ACT Research Report Series 2000-3

Validating Two-Stage Course Placement Systems When Data Are Truncated

Jeff L. Schiel

Matt Harmston

ACT

February 2000

For additional copies write:
ACT Research Report Series
PO Box 168
Iowa City, Iowa 52243-0168

Validating Two-Stage Course Placement Systems When Data Are Truncated

Jeff L. Schiel
Matt Harmston

Abstract

In two-stage course placement systems, students first take a screening test. Students who score at or above the screening test cutoff score K enroll directly in a standard college course, whereas those who score below K take a placement test. Students who subsequently score at or above the placement test cutoff K' also enroll in the standard course. Consequently, students in the standard course will not have placement test scores below K'. Moreover, placement test scores are somewhat truncated above K', because students who earned scores above K on the screening test did not have to take the placement test. Hence, their placement test scores, which likely would have equaled or exceeded K', are "missing."

Previous research has only examined truncation in one-stage placement systems, in which it occurs below, but not above, the cutoff score. In this study, the effects of truncation on estimated optimal cutoffs, accuracy rates, and success rates under different combinations of logistic regression curve, test score distribution shape, and sample size were examined for two-stage placement systems. It is shown that even when data are moderately truncated in such systems (e.g., baseline truncation below K' and 80% truncation above K'), validity statistics and optimal cutoffs can be estimated with reasonable accuracy.

Validating Two-Stage Course Placement Systems When Data Are Truncated

Postsecondary institutions often use standardized test scores when deciding into which courses students ought to be placed. After selecting a cutoff score, institutional staff will permit students scoring at or above it to be placed into a standard course (e.g., pre-calculus). Students scoring below the cutoff will be placed into a lower-level, remedial course (e.g., college algebra). For the benefit of their institutions and students, institutional staff want to make correct placement decisions, of which there are two types: 1) students placed into a standard course have the necessary skills and knowledge to ultimately succeed in the course, and 2) students placed into a remedial course would not have succeeded in the standard course had they instead been placed into it. Incorrect placement decisions may negatively affect both students and institutions. For example, a student with better-than-average mathematical skills who is incorrectly placed into a remedial mathematics course may become frustrated by the expense and time required to complete an additional course, and may consider transferring to another institution.

If students, parents, or others perceive placement systems as being unfair or hastily developed, then these systems may be criticized. By establishing statistical validity evidence that relates standardized test scores or other variables to successful performance in standard courses, institutions can strengthen their respective rationales for using certain placement procedures, tests, and cutoff scores. In this way, institutions are better prepared to respond to potential criticism of their placement systems.

One method for providing course placement validity evidence uses logistic regression and decision theory to describe relationships between outcomes in standard college courses and test scores, estimate proportions of correct decisions given particular cutoff scores, and identify optimal cutoffs (ACT, 1994; Noble & Sawyer, 1997; Sawyer, 1989; Sawyer, 1996). In evaluating course placement systems, logistic regression can be used to estimate the conditional probability of success $\hat{P}$ in a standard course, given test score (or other predictor variables). Estimated probabilities can then be used with the marginal distribution of test scores to estimate other course placement validity statistics, such as the *accuracy rate* $\hat{A}$, which is the estimated proportion of correct placement decisions. The *optimal cutoff score* is the cutoff score at which $\hat{A}$ is maximized. Another validity statistic, the *success rate* $\hat{S}$, is the estimated proportion of students succeeding in the standard course, among all students who could have been placed in that course.

Because students who score below the cutoff typically do not enroll in the standard course and do not have course outcome data (e.g., grades), the data of course placement systems are truncated below the cutoff. This presents certain difficulties in estimating statistics, regardless of the method used to evaluate a placement system. For example, a logistic regression function, which is computed from the data of students who completed the standard course, must be extrapolated to test scores below the cutoff in order to estimate $\hat{P}$, $\hat{A}$, and $\hat{S}$ over the entire range of placement test scores. Thus, the statistics will be useful only to the extent that their accuracy is not adversely affected by truncation.

In general, as truncation increases, the accuracy of validity statistics decreases (Houston, 1993; Schiel & Noble, 1992; Schiel, 1998; Schiel & King, 1999). Moreover, *hard truncation*, a condition in which data are unavailable for all students below the cutoff, generally results in less accurate validity statistics than does *soft truncation*, where data are available for some, but not all, students (Schiel, 1998; Schiel & King, 1999). One instance in which soft truncation occurs is when an institution does not strictly enforce a cutoff, but permits students who score below it to enroll in a standard course. For example, students with low placement test scores may be confident that they can succeed in the standard course, or they may furnish to the institution additional information that suggests they are likely to succeed (e.g., a score on an ancillary, local placement test). Whatever their reasons for enrolling in the standard course, some of the students with scores below the cutoff will have standard course outcome data that can be included with the data of students scoring above the cutoff, thereby augmenting the sample used to estimate validity statistics.

Using computer-generated data to estimate conditional probabilities of success, Schiel (1998) found that fairly accurate estimates of $\hat{P}$ could be obtained under simulated soft truncation when the logistic regression curve was steep. In addition, distributions that were initially negatively skewed with respect to the predictor (test score) variable tended to be more resistant to the influence of truncation than did symmetrical distributions.

When examining estimated optimal cutoff scores, Schiel noted that data with steep logistic curves tended to produce reasonably accurate estimates (i.e., accurate to within 1 ACT Assessment scale score point), even with what was termed "40%" soft truncation. In general, the slope of the logistic curve and the skewness of the marginal test score distribution appeared to have little to do with the relative accuracy of the validity statistics unless soft truncation exceeded 40%.

Schiel and King (1999) studied a somewhat different definition of soft truncation. They used a chance-level score below which all student data were deleted (not to be confused with the cutoff score). Moreover, the authors specified that observations that were below the cutoff score, but nearer to it, would have a higher probability of being retained than would those that were nearer to the chance-level score. The rationale for this was that the nearer a low-scoring student was to the cutoff score, the greater was his or her likelihood of enrolling in the standard course.

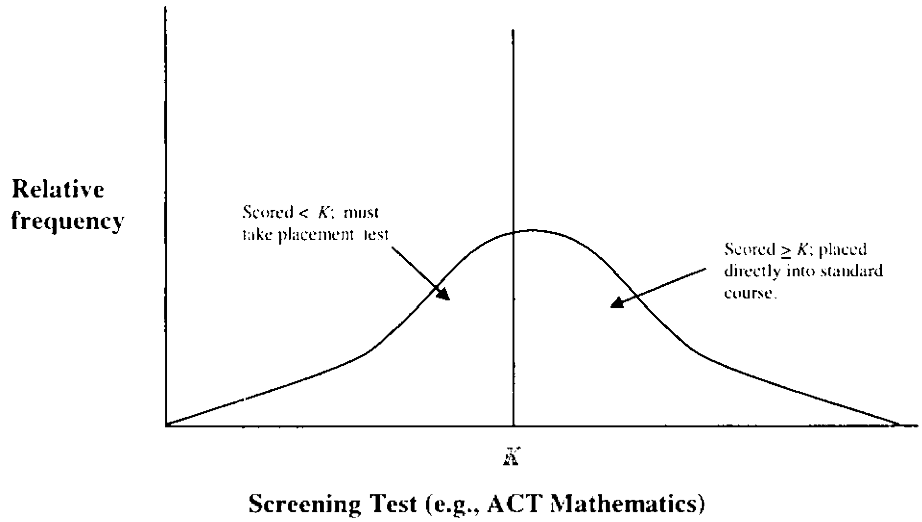
With some exceptions, Schiel and King observed that reasonably accurate estimates of $\hat{P}$ could be obtained under varied levels of soft truncation. Although increased degrees of soft truncation were associated with decreased accuracy in $\hat{P}$ and $\hat{A}$, the decrease was not unacceptably large. In addition, reasonably accurate optimal cutoff scores could often be obtained under 40% soft truncation. In some instances, the accuracy of cutoff scores was reasonable under soft truncation as high as 80%.

Truncation in a Two-Stage Placement System

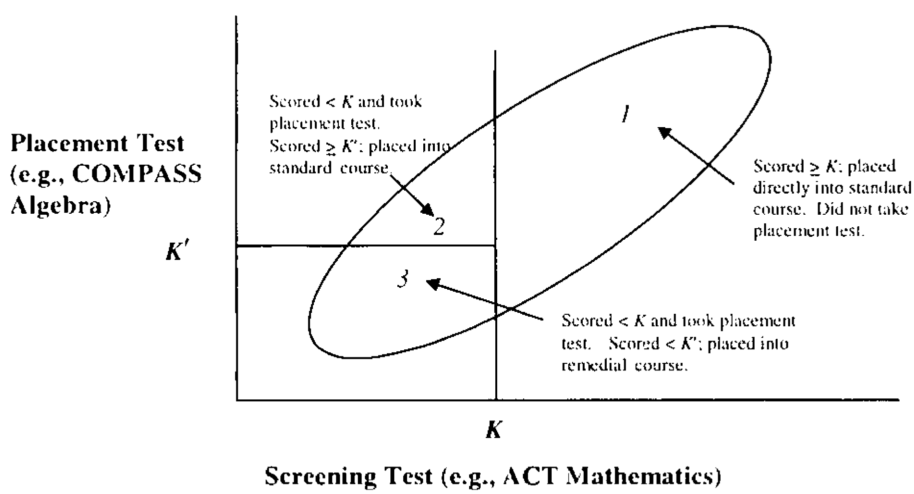
Previous research has only examined truncation as it occurs *below* a given cutoff score in one-stage placement systems. There are, however, situations in which truncation is present both below and above the cutoff. For example, in two-stage placement systems, students are required to take a screening test and, in some instances, a placement test as well. Such a system is illustrated in Figure 1.

