ACT RESEARCH REPORT

No. 82

82

January 1982

THE RELATIONSHIP BETWEEN COLLEGE FRESHMAN CLASS SIZE AND OTHER INSTITUTIONAL CHARACTERISTICS AND THE ACCURACY OF FRESHMAN GRADE PREDICTIONS

R. Sawyer
E. J. Maxey

PUBLISHED BY THE RESEARCH AND DEVELOPMENT DIVISION
THE AMERICAN COLLEGE TESTING PROGRAM
P. O. BOX 168, IOWA CITY, IOWA 52243

THE RELATIONSHIP BETWEEN COLLEGE FRESHMAN CLASS SIZE AND OTHER INSTITUTIONAL CHARACTERISTICS AND THE ACCURACY OF FRESHMAN GRADE PREDICTIONS

Prepared by the Research and Development Division
The American College Testing Program

For additional copies write:
ACT Publications
P.O. Box 168
Iowa City, Iowa 52243

(Check or money order must accompany request.) Price: $1.00

ABSTRACT

This report documents the validity of ACT test scores and self-reported high school grades in predicting college freshman grade average. The accuracy of prediction equations based on these measures is documented for institutions of differing freshman class size, affiliation, degree level, and racial/ethnic composition.

The results in this report are based on data collected from a random sample of 205 colleges that participated in the ACT Research Services in 1974-75 and 1976-77. A separate prediction equation for each college was calculated from its 1974-75 data. Each resulting prediction equation was then applied to data for the 1976-77 freshmen, and the predicted and actual grades were compared.

The relationship between predictive validity and freshman class size was further investigated in two additional studies. In the first study, prediction equations were developed and cross-validated separately for males and females in each college. In the second study, prediction equations were developed from random subsamples of the 1974-75 freshman data from each college. Both studies supplied evidence of the relationship between prediction accuracy and sample size for samples smaller than the freshman classes represented in the data base.

The predictive validity of ACT test scores and high school grades was weakly related to freshman class size at colleges with 90 or more freshmen. For example, the average mean absolute error of prediction ranged only from .51 to .54 grade units over the five size categories studied. Similarly, the average cross-validated correlation ranged from .53 to .56 over the five size categories.

Prediction accuracy was moderately related to the institutional characteristics affiliation, degree level, and racial/ethnic composition. The average mean absolute error, for example, was .49 grade units for private colleges and .55 grade units for public colleges. The average mean absolute error was .55 grade units for two-year colleges, .50 grade units for four-year colleges, and .52 units for colleges with graduate programs. For colleges with the smallest proportion of black students, the average mean absolute mean error was .51 grade units, and for colleges with the highest proportion of black students, it was .59 grade units.

Among the total group of colleges, the accuracy of separate-sex predictions was less strongly related to freshman class size than it was to the other institutional characteristics studied. The accuracy of separate-sex predictions was, however, more strongly related to freshman class size at private and four-year institutions than at other kinds of institutions.

Combined-sex equations based on simple random samples of size 50 from the base year data were almost as accurate, on the average, as equations based on all records from the colleges. These results suggest that ACT data could be used to make predictions of acceptable accuracy at colleges with as few as 50 freshmen.

THE RELATIONSHIP BETWEEN COLLEGE FRESHMAN CLASS SIZE AND OTHER INSTITUTIONAL CHARACTERISTICS AND THE ACCURACY OF FRESHMAN GRADE PREDICTIONS

Richard Sawyer
E. James Maxey

Introduction

The American College Testing Program (ACT) offers research services through which colleges can predict the freshman grades of future students (The American College Testing Program, 1981). The students' predicted grades are based on their ACT test scores, their self-reported high school grades, and, optionally, on other predictive information. The predicted grades are calculated by weighting the predictor variables in multiple regression equations that are specific to each college.

The weights in a college's prediction equation are usually calculated from data on an entire previous freshman class (or classes). Because these weights are estimates whose accuracy depends on the size of the base sample used to calculate them, and because error in estimating the weights propagates error in prediction, the freshman class size affects prediction error. It is possible, therefore, that weights calculated from very small freshman classes could be subject to large sampling errors, resulting in predictions of unacceptable accuracy.

One way to address the issue of sample size is to assume that the freshmen in a college are a random sample from a hypothetical population with postulated statistical characteristics. Under this assumption, determining the appropriate sample size for calculating prediction weights becomes a mathematical problem of relating measures of prediction accuracy to parameters of a statistical model. Sawyer (1981) took this approach and found that for equations with two predictors, a sample size of about 30 would yield prediction equations with approximately the same accuracy as equations based on larger sample sizes. For five predictors, a sample size of 65 would yield comparable accuracy, and for ten predictors, a sample size of 120 would be needed.

A potential limitation of the above approach is that the assumptions on which the formulas are based may not be true in practice. For example, students from colleges of different sizes may be samples from different populations of students, insofar as the predictability of their grades is concerned. Thus, a college's size, as an institutional characteristic that attracts certain kinds of students, could be strongly related to the validity of the ACT Assessment in predicting freshman grades. A statistical model which does not take this possibility into account might, therefore, yield incorrect conclusions about the base sample size needed for a given level of prediction accuracy.

