

■ **Analysis of Differential Prediction of Law
School Performance by Gender Subgroups**

**Vincent F. Harris
Louis A. Roussos
Peter J. Pashley**

■ **Law School Admission Council
LSAT Technical Report 98-03
January 2006**

The Law School Admission Council (LSAC) is a nonprofit corporation whose members are more than 200 law schools in the United States and Canada. It was founded in 1947 to coordinate, facilitate, and enhance the law school admission process. The organization also provides programs and services related to legal education. All law schools approved by the American Bar Association (ABA) are LSAC members. Canadian law schools recognized by a provincial or territorial law society or government agency are also included in the voting membership of the Council.

© 2006 by Law School Admission Council, Inc.

All rights reserved. No part of this report may be reproduced or transmitted in any part or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission of the publisher. For information, write: Communications, Law School Admission Council, 662 Penn Street, Box 40, Newtown, PA 18940-0040.

LSAT and LSAC are registered marks of the Law School Admission Council, Inc.

This study is published and distributed by the Law School Admission Council (LSAC). The opinions and conclusions contained in these reports are those of the authors and do not necessarily reflect the position or policy of the Law School Admission Council.

Table of Contents

Executive Summary	1
Introduction	1
Methods	2
<i>Sample</i>	2
<i>LSAT Version</i>	2
<i>Variables Used in the Study</i>	2
<i>Analysis Methods</i>	3
Results	3
<i>Descriptive Statistics</i>	3
<i>Predicting First-year Averages</i>	6
Conclusions	8
References	9
Author Note	10

Executive Summary

In the law school admission process, it is essential that the criteria used for admission are fair to all subgroups in the applicant population. One method used to evaluate the fairness of the admission process is to compare the predicted and actual first-year grade-point averages (FYA) for each law school for various subgroups of the applicant population. The purpose of the current study is to address questions of differential prediction between male and female first-year law school students based on data from the 1993, 1994, and 1995 first-year classes of 159 law schools.

Least-squares regression analyses were used to predict first-year average using three traditional predictors: undergraduate grade-point average (UGPA), scale score on the Law School Admission Test (LSAT), and the best predictive linear combination of UGPA and LSAT score. A separate analysis was conducted for each law school included in the study.

The results indicate that a regression model with both LSAT and UGPA as predictor variables performed the best. About 86.8% of schools had predicted minus actual FYA means for females that were less than one point (i.e., a tenth of a standard deviation on the FYA scale) from 0, while about 95.6% of the schools had predicted minus actual FYA means for males that fell within the same range. Only one school experienced a predicted minus actual FYA mean of more than two points. When UGPA was used as the predictor variable, systematic underprediction occurred for males and systematic overprediction occurred for females. In addition, the predicted minus actual FYA mean fell in the range of -1.0 to 1.0 at only 67.9% of the schools for females and 80.5% of the schools for males. When LSAT alone was used, the results were better than when UGPA alone was used, but not as good as when LSAT and UGPA were used in combination. With the model using LSAT alone, 94.3% of the schools had predicted minus actual FYA means for males within a point of zero. For female test takers, 83.0% of the schools had predicted minus actual FYA means that were within a point of zero. For both male and female test takers, the predicted minus actual FYA means were less than two points from zero for all but one school. Overall, the results of this study do not support the concern that the use of LSAT scores or the traditional combination of LSAT scores and UGPAs result in unfair admission decisions with regard to gender.

While considering the results of this study, the reader should keep in mind that the results refer only to subgroup behavior and not to individuals. For example, while results may suggest that UGPAs alone may overpredict FYAs for female law students on average, the performance of many individual female law students will be underpredicted based solely on their UGPAs.

Introduction

In the law school admission process, it is essential that the criteria used for admission are fair to all subgroups in the applicant population. One method used to evaluate the fairness of the admission process is to compare the predicted and actual first-year grade-point averages (FYA) for each law school for various subgroups of the applicant population. If one subgroup of the applicant population experiences either significantly more overprediction (average predicted FYA greater than the actual average FYA) or underprediction (average predicted FYA less than the actual average FYA) than some other group, then *differential prediction* is said to be occurring. The purpose of the current study is to address questions of differential prediction between male and female first-year law school students based on data from the 1993, 1994, and 1995 first-year classes.