FIGURE 1. Placement Using Screening and Placement Tests

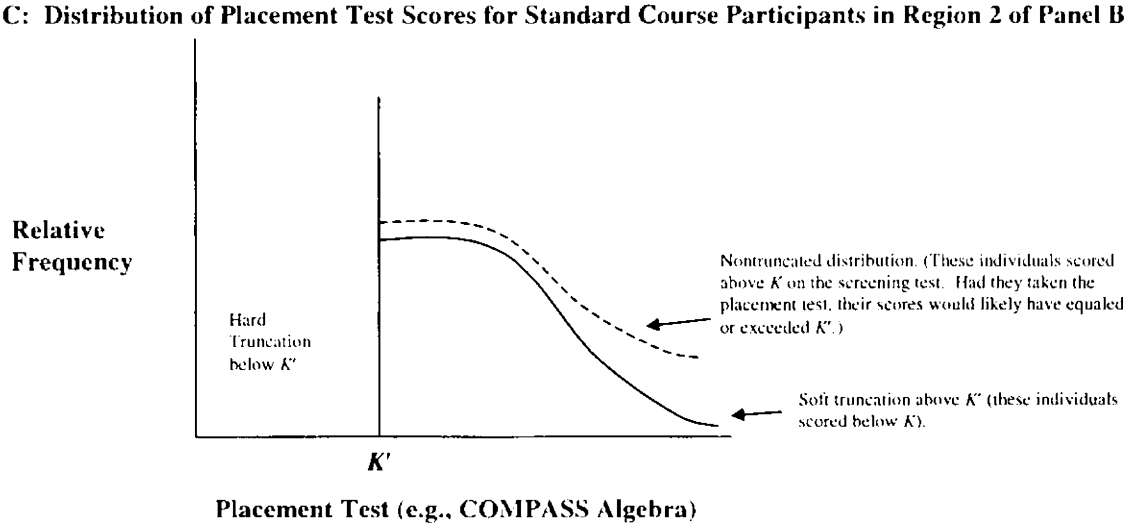
A: Placement Based on Results of Screening Test



B: Placement Based on Results of Screening Test and Placement Test



C: Distribution of Placement Test Scores for Standard Course Participants in Region 2 of Panel B



In this placement system, all incoming students are tested with the ACT Assessment (screening test), a curriculum-based test used in college admissions and placement. ACT Mathematics scores are used, for example, as an initial indicator of whether to place students into either a standard or remedial mathematics course. As shown in Panel A of Figure 1, all students scoring at or above the screening test cutoff score K are placed directly into the standard course. Those scoring below K must instead take the COMPASS Algebra test (placement test). COMPASS is a computer adaptive testing system that measures students' academic skills and knowledge in mathematics, reading, and writing.

Panel B of Figure 1 illustrates that of those students who must take the placement test, only students scoring at or above the cutoff score K' on this test can enroll in the standard course. Consequently, both hard and soft truncation of the conditional placement test score distribution (Region 2) for standard course participants may be present, as shown in Panel C. Hard truncation occurs below K', whereas soft truncation occurs above K'. The dashed curve in Panel C illustrates a nontruncated condition above K'. Note that soft truncation can also occur below K', depending on an institution's enforcement of cutoff scores.

Soft truncation (as depicted in Panel C) occurs because students who earned high scores (i.e., $\geq K$) on the screening test did not have to take the placement test. Hence, their placement test scores, which likely would have equaled or exceeded K', are "missing." Note that the relationship between the scores on the screening test and those on the placement test is imperfect. For example, most students who earn low scores on the screening test (i.e., $< K$) will also earn low scores on the placement test, but some will earn high placement test scores. If the screening test and placement test were perfectly correlated, then hard truncation would occur above K' and there would be no need for the placement test.

This study investigated the effects of truncation on the accuracy of validity statistics for two-stage placement systems. As described in the following section, the extent of simulated truncation both below and above K' was adjusted. It was expected that as truncation was increased to the point where it was relatively severe (e.g., hard truncation below K' paired with 80% soft truncation above K'), the accuracy of estimated validity statistics would decrease. However, given that truncation in a two-stage placement system differs from that in a one-stage system, it was possible that relationships between truncation severity and validity statistic accuracy would differ in the two systems.

Method

Computer-generated data representing a two-stage placement system were used in this study. The screening test was assumed to have the score scale and properties of an ACT Assessment subject area test (e.g., Mathematics). A cutoff score of 20 was selected for the screening test because of its consistency with cutoffs identified in Houston (1993) and in ACT's course placement research (ACT, internal memorandum, September 17, 1998). With certain assumptions concerning the shape of the test score distribution (e.g., negative skewness), this cutoff would place approximately 38% of ACT-tested examinees into the standard course.

It was assumed in this study that hypothetical examinees scoring below 20 on the ACT Mathematics test would take the COMPASS Algebra test. Those scoring at or above a selected cutoff score on COMPASS Algebra would be placed directly into the standard course. A target COMPASS cutoff of 32 was used for two reasons. First, this cutoff results in approximately the same percentage of examinees enrolling in the standard course as does the ACT Mathematics cutoff score of 20. Second, the cutoff is near the COMPASS cutoff used by a large state postsecondary system, which uses the ACT Assessment as a screening test. Due to the initial screening based on ACT Mathematics, this study assumed that conditional, truncated COMPASS distributions were positively skewed, as illustrated in Figure 1. Moreover, some of the placement group (nontruncated) distributions were generated to have positive skew, to mimic that exhibited by the distribution of COMPASS Mathematics scores for students nationwide.

Throughout this paper, ACT Mathematics and COMPASS Algebra scores are used as examples to facilitate discussion, as well as to provide a rationale for selecting cutoff scores. However, the results of the study are not necessarily limited to mathematics tests, or even to these two test batteries.

Generation of Placement Group Data

Nontruncated COMPASS score distributions were generated to form placement groups. A *placement group* consists of all students for whom placement decisions must be made and for whom placement test scores are available. In this study, data for 11 placement groups were generated. Placement group distributions contained standard course outcomes corresponding to the full range of COMPASS scores, including those that would have been earned by high-scoring ACT Assessment examinees had they taken COMPASS. Validity statistics from these distributions were considered "true" values to which validity statistics from truncated distributions were compared.

Placement groups were defined according to two sample sizes (100 and 500), two logistic function slopes (steep and flat), and three levels of skewness of the marginal distribution of COMPASS scores (approximately zero, medium positive, high positive). Table 1 describes the characteristics of the 11 placement groups.

TABLE 1

Placement Group Characteristics

Placement group	N	Estimated optimal cutoff	Slope	Logistic model parameter estimates	Skewness of marginal COMPASS distribution	
1	500	32	Steep	$b_0 = -3.67$, $b_1 = .12$	High pos.	(.70)
2	500	33	Steep	$b_0 = -3.53$, $b_1 = .11$	Medium pos.	(.46)
3	500	32	Steep	$b_0 = -3.71$, $b_1 = .12$	Zero	(.18)
4	500	27	Flat	$b_0 = -0.63$, $b_1 = .02$	High pos.	(.66)
5	500	30	Flat	$b_0 = -1.16$, $b_1 = .04$	Medium pos.	(.47)
6	500	39	Flat	$b_0 = -1.49$, $b_1 = .04$	Zero	(.10)
7	100	36	Steep	$b_0 = -4.84$, $b_1 = .14$	High pos.	(.70)
8	100	33	Steep	$b_0 = -6.36$, $b_1 = .19$	Medium pos.	(.20)
9	100	32	Steep	$b_0 = -3.91$, $b_1 = .13$	Zero	(-.01)
10	100	34	Flat	$b_0 = -0.47$, $b_1 = .01$	High pos.	(.82)
11	100	35	Flat	$b_0 = -1.88$, $b_1 = .06$	Zero	(-.15)

Data were also generated for a twelfth placement group of size n=100 with a flat slope and medium skewness. It was found, however, that the maximum $\hat{A}$ for this group occurred at a COMPASS score of 16. Such a low optimal cutoff score would not likely be used in actual placement systems. Moreover, the low optimal cutoff prevented the development of score intervals for purposes of truncation simulation (see the following section). For these reasons, data from this particular placement group were not analyzed.

The data generation process consisted of the following steps:

1) COMPASS scores were generated using methods similar to those in Houston (1993). Under the condition of high skewness, for example, random variables X_1 and X_2 were drawn from gamma $(1.5,\beta)$ and gamma $(3,\beta)$ distributions, respectively. The COMPASS score X was defined as $X_1 / (X_1 + X_2)$, and was

distributed as a beta (1.5,3) random variable. Because X was continuous, ranging from 0 to 1, it was multiplied by 99 and rounded to the nearest integer to obtain a COMPASS score. Table 1 shows the actual skewness for each placement group.

