Journal of Educational Measurement
Winter 1988, Vol. 25, No. 4, pp. 321–331

Gender Bias in the Prediction of College Course Performance

Robert L. McCornack
and
Mary M. McLeod
San Diego State University

Is the relationship of college grades to the traditional predictors of aptitude test scores and high school grades different for men and women? The usual gender bias of underpredicting the grade point averages of women may result from gender-related course selection effects. This study controlled course selection effects by predicting single course grades rather than a composite grade from several courses. In most of the large introductory courses studied, no gender bias was found that would hold up on cross-validation in a subsequent semester. Usually, it was counterproductive to adjust grade predictions according to gender. Grade point average was predicted more accurately than single course grades.

Women consistently get better grades in school than men. Hundreds of studies have found this at the high school level (Astin, 1971). Among high school graduates headed for college, women have the higher high school grade point average (GPA) (College Entrance Examination Board, 1985). Women continue to earn the higher GPA at the college freshman level (Astin & Panos, 1969). It is not our purpose to replicate this small but consistent mean GPA difference. The issue here is the possibility that the relation between college course grades and predictors such as aptitude test scores and/or high school grades is different for men and women. Grant and Sleeter (1986) review such concerns. Ignoring gender and predicting GPA from a single equation for both men and women usually leads to a gender bias of systematically underpredicting for women (Linn, 1973, 1978).

Gender prediction bias can be eliminated if the prediction equation includes different slopes and intercepts for men and women. Gender differences can be checked sequentially (e.g., Hogrebe, Ervin, Dwinell, & Newman, 1983). If the slopes do not differ, then the prediction equation is augmented by a dummy variable for gender, which allows for different intercepts. Different intercepts eliminate any systematic over- or underprediction. The reliability and stability of such gender effects is questionable, so verification through cross-validation is critical. Sawyer (1986) found that on cross-validation freshman GPA equations with different intercepts or gender specific equations were only moderately effective in improving predictive accuracy. Gender bias was reduced in a bare majority of the 200 colleges studied.

One possible source of gender bias is in the courses that make up the GPA. Women may select majors or courses with less strict grading standards. Students' academic choices can be influenced by achievement expectations, perceived course difficulty, and ability self-perceptions (Gallini, 1982). Women may be better at selecting courses that conform to their abilities (Decore, 1984). Strenta and Elliott (1987) document stable grading standard differentials between major fields. Gamache and Novick (1985) criticize earlier studies for not controlling the confounding effects of different majors for gender groups. They achieved partial control by analyzing students in a single college.

This study sought to exert even more control by using the grade in single courses as the criterion rather than a GPA. Will elimination of this confounding also eliminate gender bias? Just as with GPA, gender mean differences in single course grades have often been found (e.g., Boli, Allen, & Payne, 1985; Frisbee, 1984; Michael & Shaffer, 1978), especially in mathematics (Fennema & Sherman, 1977; Meece, Parsons, Kaczala, Goff, & Futterman, 1982) and statistics (Elmore & Vasu, 1986; Feinberg & Halperin, 1978). The main purpose here was to determine whether a gender bias exists in the prediction of individual college course grades from scholastic aptitude test scores and high school grades. We examined whether these gender effects were substantial enough and reliable enough to improve predictive accuracy on cross-validation.

A secondary purpose of this study was to compare accuracy when predicting GPA with accuracy when predicting individual course grades. Composites, such as GPA, tend to be easier to predict because they are more reliable. On the other hand, differential grading standards serve to reduce the predictability of a composite of several courses. Strenta and Elliott (1987) found prediction of GPA slightly more accurate than the prediction of grades in two courses. However, Goldman and Slaughter (1976) report grades in five courses slightly easier to predict than GPA. They suggest that the tendency of low ability students to choose courses with lax grading standards may partly account for the difficulty in predicting GPA.

Data Sources

The study included introductory courses ("first" and other basic courses) in a wide range of subjects at a large, urban state university enrolling over 27,000 undergraduates. All lower division courses with at least 50 students and a minimum of 15 students of each gender were included. Typically, students take these courses in their freshman or sophomore year. The 88 introductory courses included 9 in mathematics; 7 in art; 5 each in biology, English, and physics; 4 in history; 3 each in accounting, chemistry, economics, political science, psychology, and Spanish; 2 each in anthropology, drama, engineering, French, geography, music, philosophy, sociology, and telecommunications; and 1 each in astronomy, finance, German, geology, health science, Italian, journalism, linguistics, religious studies, and speech. Most courses were taught with many small sections, usually with different instructors. All courses were taught in both the fall and spring semesters of the academic year 1985–86. In the fall semester, with 30,410 enrollments in 872 sections, the average section had 35 students. Only 38 sections had more than 100 students. Although 11 of the 88 courses had a single section, one course had 69 sections. The median course had 7 sections. There were 26,709 students enrolled in the same 88 courses in the spring semester. A GPA was obtained for each of 11,153 students enrolled in one or more of these courses in the spring semester of 1986.