Questions about differential prediction are not new to Law School Admission Council (LSAC)-sponsored research nor are they unique to the Law School Admission Test (LSAT) or to the law school admission process. Several studies using LSAT data to investigate questions of differential subgroup prediction have been sponsored previously by the LSAC (e.g., Linn & Hastings, 1984; Powers, 1977; Schrader & Pitcher, 1973, 1974). Prediction issues have been the subject of research studies for other admission-testing programs such as the Scholastic Aptitude Test (e.g., Breland, 1979; Willingham, Lewis, Morgan, & Ramist, 1990) and the Graduate Management Admission Test (e.g., Braun & Jones, 1981). Numerous studies focusing on the same questions in the arena of employment testing have been reported (e.g., Houston & Novick, 1987; National Research Council, 1989; Schmidt & Hunter, 1981). Most of these studies concluded that although there is evidence of differential prediction for minorities, the evidence is generally in the direction of overprediction rather than underprediction. That is, the use of the majority regression model tends to overpredict minority performance on the criterion variable.

The present study, a companion paper to the LSAC Technical Report titled *Analysis of Differential Prediction of Law School Performance by Ethnic Subgroups* (Anthony, Reese, & Pashley, 2000), is designed to address the following two questions:

1. Do the traditional predictors, UGPA and LSAT score, demonstrate differential prediction between males and females when used either separately or in combination to predict law-school performance?
2. Do any of the predictors tend to underpredict future law school performance for either males or females?

Methods

Sample

The sample used in this study is drawn from 1993, 1994, and 1995 entering law-school classes. The study included every law school in the United States and Puerto Rico that participated in the 1996 LSAC Correlation Studies for which three years of sufficient gender data were available, which turned out to be 159 schools. The total number of law schools participating in the LSAC Correlation Studies during that year was 172, so over 92% of all such schools were included. The total pool includes approximately 88,585 law school students across three entering classes. The data from the three classes are combined to assure stability in the analyses and to remain consistent with previous studies. The data are analyzed separately for each law school.

LSAT Version

All students whose data were used in this study were tested with the most recent version of the LSAT. The current version of the test includes five 35-minute sections. One section is a variable section that contains material that is used to pretest new questions or preequate new test forms. The variable section does not contribute to a test taker's score. The other four sections contain items designed to measure analytical (or deductive) reasoning, verbal (or informal logical) reasoning, and reading comprehension. The specific item type makeup is as follows:

Item Type	Number of Items	Time
Reading Comprehension	26 to 28	35 minutes
Logical Reasoning A	24 to 26	35 minutes
Logical Reasoning B	24 to 26	35 minutes
Analytical Reasoning	22 to 24	35 minutes

The total number of scored items was 101 for all the forms analyzed in the current study. A single score derived from the sum of the total number of questions answered correctly across the four scored sections is then equated and reported on an LSAT scale that ranges from 120 to 180. A 30-minute writing sample is administered at the end of the test. This writing assessment is not scored by LSAC, but copies of the writing sample are sent to all law schools to which the test taker applies.

Variables Used in the Study

The variables analyzed in this study are those that are currently used in the LSAC Correlation Studies: first-year average (FYA), undergraduate grade-point average (UGPA), and LSAT score. LSAT score and UGPA are the predictor variables (i.e., the variables that are used to predict performance in the first year of law school). FYA, the measure of performance in the first year of law school, is the criterion variable, or the variable that LSAT and UGPA are used to predict. Only students for whom data are available on each of the three variables were included in this study. Additional operational details related to these three variables are now given:

First-year average. This variable is the average grade earned by the student in the first year of law school. First-year average is provided for each student by the individual law schools. Different law schools use different scales for first-year grades. Data analyses were conducted using FYA on the scale in which the school supplied it. In order to maintain the confidentiality of the individual schools and to allow direct comparison across law schools, FYA values were transformed to a scale having a mean of 50 and a standard deviation of 10. Results presented in this report are on the transformed 50/10 scale.

