ACT Research Report Series 94-3

ACT Composite Scores of Retested Students

Charles Wood Lanier

June 1994

ACT

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243

ACT COMPOSITE SCORES OF RETESTED STUDENTS

Charles Wood Lanier

ABSTRACT

Many students elect to take the ACT Assessment more than once. This paper is concerned with the typical changes in scores observed between students' first and second testings. Analyses were based on data from a sample of students who tested twice between 1989 and 1992. The scores of twice-tested females, males, blacks, whites, and students who tested on particular national test dates were also studied.

On average, students' Composite scores increased by 0.8 scale score units from the first to the second testing. About one-third of the students' Composite scores increased by two or more units, and about one-fifth of the students' Composite scores declined. The analyses of subgroup data showed that given the same initial test scores, males were slightly more likely to earn a particular second score or higher than were females; whites were more likely to earn a particular second score or higher than were blacks; and students who tested late in the junior year of high school and again early in the senior year were more likely to earn a particular second score or higher than were students who first tested after October of the senior year.

Other variables were also considered for their relationships to retest scores. Students' high school course work and high school rank were both practically and statistically significant predictors in nearly every model explored. This result suggests that students can improve their ACT Assessment scores by taking extensive course work in ACT-tested areas and by earning high grades.

ACT COMPOSITE SCORES OF RETESTED STUDENTS

During the past decade there has been a steady increase in the proportion of students who elect to repeat the ACT Assessment. According to internal ACT management reports, approximately 10% of ACT Assessment examinees tested at least twice between October 1983 and June 1985. This figure rose to 16% for the test dates between October 1986 and June 1988, and jumped to nearly 28% (based on slightly different criteria) for the test dates between October 1989 and June 1991. Among students who last tested in the 1992-93 academic year, the retest rate was slightly over 30%.

Although most examinees who retest improve on their initial scores, some examinees experience a decrease. As more students take the ACT Assessment multiple times, educational researchers need to explain score changes in ways that students, parents, and counselors can readily understand.

Depending on one's perspective, changes in true scores, as well as changes in observed scores, may be of interest. According to a standard model used by psychometricians, a person's observed score is equal to the sum of a "true score" and a random measurement error. The magnitude of the measurement error can be described in terms of probability. For example, the magnitude of the measurement error for the ACT Composite score is, with probability approximately .95, less than 2 points. Therefore, an observed ACT Composite score is, with probability approximately .95, within ±2 points of the true Composite score (ACT, 1989). An analogous statement can be made about changes in the scores of multiple-tested students: Assuming that measurement errors on different testing occasions are independent, an observed Composite score change is, with 95% probability, within ±3 points of the true Composite score change.

While psychometricians are concerned with true score changes and measurement errors, college-bound students and their advisors are primarily concerned with observed scores. They know that students' observed scores will be used as measures of readiness for college, and they simply want the highest scores possible. Thus, a student's goal in retesting is to improve upon the previous observed score.

There are several factors that directly influence changes in observed scores over multiple testings. For example, self-selection, measurement error, practice effects, motivation, and the possibility that learning occurred between test administrations all can influence changes in observed scores. However, many of these factors are often either very difficult or impossible to assess. This study focused on associations between factors that are more easily assessed (such as ethnicity, gender, high school course work, and high school grade point average) and score differences between a first and second administration of the ACT Assessment.

Purpose

The principal goal of this study was to estimate the probability that a student will achieve or exceed a particular Composite score on the second administration of the ACT Assessment, given the score achieved on the initial administration. Variables such as race, sex, high school grade point average, courses taken, and other background variables were also considered. An examination of the associations between these variables should address the concerns of many high school students, parents, and counselors by providing answers to the following questions:

1. Given an initial Composite score, X_1, and other background characteristics, what is the probability that a certain score, X_2, is reached or exceeded?
2. Does taking certain high school courses or receiving high grades in these courses significantly increase the probability of obtaining a higher Composite score upon retesting?
3. Do the probabilities differ by race or gender, or by dates of initial and subsequent administrations of the ACT Assessment?

Data

Data for this study consisted of the records of individuals who took the ACT Assessment more than once on test dates between October 1989 and June 1992, inclusive. The file was generated by selecting recurring Social Security numbers (SSNs) and ACT Identification Numbers from the master file of ACT records. ACT Assessment Composite scores are the focus of this study, in order to limit its scope, and because Composite scores are widely used for making college admission decisions. Furthermore, only examinees' first two test scores were studied, because few students take the ACT Assessment three or more times, and because most students are more concerned with whether to retake the ACT Assessment than with how many times to retake it. The initial Composite score is referred to as X_1, and the second score as X_2.

