Validity
IDENTIFIERS	*Cross Validation; *Least Squares Analysis; Sample Size

ABSTRACT

Least squares and Bayes methods were used in a cross validation study conducted for comparison purposes. The study applies to situations with the following conditions: predictor data are given on the same scales; criterion data may be given on different scales; and it is necessary to pool data even though criterion scale differences exist. Such a system may be needed for minority group prediction studies or graduate school prediction studies where the group sizes are small. Data for the study were taken from the files of the Validity Study Service of the College Entrance Examination Board. Also a very limited amount of data were supplied by a few American graduate schools. The Bayes method was somewhat better, but it was found that both methods yielded negative regression weights; when the absolute values of the weights were used, both the methods were improved and yielded results which were very similar in terms of evaluative statistics computed in the cross validation sample. (Author/CTM)

RB-75-15

RESEARCH BULLETIN

COMPARISON OF A BAYESIAN AND A LEAST SQUARES METHOD OF EDUCATIONAL PREDICTION

Robert F. Boldt

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
April 1975

COMPARISON OF A BAYESIAN AND A LEAST SQUARES

METHOD OF EDUCATIONAL PREDICTION

Abstract

The prediction systems under discussion apply where the following conditions obtain: Predictor data are given on the same scale, criterion scores may be given on different scales, and it is necessary to pool data even though criterion scale differences exist. Such a system may be needed for minority group or graduate student prediction where the group sizes are small. Least squares and Bayes methods are used in a cross-validation study conducted for comparison purposes. Data for the study were taken from the files of the Validity Study Service of the College Entrance Examination Board. A very limited amount of data were supplied by a few American graduate schools. The Bayes method was better, but it was found that both methods yield negative regression weights; when the absolute values of the weights were used, the methods were both improved and yielded results which were very similar in terms of evaluative statistics computed in the cross sample.

COMPARISON OF A BAYESIAN AND A LEAST SQUARES METHOD OF EDUCATIONAL PREDICTION[1]

Introduction

Very often the evaluation of the effectiveness of test scores for prediction is infeasible because those interested in such evaluation are unable to assemble a group of examinees for whom comparable criterion scores are available. In some population segments, such as minority groups, graduate students, and possibly various occupational groups, one often cannot find enough people at a single place where an acceptable criterion exists to conduct a statistical study of the predictive validity of selection instruments, or at least a study in whose results one can have confidence. It is more common to find small groups from the population of interest interspersed through a variety of locations, performing tasks that seem reasonably similar. Evaluation of the performances is made with reference to the group at a location but without reference to performances outside of that group. Thus, the groups may differ from each other in terms of average performance or in the variation in performance, but these differences may not be inferable from the corresponding statistics calculated using quantitative evaluations of performances made at each location. This type of problem was encountered in a study of the use of the Prueba de Aptitud Academica (PAA) in predicting the success of Spanish-speaking students in American universities (Gannon, Oppenheim, & Wohlhueter, 1966). In that study, efforts to accumulate usable data from six locations that apparently promised a reasonable supply of Spanish-speaking students yielded Spanish-speaking U.S. citizens in group sizes of 72, 8, and 15, and noncitizens in group sizes of 22, 23, 6, 27, and 32. Because of such low numbers of available cases, one would usually abort a study in which only separate validity studies at each location were contemplated. In a later attempt to repeat the study by Gannon et al., Carlson (1967) had to deal with group sizes as low as nine.

Similar problems were encountered in the development of the Comparative Guidance and Placement (CGP) battery of the College Entrance Examination Board (CEEB), which is a battery of psychological tests intended for guidance and placement use in American junior colleges. One research problem in the development of this battery was to choose a set of tests that would be valid for predicting success in each of a number of curricula. Although any one junior college would have a freshman class large enough to use in conducting a study, that class, when broken down by curricula, would become highly fractionated. Restrictions in class size were also necessary in order to accommodate institutions with limited testing facilities or with other problems in producing data. In this research (Educational Testing Service, 1968), the median of the average class sizes for the curricula was 69, but these average class sizes ranged down to 23 and 36, with an administratively enforced lower bound of 20. In a later reconstitution of the curricula groups these problems were alleviated somewhat, though for some new curricula groups it was necessary to use data from schools which could supply as few as 25 cases.

The problem of few cases at many locations arises also in research on black students. Cleary (1968) and subsequently Temp (1971) have encountered this problem in connection with the study of the validity of the Scholastic Aptitude Test when used for blacks and whites at racially integrated schools. The study was undertaken, in part, to determine whether one should use the same prediction system for forecasting grade point averages for blacks as for whites, or, in effect, whether the two regression lines are the same. To conduct the study, Cleary located schools with sufficient numbers of cases so that a school-by-school approach could be used; Temp's study followed the Cleary approach. In the opinion of the author, the Cleary approach is an excellent one, but whether it leads to results that are applicable to most schools, or perhaps to most blacks is open to question. If a study is limited to schools which enroll many blacks, then one has not observed schools where there are few blacks, and in other phenomena of racial mixture the nature of outcomes to be observed may depend on that mixture. A study of a large number of schools with few blacks remains of interest.

The problem of locating minority groups becomes even more difficult at a more selective level of education. In an as yet unpublished report, Schrader and Pitcher (1972) compared the regression functions for blacks and whites in American law schools. Of the five schools from which sufficient data were available for this study, one had data for only 44 blacks and the other for only 31; these cases were accumulated only by combining data across three and two years, respectively. In another unpublished study, the author (Boldt, 1971) reported a study of the validity of the Admissions Test for Graduate Schools of Business in which the group sizes ..d from 7, 10, and 12 up to 31. In this latter study, a large number of schools had been approached for data, and six thought they had sizable groups. When it came to producing usable cases, however, the very small groups listed above were all that were forthcoming.

Even without the complication of locating minority group students, the problem of conducting validity research on graduate education has been long exacerbated by a dearth of usable data. Summarizing attempts to do validity work from 1952 to 1967 on the Graduate Record Examinations (GRE), Lannholm (1968) reported 22 studies in which performance prediction research was done and the sample sizes by department are given. In these studies the combination of data by institution was avoided, but within a department it was sometimes necessary to accumulate data over several years. The average group size per department in these studies was 51, the smallest being 7 and the largest being 185. By far the largest of these groups were from the education area, the group of 185 being in secondary education, an area where the course work and majors are not interchangeable because different academic areas are involved. Indeed, as far as numbers of people are concerned, education is in a favorable position since graduate work is required for promotion or certification in many areas. But education also has many specialties with different requirements and very different course work. It is suspected that pooling people for a study because they are all in an education department may not be appropriate and, in any case, education departments are not representative of the rest of the graduate world. If the nine studies reported by Lannholm (1968) which deal with education departments are removed, the average size of the remaining group is 37 per department, the smallest still being 7 and the largest being 96. In 10 of these departments, 50 or fewer cases are involved.

Lannholm, Marco, & Schrader (1968), being aware of the sample size problem, attempted to accumulate graduate school data by inviting 32 departments in 15 different universities to participate in their research. As a result, data were received from 21 departments of 10 universities, with sample sizes ranging from 8 to 116. The average size of the groups by department was 45 even though these authors made a good attempt to get better and more numerous data than had been made by other authors.

Lannholm continued hi chronology of validity studies with his 1972 summary of GRE validity studies from 1966 to 1970. In these studies the education departments continued to provide sizable groups of cases, yielding an average of 130 cases, though group sizes as low as 12 and 27 were reported. In departments other than education, the average number of cases reported was 63 with a range from 5 to 147--it should be mentioned that one of the schools with a group size of 137 had accumulated its data over a period of 11 years. Clearly, heroic measures are needed to produce a study.

Central Prediction Approaches

Even though the problems of the availability of comparable data are well known to researchers in the social sciences, the commonly available regression models contain no provision for inclusion of sets of criterion data that do not lie on the same scale. In addition, there are problems where sets of predictor data may not lie on the same scale. For example, the grade point average at one undergraduate institution may not be comparable to that from another sending institution for predicting performance at the graduate level. Bloom and Peters (1961) were among the early researchers to investigate the problem of grade adjustment with very significant results, though results have not held up in later studies. Tucker (1963) has mentioned certain technical problems that might account for this and developed a number of formal models for central prediction. These models were developed in the context of prediction at the undergraduate level using empirically adjusted high school grade point averages as well as empirically developed adjustments for the grade scales of the receiving colleges. Tucker's models are least squares models, whereas those discussed by Potthoff (1964) in a paper sponsored by the Educational Testing Service (ETS) are models using maximum likelihood estimation under the commonly made assumptions of joint normality. Bashaw (1965a, 1965b) has reported central prediction models that are formally identical to that of Tucker, though with a somewhat different computational scheme. Linn (1966) in surveying research on grade adjustment in the sending institutions has supplied a more detailed discussion of these problems and models.

Tucker's paper includes a prediction model which is responsive to an even more general formulation of the problem of prediction in allowing, in addition to adjustments to grades from the sending and receiving institutions, for differences in the types of institutions involved, e.g., an engineering school as opposed to an institution concerned primarily with liberal arts. The research problem to which the present paper is addressed is quite a bit more determined than the problem to which Tucker's predictive model is addressed because only one kind of sending institution will be studied in a particular solution and no adjustment of grades from sending institutions will be contemplated. Nevertheless, the least squares approach applied here is in the spirit of Tucker's predictive model where the residuals whose squared sum is minimized contain no adjustment on the criterion. This is opposed to Potthoff's maximum likelihood solution which determines a transformation on the criterion scores; i.e., the grades at the receiving institutions. The least squares solution used in the present paper is described in Appendix A and was originally developed by the author for use in the Gannon et al. (1966) study and was subsequently used by Carlson (1967) and in the development of the CGP battery (ETS, 1968). Appendix B contains a least squares solution that adjusts the criterion scores and, although it is not as desirable from a prediction point of view for that reason, it has certain computational advantages that would be useful in test selection procedures.

Bayesian Approach

The approach taken by some of the researchers in central prediction is to etain an assumption of structure similar to the linear form used in multiple regression but to allow additional linear adjustments that account for the source of the particular data set. Thus, the large body of data is summarized in relatively few parameters, though more than would be required by a single regression model and less than would be used by the fitting of a whole set of individual constants to every data set. But the previously discussed central prediction procedures have no way to take advantage of partial, or vague, information available prior to the estimation study. For example, if one were to estimate the regression of college grades on ACE scores, one could examine results from a variety of studies that would indicate something about the likely range of the coefficients to be found. Reasonable ranges for the means and variances could be set up as well as for the correlation coefficients. Surely, such a study would not be the first of its kind with entirely new knowledge being made available, but it would be at least a partial affirmation of existing knowledge though applied in a slightly new context. Where several schools are involved, one would want to incorporate the notion that they are more or less similar. One would certainly not want to proceed under the assumption that all schools are uniquely different, conceivably, and that no prior information is in existence.[2]

Being aware of these problems through discussion with M. R. Novick, Lindley has developed a Bayesian approach to the type of problem of interest in the current study. Indeed, a series of Educational Testing Service Research Bulletins (Jackson, Novick, & Thayer, 1970; Lindley, 1969a, 1969b, 1970a, 1970b) have documented the development and discussed applications of the Bayesian approach to problems in educational prediction and guidance. Lindley's approach has the incorporation of vague prior information, allows for local differences in regression equations, and produces a method of proper balancing of sparse local data against data from the entire set of schools under study. The estimation procedures use a likelihood function weighted by a prior distribution, hence are more similar to Potthoff's (1964) methods than to least squares. Compared to that of Potthoff, Lindley's method would clearly seem to be the method of choice, since it employs a predictive model in the sense of Tucker and because it incorporates the desirable Bayesian features mentioned above.

Lindley's model (1970a, 1970b) incorporates the following: the likelihood is characterized by grades and predicted grades for which linear regression holds and whose discrepancies are normally and independently distributed; coefficients of linear regression have independent priors across colleges but within colleges have joint normal priors which are, for a college, characterized by a vector of means and a dispersion matrix which is exchangeable with the vectors of means and dispersion matrices of other colleges; the priors for the exchangeable vectors of means and dispersion matrices are uniform and Wishart, respectively, in form; the dispersion parameters from the likehood are exchangeable and have independent inverse chi-square priors whose central tendency parameter has a prior which is chi-square in form.

Need for Comparative Experiment

The list of assumptions above is rather long and parts of it may be untenable: the normality of errors of (b) is clearly incorrect because all of the variables are bounded, because, even if college grades were normally distributed, the test scores are not; and because the homoscedasticity implied by the independence assumption is incorrect. Furthermore, the assumption of uniform distribution on the means of (c) is quite unreasonable considering the rather narrow range that can be involved. In fact one could probably reject any assumption one could test, given enough data, and yet it is not proposed to abandon the approach. Rather, the assumptions, together with the estimation procedure can be used for data for a back sample, and the prediction system can be tested in a cross sample as is often done to observe the effects of shrink due to capitalization on chance in the back sample. If one system works better than another, it is better even if arrived at through some questionable assumptions.

Novick, Jackson, Thayer, & Cole (1972) conducted such an experiment using data from the Basic Research Service of the American College Testing (ACT) Program. Their comparison of the Bayes prediction system was made with reference to the standard least squares procedure where estimates are made at each location and independently from the others. In this study, 22 schools were chosen, and data from each provided a back and a cross sample. The back sample, which was used to develop regression parameters, was collected in 1968; the cross sample, used to evaluate the prediction system, was collected in 1969 from the same schools. The back sample sizes averaged 246, with a high of 739 and a low of 113; these sizes are considerably in excess of the troublesome ones with which one often deals, and one might expect that shrink effects usually observed in the cross sample would be at a minimum. In fact, the drop in validity averaged over schools is very small, going from .51 to .47. This suggests that capitalization on chance is not a very potent factor and leads one to expect that the use of prior information afforded in the Bayesian system would not be crucial.

To compare the Bayesian and least squares results, Novick et al. (1972) used four indices. Each of these indices uses a predicted score and an actual score. The predicted score uses a prediction function whose parameters are computed in a back sample and whose arguments are the four variables of the American College Test; there is a different set of parameters for each school, and Novick et al. (1972) presented the four indices for each college. The indices were (a) the familiar mean square error (MSE), which is the average of the squared differences between the predicted grade and the observed grade, (b) the average absolute error (AE), which is the average of the absolute values of the difference between the predicted grade and the observed grade; (c) the zero-one loss (ZOL), which is the average of a variable that is zero if the prediction is within half a standard deviation of the actual grade on the observed grade scale and one otherwise; and (d) the correlation between the observed scores and the predicted scores (COR). The averages were of the indices, giving least squares results first and Bayes second, (a) .56 and .55 for MSE, (b) .58 and .58 for AE, (c) .56 and .56 for ZOL, and (d) .47 and .48 for COR. Although the sample sizes are not given for the 19(samples, the author seriously doubts that they are large enough to detect significance for the small differences shown; if they are different, the difference is certainly not of much practical interest.