2) A logistic regression function was used to calculate $\hat{P}$ using the obtained COMPASS score (X). The "slope" parameters (β_1) were selected to be representative of those observed for the data of institutions participating in ACT's Course Placement Service. These parameters were fixed to be .12 and .03, respectively, for the steep and flat slope conditions. The "intercept" parameters were then found by solving for β_0 in the logistic function $\pi = [1 + \exp(-\beta_0 - \beta_1 X)]^{-1}$, with $\pi = .5$ and $X = 32$. These parameters were fixed to be –3.84 and -.96, respectively, for the steep and flat slope conditions. Using these slope and intercept parameters to generate data ensured that when logistic curves were subsequently fitted to the data, their inflection points (corresponding to the optimal cutoff scores) would occur near a COMPASS score of 32. Table 1 contains the (fitted) logistic parameter estimates.

3) Using the probability calculated in Step 2, a random variable Y was selected from a Bernoulli distribution with $\Pr(Y = 1) = \hat{P}$, for each value of X. Course success was represented by $Y = 1$; failure by $Y = 0$.

4) Steps 1 through 3 were repeated 1000 times. Placement groups of size n=500 or n=100 were randomly selected from the "population" of 1000 generated observations consisting of COMPASS score/course outcome (x,y) pairs. The population was intended to represent the entire freshman class at an institution

from which a placement group is selected. More than 1000 observations were generated initially to replace those that were eliminated because they were below 16 (chance level).

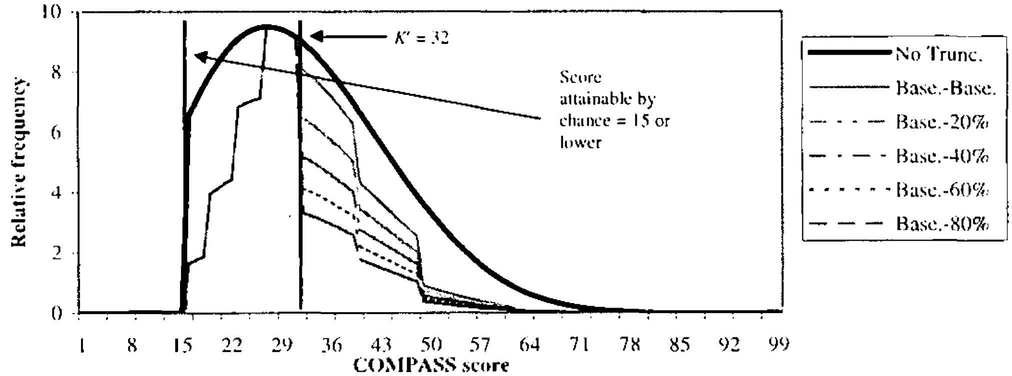
After the data were generated and placement groups selected, $\hat{P}$, $\hat{A}$, $\hat{S}$, and an optimal cutoff score were calculated for each placement group by first fitting logistic curves to the generated data. Note that most optimal cutoffs varied from the target cutoff of 32 due to random error (see Table 1).

Truncation Simulation

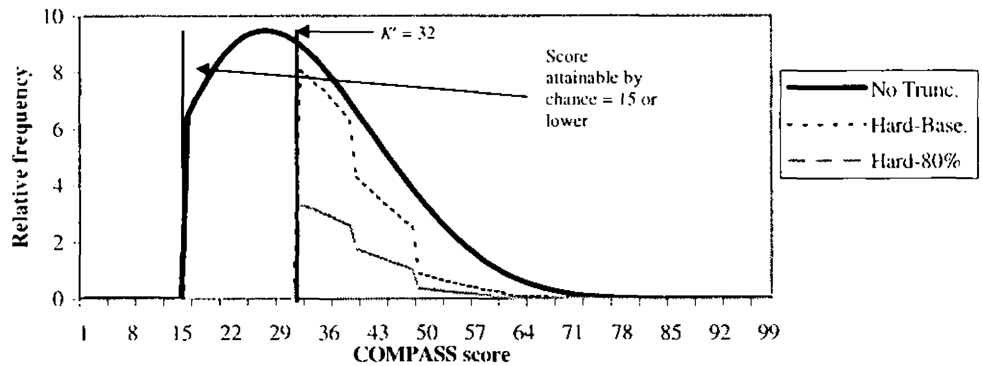
Truncation below K'. For the portion of the conditional COMPASS distribution below K', two truncation conditions were used. First, a baseline truncation condition was defined similar to that in Schiel and King (1999), where 0%, 25%, 50%, and 75% of observations were removed from 4 respective score intervals that were progressively more distant from K'. Conditions utilizing the baseline condition paired with each level of truncation above K' are illustrated in Panel A of Figure 2. The second truncation condition used was hard truncation, in which all observations below K' were deleted (see Panel B of Figure 2).

FIGURE 2. Score Distributions Under Seven Truncation Conditions

A: Baseline Truncation Below *K′* Paired With Truncation Conditions Above *K′*



B: Hard Truncation Below *K′* Paired With Truncation Conditions Above *K′*



Score intervals were defined using empirical data consisting of 26,635 observations from examinees who took both the ACT Assessment and COMPASS, and who scored below 20 on ACT Mathematics. Each of the four intervals below K' for the empirical data encompassed an approximately equal proportion of observations falling between the chance-level score and K'. The intervals were: 16-18, 19-22, 23-26, and 27-31. Intervals below K' for the generated (placement group) data varied somewhat from these, due to the slightly different shapes of the respective distributions, but were similarly intended to encompass approximately equal proportions of observations.

Truncation above K'. Above K', five truncation conditions were paired with the baseline condition below K', as illustrated in Panel A of Figure 2. The COMPASS score intervals were defined by examining a percentage polygon of the empirical COMPASS data described above. Endpoints of the intervals corresponded to slight fluctuations in the otherwise smooth curve of the polygon. The intervals were: 32-39, 40-48, 49-61, and 62-99, and they contained 46%, 32%, 14%, and 8%, respectively, of the observations above K'. Intervals for placement groups were defined so that they had widths similar to those of the empirical distribution. For placement groups with estimated optimal cutoff scores other than 32, widths for the first 3 intervals were maintained, although the locations of the intervals changed as a function of K'.

The empirical data used to define the four intervals above K' were considered to represent an "intermediate" or "typical" truncation condition, which we called a "60%" truncation condition. It seems reasonable to assume that this degree of truncation would occur above the cutoff in many two-stage placement systems. Of course, truncation could be more or less severe than this. We wanted the intervals in the (simulated) intermediate truncation condition to contain percentages of observations as noted above (46%, 32%, 14%, and 8%). In addition, we wanted truncation to proceed in 20% increments starting from a baseline truncation condition. For example, 20% of the observations from the baseline condition would be randomly selected and then removed to create a "20%" truncation condition. Twenty percent of the observations from the 20% condition would then be removed to create a "40%" condition, and so on. In order to accomplish these goals, we defined the baseline condition above K' such that 10%, 35%, 75%, and 85% of observations were removed from four respective intervals progressively more distant from K' in the nontruncated placement group distribution. An 80% truncation condition was used to examine the effects of truncation beyond the intermediate condition.

Due to the fact that placement group distributions were similar in shape but not identical, maintaining interval widths above K' and using the same amount of truncation to create baseline conditions resulted in somewhat different percentages of observations for subsequent truncation conditions. However, these differences were not substantial and therefore did not likely influence the results.

Panel B of Figure 2 illustrates that hard truncation below K' was paired with two truncation conditions above K'. These combinations were chosen to represent moderate and extreme truncation conditions.

Five hundred data sets of appropriate sample size were simulated for each combination of the 11 placement groups and 7 truncation conditions, by randomly selecting and then removing observations within each of the intervals shown in Figure 2. Table 2 contains truncation sample sizes, by placement group and truncation condition. Depending on the shape of the placement group distribution and the location of the optimal cutoff score, truncation samples varied considerably in size, ranging from 10 (Placement Group 7, Hard/80%; Placement Group 10, Hard/80%) to 272 (Placement Group 1, Baseline/baseline).

TABLE 2

Truncation Sample Sizes, by Placement Group and Truncation Condition

Placement Group	N before truncation	Hard/ baseline	Hard/ 80%	Baseline/ baseline	Baseline/ 20%	Baseline/ 40%	Baseline/ 60%	Baseline/ 80%
1	500	160	65	272	239	214	193	177
2	500	157	65	256	224	199	179	164
3	500	170	69	221	186	159	137	120
4	500	172	70	256	221	194	171	154
5	500	174	72	251	216	189	166	149
6	500	164	68	246	213	187	165	150
7	100	25	10	60	54	51	48	45
8	100	28	11	47	41	38	33	30
9	100	35	15	45	37	32	27	25
10	100	26	10	55	50	45	42	39
11	100	28	12	41	35	31	29	25

Figure A in the appendix provides additional information about the truncation process. It illustrates this process for Placement Group 1, beginning with no truncation and ending with the Hard/80% truncation condition. Figure A shows how the truncation sample sizes in Table 2 were obtained for this particular placement group.

Comparing Placement Group and Truncation Sample Validity Statistics

Logistic curves were fit to each of the 500 data sets that were simulated for each placement group/truncation condition combination. Validity statistics were calculated using the methods described in Sawyer (1996). Median validity statistics (over 500 simulations) were then calculated for each truncation condition and compared to those obtained for the respective placement groups, using procedures described in Schiel and King (1999). For example, the placement group $\hat{P}$s were subtracted from the baseline/80% truncation condition (median) $\hat{P}$s at each COMPASS score point (16-99). The (unweighted) mean difference over 84 score points ($\Delta\hat{P}$) was then calculated, and the mean of the absolute values of the differences was also calculated. Finally, estimated optimal cutoff scores were identified for the placement group (the "true" cutoff) and for each truncation condition. Differences between optimal cutoffs for each truncation condition and its corresponding true optimal cutoff were calculated.