Method

The criterion course grades of A to F were scaled from 4 to 0. Predictors were high school grade point average (HSGPA) and the Verbal (SAT-V) and Mathematics (SAT-M) scores from the Scholastic Aptitude Test (Educational Testing Service, 1948–1982). To determine gender effects, three prediction models were developed for each course: (a) the common equation with no gender effects, including only the HSGPA, SAT-V, and SAT-M predictors; (b) the different intercepts model with a dummy-coded gender predictor added to permit separate intercepts but identical slopes for HSGPA, SAT-V, and SAT-M; and (c) the gender specific model, which permitted both different intercepts and different slopes for each predictor by including gender-predictor interactions. Prediction models were developed for each course in the base semester, the fall of 1985. Predictive accuracy of each model was evaluated in the cross-validation semester, spring 1986. All but 3 of the 88 were single semester courses. The halves of the full year courses were treated as separate courses. The second half of a year course was never used to cross-validate the first half.

Predictive accuracy of each model was assessed in two ways. First, the predicted and actual grades were correlated. This is a relative measure of accuracy that ignores systematic under- or overprediction due to a change in the average grade awarded in the base and cross-validation semesters. Second, the square root of the mean squared error (RMS error) represents a combination of variable errors and systematic error due to a change in the average grade. An error was the predicted grade minus the actual grade. Changes in students and instructors in a course can influence both variable and systematic errors. Two prediction methods contained gender effects; the third did not. The primary question of gender prediction bias was addressed by testing the null hypothesis that the three prediction methods were equally accurate on the average. The predicted-actual grade correlations were normalized by the Fisher z transform. The RMS errors were log transformed. A repeated measures 3×88 analysis of variance was used with each transformed accuracy measure as the dependent variable. Computations were performed using the multiple regression and analysis of variance programs from the BMDP series (Dixon, 1983).

The secondary question, whether GPA or course grades is more predictable, was addressed by simply comparing the predicted-actual grade correlations for the 88 courses with the GPA multiple correlation in the cross-validation semester. The predictors, HSGPA and SAT scores, remained the same.

Results and Discussion

Enrollments in these large, lower division introductory courses varied from 59 to 1,491 in the base semester and from 50 to 1,130 in the cross-validation semester. The median course enrolled 124 men and 132 women in the base semester and 103 men and 119 women in the cross-validation semester. Predictor

Table 1
Median Course Values by Gender for 88 Introductory Courses in the Base and Cross-Validation Semesters

Group		HSGPA	SAT-V	SAT-M	Grade
Fall Semester 1985					
Men	Mean	2.96	432	506	2.38
	S.D.	.32	77	89	1.05
Women	Mean	3.09	428	457	2.35
	S.D.	.39	88	74	.99
Spring Semester 1986					
Men	Mean	2.94	435	505	2.39
	S.D.	.39	75	89	1.05
Women	Mean	3.09	427	462	2.36
	S.D.	.40	90	77	1.00

and grade means for each course by gender were calculated. The median of these course means with the standard deviation for the same course are reported in Table 1. These median values reflect typical gender differences. Women usually averaged much higher in HSGPA, lower in SAT-V, and much lower in SAT-M. In the base semester, women had the higher mean HSGPA in 86 of 88 courses, the lower SAT-V mean in 64 of 88 courses, and the lower SAT-M mean in 85 of 88 courses. These gender differences were similar in the cross-validation semester.

Within-course gender differences in mean grade were small. Comparing means and medians gave mixed results. Comparing medians, men had higher grades in both semesters (see Table 1). If the means in Table 2 are compared, then there was no difference in the base semester, and women had higher grades in the cross-validation semester. Women had higher course grades in 40 of 88 courses in the base semester and 49 of 88 courses in the cross-validation semester.