In smaller samples, the effects of capitalization on chance are more marked, and one would expect to observe an enhancement of the value of the Bayes approach over that of the usual regression approach in small samples.

To observe whether such might be the case, Novick et al. (1972) drew a 25% random sample of the 1968 data for use as a back sample and crossed the results into the 1969 data. The average values of the resulting goodness of fit indices are as follows, giving the least squares results first and the Bayesian results second: (a) .62 and .56 for MSE, (b) .61 and .59 for AE, (c) .59 and .56 for ZOL, and (d) .42 and .47 for COR. A further advantage of the Bayes approach was that it did not yield negative regression coefficients as did the least squares method; except in special cases one does not usually accept negative coefficients in a system in this context. The average back sample size was 61, with a range from 26 to 184, which is not as small as sometimes occurs but is small enough to indicate some superiority of the Bayesian method to which the indices above testify. Novick et al. (1972) point out that relatively more gain would be made with even smaller divisions, such as splitting the college groups by sex.

Least Squares vs. Bayes

Although the Bayes approach due to Lindley appears promising in the work of Novick et al. (1972) as described above, the question can be raised as to whether the superiority of the system as compared to the least squares system is because Bayes is better than least squares or whether a poor least squares approach was used. Most researchers would not use a standard regression in a situation like that of the 25% ACT sample. Indeed, Novick and Jackson (1974), using (probably different but) exchangeable regression coefficients across schools, but with different intercepts at each school recently reported that the mean square error in a cross sample approached that of the Bayesian method. In Gannon et al. (1966), Carlson (1967), the CGP battery (1968), and Boldt (1971), the least squares method of Appendix A was used in order to get over the problem of the small number of cases at each location. In this method, the number of parameters to be fitted is equal to the number of predictors less one, plus twice the number of schools; with a large number of schools the relative weights for the predictors would be greatly overdetermined, and one would expect the system to be quite stable with few cases per school. Since a least squares approach which features more overdetermination would be more appropriate and would seemingly be competitive with the Bayes approach, a study comparing the two was undertaken.

Sample

In spite of the researcher's ineluctable confrontation with the sample size problem when attempting research at the graduate school level, it was hoped that the present study could be accomplished using data from that source. The letter in Appendix C on GRE Board letterhead was sent to the graduate deans of 95 American graduate schools. The appropriate second page was included, depending on whether the school would be asked to contribute data on psychology or economics graduate students, or on both. Eighty-one departments of psychology and 54 departments of economics were approached; these departments were listed by Lannholm (1967) as requiring the GRE aptitude examinations as well as the advanced examination in the field of specialization. Enclosed with the letters to the deans were copies of the GRE Validity Study Questionnaire and letters to the appropriate department chairmen. These approaches were made in November 1970; four months later a follow-up letter, a second copy of the questionnaire, and a copy of the original letter to the department chairmen were sent to each nonresponding department chairman. These solicitations emphasized intended convenience in supplying data and solution to the common problems faced in doing research at the graduate level. In particular, the departments were urged to respond even though they might have few cases to supply. Fifty-seven or 70% of the psychology departments and 36 or 67% of the economics departments eventually replied. Of those who responded, 32 psychology departments and 15 economics departments either indicated an unwillingness or refusal to participate or stated that they would not be able to provide useful data. Reasons included the fact that no data were available for Ph.D.'s, that no graduate grades would be made available because of confidentiality, that GRE score requirements were not actually enforced and were available on only a few people, and that no advanced test scores would be available. Also, five economics and nine psychology departments indicated that to search the graduate files would work a hardship on their clerical staff. Two economics and five psychology departments gave both kinds of reasons of the total who responded, 25 psychology departments and 21 economics departments were usable in the study. Of these, 16 of the psychology departments and eight of the economics departments actually computed routinely a graduate grade point average.

The eight largest of the 16 psychology departments were contacted. Based on response to the GRE Validity Study Questionnaire, data for approximately 60 Ph.D.'s and 60 master's graduates should have been received, ranging from 40 to 80 for the Ph.D.'s, and from 15 to 105 for the master's. Of these, only four supplied any usable data on students for whom grade point averages were also supplied, and these gave 4, 21, 8, and 7 cases at the Ph.D. level and 35, 30, 60, and 17 at the master's level. These four schools also supplied data for 31, 7, 24, and 4 students who got both the master's and the Ph.D. degrees. One school provided data for two doctoral candidates; one school ran into difficulties when it came to actually accomplishing the clerical work, and at the last school it proved impossible to call or get a return call from the appropriate contact.

Three of the largest economics departments were contacted based on responses to the GRE Validity Study Questionnaire data for approximately 45, 40, and 30 Ph.D.'s and 75, 70, and 60 master's candidates. Actually, no data were received; one school would not respond to telephone calls, one required financial support for the clerical work, and one reported it was "plodding."

Clearly, if the GRE data were to stand on their own, the project was in trouble; the remaining schools that provided graduate grade point averages had only a few candidates and with similar attrition would not be helpful, and the schools for whom transcript analysis would be required were also schools for whom few returns could be forecasted. If the present study had been one with the main emphasis on the validity of the GRE, one would have seriously considered a course of action in which a large number of graduate schools with only a few students would be solicited for data; in doing so, one would be in the kind of situation for which the models under discussion are designed. Since the point of the present project is methodological, a better source of data was required.

Analysis of VSS Data

The files of the CEEB Validity Study Service (VSS) were examined to find data in which colleges had, on two successive occasions, participated in a validity study using comparable groups on both occasions. Data for 12 such colleges were obtained, with data for both males and females. The back sample consisted of 25 cases pulled at random (Tausworthe, 1965; Whittlesey, 1968) from data collected in the first year of participation by the colleges in the VSS. Since there were samples for males and females for all but two schools, the randomization was done using a list of males separately, a list of females separately, and then a merged list with no control on the composition by sex. Two of the schools provided data for females only; hence, the back sample for the analysis of "combined" data consists of an independent random sample from the same group used in the analysis for female cases. The average cross sample sizes are 134, 173, and 285, with lows of 48, 74, and 74, and highs of 204, 353, and 555 for the male, female, and combined groups respectively. The low for the female group is equal to the low for the combined group because the sample with 74 cases was not part of the male group, but was taken from one of the two colleges for which no data for males were available. Tables 1, 2, and 3 contain summary statistics for the schools involved for males, females, and combined groups, respectively.

As with the study of Novick et al. (1972), this analysis is concerned with the relations between a predicted grade calculated using a prediction function whose parameters are developed in the back sample and whose arguments are the test scores (V and M) and a grade point average (UR) observed in the cross sample. For each student in the cross sample, there is a predicted and an observed grade point average. The correlation between these is COR, and the difference between these is called the "residual." These residuals are used to compute AE and ZOL, as defined earlier. In addition, an average residual (AR), the variance of the residuals (VR), and the low and high residuals will be reported. All of these indices of goodness of fit of the prediction systems will be reported by school for both the Bayesian and least squares systems as applied to the data of males, females, and the combined group. They will also be presented for various combinations of the Scholastic Aptitude Test Verbal (V) and Mathematics (M) and the high school record (H). No adjustment of the H is included in the models considered.

Tables 4, 5, and 6 show the results for the index COR for males, females, and the combined sample. Notice that in these tables the Bayes and least squares predictions are about equally good with the exception of the negative entries. For these entries, the vagaries of the back sample are such that a negative multiplicative constant is needed to put the predictions on the grade point scale for the particular college because the predictions correlate negatively with the grade point average. For example, at college G the V scores of the males correlated -.01; therefore, the sign of that predictor will be reversed. But even if the back sample results indicate it, one does not accept that the correlation of V with grade point average is negative at this college or probably at any other unless the grade point scale is inverted; in practice, the negative weight would simply not be used on a priori grounds. The author is aware of this sort of reversal, having encountered it in the data for black students at the graduate schools of business (Boldt, 1971), and feels that, where such sign reversals are found, a reasonable practice to adopt would be one in which the absolute value of the multiplicative constant developed in the back sample is used in the cross sample together with an additive constant which is adjusted (as described in Appendix A) to account for the change in sign of the multiplicative constant. The additive constant would not affect Tables 4, 5, and 6 since the correlation coefficient is invariant under an additive transformation, but the sign would be reversed and one can see that, with the sign reversal the Bayesian and least squares systems are about equally good. The author wants to stress that this change in the sign was not a change suggested by the data but was intended for use prior to the collection of any data in connection with this study. A further consideration of the sign change appears in the Discussion and Conclusion section of this Bulletin.

Tables 7, 8, and 9 contain the average values of the residuals (AR) found in the cross samples. Though the entries where the negative multiplicative constant occurred are footnoted in these tables, the values entered are calculated with the adjustments referred to in the paragraph above. Examination of Tables 7, 8, and 9 reveals a slightly larger number of cells in which the average residual from the Bayes system is smaller than the average residual from the least squares system. These errors are not trivial in all cases and are highly responsive to fluctuations in the additive constant, a matter which will be referred to later.

As in Tables 7, 8, and 9, Tables 10 through 18 contain entries which are corrected for the negative correlations of predictor sets with the criterion in the back sample. However, Tables 10, 11, and 12 contain as entries the variances of the residuals, VR. Like the data in Tables 7, 8, and 9, the advantage seems to be slightly to the Bayes system in terms of the frequency of cells in which the VR is smaller for Bayes than for least squares. On examination it can be seen that cells which are footnoted c are not necessarily the ones in which the least squares system fits less well than the Bayes; the use of the absolute value of the multiplicative constant seems to have been reasonably successful. It should be noted that the index VR does not provide information about the additive part of the transformation that puts the predictions on the college scale, since VR is invariant under linear transformation of the variables.

Tables 13, 14, and 15 contain the values of ZOL, and Tables 16, 17, and 18 contain values of AE. No particular advantage for the Bayes system is observed using the ZOL measure.

Tables 19, 20, and 21 contain the parameters of the Bayes and least squares solutions. Note that in contrast to the study by Novick et al., some of the regression coefficients are negative in the Bayes system. Table 22 contains validities of the predictions using the Bayes solutions shown in Tables 19 through 21, as well as the average Bayes weights and the least squares weights. Table 22 also contains the validities that would be obtained if the Bayes solution were obtained merely by reversing the sign where the weights are negative. Note that a solution with positive weights is better than one with negative weights even for a cross sample on the same school from which the negative back sample weights were derived. This reversal of weights is treated further in the Discussion and Conclusions section of this bulletin.

Analysis of GRE Data

Despite the scarcity of data from the graduate schools, a back sample analysis of the data was conducted. In interpreting the results of this analysis, the reader should keep in mind that the returns are highly selective in the sense that the ability to supply data for the study classifies the participating institutions as atypical. Data were received from four schools and cases are identified as receiving a doctorate or as terminating with a master's degree, allowing eight classifications of students. Table 23 contains descriptive data by school for these groups and for the combined educational groups. Prediction analyses were conducted for each school using all eight classifications (Combined 8), the terminal master's only (Master's), the doctorate only (Doctorate), and the school, ignoring the degree received (Combined 4). Validity coefficients for these groups are presented for various combinations of system predictors in Table 24. These system coefficients squared give the percent of variance accounted for over and above the group means by the predictors. The computation of the coefficients for the least squares system is described in Appendix A, equation (11), and for the Bayesian system the system coefficients are calculated by combining correlations of weighted sums, the weights being the regression parameters estimated using the Bayes approach. Examination of Table 24 shows that the highest system validity coefficients are those in which the most parameters are fitted. For example, parameters are added as predictors, and one may note from Table 24 that the coefficients increase as one moves down the table. Also, the Bayesian system fits more parameters than does the least squares, especially as predictors are added. One may also note that the discrepancy between the least squares validities and the Bayesian validities increases as predictors are added; as each predictor is added, one parameter is added to the least squares system, but to the Bayesian system as many parameters are added as there are schools (four in this case). Therefore, the trends noted in Table 24 may be the results of capitalization on chance, rather than reflections of reliable trends in the data. As a check, Table 25, which gives system validity coefficients in the back sample of CEEB Validity Study Service data, is offered for comparison with Tables 4, 5, and 6. It can be seen that the least squares and Bayesian coefficients are about the same for larger numbers of predictors. Bayesian coefficients are smaller for the single predictor case (the single predictor case involves more constraint for Bayes than for least squares).[3] In Table 25 the validity of the VMH composite is about that of U alone, and in Tables 4, 5, and 6 the school coefficients are about the same as in Table 25. Similarly in Table 24 the least squares validity of VQPU is about that of U alone; thus, if a similarity with CEEB Validity Study Service data holds, one may expect system cross-validities of about .3 to .35 using the least squares predictors. An important question about the data in Table 24 is whether the large validity coefficients for the four predictor Bayes systems would hold up under cross-validation. It is the author's impression that they would not since the Bayesian system is adding only about 18 correlation points for an increase of almost 12 parameters.[4] It is, of course, a matter for additional study whether the Bayesian system produces results that would stand up with such a paucity of data, but in the author's opinion it would be extremely optimistic to accept the validities in the range of .5 from the bottom line of Table 24. It may be reasonable to expect, however, validities in the mid-thirties, provided the undergraduate average is included.

A further reason to question the validity of the regression composites in a cross sample comes from examination of the regression coefficients. In the Bayesian solution for the Combined 8 groupings, only for U alone and for Q and U together were the regression coefficients positive for all schools. For the doctorate groups, the regression coefficients were positive for all schools only when U was used alone. For the combined grouping, regression coefficients were all positive only when using P or U alone, Q and U together, and P and U together. This means that, for all other combinations of variables, one would be using negative regression weights in a new sample, and these combinations include, for example, the four variable predictor set that yielded the back sample system coefficient of .52 in Table 24. The author does not accept the conclusion that negative coefficients are correct for other samples but considers them an accident of these data.

Discussion and Conclusions

Problems in data collection have been discussed in some detail, where only a few cases are available with criterion data on a common scale, or at least on a scale which is known to be common. Minority group research is given as an example of a situation in which such problems arise, and the problems increase when graduate student populations are involved. The difficulties encountered in this study definitely make it clear that improving the state of knowledge in these very important areas will require the cooperation and effort and even some trouble on the part of many institutions that could provide data. To gain the cooperation of institutions probably requires convincing them that the solution of the problem to which the study is addressed is one in which they have an interest. Unfortunately, since the methodology under study is one designed to deal with small samples at many places, the sample size itself may preclude the development of a perception by the parent institution that the group for which the data are collected constitute in themselves a cause for concern, or a reason to take the trouble. It is hoped that some effective approach to the data collection problem will be found in the context of graduate education.