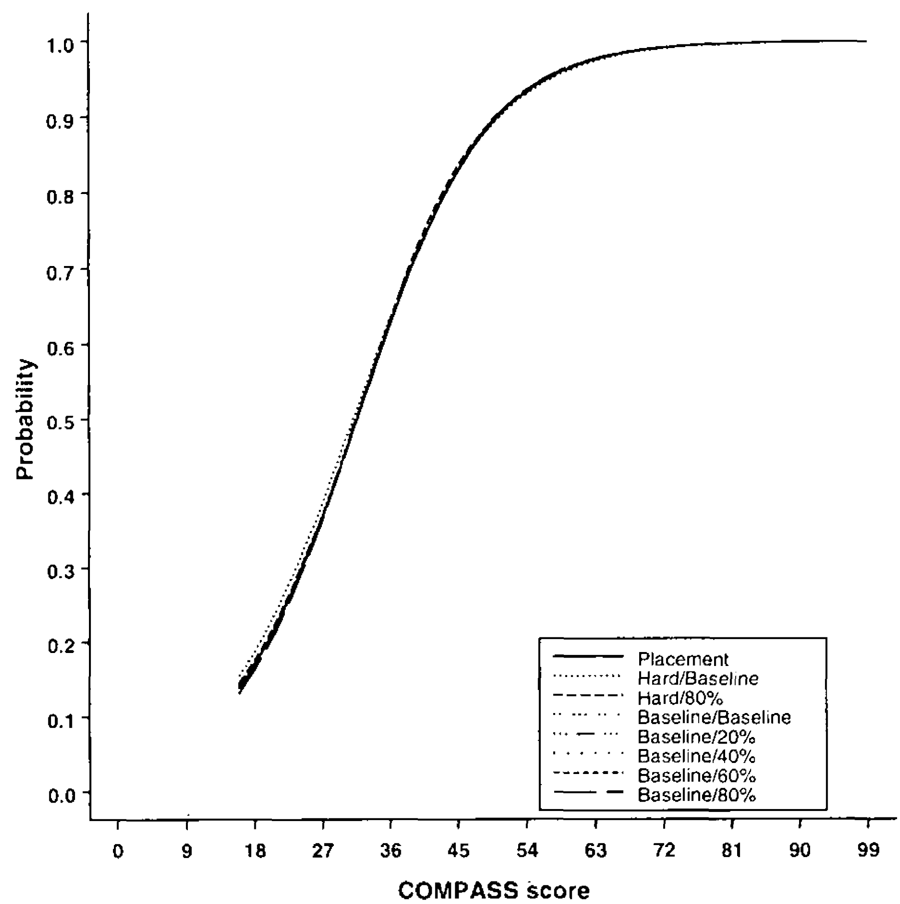
Results

Estimated Probabilities of Success

Figure 3 illustrates the effects of truncation on $\hat{P}$ for Placement Group 3 (steep slope, zero skewness, n=500). Of all the placement groups, this one was least affected by truncation with respect to estimating $\hat{P}$. The solid curve in the figure represents probabilities for the nontruncated placement group. Probabilities for the seven truncation conditions are represented by dashed or dotted curves, which are nearly identical. Clearly, truncation had little effect on estimating $\hat{P}$ for this placement group.

FIGURE 3. Effects of Truncation on Estimated Conditional Probability of Success

(Placement Group 3: Steep Slope, Zero Skewness, N=500)



Contrast the logistic regression curves for Placement Group 3 with those of Group 10, which are displayed in Figure 4. Under these conditions (flat slope, high skewness, n=100), $\hat{P}$ was relatively poorly estimated. The hard/80% truncation condition (which included only 10 observations) had the least accurate estimates of $\hat{P}$ in this figure.

FIGURE 4. Effects of Truncation on Estimated Conditional Probability of Success

(Placement Group 10: Flat Slope, High Skewness, N=100)

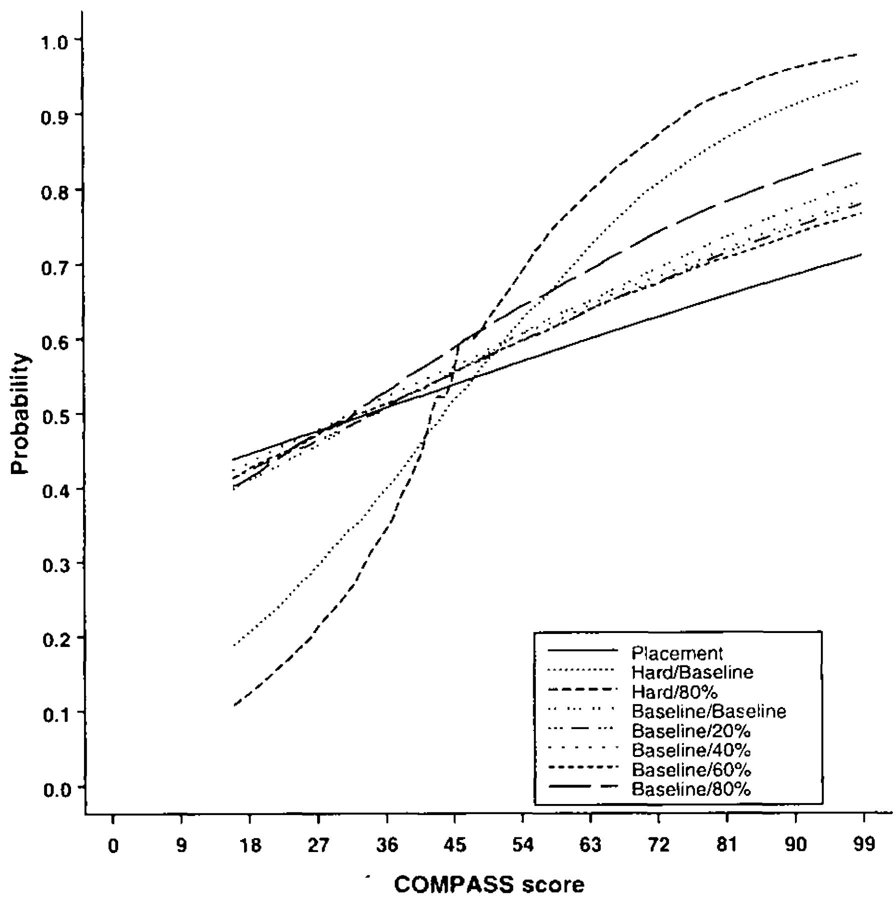


Table 3 summarizes the effects of truncation on $\hat{P}$ for all placement groups. Consistent with previous truncation research, placement groups with large samples yielded more accurate estimates of $\hat{P}$ than did those with small samples. Irrespective of sample size, steep slope conditions produced more accurate estimates than did flat slope conditions. This finding is also consistent with previous research. With respect to $\hat{P}$, the three groups least affected by truncation were 1, 2, and 3; the three groups most affected were 8, 11, and 10. Mean $|\Delta\hat{P}|$ ranged from .0003 (Group 3; baseline/60%) to .209 (Group 10, hard/80%). The relationship between extent of truncation and accuracy of $\hat{P}$ was similar to that identified in previous research, in that increased truncation was associated with decreased $\hat{P}$ accuracy.

TABLE 3

Effects of Truncation on Estimated Probability of Success, by Placement Group and Truncation Condition