The primary purpose of this report is to present and interpret empirical evidence on the relationship between prediction accuracy and freshman class size, for a national sample of colleges that use the ACT Assessment. The significance of this relationship can be more easily assessed, however, if it is done so in the context of other college characteristics. Moreover, the relationship of prediction accuracy with size could itself depend on these other college characteristics. For these reasons, the relationship between prediction accuracy and size is also reported for separate subgroups of colleges defined by their affiliation, highest degree level offered, and racial/ethnic composition.

At the time of this study, ACT required of each college participating in its predictive research services a minimum sample size of 90 student records. Thus, no direct evidence on the predictive validity of the ACT Assessment at colleges with fewer than 90 freshmen was available. An additional purpose of this study, therefore, was to estimate through indirect evidence the accuracy of predictions based on fewer than 90 freshmen. This was done by developing and cross-validating separate-sex prediction equations in each college and by developing prediction equations from random subsamples of each college's freshman class.

Earlier Results

The *Technical Report for the ACT Assessment Program* (1973) contains the results of a study in which cross-validated correlations were computed for a sample of 50 colleges, 10 in each of five size categories. The predictors were the four ACT Assessment subtest scores. The mean cross-validated correlations from prediction equations obtained from data one year old ranged from .41 in the smallest size category (100-249 students) to .46 in the largest size category (over 1,000 students). Thus prediction accuracy only varied moderately over a fairly wide range of base year sample sizes.

Novick, Jackson, Thayer, and Cole (1972) reported a cross-validation study of a prediction method due to Lindley (1970). In part of their study, they computed cross-validated mean absolute errors and correlations for predictions based on ordinary multiple regression equations. Their data consisted of the four ACT subtest scores and the grade averages from two successive classes of freshmen at 22 community colleges. The college enrollments ranged from 105 to 735 freshmen, with a mean of about 246. The cross-validated mean absolute errors ranged from about .46 to .75, with a mean of .59. The cross-validated correlations ranged from about .33 to .75, with a mean of .47. These data produced no discernible relationship between prediction accuracy and college size. When institutional prediction equations were computed from a 25% random sample drawn from each college, the cross-validated mean absolute errors ranged from .50 to .80, with a mean of .61, and the cross-validated correlations ranged from about .27 to .56, with a mean of .42. Thus, a substantial reduction in sample size caused only a moderate increase in prediction error.

Miller and Kunce (1973) studied predictions of vocational rehabilitation and concluded that prediction equations should be based on sample sizes at least ten times the number of predictors. Halinski and Feldt (1970), on the basis of a Monte Carlo study, also recommended a minimum subject-to-variable ratio of 10. Their recommendations were made in the context of random sampling from an infinite population. Snee (1977) recommended that the number of subjects need only exceed the number of variables by 15 or more in order to permit meaningful interpretation of a model. He cautioned, however, that highly correlated or historical data might require larger sample sizes.

Only a few published reports have dealt with the relationship between prediction accuracy and college characteristics other than size. Ford and Campos (1977) reported base year correlations between freshman grade average, SAT scores, and high school rank for two-year and four-year colleges. They found a median multiple correlation of .53 for two-year colleges, as compared to a median multiple correlation of .58 for all colleges in the data base.

The *Technical Report for the ACT Assessment Program* reports validity data for four types of colleges defined by the highest degree level offered. The multiple correlations reported are for the four subtests of the ACT Assessment and high school grades in predicting overall freshman grade averages. The median multiple correlations are .52 for two-year colleges, .63 for four-year colleges, .61 for colleges with master's degree programs, and .57 for colleges with doctoral programs. The median multiple correlation for two-year colleges is quite similar to that reported by Ford and Campos.

There is a large body of published research and opinion on differential validity for racial/ethnic groups. Linn (1978) clarified and summarized recent thinking on this issue. Breland and Minsky (1978) reviewed and summarized published reports and papers on the validity of various college entrance measures for several different populations, including populations defined by racial/ethnic characteristics. The present paper, however, is not intended to address either the issue of differential validity or of selection bias for individual students. Rather, it is concerned with the relationship between prediction accuracy and racial/ethnic composition, as an institutional characteristic, together with size, affiliation, and degree level.

The ACT Assessment Program

The ACT Assessment Program is a comprehensive evaluative, guidance, and placement service for students and educators involved in the transition from high school to college. The four academic tests of the ACT Assessment measure developed abilities in the subject areas traditionally identified with college and high school programs: English, mathematics, social studies, and natural sciences. ACT test scores are reported on a standard scale that ranges from 1 to 36. More detailed descriptive and technical information about ACT test scores can be found in the *Technical Report for the ACT Assessment Program.*

When students register for the ACT Assessment Program, they report the last grade received prior to the senior year of high school in each of the above four subject areas, as well as various demographic and background information. For a technical discussion of the psychometric characteristics of the self-reported high school grades and a description of the other measures, see the *Technical Report*.

An important part of the ACT Assessment is the predictive research services, through which colleges can measure the local predictive validity of the ACT Assessment. These research services summarize the relationships between the ACT scores, high school grades and college grades of students at an institution. These services can also be used to generate weights for predicting the college grades of future applicants.

Data Base

This study is based on student records submitted by institutions through their participation in ACT's predictive research services. To reflect colleges' typical frequency of participation, the prediction equations were calculated from grades two years older than the grades being predicted. At the time this data base was constructed, the most currently available grades were for 1976-77 freshmen; therefore, the prediction equations are based on 1974-75 freshman grades.