The present study asks whether the Bayesian system and estimation procedures, used by Novick et al. (1972), would prove superior to the least squares system used here (Appendix A) which, in contrast to the usual regression system, allows pooling of data across colleges. It was found that, if negative multiplicative parameters are developed using the least squares system are converted to positive and if the additive constants are adjusted accordingly, the least squares system and the Bayes system are about equally good in a cross-validation study of prediction of first-year college grades. Five indices were used to indicate the fit of the prediction in the cross sample: the validity (COR), the average residual (AR), the variance of the residuals (VR), the zero-one loss (ZOL), and the average error (AE). The difference between estimation methods produced very little variation in these figures of merit.

The least squares estimates were subjected to sign reversals where negative multiplicative parameters occurred (the entries where this happened are noted in Tables 4, 5 and 6). Clearly, had the reversal in sign not occurred, the values of COR would have been negative--prediction would have been backward. Other indices of fit in the back sample with the exception of AR would have suffered as well. But in the tables the Bayes and least squares results appear to be about equally as good. Therefore, before the sign change Bayes was better.

Realizing that the treatment of the two methods had not been entirely symmetric, the author examined the Bayesian regression coefficients to see if some of them were negative and might be changed. Some were indeed found, their signs reversed, and the results of that reversal are presented in Table 22. It can be seen that reversal of the negative signs in the Bayes formulae improved the Bayes predictions, also, as the author expected. Symmetric treatment, treatment which leaves the signs alone in both systems or treatment which changes the signs in both systems, seems to leave the Bayes system with a slight advantage.

Readers will undoubtedly differ on whether the sign changes are acceptable. The author justifies them on several grounds. First, experience shows that when aptitude and grades are used to predict later grades, the regression weights are overwhelmingly positive when samples are adequate. Second, the back samples are small. Third, the changes were possible without post hoc reliance on cross sample criterion data, a necessary characteristic of any acceptable estimation procedure. Fourth, positive weights make better policy sense. These points are amplified below.

The least squares system used here is particularly prone to sign change errors by college as examination of Appendix A reveals. There it can be seen that the prediction formula is of the form Az + B where z is a linear function of the predictors; the parameters of z do not depend on the particular college. Therefore, the coefficients of Az + B are estimated by the formula for regression of grades on z at particular colleges. These weights are based on only 25 cases so if one obtains a negative multiplicative constant for one college and if one must recommend a prediction formula for that college, one has to believe that aptitudes and grades predict backwards at that college to recommend the use of the negative value (in the present data all the scales are arranged so that "good" grade point averages are large ones). Therefore, in substantive terms a negative value for A means that increasing Verbal, Math, and undergraduate grade point average implies decreasing grade point averages. The author declines to accept that the colleges here are that strange; therefore the sign reversal or a zero weight is indicated. Zero might be considered as a solution because one might think that a negative weight estimates a zero; the zero might also be accepted as the estimate for a least squares objective function like that of Appendix A but with the explicit constraint that multiplicative weights must be nonnegative; but in other studies the predictors used here yield positive correlations with college grades with overwhelming frequency and the use of a multiplicative weight of zero would limit one to cross validities of zero--one knows one can do better.

Accepting that positive weights are needed, what should their magnitude be? Intuition says that the variance of the predictions should probably be somewhat less than that of the criterion; as a basis for producing that variance one might assume that the correlation coefficient was about right in size but wrong in sign. If so, the weight should be reversed in sign as was done with the results here. But other solutions in the form of other least squares objective functions were sought; only one appeared worth checking empirically. In this alternative least squares method the quantity w for a college was taken as proportional to the inverse of the multiplicative constant for that college thus ensuring the existence of a relative minimum for the objective function in the region where the multiplicative constants for the colleges are all positive. In this case the estimator of the multiplicative constant for a college turns out to be the ratio of the standard deviation of the grades at the college to the standard deviation of the z's at the college; the estimator equates the variance of the predictions to that of the criterion. As one might expect from the intuition mentioned above, this estimate does not work as well as the simple sign change that was actually used. Results obtained using the ratio of the standard deviations were not tabled since the estimate is of no further interest.

The discussion above deals with the least squares estimates. Bayes weights were also reversed to attain a symmetry of treatment of the two methods, since in the author's judgment positive weights would improve prediction, in part because the overwhelming body of evidence indicates that such weights should be positive. The other part of the reason for rejecting negative weights can be appreciated if one supposes that a negative weight for math was obtained at a college, and that that weight was incorporated into the admissions policy. Then one would almost certainly encounter a candidate whose grades and verbal scores were good but whose math was so high it offset his grades and verbal ability and led to his rejection. To explain to the parents and other institutional officials that the high math score makes him lose the admissions race would seem to the author to be exceedingly difficult. Especially would this be so where the weights arise from limited quantities of data for the college, a situation for which the current methods are intended. For political reasons one would want to be absolutely sure that the negative weight is correct, and this certainty would be needed in a situation where the great bulk of evidence suggests that the weights are positive and that the negative weights are a product of instability due to the limited sample size. The author's judgment was that the negative weights are probably not right, and that prediction would be improved if the signs were reversed. That proved to be the case.

The sign reversals used in the present study could be applied in prediction contexts less familiar than that of educational prediction where such well-known variables as V, M, and H are used. If one were using the least squares method, one would merely reverse the signs of the few schools which seemed to work backwards from the rest. If one were using the Bayes method, one would reverse the sign of the predictors where only a few differed in sign from the others. In either case, if more than only a few parameters showed difference in sign, one might suspect that institutions are being grouped which should not be. Where such sign reversals are done in the least squares system, the additive constant is determined by choosing a value so that the average back sample residual for a school is zero. This change is rather simple, but the change in additive constant for the Bayes model is more complicated and has yet to be worked out.

It is interesting to note that the arbitrary sign changes are much in agreement with that part of the philosophy of Bayesians which says that prior knowledge should reasonably be expected to influence the inferences one makes from sets of data. In this case one would prefer, of course, that the Bayes system be set up so that the occurrence of negative weights is unlikely or simply impossible. But Lindley and o er statisticians write for many applications and the generality of their methods would be limited if they included priors where negative coefficients were not possible. It is reasonable to suppose that measurement specialists might seek prior distributions which are appropriate to their special context of application and that such changes would improve estimation in that context, even though the results would not be as generally applicable as would Lindley's. Such change and improvement in the models for use in the data of interest to educational measurement would constitute progress in the science of the subject.

References

Bashaw, W. L. A FORTRAN program for a central prediction system. Educational and Psychological Measurement, 1965, 25, 201-204. (a)

Bashaw, W. L. Central-prediction-system models. American Educational Research Journal, 1965, 2, 151-162. (b)

Bloom, B. S., & Peters, F. R. Academic prediction scales. New York: The Free Press of Glencoe, 1961.

Boldt, R. F. Interim report: ATGSB Negro validity study. Report to the Research Committee of the Council of the ATGSB, 1971.

Carlson, A. B. Further results on the validity of the Prueba de Aptitud Academica in United States colleges and universities. Statistical Report 67-67. Princeton, N.J.: Educational Testing Service, 1967.

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Educational Testing Service. Progress Report: Comparative Guidance and Placement Program. Prepared for the College Entrance Examination Board, Princeton, N.J., September 1968.

Gannon, F. B., Oppenheim, D., & Wohlhueter, J. F. A validity study of a Spanish Language Scholastic Aptitude Test in United States colleges and universities. Statistical Report 66-2. Princeton, N.J.: Educational Testing Service, 1966.

Jackson, P. H., Novick, M. R., & Thayer, D. T. Bayesian inference and the classical test theory model II: Validity and prediction. Research Bulletin 70-32. Princeton, N.J.: Educational Testing Service, 1970.

Lannholm, G. V. Graduate schools requiring or recommending GRE scores for initial admission to graduate study. Graduate Record Examinations Special Report 67-3. Princeton, N.J.: Educational Testing Service, 1967.

Lannholm, G. V. Review of studies employing GRE scores in predicting success in graduate study, 1952-1967. Graduate Record Examinations Special Report 68-1. Princeton, N.J.: Educational Testing Service, 1968.

Lannholm, G. V. Summaries of GRE validity studies 1966-1970. Graduate Record Examinations Special Report 72-1. Princeton, N.J.: Educational Testing Service, 1972.

Lannholm, G. V., Marco, G. L., & Schrader, W. B. Cooperative studies of predicting graduate school success. Graduate Record Examinations Special Report 68-3. Princeton, N.J.: Educational Testing Service, 1968.

Lindley, D. V. A Bayesian solution for some educational prediction problems. Research Bulletin 69-57. Princeton, N.J.: Educational Testing Service, 1969. (a)

Lindley, D. V. A Bayesian solution for some educational prediction problems, II. Research Bulletin 69-91. Princeton, N.J.: Educational Testing Service, 1969. (b)

Lindley, D. V. The estimation of many parameters. Research Bulletin 70-16. Princeton, N.J.: Educational Testing Service, 1970. (a) Also, V. P. Godambe and D. A. Sprott (Eds.), Foundations of statistical inference. Toronto: Holt, Rinehart and Winston of Canada, Ltd, 1971. Pp. 435-453.

Lindley, D. V. A Bayesian solution for some educational prediction problems, III. Research Bulletin 70-33. Princeton, N.J.: Educational Testing Service, 1970. (b)

Linn, R. L. Grade adjustments for prediction of academic performance: A review. Journal of Educational Measurement, 1966, 3, 313-329.

Novick, M. R., & Jackson, P. H. Further cross-validation analysis of the Bayesian m-group regression method. American Educational Research Journal, 1974, 11(1), 77-85.

Novick, M. R., Jackson, P. H., Thayer, D. T., & Cole, N. S. Estimating multiple regressions in m-groups: A cross-validation study. British Journal of Mathematical and Statistical Psychology, 1972, 25, 33-50.

Potthoff, R. F. The prediction of college grades from College Board scores and high school grades. Institute of Statistics Mimeo Series 419. Chapel Hill: University of North Carolina, 1964.

Schrader, W. B., & Pitcher, B. Predicting law school grades for black American law students. Draft report to the Test Development and Research Committee of the Law School Aptitude Test Council, 1972.

Tausworthe, R. C. Random numbers generated linear recurrence module two. Mathematical Computations, 1965, 19, 201-209.

Temp, G. Test bias: Validity of the SAT for blacks and whites in thirteen integrated institutions. Research Bulletin 71-2 and CEEB Research and Development Report 70-71, No. 6. Princeton, N.J.: Educational Testing Service, 1971.

Tucker, L. R. Formal models for a central prediction system. Psychometric Monograph No. 10. Richmond, Va.: William Byrd Press, 1963.

Whittlesey, J. A comparison of the correlational behavior of random number generators for the IBM 360. Committee of the ACM, 1968, 11, 641-644.

Footnotes

[1]This research was supported by the Graduate Record Examinations Board. Its conclusions are those of the author and not necessarily those of the Board or Board members.

[2]In practice, even researchers who use classical statistical methodology use available prior information quite often. For example, if a negative correlation of verbal test scores with college grades were found, the researcher would examine the computation processes very closely, as he would the sample, probably until some reason were found to judge one or the other pathological. And his behavior is not entirely unreasonable because he "knows" that if a new sample were correctly drawn and correctly analyzed, the resulting validity would be positive. He would behave similarly if the analysis produced a validity of .95, though he might be somewhat less reticent to record the result.

[3]For the single predictor cases, the least squares and the Bayes systems fit twice as many parameters as there are schools, but the Bayes parameters are interconnected through the prior distribution. There being no interconnection of the least squares parameters, one may regard the least squares validities as being more subject to shrink in a cross sample. But as predictors are added, the Bayes system adds one parameter per school per predictor while the least squares system adds only one parameter per predictor. Even though the Bayes parameters have some interconnection through the prior distribution, it seems to the author that for larger numbers of predictors the Bayes system must be more subject to shrink.

[4] It is common to compare correlation gains to the number of parameters fitted. This is done because the potential for validity shrink in a new sample is greatly increased as the number of parameters is increased. Although this practice has grown up in a least squares context, it is supposed that the same comparison would apply with the Bayesian system. A gain of 18 correlation points for 12 parameters is quite small from this point of view.

Table 1

Summary Statistics on High School Record (H), Undergraduate Record (UR), CEEB Verbal (V), and Math (M) for Undergraduate Male Cross Sample

	A	B	C	D	E	F	G	H	I	J	K	L
No. Cases	179	48	116	204	151	109	93		129	112		202
Means												
V[a]	51	30	57	52	47	60	47		43	39		50
M[a]	56	34	60	56	51	63	49		46	41		54
H	3.02	2.57	3.10	3.05	2.62	3.13	2.84		2.49	1.94		2.91
UR	2.54	2.03	2.52	2.80	2.20	2.05	2.31		2.01	1.80		2.53
Standard Deviations								WOMEN ONLY			WOMEN ONLY	
V	9	6	9	11	9	9	9		9	8		9
M	8	6	9	12	10	8	10		8	8		4
H	.44	.52	.59	.53	.62	.43	.40		.56	.55		.41
UR	.61	.68	.70	.47	.62	.81	.58		.84	.69		.56
Correlations												
V M	.34	.45	.38	.68	.64	.52	.55		.56	.57		.50
V H	.14	.42	.17	.47	.50	.14	.30		.29	.36		.48
V UR	.39	.46	.31	.15	.40	.20	.34		.33	.41		.36
M H	.32	.22	.40	.49	.59	.20	.37		.33	.28		.45
M UR	.31	.19	.16	.08	.43	.14	.30		.33	.31		.32
H UR	.63	.31	.43	.33	.54	.48	.58		.53	.44		.55

[a] Reported on one-tenth College Board scale.

Table 2

Summary Statistics for Undergraduate Female Cross Sample

	A	B	C	D	E	F	G	H	I	J	K	L
No. Cases	199	93	120	220	116	142	138	344	155	128	74	353
Means												
V[a]	50	29	57	56	49	59	48	50	44	41	50	50
M[a]	52	31	57	55	47	58	47	49	45	41	49	50
H	3.23	2.63	2.45	3.38	2.91	3.26	3.06	2.78	2.95	2.64	3.04	3.20
UR	2.65	1.87	2.75	2.98	2.46	2.15	2.68	2.46	2.36	2.28	2.73	2.75
Standard Deviations												
V	9	5	8	8	9	8	10	8	8	10	8	9
M	9	6	9	9	9	9	9	8	9	8	9	9
H	.40	.49	.40	.39	.65	.47	.44	.62	.60	.67	.39	.46
UR	.59	.57	.55	.43	.70	.62	.54	.70	.75	.78	.51	.52
Correlations												
V M	.49	.43	.50	.47	.59	.47	.59	.52	.65	.53	.50	.55
V H	.36	.33	.25	.37	.49	.28	.33	.42	.46	.39	.32	.51
V UR	.43	.50	.54	.24	.53	.27	.57	.40	.56	.55	.40	.52
M H	.30	.27	.38	.40	.58	.47	.43	.37	.44	.49	.28	.53
M UR	.34	.28	.47	.27	.50	.3[illegible]	.44	.43	.52	.52	.40	.46
H UR	.68	.50	.43	.44	.66	.48	.62	.60	.72	.64	.39	.65

[a]Reported on one-tenth College Board scale.