After incomplete records were deleted (those with missing or invalid test scores or high school GPAs), over 500,000 records remained. Two percent of the records were selected (by systematic random sampling) to provide the data used in the analyses. Of the 10,184 student records selected, 4,304 were for males, 5,880 were for females, 1,078 were for blacks, and 7,978 were for whites.

Student records were also grouped by typical testing patterns (see Figure 1). Group #1 includes only students who tested and then retested during the spring term of the junior year in high school. Group #2 includes students who initially tested during the spring term of the junior year, and then retested in the fall term of the senior year. Students in Group #3 initially tested in October of the senior year and then retested later in the senior year. The students in Group #4 tested and retested after October of the senior year. Of the 10,184 records in the total sample, 1,942 were in Group #1, 4,371 were in Group #2, 1,076 were in Group #3, and 720 were in Group #4. The records in these four groups collectively account for about 80% of the total sample.

The mean initial Composite score of 20.2 for multiple-tested students was slightly below the mean Composite score of 20.6 for all students who graduated from high school in 1990, 1991, or 1992. The average initial score, average second score, and range of scores for the sample used in this study are reported in Table 1.

Table 1
Means, Standard Deviations, Minimums, and Maximums of Initial and Second ACT Assessment Composite Scores, by Student Group

Student group	n	Initial score				Second score			
		Mean	S.D.	Minimum	Maximum	Mean	S.D.	Minimum	Maximum
Females	5,880	20.1	3.9	7	33	20.9	4.1	7	34
Males	4,304	20.3	4.3	8	34	21.3	4.5	10	35
Blacks	1,078	16.8	3.2	7	30	17.5	3.3	7	30
Whites	7,978	20.8	3.9	8	34	21.6	4.1	11	35
Group #1	1,942	21.6	3.8	11	32	22.2	4.0	12	34
Group #2	4,371	20.6	4.0	10	34	21.4	4.1	11	34
Group #3	1,076	18.8	3.7	7	32	19.5	3.9	7	32
Group #4	720	17.2	3.6	10	31	17.8	3.6	10	31
Total	10,184	20.2	4.0	7	34	21.0	4.2	7	35

For the total sample, the change from initial score to second score ranged from -9 to +10 scale score points. The distribution of the observed score changes had mean of +0.8 and a median and mode equal to +1.0. A frequency distribution for observed score change is presented in Table 2.

Table 2
Frequency Distribution for the Change
Observed in Composite Scores on Retesting

Change score	Frequency	Percent
+3 or more	1,529	15.0
+2	1,809	17.8
+1	2,450	24.1
0	2,254	22.1
-1	1,352	13.3
-2	578	5.7
-3 or less	212	2.1
Total Group	10,184	100.1*

* Total exceeds 100 percent due to rounding.

Method

A logistic regression model (Hosmer and Lemeshow, 1989) was used to estimate the probability of scoring at least x_2, given an initial score of x_1:

$$P[X_2 \geq x_2 \mid X_1 = x_1] = \frac{1}{1 + \exp\{-a_{x_2} - b(x_1)\}}, \quad 7 \leq x_1 \leq 34$$

where x_2 is fixed, and where the full range of values (occurring in the sample) for X_1 is used to predict the probability of obtaining a score of x_2 or higher.

The values assigned to X_2 were x_2 = 17, 18, 21, 24, 27, and 30. They were chosen for the following reasons. During 1993, the National Collegiate Athletic

Association (NCAA) used a cutoff score of 17 to determine first-year athletes' academic eligibility. (A cutoff score of 18 was used until fall, 1992.) A minimum Composite or subject area score of 21 is often required for enrollment in standard first-year courses. Other scores in the twenties are used for placement in more advanced first-year courses. A Composite score of 30 is typical of students who enroll in highly selective institutions or who receive certain academic scholarships. Therefore, this range of values covers most scores used for college admission and placement, and academic and athletic scholarships. Interpolation could be used to estimate results for intermediate values.

Note that this model allows the constant term (a_{x_2}) in the exponent to vary with given values of X_2, but constrains the slope term (b) to be fixed with respect to X_2. This type of model can be estimated by SAS PROC LOGISTIC (SAS Institute Inc., 1990) with one pass through the data. An iteratively reweighted least squares procedure computes maximum likelihood estimates ($\hat{a}_{x_2}$, $\hat{b}$) of the parameters. These estimated parameters were used to construct the estimated probability curves displayed in Figure 2.

The model is based on the assumption that the slope parameter b remains constant across all values of X_2. The appropriateness of this assumption was tested with a chi-square score statistic, and the resulting p-value was 0.17. Therefore, observed differences in the slopes associated with different values of X_2 could be reasonably attributed to chance. For further details on testing the equal slopes assumption, see Lanier (1993).