Undergraduate grade-point average. The average grade earned by each student during his or her undergraduate study is computed by the Law School Data Assembly Service (LSDAS), or according to LSDAS procedures, following the computing options selected for the undergraduate school the student attended. Grades computed in this manner are expressed on a scale of 0.00 to 4.33. The UGPA used in these studies are the same as those used in the LSAC Correlation Studies carried out for the individual law schools.

LSAT scores. Only LSAT scores reported on the 120 to 180 score scale were used in this study. For students who present multiple LSAT scores, a single arithmetic average of the multiple scores was used. If any student took the test more than three times, only the most recent three scores were averaged.

Analysis Methods

The methods used to predict FYA in the current study are the same as those used in the ongoing predictive validity studies for individual schools that participate in the LSAC Correlation Studies. Least-squares regression analyses were used to predict first-year average using three traditional predictors: UGPA, LSAT score, and the best predictive linear combination of UGPA and LSAT score. The regression equations were used to obtain a predicted FYA value for each student at a particular law school. Next, the mean predicted FYA value was obtained for male and female subgroups separately. Actual observed FYA mean values for female and male student subgroups were also calculated. Then, the observed mean for females was subtracted from the predicted mean for females, and the same was done for the males within each law school. These two differences then tell us for a given school the average amount of overprediction or underprediction that occurred for either males or females. By the use of these calculations, the current study seeks to evaluate the amount of differential prediction for males and females. Again, note that the analyses were carried out separately for each law school. Thus, three regression lines were estimated for each school. Also, the analyses used data pooled from the three years under investigation to assure that stable regressions within schools were achieved. LSAC Correlation Studies also use three years of data to obtain stable results.

Results

The results from this study are presented in two sections. The first section includes descriptive data about female and male first-year law students. The second section presents the results of applying the prediction equations derived using the total groups data (i.e., male and female first-year students combined) to female and male students separately.

Descriptive Statistics

Descriptive statistics for the first-year students within the law schools used in this study are presented in Tables 1–3 and in Figures 1 and 2a–2c.

TABLE 1

Number and percentage of female and male first-year students among schools that participated in the 1996 LSAC correlation studies

Entering Class	Total	Number of Schools	Male Count	Male Pct.	Female Count	Female Pct.
1993	35,746	166	19,957	55.83	15,752	44.07
1994	38,693	171	21,718	56.13	16,938	43.78
1995	36,655	166	20,225	55.18	16,388	44.71
Pooled Data	111,094	172	61,900	55.72	49,078	44.18

TABLE 2

Number and percentage of female and male first-year students among schools that participated in the current study

Entering Class	Total	Number of Schools	Male Count	Male Pct.	Female Count	Female Pct.
1993	28,925	155	16,113	55.71	12,812	44.29
1994	30,937	159	17,328	56.01	13,609	43.99
1995	28,723	154	15,814	55.06	12,909	44.94
Pooled Data	88,585	159	49,255	55.60	39,330	44.40

TABLE 3
A summary of the number of included law schools by average size of gender group

Gender Group	Size of Group				
	< = 50	50-100	100-150	150-200	> 200
Male	13	69	59	11	7
Female	26	93	31	6	3

Table 1 describes the overall gender subgroup breakdown among the 172 schools that participated in the 1996 LSAC Correlation Studies. Table 2 provides similar information for the 159 schools included in the current study. A comparison of the two tables shows that this study is very representative of the LSAC Correlation Studies sample. Of the 88,585 students who reported male or female as their gender at the 159 schools represented across the three years, 39,330 (44.4 percent) are female and 49,255 (55.6 percent) are male. The percentage of females participating in the LSAC Correlation Studies is consistent with the 44.3% reported by the American Bar Association (Morgan & Snyder, 1997) for all law schools over that three-year period. According to Table 1, the percentage of females and males have fluctuated only slightly across the three years of this study. The percentages of male and female first-year law students for each year of the study is quite close to the percentage of males and females for the pooled data. On average, the number of males is about 25% more than the number of females.