Table 3

Summary Statistics for Undergraduate Cross Sample, Both Sexes Combined

	A	B	C	D	E	F	G	H	I	J	K	L
No. Cases	378	141	236	424	267	251	231	SEE TABLE 2	284	240	SEE TABLE 2	555
Means												
V[a]	51	29	57	54	48	60	48		44	40		50
M[a]	54	32	58	55	50	60	48		46	41		51
H	3.13	2.61	3.27	3.22	2.75	3.20	2.97		2.74	2.31		3.10
UR	2.60	1.93	2.64	2.90	2.31	2.10	2.53		2.20	2.09		2.67
Standard Deviations												
V	9	6	9	10	9	8	9		9	9		9
M	9	6	9	10	9	9	9		8	8		9
H	.44	.50	.53	.49	.65	.46	.44		.63	.71		.39
UR	.60	.61	.64	.46	.70	.71	.58		.81	.76		.51
Correlations												
V M	.42	.45	.43	.59	.58	.47	.56		.60	.54		.52
V H	.23	.35	.20	.46	.50	.21	.32		.38	.38		.47
V UR	.4[illegible]	.49	.40	.22	.47	.23	.46		.45	.50		.44
M H	.23	.23	.30	.42	.51	.31	.35		.34	.33		.41
M UR	.30	.26	.25	.15	.40	.23	.31		.40	.41		.36
H UR	.66	.41	.46	.41	.61	.48	.63		.65	.60		.63

[a] Reported on one-tenth College Board scale.

Table 4

Values of COR[a] in the Cross Sample: Males

College	System Type	V	M	H	V,M	V,H	M,H	V,M,H	Sample Size
		Predictor Combination							
A	Bayes	.39	.31	.63	.43	.60	.62	.69	179
	L.S.[b]	.39	.31	.63	.43	.69	.64	.69	
B	Bayes	.46	.19	.31	.44	.48	.27	.40	48
	L.S.	.46	-.19[c]	.31	.36	.38	.33	.38	
C	Bayes	.31	.16	.43	.29	.48	.41	.47	116
	L.S.	.31	.16	.43	.27	.47	.42	.46	
D	Bayes	.15	.08	.33	.14	.29	.31	.30	204
	L.S.	.15	.08	.33	.12	.32	.30	.30	
E	Bayes	.40	.43	.54	.46	.56	.56	.57	151
	L.S.	.40	.43	.54	.46	.56	.56	.56	
F	Bayes	.20	.14	.48	.19	.48	.48	.50	109
	L.S.	.20	.14	.49	.20	.50	.48	.49	
G	Bayes	.34	.30	.58	.36	.50	.58	.60	93
	L.S.	-.34[c]	.30	.58	.36	.60	.58	.60	
H	Bayes	WOMEN ONLY							
	L.S.								
I	Bayes	.33	.33	.52	.37	.56	.46	.56	129
	L.S.	.33	-.32[c]	.53	-.37[c]	.56	.55	.56	
J	Bayes	.41	.31	.44	.38	.48	.48	.50	112
	L.S.	.41	.31	.44	.40	.49	.47	.49	
K	Bayes	WOMEN ONLY							
	L.S.								
L	Bayes	.36	.32	.55	.39	.56	.56	.56	202
	L.S.	.36	.32	.55	.39	.56	.55	.56	

[a]COR is the correlation of predicted scores with observed scores.

[b]L.S. stands for least squares.

[c]Multiplicative parameter was negative.

Table 5

Values of COR[a] in the Cross Sample: Females

College	System Type	V	M	H	V,M	V,H	M,H	V,M,H	Sample Size
		Predictor Combination							
A	Bayes	.43	.34	.68	.46	.68	.66	.69	199
	L.S.[b]	.43	.34	.68	.46	.70	.69	.70	
B	Bayes	.49	.28	.50	.49	.55	.51	.55	93
	L.S.	.50	.28	.50	.49	.58	.52	.58	
C	Bayes	.54	.47	.43	.58	.62	.40	.59	120
	L.S.	.54	.47	.43	.58	.59	.51	.60	
D	Bayes	.24	.27	.44	.29	.30	.42	.43	220
	L.S.	-.24[c]	-.27[c]	.44	-.29[c]	.43	.45	.43	
E	Bayes	.53	.50	.66	.58	.70	.67	.70	116
	L.S.	.53	.50	.66	.58	.70	.67	.70	
F	Bayes	.27	.37	.48	.36	.50	.45	.47	142
	L.S.	.27	.37	.48	.36	.49	.50	.50	
G	Bayes	.57	.44	.62	.58	.73	.64	.72	138
	L.S.	.57	.44	.62	.58	.73	.65	.72	
H	Bayes	.40	.43	.60	.47	.62	.63	.63	344
	L.S.	.40	.43	.60	.47	.62	.63	.63	
I	Bayes	.56	.52	.72	.60	.76	.75	.77	155
	L.S.	.56	.52	.72	.60	.76	.75	.77	
J	Bayes	.55	.52	.66	.60	.70	.66	.71	128
	L.S.	.55	.52	.64	.60	.71	.67	.71	
K	Bayes	.40	.40	.39	.46	.47	.46	.50	74
	L.S.	.40	.40	.39	.46	.47	.47	.49	
L	Bayes	.52	.46	.65	.56	.68	.66	.68	353
	L.S.	.52	.46	.65	.56	.68	.66	.68	

[a]COR is the correlation of predicted scores with observed scores.

[b]L.S. stands for least squares.

[c]Multiplicative parameter was negative.

Table 6

Values of COR[a] in the Cross Sample: Combined Males and Females

College	System Type	V	M	H	V,M	V,H	M,H	V,M,H	Sample Size
		Predictor Combination							
A	Bayes	.40	.30	.66	.42	.70	.66	.70	378
	L.S.[b]	.40	-.30[c]	.66	.42	.70	.67	.71	
B	Bayes	.49	.26	.41	.49	.52	.45	.50	141
	L.S.	.49	.26	.41	.49	.50	.43	.50	
C	Bayes	.40	.25	.46	.41	.51	.42	.53	236
	L.S.	.40	-.25[c]	.46	.41	.54	.47	.54	
D	Bayes	.22	.15	.41	.22	.37	.38	.39	424
	L.S.	.22	.15	.41	.22	.39	.40	.39	
E	Bayes	.47	.40	.61	.49	.62	.59	.63	267
	L.S.	.47	.40	.61	.49	.64	.62	.64	
F	Bayes	.23	.23	.48	.26	.49	.48	.49	251
	L.S.	.23	.23	.48	.25	.49	.48	.49	
G	Bayes	.46	.31	.63	.46	.63	.64	.69	231
	L.S.	.46	.31	.63	.46	.69	.64	.68	
H	Bayes	.40	.43	.60	.45	.62	.64	.63	344
	L.S.	.40	.43	.60	.44	.62	.62	.62	
I	Bayes	.45	.40	.65	.48	.68	.67	.68	284
	L.S.	.45	.40	.65	.47	.68	.67	.69	
J	Bayes	.50	.41	.60	.51	.66	.55	.65	240
	L.S.	.50	.41	.60	.52	.66	.62	.66	
K	Bayes	.40	.40	.39	.45	.48	.49	.50	74
	L.S.	.40	.40	.39	.44	.47	.44	.48	
L	Bayes	.44	.36	.63	.46	.57	.60	.65	555
	L.S.	.44	.36	.63	.46	.65	.64	.65	

[a] COR is the correlation of predicted scores with observed scores.

[b] L.S. stands for least squares.

[c] Multiplicative parameter was negative.

Table 7

Values of AR[a] in the Cross Sample: Males

School	System Type	Predictor Combination: V	M	H	V,M	V,H	M,H	V,M,H
A	Bayes	-.20	-.06	-.20	-.16	-.27	-.19	-.20
	L.S.[b]	-.38	-.19	-.27	-.23	-.30	-.25	-.28
B	Bayes	.34	.31	.27	.34	.32	.31	.29
	L.S.	.38	.37[c]	.37	.37	.37	.37	.37
C	Bayes	.02	.06	.09	-.02	-.02	.06	.01
	L.S.	-.06	.06	.11	-.02	.07	.09	.06
D	Bayes	.23	.31	.16	.27	.18	.19	.21
	L.S.	.18	.15	.12	.21	.15	.16	.16
E	Bayes	-.12	-.05	-.09	-.06	-.08	-.06	-.07
	L.S.	-.13	-.01	-.12	-.05	-.11	-.09	-.10
F	Bayes	-.35	-.34	-.47	-.34	-.47	-.45	-.46
	L.S.	-.32	-.28	-.47	-.27	-.45	-.44	-.44
G	Bayes	-.02	.08	-.02	.04	-.02	.00	.00
	L.S.	-.06[c]	-.02	-.05	-.05	-.06	-.04	-.05
H	Bayes			WOMEN ONLY				
	L.S.							
I	Bayes	.06	.11	-.01	.09	-.06	.02	-.01
	L.S.	.16	.17[c]	.04	.17[c]	.02	.07	.04
J	Bayes	.03	.07	.21	.14	.27	.31	.29
	L.S.	.09	.20	.24	.21	.28	.28	.30
K	Bayes			WOMEN ONLY				
	L.S.							
L	Bayes	.10	.20	.14	.16	.14	.19	.15
	L.S.	.08	.18	.21	.14	.18	.24	.20

[a] AR is the average difference between predicted and observed grades.

[b] L.S. stands for least squares.

[c] Multiplicative parameter was negative.

Table 8

Values of AR[a] in the Cross Sample: Females

School	System Type	V	M	H	V,M	V,H	M,H	V,M,H
		Predictor Combination						
A	Bayes	.13	-.03	-.02	.08	.02	-.11	-.03
	L.S.[b]	.11	.03	-.07	.05	-.01	-.10	-.02
B	Bayes	.11	.07	.07	.15	.15	.11	.14
	L.S.	.12	.15	.15	.15	.20	.19	.20
C	Bayes	.21	.19	.23	.20	.29	.26	.27
	L.S.	.30	.22	.28	.29	.33	.28	.33
D	Bayes	.28	.35	.21	.27	.22	.23	.20
	L.S.	.25[c]	.26[c]	.17	.26[c]	.20	.20	.20
E	Bayes	.04	.08	.14	.06	.11	.11	.10
	L.S.	.05	.11	.13	.07	.11	.14	.11
F	Bayes	-.30	-.29	-.31	-.28	-.28	-.28	-.27
	L.S.	-.16	.22	.28	-.17	-.23	-.25	-.22
G	Bayes	.24	.27	.16	.24	.15	.15	.16
	L.S.	.21	.22	.10	.21	[illegible]	.13	.14
H	Bayes	-.07	-.01	-.05	[illegible]	-.08	-.06	-.10
	L.S.	-.05	-.04	-.06	[illegible]	-.09	-.06	-.09
I	Bayes	.09	.09	.06	.08	.00	-.02	-.01
	L.S.	.04	.07	.02	.04	-.05	-.01	-.05
J	Bayes	.20	.19	.19	.23	.21	.21	.22
	L.S.	.26	.30	.22	.27	.21	.24	.22
K	Bayes	.16	.15	.08	.10	-.13	-.46	-.32
	L.S.	-.10	-.36	-.15	-.35	-.27	-.39	-.34
L	Bayes	.20	.25	.08	.14	.09	.11	.09
	L.S.	.17	.19	.10	[illegible]	.10	.11	.10

[a] AR is the average difference between predicted and observed grades.

[b] L.S. stands for least squares.

[c] Multiplicative parameter was negative.

Table 9

Values of AR[a] in the Cross Samples: Combined Males and Females

School	System Type	V	M	H	V,M	V,H	M,H	V,M,H
		Predictor Combination						
A	Bayes	.08	.05	.03	.07	.04	.04	.04
	L.S.[b]	.04	-.01[c]	.02	.03	.04	.02	.03
B	Bayes	.04	.01	-.02	.06	.04	.00	.03
	L.S.	.04	.05	.03	.05	.05	.03	.05
C	Bayes	.06	.12	.15	.07	.13	.17	.13
	L.S.	.08	.12[c]	.17	.10	.16	.18	.16
D	Bayes	.26	.25	.20	.24	.17	.17	.18
	L.S.	.20	.19	.17	.19	.17	.17	.17
E	Bayes	-.04	.04	.01	-.02	-.05	.00	-.02
	L.S.	-.04	.09	.01	-.01	-.01	.02	.00
F	Bayes	-.53	-.50	-.54	-.54	-.55	-.56	-.55
	L.S.	-.47	-.52	-.57	-.47	-.53	-.57	-.53
G	Bayes	.24	.29	.25	.26	.25	.24	.25
	L.S.	.32	.34	.27	.32	.29	.28	.29
H	Bayes	-.13	-.14	-.16	-.14	-.16	-.17	-.18
	L.S.	-.22	-.22	-.22	-.22	-.23	-.23	-.23
I	Bayes	-.01	-.05	.00	.00	.04	-.02	.02
	L.S.	.02	-.00	.01	.03	.05	.02	.05
J	Bayes	.18	.16	.30	.21	.37	.33	.31
	L.S.	.29	.23	.39	.30	.40	.40	.40
K	Bayes	.11	.15	-.09	.09	-.10	-.15	-.23
	L.S.	-.07	-.07	-.09	-.12	-.28	-.18	-.29
L	Bayes	.19	.24	.19	.19	.21	.21	.18
	L.S.	.20	.24	.20	.21	.18	.21	.19

[a]AR is the average difference between predicted and observed grades.

[b]L.S. stands for least squares.

[c]Multiplicative parameter was negative.