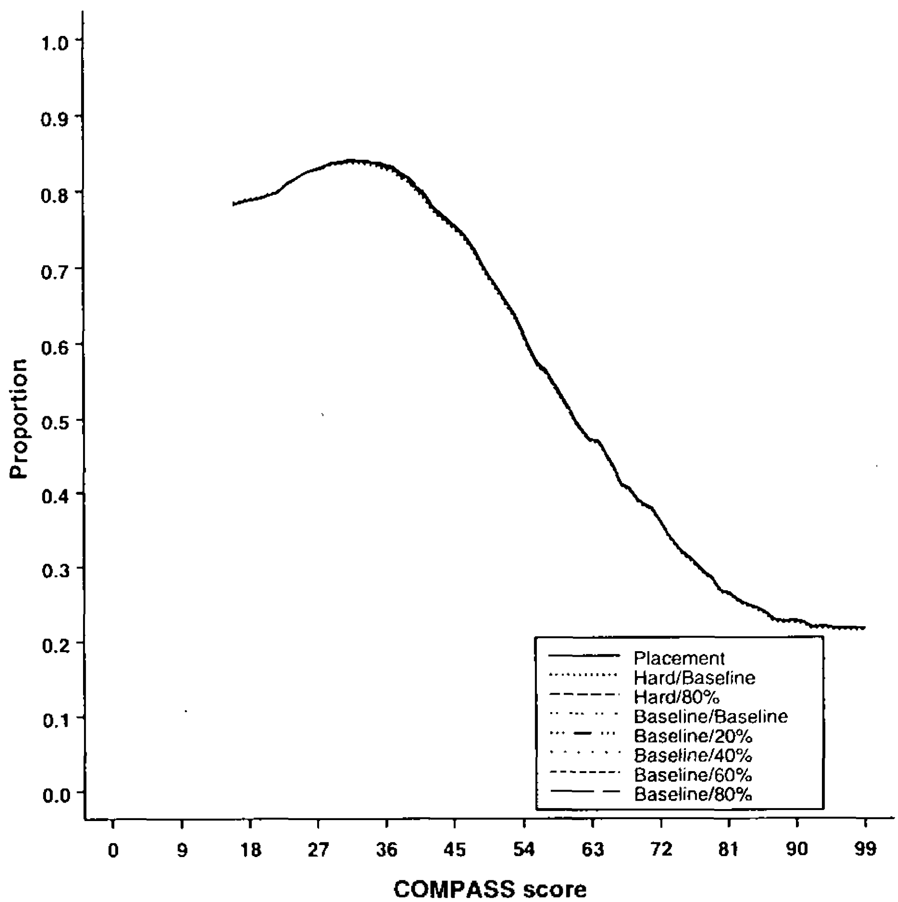
Placement group					Truncation						
No.	Slope	Skew.	N	Mean	Hard/ baseline	Hard/ 80%	Baseline/ baseline	Baseline/ 20%	Baseline/ 40%	Baseline/ 60%	Baseline/ 80%
1	Steep	High	500	$\Delta\hat{P}$	.0031	-.0003	.0005	-.0013	-.0006	-.0021	-.0027
				$\|\Delta\hat{P}\|$	.0031	.0036	.0015	.0013	.0007	.0021	.0028
2	Steep	Med.	500	$\Delta\hat{P}$	-.0031	-.0063	-.0036	-.0071	-.0085	-.0082	-.0079
				$\|\Delta\hat{P}\|$	.0104	.0060	.0096	.0094	.0117	.0110	.0106
3	Steep	Zero	500	$\Delta\hat{P}$	.0037	.0003	.0029	.0009	.0000	.0000	.0000
				$\|\Delta\hat{P}\|$	.0053	.0029	.0008	.0009	.0006	.0003	.0019
4	Flat	High	500	$\Delta\hat{P}$	-.0101	-.0084	-.0059	-.0125	-.0160	-.0104	-.0066
				$\|\Delta\hat{P}\|$	.0101	.0059	.0099	.0139	.0178	.0112	.0097
5	Flat	Med.	500	$\Delta\hat{P}$	.0048	.0058	.0053	.0042	.0027	.0055	.0051
				$\|\Delta\hat{P}\|$	.0364	.0396	.0100	.0084	.0056	.0068	.0053
6	Flat	Zero	500	$\Delta\hat{P}$	-.0189	.0008	-.0194	.0033	.0078	.0135	.0203
				$\|\Delta\hat{P}\|$	.0230	.0222	.0173	.0191	.0173	.0182	.0204
7	Steep	High	100	$\Delta\hat{P}$	-.0108	-.0067	-.0168	-.0116	-.0144	-.0179	-.0284
				$\|\Delta\hat{P}\|$	.0464	.0622	.0067	.0117	.0144	.0184	.0291
8	Steep	Med.	100	$\Delta\hat{P}$	-.0170	.0079	-.0229	.0105	.0113	.0121	.0132
				$\|\Delta\hat{P}\|$	.0475	.0575	.0193	.0211	.0221	.0216	.0229
9	Steep	Zero	100	$\Delta\hat{P}$	-.0084	.0070	-.0205	.0076	.0110	.0147	.0145
				$\|\Delta\hat{P}\|$	.0145	.0429	.0077	.0076	.0110	.0147	.0145
10	Flat	High	100	$\Delta\hat{P}$	.0443	.0368	.0629	.0275	.0362	.0263	.0695
				$\|\Delta\hat{P}\|$	.1507	.2090	.0469	.0340	.0381	.0307	.0754
11	Flat	Zero	100	$\Delta\hat{P}$	.0375	.0231	.0325	.0266	.0264	.0292	.0351
				$\|\Delta\hat{P}\|$	.0375	.0345	.0286	.0350	.0406	.0393	.0525

Estimated Accuracy Rates

Figure 5 displays the effects of truncation on $\hat{A}$ for Placement Group 3 (steep slope, zero skewness, n=500), whose estimates were more accurate overall than those of other groups. The maximum $\hat{A}$ for this placement group (corresponding to the "true" cutoff score) occurred at a COMPASS score of 32. For all truncation conditions except one (hard/baseline), the estimated optimal cutoff score was equivalent to the true optimal cutoff. Although not discernible in the figure, the optimal cutoff for the Hard/baseline condition was underestimated by one COMPASS score point.

FIGURE 5. Effects of Truncation on Estimated Accuracy Rate

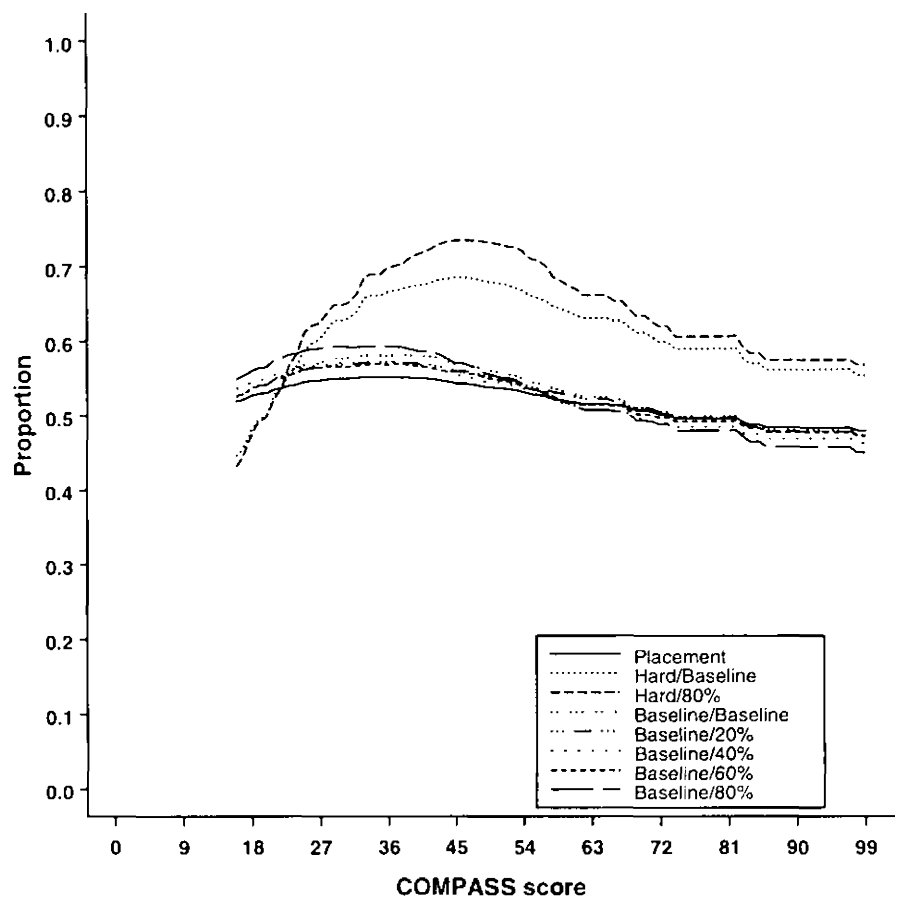
(Placement Group 3: Steep Slope, Zero Skewness, N=500)



The effects of truncation on $\hat{A}$ for a placement group with relatively inaccurate estimates (Group 10; flat slope, high skewness, n=100) are shown in Figure 6. Locations of maximum $\hat{A}$ for the Hard/baseline and Hard/80% conditions (at COMPASS scores of 46 and 47, respectively) were considerably different from those for the other truncation conditions (between scores of 35 and 38).

FIGURE 6. Effects of Truncation on Estimated Accuracy Rate

(Placement Group 10: Flat Slope, High Skewness, N=100)



The effects of truncation on $\hat{A}$ are summarized for all placement groups in Table 4. As was found for $\hat{P}$, more precise estimates of $\hat{A}$ were associated with large sample placement groups. Steep slope placement groups generally had more precise estimates of $\hat{A}$ s than did flat slope placement groups, irrespective of sample size, but there were some exceptions. For example, $\hat{A}$ s for Group 4 (flat slope, high skewness, n=500) were somewhat more precise than those for Group 2 (steep slope, medium skewness, n=500).

The three placement groups with the most precise estimates of $\hat{A}$ were 1, 3, and 4; the least precise estimates were found for Groups 7, 10, and 11. Mean $|\Delta\hat{A}|$ ranged from .0003 (Placement Group 3; baseline/40% and baseline/60%) to .1216 (Group 10; hard/80%).