Table 3 describes the number of included law schools at various gender group sizes (averaged over the three years). There are 93 schools that have average female class sizes between 50 and 100 students. This is the largest number of schools at any of the class size groupings for females. Sixty-nine schools were found to have average male class sizes between 50 and 100 students. This, too, is the largest number of schools at any of the class size groupings for males. At this interval, where males and females both have the largest number of schools represented, females have far more schools represented. This is the largest class size interval where females have the larger number of schools represented. Males were found to have a larger number of schools represented at all class size intervals greater than 100 students and females have a larger number of schools represented at all class size intervals less than or equal to 100 students.

Continuing to explore the first-year law students in this study, Figure 1 presents the distribution of the law schools according to the percentage of females. The vast majority (76.7%) of the participating schools had between 42% and 54% females in the three first-year classes included in the present study. Thus, in agreement with the data presented in Tables 1 and 2, the data here indicate that most schools had slightly more male students than female students.

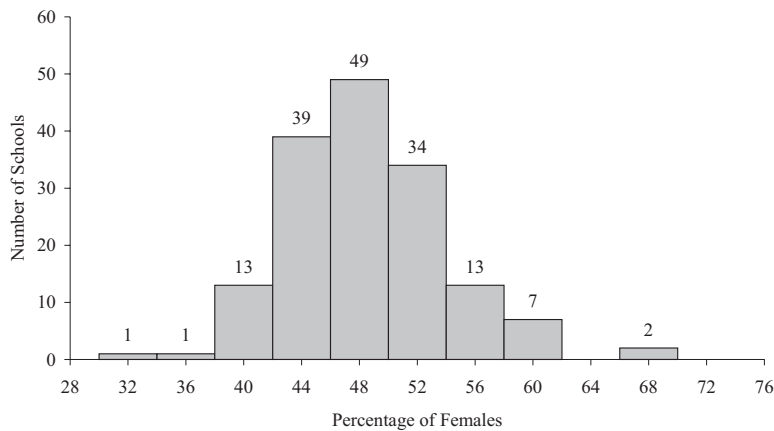


FIGURE 1. *Frequency distribution of percentage of females at the participating law schools*

In describing the included law students, performance issues were investigated in Figures 2a, 2b and 2c. The histograms in these figures represent the distribution of the law schools according to the mean difference (calculated at each school) between males and females on three indexes: LSAT score, UGPA, and FYA. A positive number indicates that males outperform females on the studied index. A negative number can be interpreted as females outperforming males on the studied index.

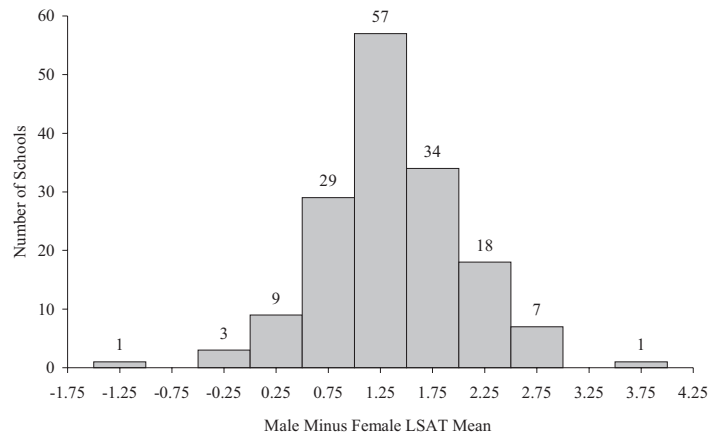


FIGURE 2a. Frequency distribution of differences between male and female LSAT means at the participating law schools

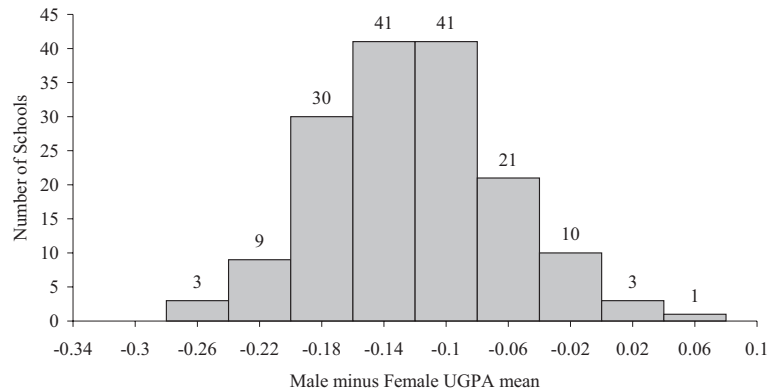


FIGURE 2b. Frequency of differences between male and female UGPA means at the participating law schools