Table 10

Values of VR[a] in the Cross Sample: Males

College	System Type	V	M	H	V,M	V,H	M,H	V,M,H
		Predictor Combination						
A	Bayes	.31	.34	.23	.30	.25	.23	.20
	L.S.[b]	.36	.33	.24	.30	.21	.23	.21
B	Bayes	.38	.44	.42	.39	.36	.43	.39
	L.S.	.37	.45[c]	.42	.42	.40	.42	.41
C	Bayes	.45	.50	.41	.46	.38	.43	.39
	L.S.	.51	.51	.41	.51	.39	.43	.40
D	Bayes	.27	.30	.25	.31	.31	.26	.29
	L.S.	.27	.22	.23	.26	.27	.23	.27
E	Bayes	.38	.37	.33	.36	.32	.32	.32
	L.S.	.38	.38	.33	.37	.32	.32	.32
F	Bayes	.63	.64	.51	.63	.51	.50	.50
	L.S.	.63	.66	.50	.66	.49	.50	.49
G	Bayes	.30	.31	.24	.30	.27	.25	.23
	L.S.	.34[c]	.32	.30	.32	.31	.29	.30
H	Bayes			WOMEN ONLY				
	L.S.							
I	Bayes	.63	.65	.53	.63	.50	.58	.51
	L.S.	.68	.69[c]	.56	.70[c]	.53	.56	.53
J	Bayes	.40	.43	.40	.41	.38	.39	.37
	L.S.	.40	.46	.40	.41	.38	.39	.38
K	Bayes			WOMEN ONLY				
	L.S.							
L	Bayes	.27	.28	.22	.27	.22	.22	.22
	L.S.	.27	.28	.26	.26	.26	.26	.26

[a]VR is the variance of the residuals.

[b]L.S. stands for least squares.

[c]Multiplicative parameter was negative.

Table 11

Values of VR[a] in the Cross Sample: Females

College	System Type	V	M	H	V,M	V,H	M,H	V,M,H
		Predictor Combination						
A	Bayes	.23	.31	.19	.27	.18	.19	.18
	L.S.[b]	.29	.32	.18	.28	.18	.19	.17
B	Bayes	.26	.30	.25	.26	.24	.25	.23
	L.S.	.27	.30	.26	.26	.22	.25	.22
C	Bayes	.21	.23	.25	.20	.19	.25	.20
	L.S.	.22	.25	.24	.22	.20	.22	.18
D	Bayes	.18	.18	.16	.18	.20	.17	.17
	L.S.	.18[c]	.19[c]	.17	.18[c]	.15	.15	.15
E	Bayes	.36	.37	.28	.34	.26	.27	.26
	L.S.	.36	.37	.29	.34	.27	.28	.26
F	Bayes	.36	.34	.30	.34	.29	.31	.30
	L.S.	.37	.34	.30	.34	.29	.29	.29
G	Bayes	.20	.23	.18	.19	.14	.17	.14
	L.S.	.22	.23	.20	.20	.14	.18	.14
H	Bayes	.42	.40	.31	.39	.33	.31	.30
	L.S.	.47	.41	.38	.43	.35	.35	.34
I	Bayes	.40	.43	.28	.38	.24	.25	.23
	L.S.	.38	.41	.30	.36	.28	.29	.27
J	Bayes	.44	.48	.36	.42	.31	.35	.31
	L.S.	.47	.50	.36	.44	.30	.34	.30
K	Bayes	.22	.23	.23	.21	.22	.37	.25
	L.S.	.27	.44	.25	.31	.23	.26	.23
L	Bayes	.20	.21	.16	.19	.15	.16	.15
	L.S.	.21	.22	.17	.20	.16	.16	.16

[a]VR is the variance of the residuals.

[b]L.S. stands for least squares.

[c]Multiplicative parameter was negative.

Table 12

Values of VR[a] in the Cross Sample: Combined Males and Females

		Predictor Combination						
College	System Type	V	M	H	V,M	V,H	M,H	V,M,H
A	Bayes	.30	.33	.22	.30	.20	.23	.20
	L.S.[b]	.31	.34[c]	.23	.31	.20	.22	.20
B	Bayes	.31	.35	.32	.31	.28	.30	.28
	L.S.	.33	.36	.32	.32	.29	.31	.29
C	Bayes	.34	.38	.32	.34	.30	.33	.29
	L.S.	.37	.38[c]	.32	.36	.29	.32	.29
D	Bayes	.23	.22	.19	.24	.25	.21	.22
	L.S.	.27	.24	.19	.31	.23	.20	.23
E	Bayes	.38	.42	.31	.37	.30	.33	.29
	L.S.	.40	.41	.33	.37	.31	.33	.31
F	Bayes	.48	.48	.39	.48	.38	.39	.39
	L.S.	.51	.48	.42	.51	.40	.41	.40
G	Bayes	.27	.31	.22	.27	.21	.21	.19
	L.S.	.31	.31	.21	.24	.20	.20	.20
H	Bayes	.42	.42	.32	.40	.30	.29	.30
	L.S.	.42	.44	.31	.40	.30	.30	.30
I	Bayes	.54	.58	.41	.52	.37	.38	.37
	L.S.	.53	.56	.41	.52	.37	.40	.37
J	Bayes	.44	.51	.38	.43	.34	.41	.34
	L.S.	.45	.56	.39	.43	.34	.37	.34
K	Bayes	.22	.22	.22	.21	.20	.20	.19
	L.S.	.22	.22	.22	.21	.20	.21	.20
L	Bayes	.24	.26	.18	.24	.21	.20	.18
	L.S.	.25	.27	.19	.25	.19	.19	.19

[a] VR is the variance of the residuals.

[b] L.S. stands for least squares.

[c] Multiplicative parameter was negative.

Table 13

Values of ZOL[a] in the Cross Sample: Males

		Predictor Combination						
School	System Type	V	M	H	V,M	V,H	M,H	V,M,H
A	Bayes	.36	.42	.45	.40	.37	.45	[illegible]
	L.S.[b]	.32	.34	.40	.36	.39	.41	.41
B	Bayes	.35	.25	.23	.31	.31	.31	.25
	L.S.	.33	.23	.25	.25	.27	.25	.27
C	Bayes	.49	.39	.41	.49	.46	.41	.42
	L.S.	.37	.39	.40	.45	.43	.39	.41
D	Bayes	.32	.31	.39	.32	.38	.39	.40
	L.S.	.35	.37	.40	.37	.43	.40	.41
E	Bayes	.42	.41	.42	.40	.40	.45	.43
	L.S.	.41	.44	.39	.42	.42	.43	.40
F	Bayes	.47	.48	.46	[illegible]	[illegible]	[illegible]	[illegible]
	L.S.	.45	.45	[illegible]	[illegible]	[illegible]	.44	.46
G	Bayes	.37	.33	.41	.33	.37	.39	.41
	L.S.	.33	.31	.35	.33	.35	.37	.35
H		WOMEN ONLY						
I	Bayes	.40	.42	.42	.41	.42	.43	.43
	L.S.	.43	.43	.43	.43	.43	.40	.42
J	Bayes	.46	.46	.46	.46	.43	.44	.40
	L.S.	.39	.42	.44	.44	.42	.44	.42
K		WOMEN ONLY						
L	Bayes	.44	.40	.44	.38	.45	.39	.40
	L.S.	.43	.41	.38	.41	.42	.33	.39

[a] ZOL is the average of a variable.

[b] L.S. stands for least squares.

Table 14

Values of ZOL[a] in the Cross Sample: Females

		Predictor Combination						
School	System Type	V	M	H	V,M	V,H	M,H	V,M,H
A	Bayes	.38	.42	.50	.37	.52	.51	.51
	L.S.[b]	.37	.43	.49	.37	.51	.48	.48
B	Bayes	.44	.40	.43	.44	.48	.47	.45
	L.S.	.43	.43	.45	.44	.41	.47	.41
C	Bayes	.33	.37	.28	.37	.38	.27	.35
	L.S.	.30	.30	.28	.35	.36	.30	.36
D	Bayes	.30	.26	.36	.30	.33	.33	.35
	L.S.	.35	.35	.35	.35	.35	.33	.35
E	Bayes	.47	.49	.50	.50	.53	.54	.53
	L.S.	.47	.47	.49	.48	.54	.52	.55
F	Bayes	.42	.42	.40	.42	.39	.44	.42
	L.S.	.46	.46	.42	.46	.44	.42	.42
G	Bayes	.41	.33	.41	.43	.51	.41	.50
	L.S.	.41	.34	.44	.43	.53	.46	.51
H	Bayes	.41	.42	.51	.42	.47	.47	.50
	L.S.	.35	.40	.43	.38	.46	.46	.47
I	Bayes	.39	.42	.52	.40	.58	.55	.58
	L.S.	.39	.44	.54	.43	.54	.57	.54
J	Bayes	.44	.45	.48	.48	.55	.48	.55
	L.S.	.42	.44	.48	.44	.55	.48	.54
K	Bayes	.32	.30	.34	.34	.46	.28	.31
	L.S.	.39	.35	.38	.26	.38	.34	.36
L	Bayes	.41	.39	.47	.44	.48	.48	.50
	L.S.	.42	.41	.47	.42	.50	.48	.48

[a] ZOL is the average of a variable.

[b] L.S. stands for least squares.

Table 15

Values of ZOL[a] in the Cross Sample: Combined Males and Females

		Predictor Combination						
School	System Type	V	M	H	V,M	V,H	M,H	V,M,H
A	Bayes	.39	.39	.50	.39	.52	.50	.52
	L.S.[b]	.40	.37	.51	.40	.52	.49	.52
B	Bayes	.40	.41	.38	.42	.41	.41	.41
	L.S.	.42	.41	.37	.41	.38	.38	.39
C	Bayes	.42	.37	.36	.42	.39	.33	.40
	L.S.	.39	.32	.35	.42	.41	.36	.39
D	Bayes	.32	.33	.39	.35	.36	.38	.38
	L.S.	.35	.32	.39	.33	.38	.39	.39
E	Bayes	.45	.41	.48	.43	.51	.46	.50
	L.S.	.43	.38	.46	.43	.49	.46	.49
F	Bayes	.35	.32	.32	.35	.32	.32	.32
	L.S.	.41	.31	.30	.42	.34	.31	.34
G	Bayes	.38	.32	.40	.37	.41	.40	.39
	L.S.	.31	.32	.41	.33	.40	.41	.40
H	Bayes	.42	.35	.48	.40	.51	.50	.52
	L.S.	.38	.35	.48	.39	.50	.50	.50
I	Bayes	.41	.40	.48	.42	.50	.48	.49
	L.S.	.40	.39	.48	.42	.49	.49	.50
J	Bayes	.48	.43	.46	.48	.41	.44	.47
	L.S.	.41	.41	.42	.43	.40	.40	.41
K	Bayes	.32	.35	.41	.34	.46	.41	.46
	L.S.	.35	.35	.41	.39	.43	.42	.42
L	Bayes	.39	.38	.44	.39	.42	.42	.46
	L.S.	.40	.37	.43	.39	.45	.42	.44

[a] ZOL is the average of a variable.

[b] L.S. stands for least squares.

Table 16

Values of AE[a] in the Cross Sample: Females

School	System Type	Predictor Combination						
		V	M	H	V,M	V,H	M,H	V,M,H
A	Bayes	.44	.44	.35	.42	.35	.35	.34
	L.S.[b]	.44	.44	.35	.43	.34	.35	.34
B	Bayes	.39	.42	.40	.40	.41	.40	.40
	L.S.	.40	.43	.43	.40	.40	.41	.40
C	Bayes	.42	.43	.46	[illegible]	.42	.48	.42
	L.S.	.47	.46	.48	.45	.45	.46	.44
D	Bayes	.42	.47	.37	.42	.40	.38	.37
	L.S.	.40	.41	.36	.40	.36	.36	.36
E	Bayes	.46	.48	.43	.45	.41	.42	.41
	L.S.	.46	.48	.44	.45	.42	.43	.42
F	Bayes	.50	.49	.48	.48	.46	.46	.46
	L.S.	.46	.46	.46	.44	.44	.45	.44
G	Bayes	.40	.45	.37	.39	.33	.35	.33
	L.S.	.40	.44	.37	.38	.32	.35	.32
H	Bayes	.51	.51	.42	.49	.45	.44	.42
	L.S.	.56	.52	.49	.53	.47	.47	.47
I	Bayes	.53	.53	.42	.51	.37	.39	.37
	L.S.	.50	.53	.42	.49	.41	.41	.40
J	Bayes	.55	.56	.48	.53	.45	.48	.45
	L.S.	.57	.60	.49	.56	.45	.48	.45
K	Bayes	.41	.42	.40	.39	.38	.62	.49
	L.S.	.42	.59	.40	.54	.41	.51	.45
L	Bayes	.40	.42	.33	.38	.32	.33	.31
	L.S.	.39	.40	.34	.38	.33	.33	.33

---

[a] AE is the average of the absolute value of the difference between the predicted grade and observed grade.

[b] L.S. stands for least squares.

Table 17

Values of AE[a] in the Cross Sample: Males

		Predictor Combination						
School	System Type	V	M	H	V,M	V,H	M,H	V,M,H
A	Bayes	.49	.46	.42	.45	.47	.41	.40
	L.S.[b]	.58	.49	.46	.48	.45	.44	.45
B	Bayes	.59	.62	.60	.60	.57	.62	.59
	L.S.	.60	.65	.63	.63	.62	.63	.62
C	Bayes	.50	.54	.51	.50	.47	.51	.48
	L.S.	.55	.55	.51	.53	.48	.51	.49
D	Bayes	.44	.48	.38	.46	.41	.39	.40
	L.S.	.42	.40	.37	.42	.38	.38	.38
E	Bayes	.50	.48	.47	.49	.46	.45	.45
	L.S.	.51	.49	.48	.49	.47	.46	.46
F	Bayes	.63	.64	.63	.63	.63	.62	.63
	L.S.	.63	.64	.64	.64	.62	.62	.62
G	Bayes	.45	.47	.41	.46	.44	.42	.40
	L.S.	.50	.48	.47	.49	.48	.46	.47
H				WOMEN ONLY				
I	Bayes	.64	.64	.59	.64	.57	.61	.58
	L.S.	.66	.67	.60	.67	.59	.60	.59
J	Bayes	.48	.51	.52	.51	.53	.54	.53
	L.S.	.50	.54	.53	.52	.53	.53	.54
K				WOMEN ONLY				
L	Bayes	.42	.46	.39	.43	.40	.42	.40
	L.S.	.42	.45	.44	.43	.43	.47	.44

[a]AE is the average of the absolute value of the difference between the predicted grade and observed grade.

[b]L.S. stands for least squares.