TABLE 4

Effects of Truncation on Estimated Accuracy Rate, by Placement Group and Truncation Condition

Placement group					Truncation						
No.	Slope	Skew.	N	Mean	Hard/ baseline	Hard/ 80%	Baseline/ baseline	Baseline/ 20%	Baseline/ 40%	Baseline/ 60%	Baseline/ 80%
1	Steep	High	500	$\Delta\hat{A}$	-.0028	.0020	-.0007	.0008	.0001	.0012	.0012
				$\|\Delta\hat{A}\|$	.0037	.0027	.0009	.0014	.0006	.0025	.0031
2	Steep	Med.	500	$\Delta\hat{A}$	-.0035	-.0028	-.0007	.0004	-.0000	.0008	-.0002
				$\|\Delta\hat{A}\|$	.0059	.0042	.0074	.0080	.0094	.0098	.0084
3	Steep	Zero	500	$\Delta\hat{A}$	-.0032	-.0015	-.0002	-.0007	-.0002	.0003	.0001
				$\|\Delta\hat{A}\|$	.0037	.0018	.0005	.0008	.0003	.0003	.0006
4	Flat	High	500	$\Delta\hat{A}$	.0018	.0047	-.0009	-.0020	-.0021	-.0020	-.0031
				$\|\Delta\hat{A}\|$	.0049	.0063	.0040	.0043	.0063	.0029	.0034
5	Flat	Med.	500	$\Delta\hat{A}$	.0204	.0213	.0033	.0032	.0016	.0004	-.0012
				$\|\Delta\hat{A}\|$	.0213	.0223	.0035	.0032	.0019	.0034	.0039
6	Flat	Zero	500	$\Delta\hat{A}$	.0113	.0100	-.0082	-.0098	-.0097	-.0101	-.0101
				$\|\Delta\hat{A}\|$	.0189	.0183	.0087	.0106	.0115	.0142	.0177
7	Steep	High	100	$\Delta\hat{A}$	.0391	.0471	.0038	.0063	.0074	.0083	.0100
				$\|\Delta\hat{A}\|$	.0433	.0552	.0077	.0130	.0149	.0195	.0278
8	Steep	Med.	100	$\Delta\hat{A}$	.0314	.0397	.0020	-.0008	.0002	-.0023	-.0016
				$\|\Delta\hat{A}\|$	.0387	.0508	.0080	.0119	.0109	.0134	.0131
9	Steep	Zero	100	$\Delta\hat{A}$	.0077	.0117	-.0042	-.0043	-.0064	-.0077	-.0080
				$\|\Delta\hat{A}\|$	.0091	.0140	.0050	.0073	.0112	.0143	.0162
10	Flat	High	100	$\Delta\hat{A}$	.0897	.1146	.0122	.0080	.0017	.0047	.0071
				$\|\Delta\hat{A}\|$	.0957	.1216	.0139	.0100	.0135	.0091	.0246
11	Flat	Zero	100	$\Delta\hat{A}$	-.0131	-.0035	.0023	.0041	.0052	.0023	.0067
				$\|\Delta\hat{A}\|$	.0263	.0214	.0218	.0260	.0289	.0297	.0358

Optimal Cutoff Scores

Estimated optimal cutoff scores are displayed in Table 5, by placement group and truncation condition. The difference between the estimated optimal cutoff for a particular truncation condition and the true cutoff (shown in the "None" column for each placement group) is displayed in parentheses beneath the corresponding cutoff.

TABLE 5

Estimated Optimal Cutoff Scores, by Placement Group (Difference from "True" Cutoff)

Placement group				Truncation							
No.	Slope	Skew.	N	None	Hard/ baseline	Hard/ 80%	Baseline/ baseline	Baseline/ 20%	Baseline/ 40%	Baseline/ 60%	Baseline/ 80%
1	Steep	High	500	32	32	32	32	32	32	32	33
					(0)	(0)	(0)	(0)	(0)	(0)	(1)
2	Steep	Med.	500	33	33	33	33	34	34	34	34
					(0)	(0)	(0)	(1)	(1)	(1)	(1)
3	Steep	Zero	500	32	31	32	32	32	32	32	32
					(-1)	(0)	(0)	(0)	(0)	(0)	(0)
4	Flat	High	500	27	26	25	26	27	25	27	25
					(-1)	(-2)	(-1)	(0)	(-2)	(0)	(-2)
5	Flat	Med.	500	30	33	32	30	30	31	30	29
					(3)	(2)	(0)	(0)	(1)	(0)	(-1)
6	Flat	Zero	500	39	43	43	38	36	35	38	36
					(4)	(4)	(-1)	(-3)	(-4)	(-1)	(-3)
7	Steep	High	100	36	37	39	36	36	37	37	37
					(1)	(3)	(0)	(0)	(1)	(1)	(1)
8	Steep	Med.	100	33	36	36	33	32	32	32	32
					(3)	(3)	(0)	(-1)	(-1)	(-1)	(-1)
9	Steep	Zero	100	32	33	34	31	31	30	31	31
					(1)	(2)	(-1)	(-1)	(-2)	(-1)	(-1)
10	Flat	High	100	34	46	47	38	36	36	35	35
					(12)	(13)	(4)	(2)	(2)	(1)	(1)
11	Flat	Zero	100	35	31	29	34	34	37	33	35
					(-4)	(-6)	(-1)	(-1)	(2)	(-2)	(0)

Optimal cutoff scores were estimated very accurately for Placement Groups 1, 2, and 3 over all truncation conditions, deviating no more than one COMPASS score point from the true cutoff. These results are well within one standard error of measurement (SEM) for COMPASS, which ranges from about five to six for the Writing Skills, Reading, and Algebra tests. Cutoffs were accurately estimated for size n=100 placement groups when the logistic curve was steep. Interestingly, the results for Group 11 were more accurate than those for Group 6; these groups differed only in their initial sample sizes (100 and 500, respectively). Generally, one would expect more accurate cutoff estimates to be associated with large placement groups.

The only placement groups yielding somewhat inaccurate cutoff estimates (i.e., 6 or more points above or below the true cutoff) were Groups 10 and 11, both of which had flat logistic curves and small sample sizes. The Hard/80% condition produced an optimal cutoff that overestimated the Placement Group 10 true cutoff by 13 scale score points; a 12-point overestimate and a 6-point underestimate were produced by the Hard/baseline and Hard/80% conditions in Groups 10 and 11, respectively.

Estimated Success Rates

The most accurate estimates of $\hat{S}$, as measured by mean $\left|\Delta\hat{S}\right|$, were found for Placement Groups 1, 3, and 9. These groups had steep logistic curves in common, but differed in initial sample size; Groups 1 and 3 contained 500 observations, whereas Group 9 contained 100. The least accurate estimates of $\hat{S}$ were found for Groups 7, 10, and 11, all of which initially contained 100 observations. Two of these groups had flat logistic curves.

The characteristics associated with accurate estimates of $\hat{S}$, with a few exceptions, were similar to those associated with accurate $\hat{A}$ and $\hat{P}$: large placement group samples and steep logistic curves. Mean $|\Delta\hat{S}|$ ranged from .0001 (Group 3; baseline/60%) to .1784 (Group 10; hard/80%). These statistics are summarized, by placement group, in Table A in the appendix.

Discussion

It was shown in this study that validity statistics and optimal cutoff scores can be estimated with reasonable accuracy from the truncated data of two-stage course placement systems. For example, optimal cutoff scores were under- or overestimated by no more than 4 COMPASS score points, over all combinations of distribution shape and logistic regression curve, even when baseline truncation below K' was paired with 80% truncation above K'. It was only when hard truncation was paired with either baseline or 80% truncation that optimal cutoff score estimates differed substantially from true cutoffs. Moreover, substantial differences in estimated optimal and true cutoffs were found only for placement groups having flat logistic curves and small sample sizes under these two truncation conditions. Consistent with previous research (Schiel, 1998; Schiel & King, 1999), more accurate estimates of validity statistics and optimal cutoffs were associated with large sample, steep logistic curve placement groups.

The three least accurate estimated optimal cutoff scores (within 12, 13, and 6 score points of corresponding true cutoffs) occurred for truncation samples containing 26, 10, and 12 observations, respectively. It is unlikely that a postsecondary institution would use logistic regression and decision theory to evaluate test score/course outcome relationships for such small samples, because the accuracy of estimated logistic regression parameters declines significantly for very small sample sizes (Houston, 1993). Optimal cutoff scores over- or underestimated to this extent therefore have a small likelihood of occurring in practice. A more typical over- or underestimate, given the results of this study, would be about four COMPASS score points.

What are the practical implications of an institution employing an optimal cutoff score for a standard course that is over- or underestimated by four COMPASS score points? One way to answer this question is by examining accuracy rates. Considering Placement Group 10 (flat slope, high skewness, n=100) as an example, the Baseline/baseline truncation condition, based on 55 observations, yielded a median estimated optimal cutoff of 38. The median $\hat{A}$ corresponding to this cutoff indicated that 58.2% of students would be correctly placed if it were used. The true cutoff for Group 10 was 34; the corresponding $\hat{A}$, expressed as a percentage, was 58.1. Thus, in this instance, there would be no substantive effect of using a cutoff score that was underestimated by four COMPASS score points. Note that absolute differences between the median $\hat{A}$ s for the other placement group that had four-point over- or underestimates (Group 6; flat slope, zero skewness, n=500) were .005 or less, similarly suggesting no substantive effect of using a cutoff score within 4 score points of the true cutoff.

Postsecondary institutions that experience moderate truncation (i.e., baseline/baseline to baseline/80%) in two-stage course placement systems can expect to estimate validity statistics and optimal cutoff scores with reasonable accuracy. It is only when truncation is extreme (i.e., hard/baseline or hard/80%), logistic regression curves are flat, and sample sizes are very small (e.g., about 25 or less) that institutions risk obtaining optimal cutoff scores that differ substantively from those of nontruncated placement group distributions. One might consider a "substantive" difference to be two or more percentage points between the $\hat{A}$ corresponding to a true cutoff and the $\hat{A}$ for an estimated optimal cutoff. In a placement group consisting of about 100 students, for example, a two-percentage point decrease in $\hat{A}$ would mean that about 2 students would be incorrectly placed as a result of estimation error.