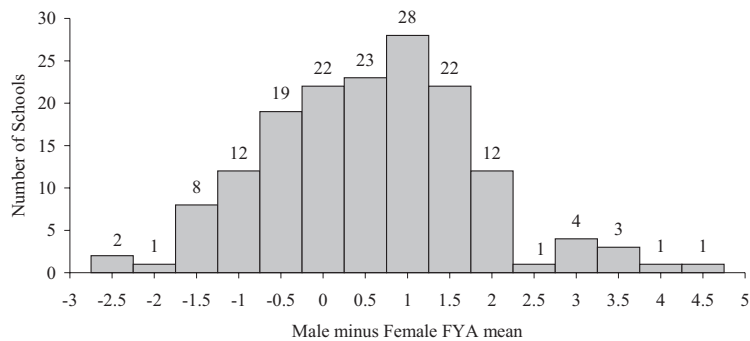


FIGURE 2c. Frequency distribution of differences between male and female FYA means at the participating law schools

Figure 2a presents a histogram of LSAT mean difference between males and females. There are 159 values, one for each school. The histogram shows that most (75.5%) of the schools observed mean differences between 0.50 and 2.00. The total percentage of schools in which the male mean was higher was 97.5. Of the schools in which the male mean was higher, 83.2% of them experienced at most a two-point mean difference between males and females. So, while most of the differences favored males, the magnitude of most of the differences was small. Figure 2b is a male minus female UGPA histogram. Whereas males tended to have higher LSAT means, Figure 2b suggests that females tend to have higher UGPA means at the participating

schools. One hundred fifty-five schools, or 97.5%, showed a higher mean UGPA for women than for men. Looking at the graph, it can be seen that 133 (83.6%) of the schools have differences between male and female UGPA means that are greater than -0.20 but less than or equal to -0.04 . So, while the direction of the differences is systematic, the magnitude of the differences is not very substantial. Figure 2c is the last histogram that investigates performance. It looks at the difference between male and female mean FYA by participating school. Since 8 of the 22 schools at the interval whose midpoint is zero have a positive male / female difference, approximately 64.8% of the schools in the study were found to have a mean FYA that was higher for males than for females. The overwhelming majority of the schools (84.3%) had differences in male mean and female mean between -1.75 and 1.75 . Considering that the data are on a standardized scale with a mean of 50 and a standard deviation of 10, such differences are considered small, and no schools had mean differences of 5 or more points. Only 8 schools had differences of 3 points or more, and all these schools had FYA means in favor of males.

Predicting First-year Averages

Concern about differential prediction is an outgrowth of a larger concern about how to most fairly and accurately evaluate test scores and undergraduate grade reports included in law-school application materials. One method to address the question of differential prediction is to determine how accurately LSAT scores, UGPA, and a combination of LSAT score and UGPA predict performance in law school for female and male law school students. Figures 3, 4, and 5 examine the differences between predicted and actual FYA for males and females. Prediction equations that used LSAT, UGPA, and a combination of LSAT and UGPA were examined. The mean actual first-year average earned for students at a participating school was subtracted from the average predicted first-year average for students at the school. A negative difference should be interpreted as underpredicted performance in law school; likewise, a positive difference means the regression equation overpredicted performance. For reporting the results from this study, all of the first-year averages have been converted to a scale where the mean for the total group (i.e., males and females combined) at a law school is set to 50 and the standard deviation to 10. The conversion is made to preserve the confidentiality of the data and to allow comparisons across law schools.

Figures 3 to 5 present the primary results of the current study, the distributions across the participating law schools of the difference between mean predicted FYA and actual mean FYA, calculated separately for males and females, with each figure corresponding to a different predictor.