Next, logistic regression models based on the equal-slopes model were estimated for groups defined by sex, race, and testing pattern. Models were then compared for females and males, blacks and whites, and students who repeated the ACT Assessment within the time frames displayed in Figure 1. Comparisons between coefficients for different groups were made by testing the hypothesis H_o: $a_{x_2,i} = a_{x_2,j}$ and $b_i = b_j$, where i and j represent the two groups being compared. For details about the statistical procedure used to test this hypothesis, see Lanier (1993).

The models discussed so far include only a single explanatory variable: initial test score, X_1. Modeling the probability of reaching or exceeding a given second score may be enhanced by considering other explanatory variables. Students' high school course work, accomplishments, plans for the future, and variables related to other background information were used when fitting logistic regression models to the data. Results from earlier comparisons were also used to develop multiple variable logistic models.

Results

The estimated conditional probability curves in Figure 2, Figure 3, Figure 4, and Figure 5 pertain to the total group, females and males, blacks and whites, and students following different testing patterns, respectively. From an examination of the plots, it is clear that differences exist between females and males, blacks and whites, and groups of students following different testing patterns.

Figure 3 shows that differences between females and males are moderate but consistent across all levels of X_2 except where X_2=17 or 18. Given identical initial scores, males are more likely to reach or exceed a given second score than are females. The largest difference occurs when the initial score is 29; females achieve a score of at least 30 with an estimated probability of .48, while males achieve that same score with an estimated probability of .58.

Chi-square test statistics indicate that differences in the logistic regression coefficients are statistically significant (i.e., that the observed differences can not be reasonably attributed to chance). For all levels of X_2 except X_2=17 and X_2=18, there existed a statistically significant difference between females and males ($p<.025$). Moreover, an overall test between females and males (i.e., the set of all coefficients for females was compared to the set of all coefficients for males) revealed that a statistically significant difference existed between the sexes ($p<.001$).

It is clear from Figure 4 that, given identical initial scores, whites were more likely to reach or exceed a given score on retesting than were blacks. For example, given an initial Composite score of 26, the estimated probability is .55 that whites will achieve a score of 27 or higher on retesting. For blacks with an initial score of 26, the probability of achieving a score of 27 on retesting is estimated to be .28. Differences exist at all levels.

Differences between blacks and whites in their logistic regression coefficients were statistically significant, but not as consistently so as the differences between females and males. Whites were more likely to reach or exceed a second score of 17 ($p<.03$), 18 ($p<.03$), 21 ($p<.001$), or 27 ($p<.01$) than blacks. However, there were no statistically significant differences between the logistic regression coefficients for whites and blacks at the .05 level when the second score criterion was $X_2=24$ or $X_2=30$. A likely reason is that the sample size for blacks was much smaller than for the gender groups.

A comparison among the groups of students exhibiting the four most common testing patterns was more complex. According to Figure 5, the estimated conditional probabilities for Group #2 are consistently higher than those for the other groups across all levels of X_2. Even more noticeable is the fact that the estimated probabilities for Group #4 are consistently lower than those for all other groups. There may be some benefit to taking the ACT Assessment late in the junior year of high school and again in the fall of the senior year (i.e, being in Group #2). Students who take the test late in their senior year and repeat soon after (i.e., those in Group #4) are not as likely to improve at the same rate as those in the other groups.

A chi-square statistic was used to test the statistical significance of differences among the slope coefficients for the four groups. The alpha level used for these tests was different because multiple (six) comparisons were made. For a difference between groups to be considered statistically significant at the .01 level, the test between groups was required to be statistically significant at a level of 0.01/6 = 0.00167. The only comparison that gave a chi-square statistic significant at this level was the test at X_2=18 for Group #2 versus Group #4 (p<.001). A comparison of the entire set of coefficients for each group with the corresponding set of coefficients for each other group yielded no statistically significant results.

Multiple-Predictor Models

The next step was to estimate logistic regression models, based on multiple predictor variables, that best described the data for the different subgroups. A stepwise logistic regression procedure (SAS PROC LOGISTIC) was used to select the variables in each model. The procedure selects variables according to their statistical significance. An entry and exit p-value of .01 was used to select variables (see Lanier (1993) for a detailed description).