Because the data in the study were collected from colleges participating in ACT's predictive research services, in some respects they are not representative of students nationally:

- Colleges using the ACT Assessment are located mainly in the Rocky Mountains, Great Plains, South, and Midwest, with comparatively fewer in the East/Northeast and West Coast.

- Privately-controlled institutions are relatively underrepresented among colleges that use the ACT Assessment, and publicly-controlled institutions are overrepresented.

- Participation in ACT's research services is voluntary; therefore, the data base is self-selected even among colleges that use the ACT Assessment Program.

The results of the study cannot be claimed to reflect precisely the results that would be obtained if data from all colleges in the nation could somehow be collected. One should be cautious, therefore, in applying the results to institutions which do not use the ACT Assessment or do not participate in ACT's predictive research services. Nevertheless, this study will be useful to suggest major trends and to extend knowledge in this area beyond the results available to date.

Most colleges using ACT's predictive research services choose to predict first-semester freshman grades. Colleges do, however, have the option of predicting first-year freshman grades. Although ACT does not maintain records of individual colleges' choices of criteria, it is estimated that over 60% of the colleges in the study data base reported first-semester grades for the academic year of record. There is no evidence to suggest that the predictive validity of the ACT Assessment differs significantly for these two criteria.

Sample Design

To reduce the computational costs of this study, weights were calculated and prediction equations were cross-validated on records from a probability sample of colleges in the data base described above. Because results on prediction accuracy are reported separately by college affiliation, degree level, racial/ethnic composition, and size, these variables were used to stratify the sample of colleges. The strata were defined by:

- The affiliation of a college: public or private.

- The level of a college, as determined by the highest degree level it offers.
  Two-Year Maximum. At least two, but less than four years of work beyond Grade 12; includes junior colleges, technical institutes, normal schools
  Four-Year Maximum. Only the bachelor's or first professional degree—includes those institutions offering courses of study leading to the customary Bachelor of Arts or Bachelor of Science degree and all those degrees which entitle the possessor to enter the profession indicated
  Graduate level. Master's or second professional degree and/or Doctor of Philosophy or equivalent degrees

- The racial/ethnic composition of a college, as determined by the percentage of students who indicated their racial/ethnic background as "Afro-American/ Black" when they registered for the ACT Assessment.

Low: 0%-25% "Afro-American/Black"

Middle: 25%-50% "Afro-American/Black"

High: 50%-100% "Afro-American/Black"

About 20% of all students do not report their racial/ethnic background when they write the ACT Assessment. The racial/ethnic categories above should therefore be considered an ordinal measure, rather than an indication of the actual percentage of blacks in a college.

- The *size* stratum for a college, as determined by the number of students for which the college reported 1976-77 freshman grades.

Category 1: 90-100 students

Category 2: 101-200 students

Category 3: 201-500 students

Category 4: 501-1000 students

Category 5: 1001 or more students

The size category of a college was determined by the number of freshman grades submitted in 1976-77 rather than 1974-75. The use of the later year was dictated by the requirements for a separate study which used the same data base. There are, however, only minor differences between the counts for the two years.

At the time these data were collected, ACT required a minimum of 100 records from colleges participating in its predictive research services. The computer program which calculated prediction equations for the ACT predictive research services, however, accepted institutional data bases with as few as 90 valid records. This was done to avoid penalizing small colleges which may have inadvertently submitted a few invalid records. Because we were especially interested in prediction accuracy for small colleges, the first size category was defined to include colleges with 90-100 records. The total numbers of colleges and students in this and the other size categories are displayed in Table 1.

TABLE 1

Summary of Data Base and Sample for Cross-Validation Study

Category	**Number of colleges in data base**	**Number of colleges in sample**			**Number of 1976-77 student records in sample**		
		Total group predictions	Males predictions[a]	Females predictions[a]	Total group	Males	Females
Base Sample Size (1974-75)							
100 or less	129	15	78	82	2,544	4,770	5,184
101-200	196	76	40	45	11,007	5,471	6,801
201-500	150	50	37	40	15,951	12,544	14,235
501-1,000	68	35	20	19	29,603	14,545	14,720
1,000 or more	51	29	6	8	55,773	6,489	10,513
Affiliation							
Public	297	124	118	118	91,503	37,568	42,868
Private	197	81	63	76	23,375	6,251	8,585
Degree Level							
2-Year Maximum	181	70	56	66	19,755	7,321	9,693
4-Year Maximum	136	53	50	51	12,403	4,252	5,357
Graduate	177	82	75	77	82,720	32,246	36,403
Proportion of Black Students							
Low	415	177	154	167	100,642	38,562	44,546
Middle	70	23	22	22	12,752	4,743	6,089
High	9	5	5	5	1,484	514	818
Total	494	205	181	194	114,878	43,819	51,453

[a]For the separate-sex predictions, the ranges under "Base Sample Size" refer to the number of records used to calculate the separate-sex prediction equations. Thus, separate-sex prediction equations for males were developed at 78 colleges with 100 or fewer males and were cross-validated on a total of 4,770 records for males.

The number of records a college submits to ACT's predictive research services for a given year need not be exactly the same as its freshman class size that year. For example, colleges with fewer than 100 freshmen may pool their current data with data from previous years, and colleges with more than 100 freshmen may submit a random sample of 100 or more records. ACT does not maintain records of the sampling methods used by individual participants in its predictive research services. Comparison with institutional record counts in other ACT research services, however, indicates that for about 70% of all colleges, the difference between the number of records submitted and the actual freshman class size could reasonably be accounted for by factors such as attrition and the addition of new student records.