Table 18

Values of AE[a] in the Cross Sample: Combined Males and Females

School	System Type	V	M	H	V,M	V,H	M,H	V,M,H
		Predictor Combination						
A	Bayes	.44	.46	.37	.44	.35	.38	.36
	L.S.[b]	.45	.50	.38	.45	.36	.37	.35
B	Bayes	.44	.47	.46	.44	.43	.45	.44
	L.S.	.45	.48	.46	.45	.44	.45	.44
C	Bayes	.45	.50	.48	.45	.45	.49	.45
	L.S.	.48	.57	.48	.47	.45	.48	.45
D	Bayes	.43	.43	.37	.43	.39	.37	.37
	L.S.	.43	.42	.36	.44	.38	.36	.38
E	Bayes	.49	.52	.44	.49	.43	.45	.42
	L.S.	.49	.53	.46	.49	.44	.46	.44
F	Bayes	.66	.64	.64	.66	.65	.67	.65
	L.S.	.63	.65	.69	.63	.65	.68	.65
G	Bayes	.47	.52	.43	.47	.42	.42	.41
	L.S.	.52	.54	.43	.52	.44	.43	.44
H	Bayes	.52	.54	.44	.51	.43	.43	.43
	L.S.	.54	.56	.45	.53	.44	.45	.44
I	Bayes	.59	.61	.51	.58	.48	.49	.48
	L.S.	.59	.60	.57	.58	.49	.50	.48
J	Bayes	.53	.58	.53	.53	.37	.56	.52
	L.S.	.57	.62	.57	.56	.56	.57	.56
K	Bayes	.40	.42	.38	.39	.36	.38	.38
	L.S.	.39	.39	.38	.39	.40	.38	.41
L	Bayes	.42	.45	.38	.42	.40	.39	.37
	L.S.	.43	.46	.39	.43	.38	.39	.38

[a]AE is the average of the absolute value of the difference between the predicted grade and observed grade.

[b]L.S. stands for least squares.

Table 1[illegible]

[illegible] Squares Parameters: Males

Predictor Set	Parameter	Schools A	B	C	D	E	F	G	H	I	J	K	L	Average Bayes	Least Squares
V	V^a	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		.03	.03		.03	.03	1.0
	α^b	1.41	.90	1.02	1.21	1.09	.86	1.11		.84	.84		1.11	1.04	
	A^c	.0512	.0342	.0526	.0264	.0318	.0086	-.0007		.0059	.0242		.0260		
	B^d	.3136	.6107	-.4179	1.2413	.8228	1.8480	2.3957		1.5944	.8429		1.1388		
M	M^a	.03	.02	.03	.03	.03	.02	.02		.02	.03		.03	.03	1.0
	α^b	.99	.94	.85	1.04	.72	.85	1.11		.92	.51		.90	.88	
	A^c	.0205	-.0087	.0306	.0041	.0412	.0337	.0074		-.0046	.0506		.0256		
	B^d	1.5809	1.9168	.6214	2.4326	.0824	.2107	1.9622		2.0540	-.4116		.9783		
H	H^a	.71	.46	.66	.72	.72	.72	.53		.57	.73		.83	.66	1.0
	α^b	.60	.59	.39	.55	.41	.25	.81		.60	.25		-.28	.41	
	A^c	.5849	.2726	.6900	.6346	.7430	.8411	.1526		.4199	.7810		1.2567		
	B^d	1.0498	.9597	.2803	.7295	.3665	-.1156	1.9258		.9188	.1314		-1.3371		
VM	V^a	.02	.02	.02	.02	.02	.01	.01	WOMEN ONLY	.01	.01	WOMEN ONLY	.02	.02	.6323
	M^a	.02	.01	.02	.01	.02	.02	.01		.01	.03		.02	.02	.7748
	α^b	.45	.80	.27	.72	.15	.65	1.35		.95	-.10		.63	.59	
	A^c	.0264	.0121	.0406	.0175	.0347	.0323	.0030		-.0010	.0384		.0222		
	B^d	.7756	1.1110	-.8079	1.2595	-.1795	-.4751	2.1505		1.9004	-.5126		.7584		
VH	V^a	.05	.04	.01	.02	.02	.00	-.01		.02	.01		.01	.02	.0192
	H^a	.47	.26	.54	.55	.60	.74	.38		.55	.72		.91	.57	.9998
	α^b	-.94	-.03	.10	-.19	-.24	.23	1.50		.04	-.28		-.60	-.05	
	A^c	.5809	.2846	.6312	.6255	.6889	.7876	.0910		.4422	.7562		1.0267		
	B^d	.5231	.7655	-.1896	.1208	-.1198	-.8788	2.0187		.5221	-.4379		-1.6304		
MH	M^a	.02	-.01	.01	.01	.02	.01	.01		-.01	.03		.01	.01	.0148
	H^a	.55	.44	.60	.63	.58	.71	.39		.51	.64		.88	.59	.9999
	α^b	.13	.93	-.27	.34	-.11	-.06	.94		1.27	-.98		-.66	.15	
	A^c	.5358	.2246	.6922	.4981	.6518	.7542	.1582		.3585	.7716		1.0822		
	B^d	.7287	.9757	-.3264	.7185	.0808	-.5731	1.7832		.8045	-.3652		-1.7247		
VMH	V^a	.01	.01	.01	.01	.01	.01	.01		.01	.01		.01	.01	.0165
	M^a	.01	.00	.01	.00	.01	.00	.00		.00	.02		.01	.00	.0084
	H^a	.57	.43	.55	.57	.58	.63	.46		.52	.64		.71	.57	.9998
	α^b	-.19	.25	-.22	-.06	-.25	-.14	.57		.31	-.69		-.49	-.09	
	A^c	.5518	.2581	.6408	.5640	.6491	.7601	.1025		.3986	.7520		.9693		
	B^d	.4361	.7990	-.4509	.1715	-.1919	-1.0484	1.9420		-.5392	-.6237		-1.7349		

[a] Lines where V, M, or H appear give the multiplicative Bayesian coefficient for the parameter V, M, or H.

[b] School parameters on lines marked α are Bayes additive constants.

[c] School parameters on lines marked A are least squares multiplicative constants.

[d] School parameters on lines marked B are least squares additive constants.

Table 29

Bayes and Least Squares Parameters: Females

Predictor Set	Parameter	A	B	C	D	E	F	G	H	I	J	K	L	Average Bayes	Least Squares
		Schools													
V	V^a	.03	.03	.03	.02	.03	.02	.04	.02	.03	.03	.04	.03	.03	1.0
	α^b	.85	.88	.72	1.41	.90	.98	.67	1.35	.73	.83	.71	1.28	.94	
	A^c	.0381	.0241	.0473	-.0062	.0322	.0326	.0483	.0044	.0516	.0206	.0567	.0173		
	B^d	.6256	1.0550	-.2513	1.0551	.9246	.3787	.1606	2.2946	.0363	1.1692	-.0022	1.7191		
M	M^a	.03	.03	.02	.03	.03	.02	.03	.03	.03	.03	.04	.03	.03	1.0
	α^b	.93	.85	1.16	1.25	.92	1.31	1.05	1.05	.89	1.03	.72	1.11	1.02	
	A^c	.0345	.0322	.0166	-.0002	.0398	.0149	.0274	.0216	.0426	.0201	.0781	.0203		
	B^d	.9979	.7366	1.5891	2.7318	.4566	1.5107	1.1781	1.4262	.3508	1.1610	-.7115	1.5412		
H	H^a	.73	.78	.72	.72	.70	.73	.76	.63	.77	.71	.71	.70	.72	1.0
	α^b	.30	-.25	.05	.34	.29	.10	.18	.75	.02	.22	.51	.43	.25	
	A^c	.9463	.8694	.6674	.8027	.5574	.7066	1.1246	.2708	1.1507	.6188	.9607	.5175		
	B^d	-.3377	-.5225	.1797	.1050	.7011	.1262	-.8594	1.7693	-1.0509	.4248	-.0446	.9951		
VM	V^a	.03	.02	.02	.02	.02	.02	.03	.02	.03	.02	.03	.02	.02	.8596
	M^a	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.5110
	α^b	.66	.68	.44	.99	.65	.60	.53	.88	.53	.59	.51	.89	.66	
	A^c	.0350	.0239	.0455	-.0030	.0282	.0261	.0334	.0112	.0416	.0205	.0578	.0173		
	B^d	.1650	.7548	-1.0881	2.9457	.5144	.2114	.2974	1.7679	-.2189	.8517	-.8396	1.3985		
VH	V^a	.03	.01	.04	-.02	.01	.01	.03	.01	.03	.01	.04	.01	.02	.0320
	H^a	.63	.77	.52	.79	.49	.62	.70	.38	.83	.61	.35	.51	.60	.9995
	α^b	-.78	-.64	-1.48	1.40	.25	-.10	-.86	1.14	-1.54	-.13	-.25	.61	-.20	
	A^c	.7414	.6893	.7433	.3881	.4389	.5492	.7787	.2775	1.0213	.5839	.7662	.3701		
	B^d	-.9267	-.7787	-1.4872	.7843	.3787	-.4543	-1.0215	1.3426	-2.0519	-.2410	-.5492	.8718		
MH	M^a	.02	.01	.00	.00	.01	-.01	.01	.02	.02	.01	.07	.01	.01	.0192
	H^a	.71	.81	.63	.80	.49	.72	.88	.35	.90	.63	.41	.50	.65	.9998
	α^b	-.73	-.62	.45	.28	.22	.40	-.57	.63	-1.37	.15	-1.32	.40	-.17	
	A^c	.8832	.7579	.6048	.4871	.5016	.5753	.8448	.2846	1.0654	.5929	1.0142	.4524		
	B^d	-.9745	-.7559	-.2562	.6285	.4035	-.1065	-.7885	1.4627	-1.6943	.0143	-.9054	.7622		
VMH	V^a	.02	.02	.02	.00	.01	.01	.02	.01	.02	.01	.03	.01	.02	.0294
	M^a	.01	.00	.00	.00	.01	-.01	.00	.01	.01	.00	.04	.01	.01	.0082
	H^a	.58	.77	.55	.81	.44	.66	.76	.37	.81	.60	.27	.47	.59	.9995
	α^b	-.70	-.62	-.50	.10	.05	-.09	-.78	.19	-1.13	-.27	-.78	.19	-.36	
	A^c	.7360	.6636	.7265	.3450	.4269	.5250	.7244	.2840	.9844	.5758	.7868	.3656		
	B^d	-1.0990	-.8108	-1.6227	.8950	.3173	-.4998	-.9667	1.2345	-2.1414	-.3476	-.7872	.7897		

[a] ...es where V, M, or H appear give the multiplicative Bayesian coefficient for the parameter V, M, or H.

[b] School parameters on lines marked α are Bayes additive constants.

[c] ...l parameters on lines marked A are least squares multiplicative constants.

[d] ...l parameters on lines marked B are least squares additive constants.

Table 21

Bayes and Least Squares Parameters: Combined Males and Females

ictor et	Parameter	A	B	C	D	E	F	G	H	I	J	K	L	Average Bayes	Least Squares
		Schools													
V	V^a	.03	.03	.03	.03	.03	.03	.03	.02	.03	.03	.02	.03	.03	1.0
	α^b	1.25	1.10	.89	1.15	.95	.90	1.05	1.40	.94	.54	1.45	1.23	1.07	
	A^c	[illegible]	.0168	.0507	.0367	.0347	.0414	.0086	.0212	.0340	[illegible]	.0315	[illegible]		
	B^d	1.7072	1.3892	-.3355	.7165	.6825	.1089	1.8045	1.6288	.6955	-.3404	1.2269	1.7637		
M	M^a	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	.02	1.0
	α^b	1.67	1.32	1.54	1.64	1.36	1.51	1.35	1.75	1.35	1.16	1.70	1.54	1.49	
	A^c	-.0085	.0166	-.0141	.0247	.0204	.0217	.0236	.0125	.0465	.0051	.0328	.0115		
	B^d	3.0282	1.3518	3.3576	1.3501	1.2027	1.3248	1.0676	2.0607	.0837	1.6517	1.2031	1.8388		
H	H^a	.60	.60	.60	.60	.60	.61	.61	.60	.60	.61	.60	.60	.60	1.0
	α^b	.68	.37	.51	.76	.65	.70	.48	.94	.55	.39	.99	.61	.64	
	A^c	.5134	.3949	.5748	.6315	.3846	1.0850	.7895	.6937	.5603	.7917	.5098	.5264		
	B^d	.9698	.8654	.5890	.6938	1.2440	-.8006	-.0831	.7545	.6541	-.1230	1.2753	.8396		
VM	V^a	.02	.02	.03	.02	.03	.03	.02	.02	.02	.03	.02	.02	.02	.9693
	M^a	.00	.01	.00	.01	.01	.01	.01	.01	.01	.01	.01	.01	.01	.2458
	α^b	1.32	1.01	.86	.85	.76	.63	.88	1.30	.60	.35	1.38	1.17	.93	
	A^c	.0123	.0167	.0430	.0384	.0310	.0403	.0106	.0213	.0335	.0422	.0302	.0135		
	B^d	1.8026	1.2701	-.4504	.1692	.4936	-.3426	1.5981	1.3964	.3773	-.2766	1.0255	1.6585		
VH	V^a	.02	.02	.01	.02	.02	.01	.00	.01	.02	.03	.02	-.01	.01	.0287
	H^a	.52	.48	.52	.54	.32	.77	.65	.55	.51	.65	.32	.56	.53	.9996
	α^b	.16	.07	.38	-.32	.45	-.42	.36	.36	-.31	-.91	.82	1.07	.14	
	A^c	.5098	.4337	.5134	.5981	.3539	.8043	.4706	.6056	.5481	.7188	.5419.	.3802		
	B^d	.2282	.3809	-.0348	-.1265	.8573	-1.3161	.2023	.1439	-.0360	-.7949	.5846	.7641		
MH	M^a	.00	.01	-.01	.01	.00	.01	.00	.02	.02	-.01	.03	-.01	.01	.0100
	H^a	.54	.53	.57	.58	.47	.81	.67	.58	.51	.76	.32	.60	.58	.9999
	α^b	.84	.16	.99	.25	1.23	-.50	.06	.17	-.08	.45	.66	.92	.43	
	A^c	.5043	.3822	.5653	.6061	.3512	1.0183	.7124	.6952	.5488	.7629	.5541	.4294		
	B^d	.7283	.7711	.2961	.4431	1.1465	-1.2038	-.2064	.4097	.4289	-.3821	.9551	.9107		
VMH	V^a	.02	.02	.01	.02	.01	.01	.01	.01	.02	.02	.01	.01	.01	.0273
	M^a	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.0042
	H^a	.53	.53	.53	.53	.52	.53	.53	.53	.53	.53	.52	.53	.53	.9996
	α^b	.23	.08	.21	.11	.33	.05	.21	.26	-.02	.03	.35	.44	.19	
	A^c	.5069	.4224	.5150	.5335	.3443	.8029	.4678	.6109	.5392	.7097	.5424	.3599		
	B^d	.1624	.3787	-.1319	-.1979	.8400	-1.4473	.1483	.0364	-.0718	-.8514	.5227	.7976		

es where V, M, or H appear give the multiplicative Bayesian coefficient for the parameter V, M, or H.

ool parameters on lines marked α are Bayes additive constants.

ool parameters on lines marked A are least squares multiplicative constants.

ool parameters on lines marked B are least squares additive constants.