The variables selected were then reviewed to determine their practical significance. One way to check for practical significance, when studying coefficients of logistic regression models, is to look at how a change in the value of an explanatory variable effects the dependent variable. More specifically, one can determine how a change in the value of an explanatory variable affects the odds ratio of the probability of reaching or exceeding a given test score, given the value x of an explanatory variable.

$$OR(x) = \frac{p(x+1)/[1-p(x+1)]}{p(x)/[1-p(x)]} \quad ,$$

By adding one unit to one explanatory variable in the model and keeping all other variables constant, the effect of the single variable can be measured.

Multiple-predictor logistic regression models allow further exploration of the differences between groups discovered earlier. For example, given X_2=27, a statistically significant difference ($p<.01$) was found between the univariate models (X_1 was the only predictor used) for females and males. Therefore, separate multiple-predictor models were developed for females and males when X_2=27. Practical significance was determined for each variable that was found to be statistically significant ($p<.01$). These models for females and males and the coefficients for each statistically significant variable are presented in Table 3.

An alphabetical listing of each variable appearing in Table 3 and an explanation of each variable follows:

ENG11 - grade in eleventh grade English course;
HSGPA - high school grade point average;
MONTHS - months elapsed between first and second test date;
OTHERMA - grade in math courses other than algebra, geometry, trigonometry, or beginning calculus;
RANK - high school quartile rank;
TRIG - grade in a trigonometry course;
VA - expect (1)/ do not expect (0) to participate in intercollegiate varsity athletics during the first year of college;
X_1 - initial test score;
YRSOFM - years of mathematics courses taken; and
YRSOFNS - years of natural science courses taken.

Table 3
Logistic Regression Models for the Probability of Scoring at or Above a Given Score on Retesting, by Student Subgroup

Student subgroup	Retest Score (X_2)	Model	Variables	Coefficient
Females	27	univariate	constant	-27.0205
			** X_1	1.0403
		multivariate	constant	-29.2105
			** X_1	1.0331
			** HSGPA	0.6509
Males	27	univariate	constant	-26.3101
			** X_1	1.0231
		multivariate	constant	29.0223
			** X_1	0.9783
			** HSGPA	0.5892
			* YRSOFNS	0.1760
			* MONTHS	0.0927
Blacks	21	univariate	constant	-21.3907
			** X_1	1.0609
		multivariate	constant	-26.3681
			** X_1	1.0893
			** YRSOFM	0.5734
Whites	21	univariate	constant	-20.1518
			** X_1	1.0232
		multivariate	constant	-22.3569
			** X_1	0.9882
			** RANK	0.4939
			* YRSOFM	0.1276
			* MONTHS	0.0519
			* TRIG	0.1150
Group #2	18	univariate	constant	-17.4690
			** X_1	1.0539
		multivariate	constant	-19.6642
			** X_1	1.0443
			** RANK	0.3965
			* YRSOFM	0.1619
			** OTHERMA	0.2871

(continued on next page)

* Statistically significant (p<.01) but not practically significant
** Statistically significant (p<.01) and practically significant

Table 3 (cont.)
Logistic Regression Models for the Probability of Scoring at or Above a Given Score on Retesting, by Student Subgroup

Student subgroup	Retest Score (X_2)	Model		Variables	Coefficient
Group #4	18	univariate		constant	-17.7581
			**	X_1	1.0379
		multivariate		constant	-19.6905
			**	X_1	1.0440
			**	RANK	0.6581
			**	ENG11	0.7329
Total Group	18	univariate		constant	-17.1669
			**	X_1	1.0308
		multivariate		constant	-18.7082
			**	X_1	0.9819
			**	VA	-0.2286
			**	RANK	0.3836
			*	YRSOFM	0.1205
			*	YRSOFNS	0.0823
			*	OTHERMA	0.1505
Total Group	27	univariate		constant	-26.6576
			**	X_1	1.0308
		multivariate		constant	-30.5426
			**	X_1	1.0021
			**	HSGPA	0.4122
			*	RANK	0.4641
			*	YRSOFNS	0.1301
			*	MONTHS	0.0792

* Statistically significant ($p<.01$) but not practically significant
** Statistically significant ($p<.01$) and practically significant

Also presented in Table 3 are the multiple-predictor models for blacks and whites when X_2=21, for Group #2 and Group #4 when X_2=18, and for the total group when X_2=18 and X_2=27. When X_2=21, X_1 and YRSOFM are the only statistically and practically significant variables for blacks, and X_1 and RANK are the only statistically and practically significant variables for whites.

Discussion

Many people ask for help when making difficult educational decisions. This research should help students, parents, and counselors understand better the chances of attaining higher scores on the ACT Assessment. Testing first in the junior year, taking a strong college-preparatory curriculum, and earning high grades in courses, all increase the likelihood of reaching or exceeding a given score on retaking the ACT Assessment.