The number of colleges selected from the data base was chosen to attain pre-specified precisions in estimating the mean college cross-validation statistics (defined in the following section). Specifically, the number of colleges selected from each size stratum was chosen to yield a 95% probability that the mean college mean absolute error estimated from the sample would be within the following limits of the mean college mean absolute error computed from all records in the data base:

- ± .01 grade units of the average, over all colleges.
- ± .02 grade units of the average, over colleges in each size stratum.

There was a 95% chance that the estimated mean college P50 would be within ± .01 of the corresponding data base mean and within ± .02 of the corresponding data base mean for a particular size stratum. A similar precision was indicated for the means of the other cross-validation statistics. Sampling variances estimated from the data indicate that these expected precisions were attained.

Within each size stratum, the number of colleges selected from the substrata defined by the other characteristics was proportional to the total number of records in the substrata. Population and sample sizes for these other strata are displayed in Table 1.

Prediction Equations and Cross-Validation Statistics

Prediction equations were calculated from the 1974-75 freshman grade data using a standard eight-variable multiple linear regression:

$y = a_0$

$+a_1$ *ACT English score

$+a_2$ *ACT Mathematics score

$+a_3$ *ACT Social Studies score

$+a_4$ *ACT Natural Sciences score

$+a_5$ *high school English grade

$+a_6$ *high school mathematics grade

$+a_7$ *high school social studies grade

$+a_8$ *high school natural sciences grade

where $a_0, a_1, \ldots, a_8$ are regression weights calculated from the base year data.

Validities for high school grades alone and test scores alone were reported by Sawyer and Maxey (1979). The results reported here pertain to the standard eight-variable multiple prediction equations.

One purpose of this study was to estimate the accuracy of predictions for colleges with fewer than 100 freshmen. At most colleges, roughly half of the students are of each sex. Studying the relationship between prediction accuracy and sample size for separate-sex equations would, therefore, result in evidence about sample sizes much smaller than those obtainable from the total group of freshmen. For this reason, separate prediction equations were calculated for the males and females in each college as well as for all students in the college.

The actual 1976-77 grade averages for the students in the sample were compared with the grade averages predicted from the 1974-75 combined-sex and separate-sex equations. For each college, these comparisons were summarized in terms of five cross-validation statistics:

- P20, the proportion of students whose predicted grade averages were within 0.20 grade units of their actual averages

- P50, the proportion of students whose predicted grade averages were within 0.50 grade units of their actual averages

- P100, the proportion of students whose predicted grade averages were within 1.00 grade units of their actual averages

- MAE, the mean absolute error of prediction for students in the college

- CVR, the (cross-validated) correlation between predicted and actual grade average.

The statistic P20 measures the proportion of students for whom very accurate prediction was possible: the statistics P50 and P100 correspond to lesser degrees of accuracy. A further discussion of these statistics is given by Sawyer and Maxey (1979).

The above cross-validation statistics were computed for each college separately. The statistics from individual colleges were then summarized over the entire sample and over various subgroups of colleges. In all computations, the data were weighted by the reciprocal of the probability of selection, so as to reflect the sample design.

Results

Tables 2a, 2b, and 2c contain relative frequency distributions for the five cross-validation statistics defined above. Table 2a pertains to cross-validation statistics obtained from the use of combined sex equations. Tables 2b and 2c contain corresponding results for separate-sex equations. The numbers in these tables should be read as follows:

- In the *P20, P50,* and *P100* columns, the *Range in Statistic* is the proportion of students whose predicted scores were within certain grade units of their actual averages (for example, Table 2a indicates that in 12% of the colleges, between 30% and 40% of the students had predicted grade averages within 0.20 grade units of their actual grade averages).

- In the *MAE* column the *Range in Statistic* represents the range of MAE for some proportion of the colleges (e.g., an MAE in the range of 0.4-0.5 was found for 35% of the colleges).

- In the *CVR* column the *Range in Statistic* represents the range of CVR for some proportion of the colleges (e.g., a CVR in the range of 0.5-0.6 was found in 34% of the colleges).

In interpreting these results the reader should bear in mind the level of precision in the results, as discussed in the section on sample design.

It is seen in Table 2a that about three-fourths of the colleges had P20 in the range .2 to .3; about three-fourths had P50 in the range of .5 to .7; about nine-tenths had a P100 of .8 or higher. Three-fourths had a MAE between .4 and .6. CVR was more spread out: about 86% of the colleges had a CVR between .4 and .7, and the modal range was .5 to .6.

The distribution of these statistics using separate-sex equations (Tables 2b and 2c) shows that the freshman grade averages of males were less predictable than those of females. This is reflected in both an increase in the relative frequencies for males corresponding to larger prediction errors and in the resulting shift of the mean absolute errors.

It should be noted that the results in Tables 2b and 2c pertain to the use of separate-sex equations rather than to the differential effects of combined-sex equations. When the combined-sex equations were applied to males, however, the cross-validation statistics were quite similar to those for the separate-sex equations in colleges with 201 or more males. In colleges with 200 or fewer males, predictions from combined-sex equations were, on the average, *slightly more accurate* than predictions from the separate-sex equations. The separate-sex equations for females also resulted in no overall average improvement in prediction accuracy. This would suggest that in predicting college grade average, there is typically little or no benefit in calculating separate-sex multiple regression equations. Some other prediction method, however, such as a combined-sex equation with adjusted intercept, might offer improved prediction.