Table 22

Cross Sample Validities Obtained Using Various Regression Weights

School	Predictor Set[a]	Least Squares	Bayes	Average[b] Bayes	Absolute Value[c] of Bayes
Males		LS	B	$\bar{B}$	$\lvert B \rvert$
School F	(V, H)	.50	.48	.49	.48
School G	(V, H)	.60	.50	.59	.60
School B	(M, H)	.33	.27	.33	.34
School I	(M, H)	.55	.46	.55	.55
School B	(V, M, H)	.38	.40	.38	.41
School I	(V, M, H)	.56	.56	.56	.56
Females					
School D	(V, H)	.43	.30	.43	.43
School C	(M, H)	.51	.40	.52	.45
School D	(M, H)	.42	.45	.43	.44
School F	(M, H)	.49	.50	.48	.47
School B	(V, M, H)	.58	.55	.58	.56
School C	(V, M, H)	.60	.59	.60	.61
School D	(V, M, H)	.43	.43	.44	.46
School F	(V, M, H)	.50	.47	.50	.51
Combined					
School G	(V, H)	.69	.63	.69	.64
School L	(V, H)	.65	.57	.65	.65
School C	(M, H)	.47	.42	.47	.47
School E	(M, H)	.62	.59	.62	.62
School J	(M, H)	.62	.55	.62	.62
School L	(M, H)	.64	.60	.64	.64
School A	(V, M, H)	.71	.70	.70	.70
School B	(V, M, H)	.50	.50	.50	.50
School C	(V, M, H)	.54	.53	.54	.54
School E	(V, M, H)	.64	.63	.64	.64
School G	(V, M, H)	.68	.69	.69	.69
School J	(V, M, H)	.66	.65	.66	.66
School L	(V, M, H)	.65	.65	.65	.65

[a] Only those sets are used where negative weights were obtained in the Bayes Solution.

[b] The Bayes weights used were those obtained by averaging over schools.

[c] The absolute value of the Bayes weights were used.

Table 23

Summary Statistics for Graduate Student Data

	Master's				Doctorate				Combined			
	School				School				School			
	[illegible]	[illegible]	Y	[illegible]	[illegible]	[illegible]	Y	[illegible]	[illegible]	[illegible]	Y	[illegible]
No. Cases	34	30	60	17	35	28	32	11	69	58	92	28
Means												
GRE-V	579	605	576	608	600	617	603	639	589	611	586	620
GRE-Q	581	588	560	577	582	591	574	572	582	590	565	575
GRE-Psych.	581	595	601	584	596	614	642	655	588	604	615	612
U[a]	3.22	3.05	3.06	3.06	3.19	2.44	3.08	3.13	3.21	3.00	3.06	3.09
C[b]	3.73	3.60	3.59	3.59	3.81	3.61	3.73	3.72	3.77	3.60	3.64	3.64
Standard Deviations												
V	68	74	75	83	72	71	85	66	71	73	80	74
Q	87	94	95	101	82	83	71	78	84	89	88	93
P	50	66	78	44	56	51	71	69	54	60	78	65
U	.27	.50	.43	.40	.30	.45	.41	.48	.29	.48	.42	.43
C	.13	.26	.25	.30	.09	.21	.21	.22	.12	.24	.25	.28
Correlations												
VQ	.17	-.15	.29	.46	.28	.38	.34	.40	.22	.08	.30	.43
VP	.34	.46	.27	.60	.51	.47	.55	.80	.44	.47	.39	.64
VU	.00	.00	.13	.30	-.07	.33	.47	-.16	-.05	.13	.25	.13
VC	.09	-.08	.04	.53	-.04	.10	.29	-.28	.07	-.01	.16	.33
QP	-.02	-.09	.23	.42	.14	.07	.17	.38	.06	-.02	.22	.31
QU	-.23	-.44	-.06	.24	-.54	.35	.00	.31	-.39	-.10	-.04	.26
QC	.05	-.09	.26	-.02	.01	.00	.12	.17	.03	-.05	.24	.02
PU	.14	-.01	.09	.25	.10	-.06	.39	.04	.11	-.05	.18	.16
PC	.22	.33	.08	.39	.14	-.13	.28	-.27	.21	.16	.20	.20
UC	.23	.22	.43	.38	.20	.18	.21	.62	.18	.20	.36	.46

[a] Undergraduate average.

[b] Graduate grade point average.

Table 24

System Validity Coefficients for GRE Data in the Back Sample

Predictors	Least Squares: Combined 8	Least Squares: Combined 4	Least Squares: Doctorate	Least Squares: Master's	Bayes: Combined 8	Bayes: Combined 4	Bayes: Doctorate	Bayes: Master's
V	.23	.18	.23	.24	.22	.17	.22	.23
Q	.17	.17	.10	.19	.12	.16	.07	.18
P	.24	.19	.22	.25	.23	.18	.22	.25
U	.35	.33	.31	.36	.31	.32	.27	.34
VQ	.27	.19	.24	.28	.29	.23	.25	.33
VP	.25	.20	.24	.26	.31	.23	.27	.32
VU	.36	.35	.31	.38	.36	.35	.30	.39
QP	.25	.21	.23	.26	.29	.23	.24	.31
QU	.37	.35	.31	.40	.35	.37	.28	.40
PU	.36	.36	.31	.39	.37	.36	.34	.41
VQP	.26	.21	.25	.29	.37	.27	.30	.40
VQU	.37	.36	.31	.40	.41	.40	.32	.47
VPU	.36	.36	.32	.39	.41	.38	.36	.45
QPU	.37	.37	.31	.41	.42	.39	.33	.46
VQPU	.38	.37	.32	.41	.46	.42	.35	.52

Table 25

System Validity Coefficients for CEEB VSS Data in the Back Sample

	Least Squares			Bayes		
Predictors	M	F	C	M	F	C
V	.35	.50	.39	.27	.46	.35
M	.37	.43	.31	.30	.39	.23
H	.58	.63	.56	.55	.60	.54
VM	.39	.53	.40	.39	.48	.38
VH	.59	.67	.59	.59	.69	.60
MH	.58	.64	.57	.59	.67	.59
VMH	.59	.67	.59	.58	.70	.58

APPENDIX A

In the present study, one hopes to produce numerical weights which can be used in combination with predictor scores to compute predictions of grade point averages that would be achieved in graduate school. However, it is known ahead of time that there will not be enough cases to do a separate, stand-alone study at each school; pooling of data will be necessary, and this pooling will entail the use of some convention to relate the weights used for different schools. In this Appendix, the weights used at the different schools will be proportional, i.e., that the ratio of weights will be preserved. In addition to the proportional adjustment of weights at each school, a shift of means will also be incorporated. The hope of this appendix as well as of Appendix B is that, except for differences in difficulty and reliability, the grades measure the same thing.

For estimation purposes, it is assumed that data are available for samples of students from each of a number of institutions. The weights to be used would be chosen so as to minimize the sums of squares of errors of estimation of the observed grade point averages by the weighted sums of predictors scores. That sum of squares is written as follows:

$$\text{[illegible]} = \sum_{ij} w_i (Y_{ij} - B_i - \sum_{g} b_{ig} X_{ijg})^2 \tag{1}$$

where

- i is a subscript indicating the school;
- j is a subscript indicating student within school;
- g is a subscript indicating the predictor variable;
- w_i is an arbitrary weight which was taken as unity in the present study;

Y_{ij} is the grade point average of the <u>jth</u> student at the <u>ith</u> school;

B_i allows for a shift of means;

b_{ig} is the weight used for the <u>gth</u> variable at the <u>ith</u> school;

X_{ijg} is the score achieved on the <u>gth</u> predictor by the <u>jth</u> person at the <u>ith</u> school.

It is a well-known result in least squares analysis that the value of B_i that minimizes ϕ is the mean of the rest of the values in the parentheses of (1). That is,

$$B_i = \bar{Y}_{i.} - \sum_g b_{ig} X_{i.g} \qquad (2)$$

where the bar-dot notation is the familiar one indicating the averaging process. If the right hand side of (2) is substituted into (1), the effect is to replace the observed predictor and criterion scores by their deviations from school means. Then (1) can be rewritten

$$\phi = \sum_{ij} w_i (y_{ij} - \sum_g b_{ig} x_{ijg})^2 + 2 \sum_{ig} \lambda_{ig} (b_{ig} - a_i \beta_g)$$

$$+ \theta \, (P - \sum_{gg'} \beta_g \beta_{g'} r_{gg'}) \qquad (3)$$

including all of the desired constraints. The quantities λ are Lagrange multipliers included to incorporate the constraints that the weights will be proportional. Note that the constraint includes a product of the value a_i which is the constant of proportionality for school i, and the weight β_g for the predictor variable. Then the a's could all be multiplied by some number and the β's divided by the same number and equation (3) would remain essentially unaffected. The choice of the scale of the a's and β's is immaterial, but for numerical purposes one must be chosen. This choice is made according to the relationship in parentheses multiplied by θ in equation (3). θ is also a Lagrange multiplier used to enforce the constraint. P is positive, and the r's form a matrix of full rank.

To obtain (4), (5), (6), (7), and (8), rearrange the result of differentiating (3) with respect to b_{ig}, a_i, β_g, λ_{ig}, and θ respectively.

$$\sum_j w_i (y_{ij} - \sum_g b_{ig} x_{ijg})\, x_{ijg'} - \lambda_{ig'} = 0 \;. \qquad (4)$$

$$\sum_g \lambda_{ig} \beta_g = 0 \;. \qquad (5)$$

$$\sum_i \lambda_{ig} a_i + \theta \sum_{g'} \beta_{g'} r_{gg'} = 0 \;. \qquad (6)$$

$$b_{ig} = a_i \beta_g \;. \qquad (7)$$

$$\sum_{gg'} \beta_g \beta_{g'} r_{gg'} = P \;. \qquad (8)$$

Multiplying (6) by β_g, summing over g, and using (5) and (8) it can be seen that θ equals zero. Then multiplying (4) by $\beta_{g'}$, summing over g', and using (5) and (7) yields

$$a_i = [\sum_j y_{ij} (\sum_{g'} \beta_{g'} x_{ijg'})] / [\sum_g \sum_{g'} \beta_g \beta_{g'} (\sum_j x_{ijg} x_{ijg'})] \;. \qquad (9)$$

Multiplying (4) by a_i, summing over i, using (6) and (7), and remembering that θ equals zero, obtain

$$\sum_g \beta_g \sum_i w_i a_i^2 \sum_j x_{ijg} x_{ijg'} = \sum_i a_i w_i \sum_j y_{ij} x_{ijg'} \;. \qquad (10)$$

Equations (8), (9), and (10) provide the iterations by which a solution is found. In the program which was used in the present study $r_{gg'}$ was taken as unity if $g=g'$, zero otherwise, and P was taken as unity. A starting point for the iterations is to set all the β's equal to unity divided by the square root of the number of predictors. Then (9) can be used to find the a's , the resulting a's can be used in (10) to get β's which can be normed to satisfy (8) and the results substituted in (9) and so on. When the β's and a's cease to change from iteration to iteration, equations (7) and (2) may be used to recover the b's and β's .

One can develop an analog to the multiple correlation coefficient in that Φ is a sum of squares of residuals that can be subjected to a percentage comparison with the sum of squares around institutional means. A system coefficient that has the desired property is

$$R = \sqrt{1 - (\Phi / \sum_{ij} w_i y_{ij}^2)} \quad . \tag{11}$$

The denominator of the quantity in parentheses in (11) is the sum of squares of residuals after fitting the institutional means; the numerator of that quantity is the sum of squares of residuals after fitting the institutional means and the predictors. Unity minus that fraction can be interpreted as the percent of variance attributable to the predictors, and the square rooting completes the analogy to the multiple correlation coefficient.

On occasion, in some sets of data where a school is represented by very few cases, one will occasionally find a multiplicative constant, a , to be negative while the others are positive. This may be due to a reversal of the grade point scale from that school or to some real difference between that school and the others in the analysis. Most likely, if the criterion scale is not reversed relative to the others, the reversal in sign is not to be believed, in the opinion of the author. If a prediction on the grade point scale is to be recommended for that school, one should use the same sign displayed by the rest of the schools as a sign for the multiplicative constant for the school whose sign differs, then using (2) to adjust B for that school. However, with so little data an attempt to put the predictions on the scale for the particular school would probably await the accumulation of more data, using only the relative weights, the β's , to get predictions which are in good order but not on the grade point scale.

APPENDIX B

As in the method of Appendix A, the present method accomplishes a pooling across institutions and does so by linear adjustments which differ at each institution. However, the adjustment is applied to the criterion score rather than to the prediction. It is assumed that the samples available are the same as those in Appendix A, and that the symbols i , j , g , Y_{ij} , w_i , and X_{ijg} all have the same interpretation. The sum of squares to be minimized, however, is

$$\psi = \sum_{ij} w_i \left(A_i Y_{ij} - M_i - \sum_g h_g X_{ijg}\right)^2 , \qquad (1)$$

where the M's allow for adjustment of the means, and the A's adjust the criterion scores. The symbol h_g stands for the regression weights and only requires a single subscript since the step of partitioning the weight into regression coefficients and constants of proportionality is already, in a sense, accomplished. As in the method of Appendix A, the value of M_i is equal to the mean of the rest of the values in the parentheses of (1). That is,

$$M_i = A_i \bar{Y}_{i.} - \sum_g h_g \bar{X}_{i.g} \qquad (2)$$

If the right-hand side of (2) is substituted into (1), the effect is to replace the observed predictor and criterion scores by their deviations from school means. Then (2) can be rewritten

$$\psi = \sum_{ij} w_i \left(A_i y_{ij} - \sum_g h_g x_{ijg}\right)^2 + 2\gamma \left(Q - \sum_i w_i A_i \sum_j y_{ij}^2\right) \qquad (3)$$

where γ is a Lagrange multiplier which imposes a scale constraint on the criterion scale. The constraint is needed because, as examination of the squared quantity in (3) shows, a trivial minimum of ψ can be obtained by defining all parameters equal to zero. The Lagrange constraint ensures that such a solution will not be obtained.