One result of the multiple-predictor logistic regression portion of this study was particularly interesting. During the period of this study, the NCAA used an ACT Assessment score of 18 as a cutoff score for academic eligibility. The variable VA was an indicator variable used to determine whether or not students planned to participate in collegiate varsity athletics. In the model fitted for the total group, with X_2=18 as the criterion for success, the variable VA was found to be a statistically and practically significant contributor to the model.

Table 3 shows that the coefficients for all the variables except VA have the same sign. The exception for VA occurred because it is negatively related to the probability of reaching or exceeding a score of 18 upon retesting, while all other variables have a positive relationship with this probability. The above result demonstrates that the likelihood of reaching or exceeding a second score of 18 is diminished for students who are potential collegiate athletes. Clearly, students regarded as potential collegiate athletes influence the model. Since the data for this study were collected, NCAA recently changed the eligibility cutoff score to 17. In September, 1994, the NCAA will implement a sliding scale based on GPAs and cutoff scores. Replication and variation of this study with more current data could yield interesting results.

Two issues addressed by this research are those of gender and race. Males are somewhat more likely than females to reach or exceed a particular second score X_2, given the same first score, except when X_2=17 and X_2=18. Differences between blacks and whites were also found to exist at several particular levels of X_2. Future research may help sort out the causes for differences between gender and racial/ethnic groups.

Earning high grades in high school courses and taking the ACT Assessment for the first time in the eleventh grade both appear to be helpful in increasing the ACT Composite score through retesting. The variable RANK (quartile rank in high school class) is dependent on HSGPA. Both of these variables were shown to be of practical significance in every logistic model explored except when blacks were modeled at X_2=21. Students who test for the first time after October of the senior year have consistently lower probabilities of reaching a given second score for all levels of X_2 than do students who first test in the junior year. More research is needed in this area to determine what other characteristics may be associated with increased probabilities of obtaining higher Composite scores on retesting.

References

ACT, (1989). ACT Assessment Program Technical Manual. Iowa City, Iowa. Author.

Hosmer, D. W., and Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley and Sons.

Lanier, C. W. (1993). Examination of Observed Scores on the ACT Assessment. Unpublished master's thesis, University of Iowa, Iowa City.

SAS Institute Inc. (1990). SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2. Cary, North Carolina: Author.