TABLE 2a

Distribution of Proportions of Cross-Validation Statistics over Colleges
(Based on Combined-Sex Equations)

Range in statistic	**Cross-validation statistic** P20	P50	P100	MAE	CVR
0.0-0.1	.00	.00	.00	.00	.01
0.1-0.2	.14	.00	.00	.00	.00
0.2-0.3	.74	.00	.00	.00	.01
0.3-0.4	.12	.01	.00	.07	.06
0.4-0.5	.01	.17	.00	.35	.22
0.5-0.6	.00	.46	.00	.40	.34
0.6-0.7	.00	.30	.00	.14	.30
0.7-0.8	.00	.07	.10	.04	.06
0.8-0.9	.00	.00	.53	.00	.00
0.9-1.0	.00	.00	.37	.00	.00
Median	.24	.57	.88	.52	.56
Mean	.25	.57	.88	.52	.55

TABLE 2b

Distribution of Proportions of Cross-Validation Statistics over Colleges
(Based on Separate-Sex Equations for Males)

Range in statistic	**Cross-validation statistic** P20	P50	P100	MAE	CVR
0.0-0.1	.03	.00	.00	.00	.01
0.1-0.2	.32	.01	.00	.00	.03
0.2-0.3	.57	.01	.00	.00	.06
0.3-0.4	.08	.08	.00	.03	.15
0.4-0.5	.00	.30	.01	.24	.21
0.5-0.6	.00	.37	.01	.32	.27
0.6-0.7	.00	.17	.05	.28	.21
0.7-0.8	.00	.05	.23	.09	.05
0.8-0.9	.00	.00	.47	.03	.00
0.9-1.0	.00	.00	.23	.01	.00
Median	.23	.53	.84	.58	.51
Mean	.22	.52	.83	.58	.49

TABLE 2c

Distribution of Proportions of Cross-Validation Statistics over Colleges
(Based on Separate-Sex Equations for Females)

Range in statistic	Cross-validation statistic P20	P50	P100	MAE	CVR
0.0-0.1	.01	.00	.00	.00	.01
0.1-0.2	.17	.00	.00	.00	.01
0.2-0.3	.58	.01	.00	.00	.01
0.3-0.4	.22	.04	.00	.10	.05
0.4-0.5	.01	.14	.00	.38	.19
0.5-0.6	.00	.38	.00	.33	.32
0.6-0.7	.00	.30	.02	.13	.32
0.7-0.8	.00	.12	.10	.03	.09
0.8-0.9	.00	.01	.48	.02	.01
0.9-1.0	.00	.00	.40	.00	.00
Median	.25	.58	.89	.51	.57
Mean	.25	.58	.88	.52	.56

Tables 3a, 3b, and 3c contain mean cross-validation statistics for subgroups of colleges defined by their 1974-75 sample size (Base N), affiliation, highest degree level, and racial/ethnic composition. Within categories of these institutional characteristics, mean college cross-validation statistics are also given for two further subcategories defined by Base N. Results are not given for Base N subcategories of two of the racial/ethnic categories because of the small sample sizes in these two categories.

Note that the Base N categories and subcategories of Tables 3b and 3c pertain to the number of records used to develop the separate-sex equations. Therefore, the Base N for a given college in Table 3a is roughly twice its Base N in Tables 3b and 3c.

There was little discernible variation with respect to Base N in the average of the cross-validation statistics for the combined-sex equations. The average MAE varied from .51 to .54 grade units across the five size categories; this difference barely exceeds what could reasonably be expected from sampling error. The average of P20 was .24 to .26 across all Base N categories; the average of P50 ranged from .56 to .59; the average P100, from .87 to .89; and the average CVR, from .53 to .56.

According to all five cross-validation criteria, the average prediction accuracy for students enrolled in private colleges was better than the prediction accuracy for students enrolled at public colleges. For example, the average MAE for private colleges was .49, compared to .55 for public colleges. There was virtually no variation in the statistics P20, P50, P100, and MAE with regard to the Base N subcategories of colleges of the same affiliation. This finding is in agreement with the data reported by Novick, et al. The average CVR for public colleges did vary somewhat with sample size: public colleges with 90-200 student records had an average CVR of .50, compared to .54 for colleges with 201 or more student records. The corresponding CVRs for private colleges showed less sensitivity to sample size.

On the average, grade predictions were slightly more accurate for freshmen in four-year colleges (average MAE = .50) than for freshmen in graduate-level colleges (average MAE = .52) or two-year colleges (average MAE = .55). Differences in prediction accuracy between Base N subcategories within colleges of the same degree level were smaller than the differences between degree level categories.