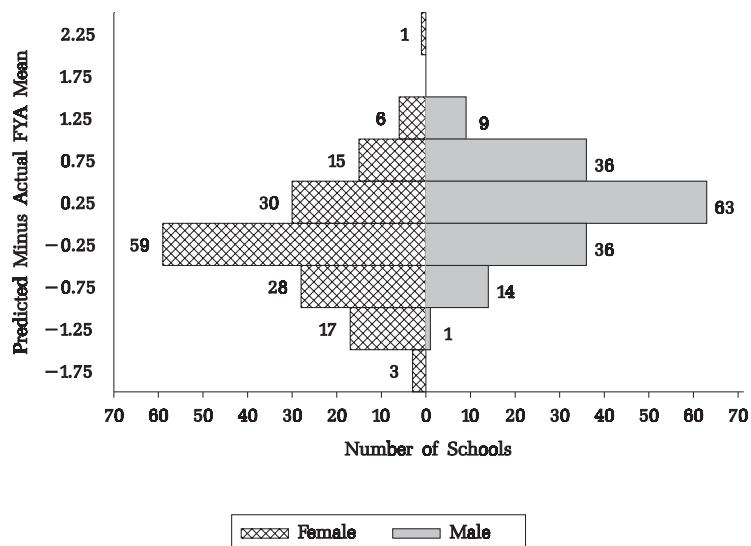


FIGURE 3. Frequency distribution of differences between predicted and actual FYA means for males and females at participating law schools using LSAT score as the predictor variable

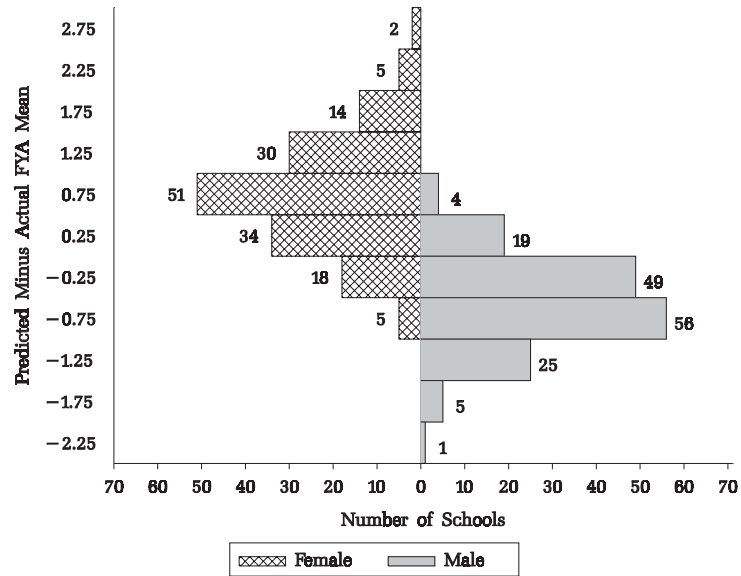


FIGURE 4. Frequency distribution of differences between predicted and actual FYA means for males and females at participating law schools using UGPA as the predictor variable

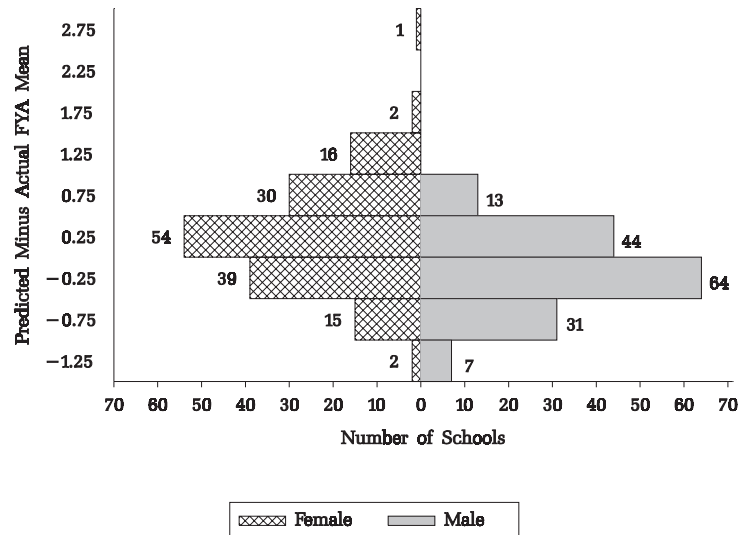


FIGURE 5. Frequency distribution of differences between predicted and actual FYA means for males and females at participating law schools using LSAT score and UGPA as predictor variables