Figure 1
Typical Patterns of Testing Among Multiple-Tested Students

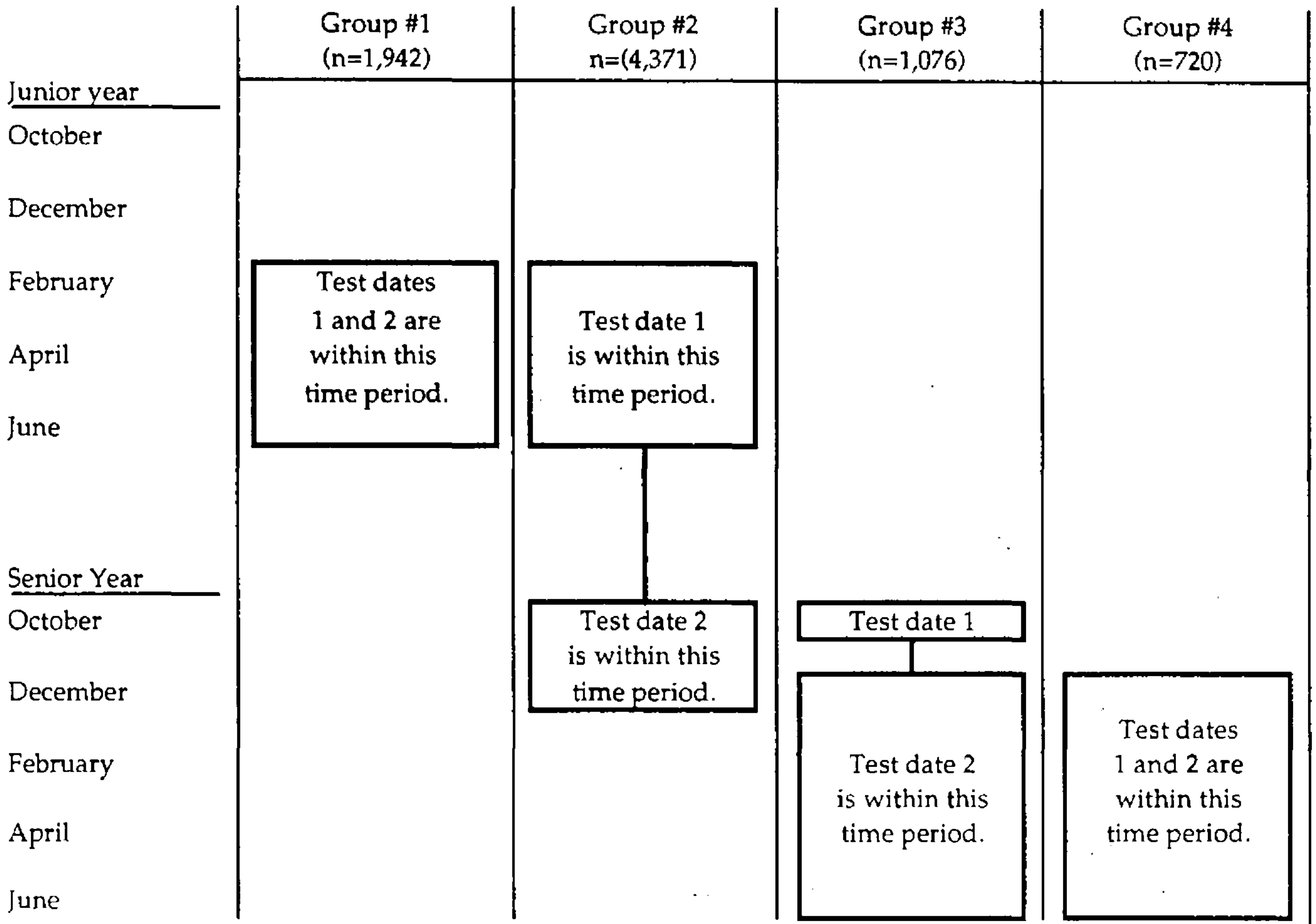

Figure 2
Probability of Obtaining an ACT Composite Retest Score (X_2), Given an Initial ACT Composite Score (X_1)