After differentiating, the following normal equations may be obtained:

$$A_i \; {}_iC_{yy} - {}_iC'_{xy} H - \gamma \; {}_iC_{yy} = 0 \quad , \tag{4}$$

$$\sum_i w_i A_i \; {}_iC_{xy} - \sum_i w_i \; {}_iC_{xx} H = 0 \quad , \tag{5}$$

$$Q - \sum_i w_i A_i \; {}_iC_{yy} = 0 \quad . \tag{6}$$

The subscript C's are defined as follows:

$${}_iC_{xx} = \left\| \sum_j x_{ijg} x_{ijg} \right\| \quad ,$$

$${}_iC_{xy} = \left\| \sum_j x_{ijg} y_{ij} \right\| \quad ,$$

$${}_iC_{yy} = \sum_j y_{ij}^2 \quad .$$

Multiplying equation (4) by w_i summing over i, and using equation (6) yields

$$\gamma = (Q - \sum_i w_i \; {}_iC'_{xy} H) \; (\sum_i w_i \; {}_iC_{yy})^{-1} \quad , \tag{7}$$

which may be substituted into (4) and the result solved for A_i to obtain

$$A_i = {}_iC_{yy}^{-1} \; {}_iC'_{xy} \; H \; + \; (Q - \sum_i w_i \, {}_iC'_{xy} \, H) \; (\sum_i w_i \, {}_iC_{yy})^{-1} \; . \tag{8}$$

The expressions for A_i from (8) may be substituted into (5) to obtain

$$[\sum_i w_i \, {}_iC_{xx} - \sum_i w_i \, {}_iC_{xy} \, {}_iC_{yy}^{-1} \, {}_iC'_{xy} \quad (\sum_i w_i \, {}_iC_{xy})(\sum_i w_i \, {}_iC_{yy})^{-1} (\sum_i w_i \, {}_iC_{xy})'] \; . \; H = \frac{Q}{\sum_i w_i \, {}_iC_{yy}} \sum_i w_i \, {}_iC_{xy} \; . \tag{9}$$

The equations (9) can then be solved for H .

The advantage of this method comes when variables are to be added as in a test selection scheme. The matrix to the left of (9) must be inverted in finding the vector H , but one could choose a method which could be worked a line at a time, such as the square root method. As variables are added successively to a problem, the repeated development of the matrix to be inverted is unnecessary, whereas the method in Appendix A requires storage of all the covariance matrices for each institution. Also, it can be shown merely by substituting in (3) for the A's and h's that $\psi = \gamma Q$, and since ψ is a sum of squares of residuals, $\sqrt{1-\gamma}$ is the analog to a multiple correlation coefficient. Reference to (7) shows that the computation of γ does not require acc to the covariance matrices by institution. Furthermore, the formula for the covariance, $C_{z\varepsilon}$, of an outside variable z , with the residuals ε , after predicting y with the variables x is:

$$C_{z\varepsilon} = [\sum_i w_i \, {}_iC_{yz} \, {}_iC_{yy}^{-1} \, {}_iC'_{xy} - (\sum_i w_i \, C_{yz})(\sum_i w_i \, {}_iC_{yy})^{-1}(\sum_i w_i \, {}_iC'_{xy})$$

$$- \sum_i w_i \, {}_iC'_{xz}] \, H - Q \, [\sum_i w_i \, {}_iC_{yz}] \, [\sum_i w_i \, {}_iC_{yy}]^{-1} \quad . \qquad (10)$$

If all possible predictor variables $\underline{x}$ included the variables x and z in (10), the distinction being that x had been used in previous calculations as predictors and z had not, then the quantities needed to compute $C_{z\varepsilon}$ would appear in matrices needed for (9) if all possible predictors were included. One needs also to keep a column whose entries are $\sum_i w_i \, {}_iC_{\underline{xy}}$. Then at each stage compute $K_z = C_z / (\sum_i w_i \, {}_iC_{zz})^{-1/2}$ for all z and choose as the next predictor to be selected the one that yields the largest value of K. More complex variables selection schemes could be formulated, but this one seems simple and is analogous to the familiar Wherry-Doolittle.

APPENDIX C

To: Dr. R. F. Boldt
Senior Research Psychologist
Educational Testing Service
Princeton, New Jersey 08534

GRE VALIDITY STUDY QUESTIONNAIRE

Institution:

Department:

Name and Title of Person to Be Contacted: ______________________________

Name and Title of Person Completing This Form: ______________________________

I. Approximately how many applicants have you had, on the average per year, for the past three years? ______________________________

II. A. Do you routinely compute a graduate grade point average? Yes No

B. If answer to A is No, would you be willing to make copies of transcripts available to GRE Board for such computations? Yes No

C. Please indicate the number of quality points used in your grading system. For example, a four point system without D's and E's and in which withdrawals (W's) are simply not counted might look as follows:

A _4_ B _3_ C _2_ D _Not used_ E _Not Used_ F _-1_ Other _Ng to 0_ Other _____

Please enter below the facts describing your system.

A _____ B _____ C _____ D _____ E _____ F _____ Other _____ Other _____

III. How many degrees were granted by your department last year?

TYPE OF DEGREE	NUMBER GRANTED*
Ph.D.	________
M.S. or M.A.	________
Other** ______________	________

*Enter the number of students who have successfully completed the final oral examination for each type degree, during the period September 1, 1968 through August 31, 1969. The intended criterion for inclusion here is whether all steps have been completed except for commencement.

**Please indicate the title of any graduate degree granted by your institution other than the ones listed above.

IV. A. Are records for students who applied for admission during the period 1965-66 and 1966-67 available now (including those rejected)?

Yes No

B. If answer to A is No, what is the earliest date for which complete records including rejectees would be available? ______________

V. For purposes of the study, certain information will be needed for each student who was granted a degree or who successfully completed a final oral examination during the period September 1, 1968 through August 31, 1969 and for those who applied for admission in 1965-66 and 1966-67 (or during the two years following the date given in question IV B, or until the present). The first group will be called "graduates" and the second group will be called "applicants." The final form in which data will be sought on graduates and applicants will be highly dependent on the responses given by the institutions contacted. It is therefore urged that you indicate below whether you can supply each type of information requested. Please do not hesitate to include extended comments on the CONTINUATION SHEETS which are at the back of this questionnaire.

Please check appropriate boxes

Can supply data for applicant	Can supply data for graduates	Cannot supply any data	
☐	☐	☐	A. Student name ______________________
☐	☐	☐	B. Birth date ______________________
☐	☐	☐	C. Sex: Male Female
☐	☐	☐	D. GRE V Score __________ None received ☐
☐	☐	☐	E. GRE Q Score __________ None received ☐
☐	☐	☐	F. Area of advanced test (if other than economics) __________
☐	☐	☐	G. Advanced area test score __________ None received ☐
☐	☐	☐	H. Undergraduate grade point average on a 4 point scale (GPA), or rank in class (and number in class), or percentile rank in class ______ ________
☐	☐	☐	I. Undergraduate Institution ______________________
☐	☐	☐	J. Date of first application of candidate for graduate work at your institution ______________________
☐	☐	☐	K. Term to which the admission decision applied __________
☐	☐	☐	L. Nature of the first admissions decision Accept Reject
☐	☐	☐	M. Date of first enrollment as a graduate student at your institution ______________________
☐	☐	☐	N. Date of last enrollment ______________________
☐	☐	☐	O. Date of award (or expected award) of degree __________
☐	☐	☐	P. Type of degree awarded (to be awarded) __________
☐	☐	☐	Q. If the degree is other than a Ph.D., is graduate work toward a doctorate suggested? (In plainer words, is he generally considered Ph.D. material?) Yes No

-68-

70 80

CONTINUATION SHEET

Institution:

Department:

ITEM NUMBER	COMMENT

Graduate Record Examinations Board
PRINCETON NEW JERSEY 08540 • AREA CODE 609 921 9000

IN AFFILIATION WITH
The Association of Graduate Schools
The Council of Graduate Schools

November 6, 1970

The members of the Graduate Record Examinations Board, which formulates policies guiding the Graduate Record Examinations (GRE), have been aware that research information supporting the use and interpretation of data on which admissions decisions are based is indeed sparse. Suitable criteria for evaluating the outcomes of graduate education need to be developed, and the relations between these criteria and information available for admissions decisions must be discovered. The Board is committed to mount systematic criterion development and admissions research efforts.

To these ends, some steps have recently been taken. The Board and the National Science Foundation have jointly begun an empirical search of data from Ph.D.'s in science areas to learn about special population effects on the relation between GRE scores and the time needed to complete the doctorate. A study has been conducted relating the Test of English as a Foreign Language (TOEFL) and the GRE with foreign student success in graduate school. The development of a biographical inventory to control the effects of motivation is being explored. The nature of the flow of graduating seniors to graduate and professional schools is being broadly surveyed under the auspices of the GRE Board, among others. However, these and other research efforts are complicated by the scarcity of clearly relevant evaluations of graduate school performance, and the limited applicability of standard research and statistical techniques in the face of the very limited amount of data available.

This letter has two purposes. The first is to inform you of some of the research efforts of the Board as described above. The second is to request your support in developing prediction techniques which are particularly applicable to graduate schools. Special new statistical techniques are needed in graduate admissions research as a consequence of the relatively small numbers of graduates produced at individual institutions (as compared with, say, recipients of baccalaureate degrees). These small numbers are further reduced at the departmental level. The techniques, being new, must be proven out on real data. Enclosed are some forms that will assist in the transfer of data from your institution to the Educational Testing Service who will conduct this project for the Board. Two studies will be executed, one in departments of psychology and the other in departments of economics. It is our hope that these departments will supply necessary information and participate in this study.

We would appreciate your forwarding the enclosed materials to the chairmen of the departments of economics and psychology at your institution. Please be assured that every attempt will be made to minimize any inconvenience for your institution and its departments and to maximize the return to you of helpful and interesting results. All data received will be held in the strictest confidence and no institutional or student identification will appear in any reports.

We hope that the criterion studies and the statistical techniques will be successful. The need for better research information in graduate admissions and graduate performance evaluation is urgent.

Sincerely,

Stephen H. Spurr

Stephen H. Spurr
Chairman

SHS/jc

Enclosures

will [illegible] separately, one in departments of psychology and the other in departments of economics. It is our hope that these departments will supply necessary information and participate in this study.

We would appreciate your forwarding the enclosed materials to the chairman of the department of economics at your institution. Please be assured that every attempt will be made to minimize any inconvenience for your institution and its departments and to maximize the return to you of helpful and interesting results. All data received will be held in the strictest confidence and no institutional or student identification will appear in any reports.

We hope that the criterion studies and the statistical techniques will be successful. The need for better research information in graduate admissions and graduate performance evaluation is urgent.

Sincerely,

Stephen H. Spurr
Chairman

SHS/jh

Enclosures

will be executed, one in departments of psychology and the other in departments of economics. It is our hope that these departments will supply necessary information and participate in this study.

We would appreciate your forwarding the enclosed materials to the chairman of the department of psychology at your institution. Please be assured that every attempt will be made to minimize any inconvenience for your institution and its departments and to maximize the return to you of helpful and interesting results. All data received will be held in the strictest confidence and no institutional or student identification will appear in any reports.

We hope that the criterion studies and the statistical techniques will be successful. The need for better research information in graduate admissions and graduate performance evaluation is urgent.

Sincerely,

Stephen H. Spurr

Stephen H. Spurr
Chairman

SHS/hw

Enclosures

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Developmental Research Division

March 1, 1971

Dear Sir:

The members of the Graduate Record Examinations Board, which formulates policies guiding the Graduate Record Examinations (GRE), have been aware that research information supporting the use and interpretation of data on which admissions decisions are based is indeed sparse. Suitable criteria for evaluating the outcomes of graduate education need to be developed, and the relations between these criteria and information available for admissions decisions must be discovered. The Board is committed to mount systematic criterion development and admissions research efforts.

In early November 1970, we contacted the Graduate School of your University partly in hopes that certain records might be made available for research purposes. Materials including the enclosed letter and questionnaire accompan the letter of initial contact and we hoped that information describing your department might be entered on the questionnaire and returned to us. We still hope so and urge you to supply us with the information solicited on the enclosed questionnaire, and to participate in the study by making certain of your records available to us.

The information on the questionnaire will be valuable to us even though it may not prove feasible for you to participate. However, we would appreciate the opportunity to try to overcome the difficulties that may at this time seem to prohibit your participation. Of course, the information from your department will be held in strictest confidence. Your department will not be identified in any way in any report of this study without your permission.

In my letter of November 6 I emphasized that the transfer of data would be arranged so that it is at a minimum of difficulty for you. I fully intended and still intend that to be the case.

Sincerely,

R. F. Boldt

R. F. Boldt
Chairman, Measurement Systems
Research Group

RFB/jh

Enclosures

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE EDUCTESTSVC

Developmental Research Division

November 6, 1970

Dear Sir:

The members of the Graduate Record Examinations Board, which formulates policies guiding the Graduate Record Examinations, have for some time been aware that research findings on the validity of our examinations are inadequate, being few and not representative. We would like to see that more studies are accomplished, but in attempting to do so encounter criterion problems and problems arising from the scarcity of suitable data. Hence we are developing plans for criterion research and also initiating methodological studies on techniques for pooling data for use in different receiving institutions. This letter requests your participation in a validation study being conducted by the Educational Testing Service for the Graduate Record Examinations Board. The study will compare a variety of complex regression systems, both Bayesian and least squares, and will be conducted in departments of economics.

The methods used will, hopefully, prove to tolerate situations where very few cases are available per school.

We would like to use data reflecting your experience and are enclosing a form on which you can indicate the type and availability of data you may have available. We do have an interest in using your data even if there are not many cases. Based on your responses we will propose a means of transferring data from your files to ours with a minimum of complication for you.

The need for better research information in graduate admissions is urgent and this study may help us provide it. We hope you will participate in this study.

Sincerely,

R. F. Boldt

R. F. Boldt
Chairman, Measurement Systems
Research Group

RFB:

Enclosure

8

EDUCATIONAL TESTING SERVICE PRINCETON, N. J. 08540

Area Code 609
921-9000
CABLE EDUCTESTSVC

Developmental Research Division

March 1, 1971

Dear Sir:

The members of the Graduate Record Examinations Board, which formulates policies guiding the Graduate Record Examinations, have for some time been aware that research findings on the validity of our examinations are inadequate, being few and not representative. We would like to see that more studies are accomplished, but in attempting to do so encounter criterion problems and problems arising from the scarcity of suitable data. Hence we are developing plans for criterion research and also initiating methodological studies on techniques for pooling data for use in different receiving institutions. This letter requests your participation in a cross-validation study being conducted by the Educational Testing Service for the Graduate Record Examinations Board. The study will compare a variety of complex regression systems, both Bayesian and least squares, and will be conducted in departments of psychology.

The methods used will, hopefully, prove to tolerate situations where very few cases are available per school.

We would like to use data reflecting your experience and are enclosing a form on which you can indicate the type and availability of data you may have available. We do have an interest in using your data even if there are not many cases. Based on your responses we will propose a means of transferring data from your files to ours with a minimum of complication for you.

The need for better research information in graduate admissions is urgent and this study may help us provide it. We hope you will participate in this study.

Sincerely,

R. F. Boldt

R. F. Boldt
Chairman, Measurement Systems
Research Group

RFB/jh

Enclosure