Figure 3 presents the results for the model that used LSAT score as the predictor of FYA. There are two sets of bars on the histogram, with the solid bars representing male data and the checkered bars representing female data. The numbers on the vertical axis are the midpoints of each predicted minus actual FYA mean category. Each category includes the lower bound number and goes up to but does not include the upper bound number. For example, the solid horizontal bar at 1.25 includes all those schools that have a predicted minus actual FYA mean for males that is at least 1.00 point but less than 1.50 points. Nine schools fall into this category. The figure shows that when LSAT alone is used as the predictor, 67.3% of the law schools had female FYA means that were underpredicted, and 67.3% of the law schools had male means that were overpredicted. Approximately 83.0% of the schools had predicted minus actual FYA means for females in the range of -1.0 to 1.0 (within one tenth of a standard deviation), while 94.3% of the law schools had predicted

minus actual male FYA means falling in the same range. No schools had a predicted minus actual FYA mean as great as 3 points. Indeed, only a single school had a predicted minus actual FYA mean as large as two points. Thus, in general, the magnitude of predicted minus actual FYA means using LSAT score as the predictor is quite small.

Figure 4 presents the results for the model that used UGPA as the predictor. The results indicate that the use of UGPA alone does not produce predicted FYA values that are as equitable as those from using LSAT alone. Figure 4 suggests that the use of UGPA alone in predicting FYA systematically overpredicts female FYA mean and underpredicts male FYA mean. Approximately 85.5% of the schools had female FYA means that were overpredicted, while 85.5% of the schools had male FYA means that were underpredicted. About 67.9% of the differences between predicted and actual FYA means for females fall in the range of -1.0 to 1.0 . For males, 80.5% fall within that range. Figure 4 shows that seven schools had predicted minus actual FYA means for females that were greater than two points and one school had a predicted minus actual FYA mean for males that was more extreme than two points. Although still quite small, the magnitudes of differences between predicted and actual FYA means for both gender groups are more extreme for the equation using UGPA alone as the predictor than for the equation using LSAT alone as the predictor. (Recall the FYA scale has a mean of 50 and a standard deviation of 10.) The pattern of over- and underprediction also seems to be more systematic for UGPA (i.e., overpredicting female FYA while underpredicting male FYA) than for LSAT score.

Figure 5 presents the results for using a combination of LSAT and UGPA as the predictor. Their combined use decreases the amount of over- and underprediction that was observed for males and females, respectively, in Figure 4. The use of LSAT and UGPA together still tends to overpredict FYA for females and underpredict FYA for males, but not to the extent that the use of UGPA alone does. Figure 5 shows about 64.2% of the schools had a female mean FYA that was overpredicted, while about 64.2% of the schools had a male FYA mean that was underpredicted. However, over 95% of the male mean differences now fall in between -1.0 and 1.0 , while 86.8% of the female mean differences now also fall within that range. Thus, even though underprediction still occurs for males and overprediction still occurs for females, the magnitudes of these differences are generally rather small. Indeed, only one school now has mean female overprediction of more than 2 points.

Conclusions

This study analyzed data from 159 law schools to determine whether evidence exists of differential prediction between males and females. Like previous LSAC studies dealing with differential prediction, this study relies on a regression-model-based definition of fairness in selection. That is, the prediction would be considered unfair if the regression equation consistently and systematically excluded members of an identifiable subgroup as a result of seriously underpredicting the performance of its members (or seriously overpredicting the performance of some other group).

The model that used LSAT and UGPA in combination appears to perform best in terms of minimizing differential prediction between males and females. This model showed no serious tendency to either overpredict or underpredict FYA for males or females. Furthermore, for a majority of schools, the predicted minus actual FYA means for both males and females fall in a small interval around zero when LSAT and UGPA were used together as predictors. Predicted minus actual FYA mean for males fell in the range of -1.0 to 1.0 (within one tenth of a standard deviation) at 95.6% of the schools. Predicted minus actual FYA mean for females fell in this same range at 86.8% of the schools.