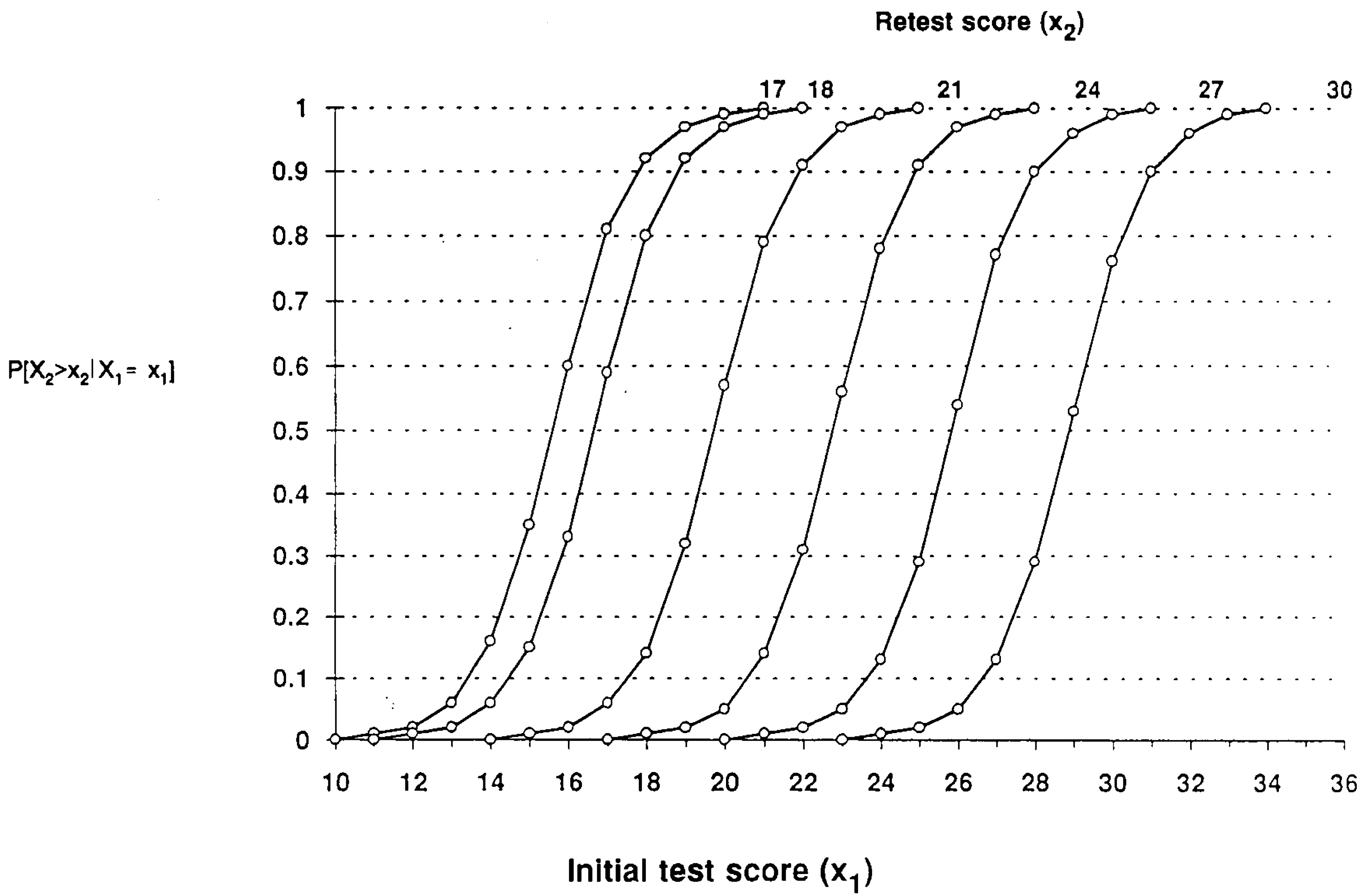

Figure 3
Probabilities of Scoring at or Above Given Score Levels on Retesting, for Given Initial Scores of Males and Females

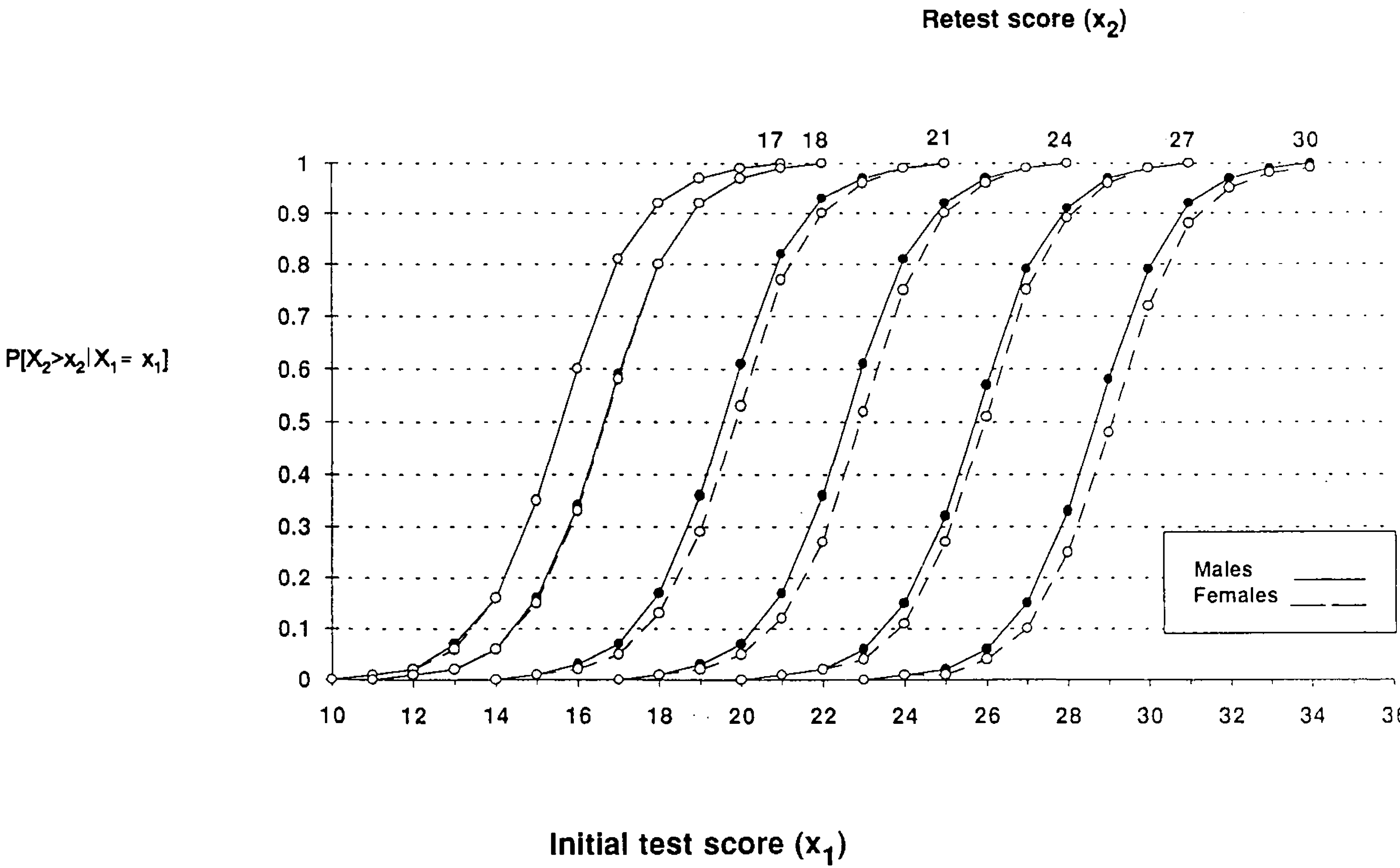

Figure 4
Probabilitles of Scoring at or Above Given Score Levels on Retesting, for Given Initial Scores of Blacks and Whites