Table 2
Means, Standard Deviations and Intercorrelations (r x 100) of Grade and Predictor Course Means by Gender for 88 Introductory Courses in the Base Semester (F85) and Cross-Validation Semester (S86)

Course Mean	Mean	S.D.	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 F85 HSGPA Men	2.97	.08	63	36	52	73	67	-20	-17	-58	69	65	24	48	70	65	-21	-15	-54
2 F85 HSGPA Women	3.11	.08		09	35	61	78	-28	-11	-62	64	72	-06	23	61	75	-12	-15	-62
3 F85 SAT-V Men	433	20			62	56	15	01	-08	-07	14	04	55	54	27	11	-14	-09	-08
4 F85 SAT-V Women	427	17				48	47	-01	09	-10	28	30	41	68	32	31	-02	15	-13
5 F85 SAT-M Men	504	26					73	-37	-32	-68	57	57	28	38	74	70	-41	-43	-63
6 F85 SAT-M Women	463	26						-39	-20	-67	57	73	-02	30	66	84	-32	-25	-67
7 F85 Grade Men	2.37	.30							75	38	-12	-25	16	06	-22	-31	69	65	40
8 F85 Grade Women	2.37	.32								34	-11	-17	01	-02	-23	-23	67	76	32
9 F85 % Women	52	16									-49	-64	15	-08	-56	-71	33	45	92
10 S86 HSGPA Men	2.95	.09										71	23	43	74	69	-12	-16	-44
11 S86 HSGPA Women	3.10	.09											01	40	73	89	-28	-21	-59
12 S86 SAT-V Men	435	20												55	37	05	02	04	15
13 S86 SAT-V Women	426	18													43	41	-04	04	-09
14 S86 SAT-M Men	505	22														79	-28	-30	-51
15 S86 SAT-M Women	462	26															-36	-31	-66
16 S86 Grade Men	2.35	.34																74	31
17 S86 Grade Women	2.40	.35																	44
18 S86 % Women	52	15																	

These results do not contradict the usual finding that women have a higher mean GPA: Courses were selected because of their elementary content and size, not because they were representative of the courses that make up the typical student's GPA. As will be reported later, women attending the cross-validation semester still had the higher mean GPA.

The intercorrelations between the gender means shown in Table 2 indicate substantial consistency in the relative positions of these introductory courses. Between-gender and between-semester correlations ranged from .63 to .72 for HSGPA, from .55 to .68 for SAT-V, and from .73 to .84 for SAT-M. Mean SAT-M scores best distinguished different courses; a similar result was reported by Strenta and Elliott (1987) for different departments. In courses in which women received higher grades, men did also (r = .75 and .74 for each semester). There was also a rather consistent course ordering between semesters for the mean grade awarded men (r = .69) and women (r = .76). Between-semester stability represents a confounding of stable instructor grading standards and the similarity of choices made by students at different ability levels. There was some tendency for courses with the higher average grade to have students with lower mean HSGPA (r = .20 and −.12 for men, −.62 and −.21 for women) and lower mean SAT-M (r = .37 and −.28 for men, −.67 and −.31 for women). These between-course correlations had the same sign but were smaller than the between-department correlations reported by Strenta and Elliott (1987).

The results of the multiple regression analyses within each of the 88 introductory courses in the base semester are summarized in Table 3. The distribution of correlations for each predictor and predictor combination is described by lowest, highest, and quartile values. The small negative gender correlation reflects the arbitrary coding of 1 = female and 0 = male, so negative values represent a higher average grade for men in the median course. HSGPA tended to be the best single predictor. The median grade correlation using HSGPA, SAT-V, and SAT-M was .37. Two factors that probably reduced this correlation were (a) students had been selected on these same variables when admitted to the university, and (b) grading standards varied between instructors of the multiple

Table 3
Range and Quartile Values Among 88 Introductory Courses of Single Predictor and Multiple Correlations with Course Grade in the Base Semester of Fall 1985

Predictor Set	Min	Q_1	Q_2	Q_3	Max
Gender (G)	-.21	-.08	-.01	.07	.31
HSGPA	-.04	.23	.28	.33	.51
SAT-V	-.20	.08	.19	.26	.57
SAT-M	-.16	.13	.19	.27	.42
HSGPA, SAT-V, SAT-M	.20	.31	.37	.41	.65
G, HSGPA, SAT-V, SAT-M,	.22	.32	.38	.42	.66
G, HSGPA, SAT-V, SAT-M, GxHSGPA, GxSAT-V, GxSAT-M	.23	.34	.39	.44	.66

Table 4
Mean, Range and Quartile Values Among 88 Introductory Courses of the Grade Correlations and RMS Errors for Three Prediction Methods in the Cross-Validation Semester