In a two-stage course placement system, truncation occurs both below and above the cutoff score on the placement test, thereby differentiating such a system from a one-stage system. One might therefore expect that the effect of truncation on estimated validity statistics in a two-stage system would differ from that occurring in a one-stage system, which is indeed the case. When the results of this study are compared with those of previous studies that examined the effects of truncation in one-stage systems, one noteworthy difference pertains to the accuracy of the respective optimal cutoff scores. In Schiel and King's (1999) one-stage system research, for example, the largest difference between a true cutoff and an estimated optimal cutoff was 17 ACT Assessment score points. This is larger than the largest difference observed in the present study (13 COMPASS score points; Group 10, hard/80%). Both of these results were obtained from placement groups with flat logistic curves and a high degree of skewness.

The difference between these results becomes intriguing when one considers that the Schiel and King result was based on a joint distribution of ACT scores and course outcomes containing considerably more observations than did the COMPASS score/course outcome distribution in the present study (330 vs. 10, respectively). One might expect a more accurate estimate to be associated with a larger sample, but this is clearly not the case. Moreover, the score scales of the two instruments differ considerably. The ACT Assessment score scale has 36 possible points, and a SEM (for the Composite) of about 1. The COMPASS score scale, on the other hand, has 84 possible points and a SEM of about 6. These characteristics suggest that the underestimation of the optimal ACT cutoff in the example from Schiel and King is considerably greater, in an absolute sense, from the underestimation occurring in the COMPASS example in the present study.

What might account for the difference between two-stage and one-stage placement system results? One possibility is the shape of the test score/course outcome distribution; the ACT Assessment/course outcome distribution in the example from the former study was highly negatively skewed, whereas the COMPASS/course outcome distribution in the present study example is highly positively skewed. In one-stage systems, higher negative skewness is associated with more accurate estimates of validity statistics. The reason for this is that when high negative skewness is present, a greater percentage of observations lie in the nontruncated region of the joint distribution. Such an association is less evident in a two-stage system, because the skewness is positive and truncation occurs in both tails of the distribution. It is possible that future research could provide additional insight into relationships between truncation and distribution shape. Nevertheless, it is important for interpretational purposes to consider that one-stage placement systems are inherently different from two-stage systems. Given the findings presented here, two-stage systems appear to be more resistant to the effects of truncation in the context of estimating course placement validity statistics.

To alleviate estimation problems that might result from soft truncation above K', institutions could consider administering the placement test to a group of students (e.g., an entering freshman class) who scored at or above K on the screening test. Course placement decisions would not have to be changed for these students. Institutions could then estimate validity statistics and optimal cutoff scores from a distribution of placement test scores and course outcomes that was not truncated above K'. A disadvantage to this approach is, of course, that institutions would have to test a larger number of students than usual with the placement test. Such an approach might be of interest to institutions whose data are severely truncated, based on small samples, and yield flat logistic curves, as they have the greatest risk of estimating inaccurate optimal cutoff scores. Most institutions, however, are not likely to benefit much from administering the placement test to students who scored at or above K, because truncation will affect their estimates only minimally.

In one-stage placement systems, $\hat{A}$ is a function of conditional probabilities estimated from the data of students who completed the standard course and the empirical distribution of test scores for the placement group, which includes the scores of students who did not complete the standard course. In two-stage systems, although both a screening test and a placement test are used, $\hat{A}$ is typically calculated just as it is in a one-stage system, using the distribution of placement test scores only. As a consequence, this statistic does not reflect the standard course outcomes of those students who scored high on the screening test, were placed directly into the standard course, and therefore did not have to take the placement test. This $\hat{A}$ could therefore differ somewhat from one that was instead based on both screening and placement test data.

Research that examines alternative ways of calculating $\hat{A}$ in two-stage placement systems would be beneficial, but would not likely alter the conclusions reached in this study, which controlled for this potential methodological problem. The placement groups in this study differed deliberately from those of actual two-stage placement systems in that they contained test scores and standard course outcomes for the full distribution of placement test scores, including those that would likely have been earned by students who earned high screening test scores and did not take the placement test. Consequently, the placement test score distributions for the placement groups, which would ordinarily be truncated with respect to the screening test in a two-stage system, were not truncated in this study. This allowed more precise comparisons between true (placement group) $\hat{A}$ s and those reflecting the effects of truncation.

Another reason that alternative methods of calculating $\hat{A}$ would not change this study's conclusions is that the effects of truncation were investigated in this study by considering *differences* between $\hat{A}$ s (and other validity statistics) calculated when truncation was and was not present. The size of such differences should remain relatively constant across different methods of calculating a particular validity statistic, provided that the same method is applied consistently under both nontruncated and truncated conditions.

References

American College Testing (1994). *ACT Assessment Course Placement Service Interpretive Guide.* Iowa City, IA: Author.

Houston, W. M. (1993). *Accuracy of validity indices for course placement systems.* Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Noble, J. & Sawyer, R. (1997). *Alternative methods for validating admissions and course placement criteria.* (AIR Professional File No. 63). Tallahassee, FL: Association for Institutional Research.

Sawyer, R. L. (1989). *Validating the use of ACT Assessment scores and high school grades for remedial course placement in college* (Research Report No. 89-4). Iowa City, IA: ACT.

Sawyer, R. L. (1996). Decision theory models for validating course placement tests. *Journal of Educational Measurement, 33*, 271-290.

Schiel, J. (1998). *Estimating conditional probabilities of success and other course placement validity statistics under soft truncation* (Research Report No. 98-2). Iowa City, IA: ACT.

Schiel, J. L. & King, J. E. (1999). *Accuracy of course placement validity statistics under various soft truncation conditions.* (ACT Research Report No. 99-2). Iowa City, IA: ACT.

Schiel, J. & Noble, J. (1992). *The effects of data truncation on estimated validity indices for course placement* (Research Report No. 92-3). Iowa City, IA: ACT.

Appendix

TABLE A

Effects of Truncation on Estimated Success Rate, by Placement Group and Truncation Condition

Placement group					Truncation						
No.	Slope	Skew.	N	Mean	Hard/ baseline	Hard/ 80%	Baseline/ baseline	Baseline/ 20%	Baseline/ 40%	Baseline/ 60%	Baseline/ 80%
1	Steep	High	500	$\Delta\hat{S}$	.0016	.0007	-.0009	-.0010	-.0006	-.0018	-.0027
				$\lvert\Delta\hat{S}\rvert$	.0016	.0017	.0009	.0010	.0006	.0018	.0027
2	Steep	Med.	500	$\Delta\hat{S}$	-.0072	-.0058	-.0083	-.0085	-.0104	-.0099	-.0095
				$\lvert\Delta\hat{S}\rvert$	.0072	.0058	.0083	.0085	.0104	.0099	.0095
3	Steep	Zero	500	$\Delta\hat{S}$	-.0004	.0009	-.0003	.0002	.0002	-.0000	.0007
				$\lvert\Delta\hat{S}\rvert$	.0012	.0009	.0003	.0002	.0002	.0001	.0007
4	Flat	High	500	$\Delta\hat{S}$	-.0113	-.0042	-.0116	-.0164	-.0210	-.0138	-.0103
				$\lvert\Delta\hat{S}\rvert$	.0113	.0042	.0116	.0164	.0210	.0138	.0103
5	Flat	Med.	500	$\Delta\hat{S}$	.0244	.0265	.0101	.0080	.0053	.0077	.0060
				$\lvert\Delta\hat{S}\rvert$	.0263	.0285	.0101	.0080	.0053	.0077	.0060
6	Flat	Zero	500	$\Delta\hat{S}$	-.0073	-.0088	-.0103	-.0094	-.0040	.0022	.0111
				$\lvert\Delta\hat{S}\rvert$	.0118	.0119	.0113	.0112	.0080	.0078	.0111
7	Steep	High	100	$\Delta\hat{S}$	.0087	.0057	-.0065	-.0118	-.0136	-.0190	-.0305
				$\lvert\Delta\hat{S}\rvert$	.0188	.0245	.0065	.0118	.0136	.0190	.0305
8	Steep	Med.	100	$\Delta\hat{S}$	-.0004	-.0045	.0071	.0084	.0086	.0090	.0096
				$\lvert\Delta\hat{S}\rvert$	.0107	.0148	.0071	.0084	.0086	.0090	.0096
9	Steep	Zero	100	$\Delta\hat{S}$	.0012	.0046	.0006	.0035	.0051	.0076	.0090
				$\lvert\Delta\hat{S}\rvert$	.0039	.0078	.0012	.0035	.0051	.0076	.0090
10	Flat	High	100	$\Delta\hat{S}$	.1229	.1641	.0573	.0418	.0489	.0382	.0965
				$\lvert\Delta\hat{S}\rvert$	.1352	.1784	.0573	.0418	.0489	.0382	.0965
11	Flat	Zero	100	$\Delta\hat{S}$	.0227	.0327	.0326	.0391	.0426	.0426	.0550
				$\lvert\Delta\hat{S}\rvert$	.0227	.0327	.0326	.0391	.0426	.0426	.0550

FIGURE A. Truncation Example

(Placement Group 1: Steep slope, high skewness, n=500)