The model that used UGPA alone was found to perform most poorly. This model had the most systematic pattern of overprediction of FYA means for females and underprediction of FYA means for males. This phenomenon is probably related to the observed differences between male and female UGPA that were examined in Figure 2b, which were very systematic and in the opposite direction than the differences between male and female LSAT score displayed in Figure 2a. Although the pattern of over- and underprediction was systematic, the magnitude of differences was still fairly small. Predicted minus actual FYA mean fell between -1.0 and 1.0 at 80.5% of schools for males and at 67.9% of schools for females.

The model that used LSAT alone performed better than the model using UGPA alone, but not as well as the model that used both predictors in combination. When LSAT was used as the sole predictor, FYA tended to be overpredicted for male students and underpredicted for female students. Again, the magnitude of the differences was rather small, with most falling between -1.0 and 1.0 . About 83% of the differences for females fell within this range and 94.3% of the differences for males fell within this range.

At least two caveats should be remembered while evaluating the results of this study. First, only differences in *average* predicted performance were analyzed. That is, individuals within a subgroup which is overpredicted on average may still be themselves underpredicted in terms of their individual law school performance. It is also possible (though highly unlikely) that subgroups who are overpredicted on average may be underpredicted at a particular point on the predictor scale (e.g., at a UGPA cutoff score one law

school might employ). Second, differential prediction is only one aspect of an overall construct validity evaluation. Other aspects, such as the strength of the correlation between a predictor and a criterion variable, should also be considered when deciding whether prediction equations are equitable and valid.

References

- Anthony, L. C., Reese, L. M., & Pashley, P. J. (2000). *Analysis of differential prediction of law school performance by ethnic groups* (LSAT Technical Report No. 98-02). Newtown, PA: Law School Admission Council.
- Braun, H., & Jones, D. (1981). *The Graduate Management Admission Test prediction bias study* (Graduate Management Admission Council Report No. 81-04, Educational Testing Service RR-81-25). Princeton, NJ: Educational Testing Service.
- Breland, H. (1979). *Population validity and college entrance measures* (Research Monograph No. 8). New York: College Entrance Examination Board.
- Houston, W. M., & Novick, M. R. (1987). Race-based differential prediction in Air Force technical training programs. *Journal of Educational Measurement*, 24, 309–320.
- Linn, R. L. & Hastings, C. N. (1984). Group differentiated prediction. *Applied Psychological Measurement*, 8, 165–172.
- Morgan, R. L., & Snyder, K., Esq. (Eds.). (1997). *ABA approved law schools: Statistical information on American Bar Association approved law schools* (1st ed.). Compiled by the Office of the Consultant on Legal Education for the American Bar Association. New York: Macmillan.
- National Research Council. (1989). Differential validity and differential prediction. In John A. Hartigan and Alexandra K. Wigdor (Eds.), *Fairness in employment testing: Validity generalization, minority issues and the general aptitude test battery* (pp. 172–188). Washington, DC: National Academy Press.
- Powers, D. E. (1977). *Comparing predictions of law school performance for Black, Chicano, and White law students* (Law School Admission Council Report No. 77-3). Newtown, PA: Law School Admission Services, Inc.
- Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128–1137.
- Schrader, W. B., & Pitcher, B. (1973). Predicting law school grades for black American law students (Report No. LSAC 73-06). In *Law School Admission Council, Reports of LSAC sponsored research: Volume II: 1970–1974*. Princeton, NJ: Law School Admission Council.
- Schrader, W. B., & Pitcher, B. (1974). Prediction of law school grades for Mexican American and black American students (Report No. LSAC 74-8). In *Law School Admission Council, Reports of LSAC sponsored research: Volume II: 1970–1974*. Princeton, NJ: Law School Admission Council.
- Willingham, W. W., Lewis, C., Morgan, R., & Ramist, L. (1990). *Predicting college grades: An analysis of institutional trends over two decades*. Princeton, NJ: Educational Testing Service.

Author Note

Vincent F. Harris is now with Bank of America. Louis A. Roussos is now with University of Illinois8.