Accuracy/Method	Mean	Min	Q_1	Q_2	Q_3	Max
Correlation:						
Common Equation	.318	.06	.24	.31	.39	.57
Different Intercepts	.313	.08	.23	.30	.39	.56
Gender Specific	.302	.01	.22	.29	.38	.54
RMS Error:						
Common Equation	.971	.62	.89	.95	1.06	1.36
Different Intercepts	.974	.64	.88	.96	1.06	1.38
Gender Specific	.983	.64	.89	.97	1.07	1.38

sections of each course. Allowing different intercepts increased the median correlation from .37 to .38, whereas gender specific equations had a median correlation of .39. These correlation increases seem less modest when expressed as gender mean differences. For instance, the multiple correlation for a (sociology) course near the median increased from .372 to .384 with different intercepts. A grade standard deviation of .92 converts the mean grade of 2.64 for 92 men and 2.40 for 111 women into a gender mean difference of .26 standard deviations.

Cross-validation results are summarized in Table 4 with low, high, quartile, and mean values. Models with gender effects tended to be *less* accurate than the common equation. The predicted-actual grade correlations for the median course declined from .31 to .30 with different intercepts, and to .29 for gender specific equations. With different intercepts, correlations declined in 55% of the courses. The gender specific models had lower correlations in 68% of the courses. Total accuracy declined for the median course as indicated by the RMS error increase from .95 to .96 with different intercepts, and to .97 for gender specific equations. RMS error increased in 59% of the courses with different intercepts and in 72% of the courses with gender specific models.

Prediction model mean differences in accuracy were transformed and evalu-

Table 5
Analysis of Variance F-tests for Prediction Method Effects on Grade Correlation (z-transform) and Log RMS Error

Accuracy Measure	Source	df	M.S. x1000	F	p
Correlation	Methods	2	7.13	12.76	<.001
	Courses	87	47.01		
	M x C	174	.56		
RMS Error	Methods	2	.66	17.30	<.001
	Courses	87	13.06		
	M x C	174	.04		

ated for statistical significance using analysis of variance (see Table 5). The overall null hypothesis was rejected for each measure of accuracy at the .001 level. The Scheffé multiple comparison test among three means for $df = 2$ and 174 at the .01 level of significance requires a criterion $F = 9.46$. The following F tests comparing pairs of models may be evaluated against this criterion value. The gender specific mean correlation was significantly lower ($F = 24.13$) than the common equation mean, and significantly lower ($F = 12.08$) than the different intercepts mean. The gender specific mean RMS error was significantly higher ($F = 32.20$) than the common equation mean, and significantly higher ($F = 17.59$) than the different intercepts mean. The common equation and different intercepts model means were not significantly different using the Scheffé test for either the correlation ($F = 2.07$) or RMS error ($F = 2.19$) measure of accuracy.

To better understand the rather infrequent circumstances under which gender did increase predictive accuracy, we identified the 10 courses in which the RMS error was reduced the most. In 7 of 10 courses there was overprediction of women's grades where women had the lower average grade. There were 2 courses in biology and telecommunications and 1 each in astronomy, geography, and political science. Astronomy illustrates the interplay of variable means (see Table 6) and regression coefficients. For the common equation model the predicted course grade was $-.82 + .46$ HSGPA $+ .11$ SAT-V $+ .20$ SAT-M. For the different intercepts model the grades were predicted from $-.75 + .62$ HSGPA $+ .13$ SAT-V $+ .11$ SAT-M $- .42$ gender. In all models the SAT predictors were scaled by 100. When different intercepts improved accuracy, the pattern involved a fairly large weight on the predictor (HSGPA) on which women had the higher average. However, because men earned higher course

Table 6
Sample Sizes, Grade and Predictor Means by Gender for Example Courses in Base and Cross-Validation Semesters

	Students	HSGPA	SAT-V	SAT-M	Grade
Astronomy:					
Base Semester					
Men	260	2.97	445	518	2.24
Women	269	3.13	443	470	1.86
C-V Semester					
Men	254	2.98	452	512	2.43
Women	230	3.10	434	463	1.96
French:					
Base Semester					
Men	50	2.95	455	508	2.16
Women	123	3.09	449	467	2.60
C-V Semester					
Men	54	2.81	447	516	1.97
Women	108	3.05	451	466	2.48

grades, women's grades would have been overpredicted without a large negative weight on the gender dummy variable. Astronomy illustrates this, and the RMS error was reduced in the cross-validation semester from .890 to .872. The correlation between predicted and actual grades increased from .366 to .403. Note that this gender bias involves overpredicting for women—just the opposite of the gender bias usually found when predicting GPA.