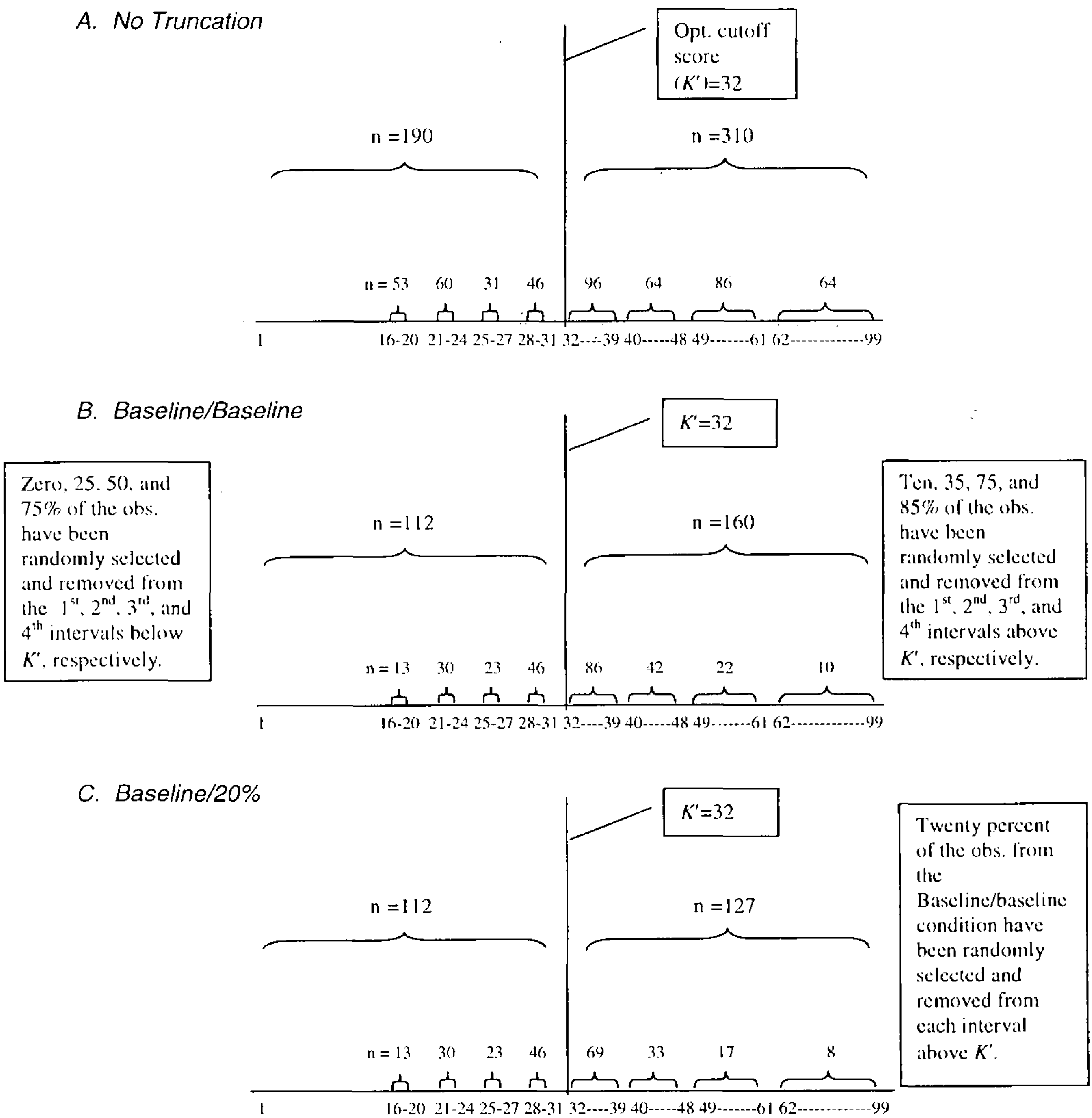


FIGURE A (continued). Truncation Example

(Placement Group 1: Steep slope, high skewness, n=500)

D. Baseline/40%

E. Baseline/60%

F. Baseline/80%

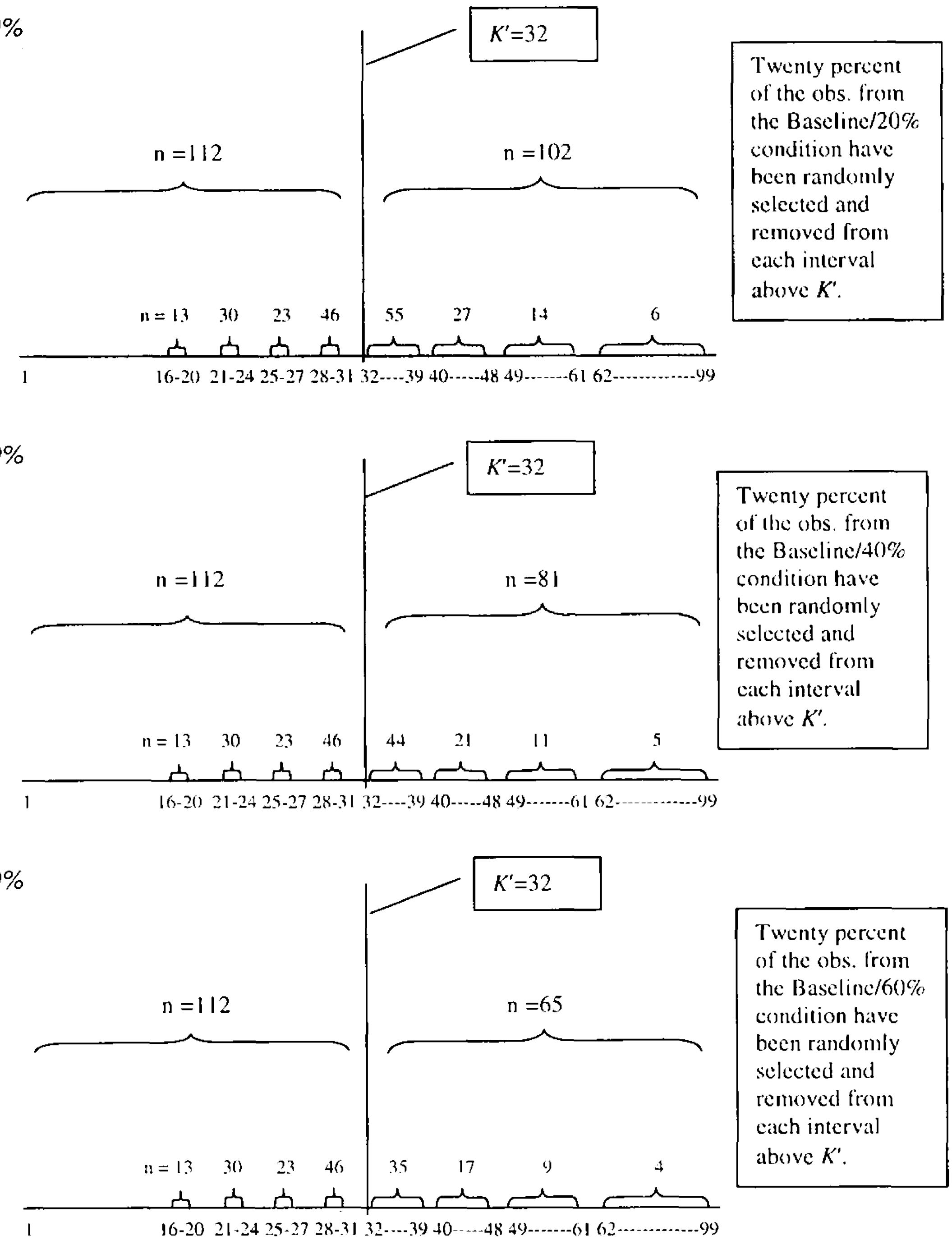
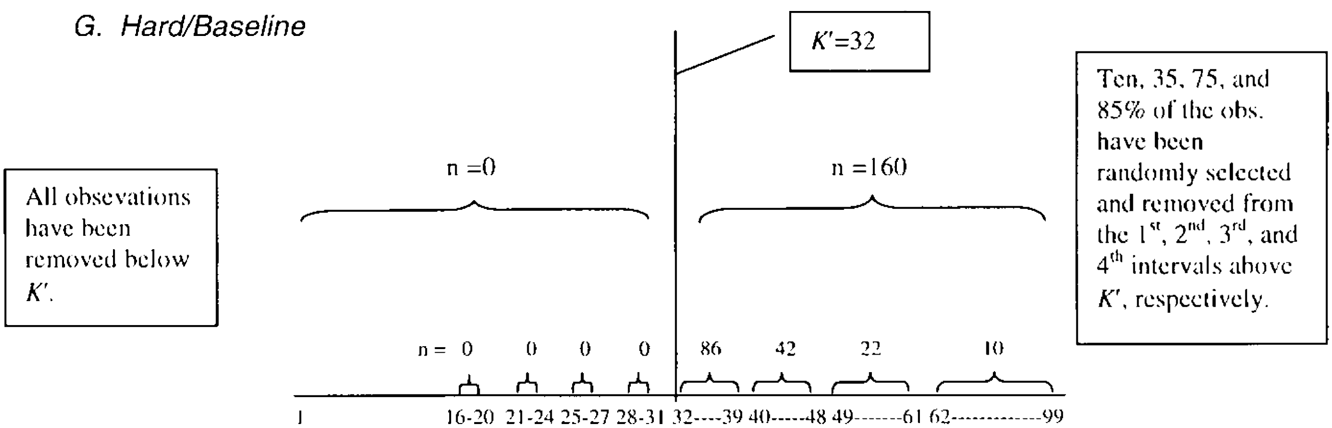


FIGURE A (continued). Truncation Example

(Placement Group 1: Steep slope, high skewness, n=500)

G. Hard/Baseline



H. Hard/80%