Differences in prediction accuracy also occurred between the three groups of colleges defined by

TABLE 3a

Mean College Cross-Validation Statistics
(Predictions Based on Combined-Sex Equations)

College category	**Cross-validation statistic** P20	P50	P100	MAE	CVR
Base N for Combined Equation					
90-100	.25	.57	.87	.52	.53
101-200	.26	.59	.89	.51	.55
201-500	.24	.56	.87	.54	.56
501-1,000	.24	.56	.87	.54	.55
1,000 or more	.25	.57	.87	.53	.56
Affiliation					
Public-Total	.24	.55	.86	.55	.53
90-200	.24	.57	.86	.55	.50
201+	.24	.55	.86	.55	.54
Private-Total	.27	.60	.90	.49	.58
90-200	.27	.60	.90	.49	.58
201+	.26	.60	.90	.49	.60
Degree Level					
2 Year Max.-Total	.24	.56	.86	.55	.49
90-200	.25	.58	.87	.53	.48
201+	.23	.53	.85	.56	.49
4 Year Max.-Total	.26	.60	.89	.50	.60
90-200	.26	.60	.89	.50	.60
201+	.26	.59	.89	.51	.61
Graduate-Total	.25	.57	.88	.52	.57
90-200	.25	.58	.89	.51	.59
201+	.25	.57	.87	.53	.57
Proportion of Black Students					
Low-Total	.25	.58	.88	.51	.56
90-200	.26	.59	.89	.50	.55
201+	.25	.57	.88	.52	.56
Middle	.23	.53	.83	.59	.53
High	.22	.52	.84	.59	.48
All Colleges	.25	.57	.88	.53	.55

racial/ethnic composition. Colleges with the lowest proportion of black students had an average MAE of .51; colleges with an intermediate proportion of black students had an average MAE of .59, as did colleges with the highest proportion of black students. The average cross-validated correlations for these three groups were .56, .53, and .48, respectively.

The relationships observed between prediction accuracy and institutional characteristics using the combined-sex equations were also true of separate-sex equations. Grade prediction was by most measures more accurate, on the average, at private than at public schools, at four-year colleges than at graduate-level colleges and two-year colleges, and at colleges with the lowest proportion of black students than at colleges with larger proportions of black students.

The prediction accuracy of separate-sex equations for males varied only slightly more with respect to Base N

TABLE 3b

Mean College Cross-Validation Statistics
(Predictions Based on Separate-Sex Equations for Males)

College category	Cross-validation statistic				
	P20	P50	P100	MAE	CVR
Base N for Male Equation					
15-100	.22	.52	.82	.59	.49
101-200	.22	.53	.84	.57	.49
201-500	.23	.52	.82	.58	.48
501+	.24	.54	.86	.55	.52
Affiliation					
Public-Total	.22	.52	.83	.59	.46
15-100	.22	.52	.83	.59	.42
101+	.22	.52	.83	.58	.47
Private-Total	.23	.53	.84	.56	.55
15-100	.22	.52	.82	.58	.54
101+	.25	.58	.89	.51	.57
Degree Level					
2 Year Max.-Total	.21	.21	.81	.61	.40
15-100	.22	.21	.81	.61	.40
101+	.21	.21	.81	.62	.42
4 Year Max.-Total	.22	.22	.85	.56	.55
15-100	.21	.21	.84	.57	.55
101+	.25	.25	.87	.53	.54
Graduate-Total	.24	.24	.83	.56	.52
15-100	.26	.26	.81	.56	.55
101+	.23	.23	.84	.57	.51
Proportion of Black Students					
Low	.22	.53	.84	.58	.50
15-100	.22	.52	.82	.59	.50
101+	.23	.54	.85	.56	.50
Middle	.23	.50	.81	.61	.46
High	.18	.48	.78	.65	.41
All Colleges	.22	.52	.83	.58	.49

than that of the combined-sex equations. For example, the average MAE for colleges with 15-100 males was .59, compared to .55 for colleges with 501 or more males. The spread in MAE for females, (.50 to .53) was similar to that of the combined-sex equations.

The importance of Base N in determining the accuracy of separate-sex predictions varied from one type of college to the other. For example, the averge MAE was .54 for females in two-year colleges with fewer than 100 females and was .55 for colleges with 101 or more females. There was a similar apparent insensitivity to sample size for males in two-year colleges. On the other hand, the average MAE for females in private colleges with 15-100 females was .52; for females in private colleges with 101 or more females, it was .43. For males in private colleges, these two average MAEs were .58 and .51, respectively. At public colleges, however, there was a spread in MAE of only .01 for males and .03 for females. It would therefore appear

TABLE 3c

Mean College Cross-Validation Statistics
(Predictions Based on Separate-Sex Equations for Females)

College category	Cross-validation statistic				
	P20	P50	P100	MAE	CVR
Base N for Female Equation					
15-100	.25	.57	.87	.53	.53
101-200	.26	.59	.89	.51	.59
201-500	.26	.59	.88	.51	.58
501+	.26	.60	.89	.50	.59
Affiliation					
Public-Total	.25	.57	.87	.54	.54
15-100	.25	.58	.86	.56	.49
101+	.25	.56	.87	.53	.56
Private-Total	.27	.60	.90	.49	.58
15-100	.25	.57	.88	.52	.55
101+	.30	.66	.93	.43	.66
Degree Level					
2 Year Max.-Total	.25	.56	.87	.54	.50
15-100	.26	.58	.87	.54	.47
101+	.24	.55	.86	.55	.54
4 Year Max.-Total	.26	.59	.89	.50	.59
15-100	.24	.57	.88	.52	.57
101+	.28	.64	.92	.46	.63
Graduate-Total	.26	.59	.88	.51	.58
15-100	.25	.58	.87	.53	.55
101+	.26	.59	.89	.51	.59
Proportion of Black Students					
Low	.26	.59	.89	.51	.56
15-100	.26	.58	.87	.52	.52
101+	.26	.60	.89	.50	.59
Middle	.21	.53	.84	.59	.54
High	.22	.53	.85	.56	.52
All Colleges	.25	.58	.88	.52	.56

that the accuracy of separate-sex predictions is moderately related to Base N at private and four-year colleges. but is related to a lesser degree at other kinds of colleges.