In 3 of these 10 courses there was the usual GPA bias of underprediction for women. The courses were in French, physics, and speech. The variable means for French are displayed in Table 6. For the common equation, the predicted course grade was $-1.60 + .90$ HSGPA $+ .27$ SAT-V $+ .02$ SAT-M. For the different intercepts model, the predicted course grade was $-1.74 + .78$ HSGPA $+ .26$ SAT-V $+ .08$ SAT-M $+ .39$ gender. Here the pattern involved a course in which women had the higher average grade; the weights on HSGPA, SAT-V, and SAT-M were ordinary in size; and women would have been underpredicted without a positive weight on the gender dummy variable. Because women have lower SAT-V and SAT-M scores and higher HSGPAs, if positive weights are given to the SAT variables, then either (a) HSGPA must be given a very large weight, or (b) a positive weight must be given to the gender dummy variable to predict women's higher grades. In the French example, using a gender weight reduced the RMS error from 1.130 to 1.116. The cross-validation correlation increased from .336 to .368.

Among the courses in which different slopes for SAT-V, SAT-M, and HSGPA improved accuracy, the predominant pattern was a heavier weight on HSGPA for women than for men and a large negative weight on the gender dummy variable to prevent women's grades from being overpredicted. When gender bias did occur when predicting course grades, it tended to involve overpredicting for women. This is just the opposite of the usual gender bias of underpredicting GPA.

The secondary purpose of this study was to compare GPA predictive accuracy with predictive accuracy of individual course grades. The predictors were HSGPA, SAT-V, and SAT-M in each case. As usual, women had the higher average GPA, 2.41 versus 2.31, for the 5,765 women and 5,388 men in the cross-validation semester. The means, standard deviations, and intercorrelations of the predictors and GPA appear in Table 7. In conformity with most research,

Table 7
Means, Standard Deviations and Intercorrelations of GPA and Predictor Variables for 11,153 Students Enrolled in One or More Introductory Courses in Spring Semester, 1986

Variable	Mean	S.D.	HSGPA	SAT-V	SAT-M	GPA
Gender	.52	.50	.178	-.053	-.269	.072
HSGPA	3.03	.39		.128	.199	.387
SAT-V	427	88			.491	.261
SAT-M	479	94				.219
GPA	2.36	.67				

GPA was significantly underpredicted for women. The multiple correlation increased from .445 to .446 when the gender dummy variable was added. This increase was significant ($F = 18.51$, with 1 and 11,148 df) at the .001 level. The GPA correlation of .445 was higher than the introductory course median or the mean correlation of .318 (see Table 4). Moreover, GPA was predicted with greater accuracy than was course grade in 89% of the 88 introductory courses. The GPA correlation is normally lowered by mixing several instructors. In this study, course grade correlations also were lowered by mixing the instructors of several sections. These results are similar to those of Strenta and Elliott (1987). They found a higher correlation of .427 with GPA for an SAT-Total predictor than grade correlations of .408 and .405 in two large ($n = 530$) courses. Goldman and Slaughter (1976) used the same predictors as those used here and report a GPA correlation of .44. In two courses with large enrollments ($n = 100$ and 68), one course correlation of .46 was higher and the other of .36 was lower than the GPA correlation. Findings were similar to ours.

How can these results be interpreted? The usual gender bias of underpredicting the GPA of women was replicated at this university. GPAs for different students represent a different mix of courses. There may be systematic differences between the mix of courses for men and women. Such differences cannot occur when predicting a single course grade. With this control, the usual gender bias was not found in predicting grades in most large introductory level courses. Prediction bias would have been found if women consistently earned higher single course grades, as their GPAs are consistently higher. A possible explanation for the higher GPA is that women select less stringently graded courses in greater numbers. Low correlations of .34 and .44 in the two semesters suggest a modest tendency of courses where women earn higher grades to have a higher proportion of women students (see Table 2). These results are consistent with, but do not prove, this explanation, because student ability and the grading standards of the course instructors are hopelessly confounded. It may be argued that women are superior students, enroll more often in certain courses, and instructors give higher grades to these superior students. However, this argument does not explain why these instructors also give higher grades to men. Similar correlations of .38 and .31 in the two semesters were found between the proportion of women students and average grade for men (see Table 2). The proportion of women was a consistent course characteristic with a between-semesters correlation of .92.