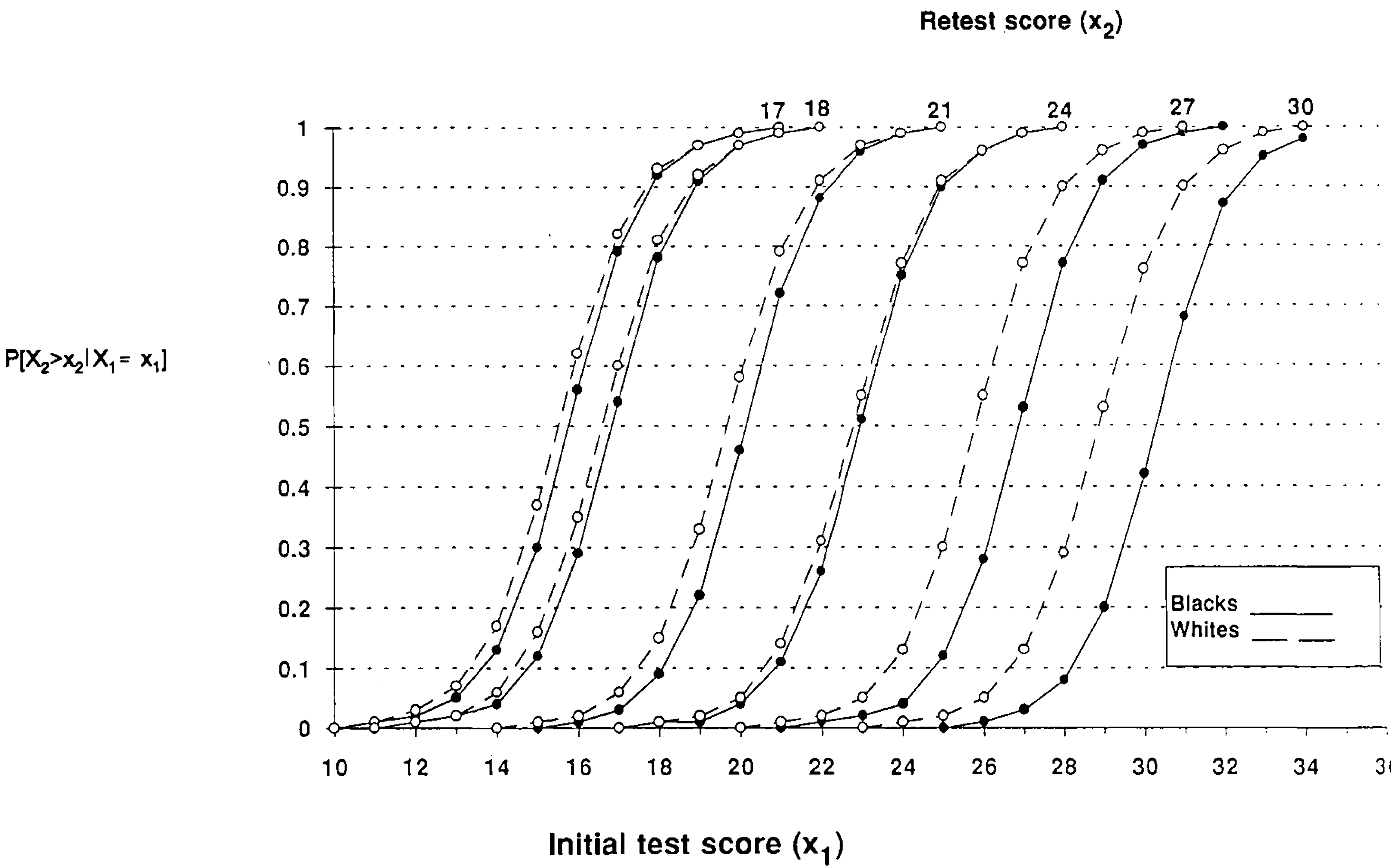

Figure 5
Probabilities of Scoring at or Above Given Score Levels on Retesting, for Given Initial Scores of Students with Different Retesting Patterns