The reason why the relationship between prediction accuracy and sample size depends on other institutional characteristics is not readily apparent. One possible explanation is that prediction accuracy is related to the variability of the predictor variables. At private colleges, for example, the variability in test scores and high school grades for a single sex may be small enough so that prediction accuracy is adversely affected at a certain sample size. In situations where there is more variability in the predictors (for example at public colleges or with a combined-sex equation), prediction accuracy might not be adversely affected at the same sample size. Validating this hypothesis will require further research.

Prediction Equations Based on Subsamples of the Base Year Data

Because every college in the data base for this study submitted 90 or more student records, it is not possible to address directly the accuracy of prediction equations for smaller colleges. It is possible to study the accuracy of grade predictions based on equations calculated from small samples of large colleges, but this may not give a true account of prediction accuracy in the smaller colleges. The reason is that a census of students from a small college may well differ in many respects from a random sample of equal size from a larger college. The degree of similarity of the two kinds of data is, however, empirically observable—in this study, for example, by comparing the prediction accuracy of equations for colleges in the 90-100 size category with the prediction accuracy of equations based on sample sizes of 100 drawn from colleges in the larger size categories. Moreover, the prediction accuracy of equations based on small samples from large colleges is of interest in its own right, as it would be less expensive for large colleges to report grades for a sample of their freshmen, rather than for the entire class.

Different samples of the base year data from a given college could yield different prediction equations and cross-validation statistics. A cross-validation statistic for a prediction equation derived from a sample of the base year data is therefore only one observation from the distribution of all possible cross-validation statistics resulting from repeated samples of the base year data.

For this study, four independent simple random subsample sizes of 25 were selected without replacement from the 1974-75 data for every college in the sample. The four resulting sets of prediction equations for a college were then cross-validated on the 1976-77 data from that college. Finally, the four resulting sets of cross-validation statistics were averaged. The average of the four cross-validation statistics for a college is an estimate of the expected value of the cross-validation statistic with respect to simple random subsample sizes of 25.

The computations described in the above paragraph were then repeated using simple random sample sizes of 50, 75, and 100. In each case four independent subsamples of a given size were selected without replacement from each college.

The within-college replication factor of four was chosen to yield a probability of 95% that a reported average MAE for any given size category would be within ± .04 of the "true" average MAE, or a 95% chance that the reported average MAE over all colleges would be within ± .01 of the "true" average. The sampling variances estimated from the data indicated that these precisions were obtained.

To reduce the substantial computational costs of this part of the study, the weights derived from sample sizes of 50, 75, and 100 were cross-validated on a subsample of each college's 1976-77 records. The subsampling fraction for a college was defined by the number of 1976-77 records: for colleges with 90-100 records, the subsampling rate was 1; for colleges with 101-200 records, it was 3/4; for colleges with 201-500 records, it was 1/3; for colleges with 501-1000 records, it was 1/8; and for colleges with 1001 or more records, it was 1/19.

The effect of the subsampling is to decrease slightly the precision of the cross-validation estimates for base year sample sizes of 50, 75, and 100. There was a 95% chance that the estimated average MAE for a given size category would be within approximately ± .045 of the "true" average MAE for the category. The corresponding tolerance for the estimated average for all colleges is approximately ± .015.

Results of Subsampling Study

Table 4a contains cross-validation statistics for combined-sex prediction equations developed from subsamples of the 1974-75 data. As one would expect, prediction accuracy increased with sample size. However, there was only a modest difference in the average accuracy of prediction equations based on sample sizes of 50, 75, or 100 and the average accuracy of prediction equations based on all records in the colleges. For example, the average college P20 for a sample size of 50 was .23, compared to an average P20 of .25 for equations based on all records. The corresponding average values of P50, P100, and MAE were .54, .85, and .57, respectively, for a sample size of 50, and .57, .88, and .53, respectively, for all records in

TABLE 4a

Mean College Cross-Validation Statistics for Prediction Equations Derived from Subsamples of Base Year Data
(Predictions Based on Combined-Sex Equations)

Size of subsample of base year data	Cross-validation statistics				
	P20	P50	P100	MAE	CVR
25	.21	.48	.79	.65	.41
50	.23	.54	.85	.57	.49
75	.24	.55	.87	.55	.52
100	.24	.56	.88	.54	.53
All records	.25	.57	.88	.53	.55

the colleges. It was not until the sample size was reduced to 25 students per college that prediction accuracy began to drop off noticeably. Therefore, a sample size of 50 students would appear to be adequate for many colleges.

Corresponding to the behavior of the other statistics, the average cross-validated r dropped off noticeably at a sample size of 25. It decreased more markedly than the other statistics, however, for sample sizes of 50, 75, and 100.