Conclusion

Is the relationship between success in college and the predictors of scholastic aptitude and high school grades the same for men and women? When success is measured by a GPA, the answer to this question appears to be "no." We confirmed the usual result. When a single equation was used to predict the cumulative GPA of lower division students at this university, a small but significant amount of underprediction occurred for women. This prediction bias may reflect gender-related course selection effects. When success is measured by the grade in a large, introductory-level course, the answer to this question seems to be "yes." In the great majority of these courses at one large, urban state university, the relationship was the same for men and women. It was counterproductive to adjust grade predictions by gender for most courses. Curiously, in those few courses in which a gender bias was found, it most often involved overpredicting for women in a course in which men earned a higher average grade. This is the reverse of the gender bias usually found when predicting GPA. In few courses could the grade be predicted as accurately as GPA. The extent to which these results would generalize to other universities is unknown.

References

Astin, A. W. (1971). *Predicting academic performance in college: Selectivity data for 2300 American colleges.* New York: Free Press.

Astin, A. W., & Panos, R. J. (1969). *The educational and vocational development of college students.* Washington, DC: American Council on Education.

Boli, J., Allen, M. L., & Payne, A. (1985). High-ability women and men in undergraduate mathematics and chemistry courses. *American Educational Research Journal, 22,* 605–626.

College Entrance Examination Board. (1985). *National college bound seniors, 1985.* New York: Author.

Decore, A. M. (1984). Vive la difference: A comparison of male-female academic performance. *Canadian Journal of Higher Education, 14,* 35–58.

Dixon, W. J. (1983). *BMDP statistical software.* Los Angeles: University of California Press.

Educational Testing Service. (1948–1982). *Scholastic aptitude test.* Princeton, NJ: Author.

Elmore, P. B., & Vasu, E. S. (1986). A model of statistics achievement using spatial ability, feminist attitudes and mathematics-related variables as predictors. *Educational and Psychological Measurement, 46,* 215–222.

Feinberg, L. B., & Halperin, S. (1978). Affective and cognitive correlates of course performance in introductory statistics. *Journal of Experimental Education, 46,* 11–18.

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American Educational Research Journal, 14,* 51–71.

Frisbee, W. R. (1984). Course grades and academic performance by university students: A two-stage least squares analysis. *Research in Higher Education, 20,* 345–365.

Gallini, J. (1982). Evidence of an adaptive level grading practice through a causal approach. *Journal of Experimental Education, 50,* 188–194.

Gamache, L. M., & Novick, M. R. (1985). Choice of variables and gender differentiated prediction within selected academic programs. *Journal of Educational Measurement, 22,* 53–70.

Goldman, R. D., & Slaughter, R. E. (1976). Why college grade point average is difficult to predict. *Journal of Educational Psychology, 68,* 9–14.

Grant, C. A., & Sleeter, C. E. (1986). Race, class and gender in education research: An argument for integrative analysis. *Review of Educational Research, 56,* 195–211.

Hogrebe, M. C., Ervin, L., Dwinell, P. L., & Newman, I. (1983). The moderating effects of gender and race in predicting the academic performance of college developmental students. *Educational and Psychological Measurement, 43,* 523–530.

Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research, 43,* 139–161.

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

Meece, J. L., Parsons, J. E., Kaczala, C. M., Goff, S. B., & Futterman, R. (1982). Sex differences in math achievement: Toward a model of academic choice. *Psychological Bulletin, 91,* 324–348.

Michael, W. B., & Shaffer, P. (1978). The comparative validity of the California State University and Colleges English Placement Test (CSUC-EPT) in the prediction of fall semester grade point average and English course grades of first-semester entering freshmen. *Educational and Psychological Measurement, 38,* 985–1001.

Sawyer, R. (1986). Using demographic subgroup and dummy variable equations to predict college freshman grade average. *Journal of Educational Measurement, 23,* 131–145.

Strenta, A. C., & Elliott, R. (1987). Differential grading standards revisited. *Journal of Educational Measurement, 24,* 281–291.

Authors

ROBERT L. McCORNACK, Director, Institutional Research, University Computing Services, San Diego State University, San Diego, CA 92182. *Degrees:* BA, University of Iowa; MA, PhD, University of Minnesota. *Specializations:* statistics, computer applications.

MARY M. McLEOD, Associate Director, Institutional Research University Computing Services, San Diego State University, San Diego, CA 92182. *Degree:* BA, San Diego State University. *Specialization:* educational research.