Freshman grades in very small colleges may be less accurately predicted than freshman grades in larger colleges for reasons other than sampling error. For example, changes in a "feeder" high school could affect a larger proportion of the freshmen at a small college than at a large one. The question of prediction accuracy for very small colleges can be answered definitively only by examining data from them. Unfortunately, such data are not available. It is still useful, in our opinion, to answer this question tentatively with such data as are available.

First, predictions based on equations developed from random subsample sizes of 100 are about as accurate as predictions based on equations developed from all records in colleges with 90-100 freshmen. (For example, the former group had an average P20 of .24, P50 of .56, and P100 of .88 while the latter had an average P20 of .25, P50 of .57, and a P100 of .88.) Moreover, the accuracy of grade predictions does not appear to be related to college size for colleges with 90 or more students. Finally, the accuracy of grade predictions based on random sample sizes of 50 does not differ markedly from this standard. Therefore, although direct evidence on the accuracy of grade predictions for colleges with fewer than 90 students is not available, it would appear that the accuracy would be quite comparable, even for colleges with as few as 50 students.

The separate-sex equations (Tables 4b and 4c) show more sensitivity to sample size than the combined-sex equations. First, the accuracy of separate-sex predictions began to drop off noticeably at 50 males or 50 females. Second, the rate of decrease in accuracy was greater at intermediate sample sizes. For example, the average MAE for predictions for males increased from .58 to .65 grade units as the sample size decreased from the entire college to 75 records. The corresponding increase in MAE for the combined-sex equation was only .02 grade units. Therefore, in developing a separate-sex equation, a sample size of 100 would be needed to maintain most of the accuracy associated with using all records from the college.

Cross-validation statistics were also calculated for subgroups of colleges defined by their total freshman class size. In accordance with our expectations, the statistics for each subgroup reflect an overall trend toward more accurate prediction with increased sample size. These differences are, for the most part, statistically nonsignificant due to the sampling error associated with the sample design.

TABLE 4b

Mean College Cross-Validation Statistics for Prediction Equations Derived from Subsamples of Base Year Data
(Predictions Based on Separate-Sex Equations for Males)

Subsample size for males in base year data	Cross-validation statistics				
	P20	P50	P100	MAE	CVR
50	.19	.44	.76	.73	.38
75	.20	.48	.81	.65	.42
100	.21	.50	.83	.62	.45
All records	.22	.52	.83	.58	.49

TABLE 4c

Mean College Cross-Validation Statistics for Prediction Equations Derived from Subsamples of Base Year Data
(Predictions Based on Separate-Sex Equations for Females)

Subsample size for females in base year data	Cross-validation statistics				
	P20	P50	P100	MAE	CVR
50	.21	.50	.82	.63	.45
75	.23	.53	.86	.58	.49
100	.24	.56	.88	.55	.53
All records	.25	.58	.88	.52	.56

Summary and Conclusions

For colleges with 90 or more freshmen, the accuracy in predicting freshman grades from ACT test scores and high school grades was weakly related to freshman class size. Prediction accuracy was moderately related to institutional affiliation, highest degree level offered, and racial/ethnic composition.

Among the total group of colleges, the accuracy of separate-sex predictions was also less strongly related to freshman class size than it was to the other institutional characteristics studied. The accuracy of separate-sex predictions was, however, more strongly related to freshman class size at private and four-year institutions than at other kinds of institutions.

Combined-sex equations based on simple random sample sizes of 50 from the base year data were almost as accurate, on the average, as equations based on all records from the colleges. On the basis of this indirect evidence, it appears that accurate combined-sex predictions could be made for students in colleges with as few as 50 freshmen.

The deterioration in the accuracy of separate-sex predictions was greater than that of combined-sex predictions when the equations were developed from samples of the base year data. On the average, a sample size of 100 was needed to develop a separate-sex equation that was as accurate as one developed from all available records in a college.

REFERENCES

The American College Testing Program. *Technical report for the ACT Assessment Program.* Iowa City, Iowa: Author, 1973.

The American College Testing Program. *Announcement of the ACT research services* (Dec. 1981 ed.). Iowa City, Iowa: Author, 1981.

Breland, H., & Minsky, S. *Population validity and college entrance measures* (Research Bulletin RB-78-19). Princeton, N.J.: Educational Testing Service, 1978.

Ford, S. F., & Campos, S. *Summary of validity data from the Admissions Testing Program Validity Study Service.* The College Entrance Examination Board, 1977. (ERIC Document Reproduction Service No. ED 148 869)

Halinski, R. S., & Feldt, L. S. The selection of variables in multiple regression analysis. *Journal of Educational Measurement,* 1970, *7,* 151-158.

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

Linn, R. L. Single-group validity, differential validity, and differential prediction. *Journal of Applied Psychology,* 1978, *63,* 507-512.

Miller, D., & Kunce, J. Prediction and statistical overkill revisited. *Measurement and Evaluation in Guidance,* 1973, *6,* 157-163.

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

Sawyer, R. Sample size and the accuracy of predictions made from multiple regression equations. Unpublished manuscript, 1981. (Available from The American College Testing Program, P.O. Box 168, Iowa City, Iowa 52243.)

Sawyer, R. L., & Maxey, E. J. *The validity over time of college freshman grade prediction equations* (ACT Research Report No. 80). Iowa City, Iowa: The American College Testing Program, 1979.

Snee, R. D. Validation of regression models: Methods and examples. *Technometrics,* 1977, *19,* 415-428.