DOCUMENT RESUME

ED 268 144 TM 860 059

AUTHOR	Angoff, William H.; Cowell, William R.
TITLE	An Examination of the Assumption that the Equating of Parallel Forms is Population-Independent.
INSTITUTION	Educational Testing Service, Princeton, NJ. Graduate Record Examination Board Program.
SPONS AGENCY	Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	ETS-GREB-83-12P; ETS-RR-85-22
PUB DATE	Jul 85
NOTE	92p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Aptitude Tests; *College Entrance Examinations; *Equated Scores; Error of Measurement; Higher Education; Intellectual Disciplines; Item Analysis; Mathematics Tests; *Sampling; Student Characteristics; Test Construction; *Testing Problems; Test Interpretation; Verbal Tests; Weighted Scores
IDENTIFIERS	Equipercentile Equating; *Graduate Record Examinations; Linear Equating Method; *Test Equivalence

ABSTRACT

Linear and equipercentile equating conversions were developed for two forms of the Graduate Record Examinations (GRE) quantitative test and the verbal-plus-quantitative test. From a very large sample of students taking the GRE in October 1981, subpopulations were selected with respect to race, sex, field of study, and level of performance (defined by GRE analytical test scores). The variance error of equated scores between the forms was calculated, and conversions were made between the test forms. Departures of the conversions, based on each of these special samples from the population conversion, were calculated and evaluated in terms of the standard error of equating at five selected raw score points on Form D3, adjusted for disparity between the means, as well as variance, skewness, and kurtosis for these populations and the corresponding total populations. The conversions for the Physical Science subpopulation were significantly different from the total population. When adjustments were made for variance, skewness, and kurtosis, these conversions fell in line for the homogeneous GRE quantitative test, but not so clearly for the heterogeneous verbal-plus-quantitative test. Conversions for all other subpopulations were acceptably within range. The assumption of population-independence for equating was supported for homogeneous but not heterogeneous tests. (Author/GDC)

GRE

GRADUATE RECORD EXAMINATIONS

AN EXAMINATION OF THE ASSUMPTION THAT THE EQUATING OF PARALLEL FORMS IS POPULATION-INDEPENDENT

William H. Angoff

and

William R. Cowell

GRE Board Professional Report GREB No. 83-12P
ETS Research Report 85-22

July 1985

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

ETS

EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

An Examination of the Assumption

That the Equating of Parallel Forms

Is Population-Independent

William H. Angoff

and

William R. Cowell

GRE Board Professional Report No. 83-i2P

July 1985

An Examination of the Assumption That the Equating of Parallel Forms Is Population-Independent

Abstract

Linear and equipercentile conversions were developed relating Forms 3DGR1 and 3DGR3 of the homogeneous GRE quantitative test and the specially constituted heterogeneous GRE verbal-plus-quantitative test, using randomly equivalent populations of about 13,500 cases taking each form, drawn from the entire candidate group tested in the regular October 1981 administration of the Graduate Record Examinations. For purposes of this study these samples of 13,500 cases were taken as representing their respective total populations, and the conversions based on them were taken as "population conversions." Empirical standard errors of equating were then developed for samples of 1,000 by drawing 100 samples of that size from each of the two base populations and calculating the variance error of equated scores on Form D1 (i.e., 3DGR1) for each successive raw score on Form D3 (i.e., 3DGR3) and fitting a second-degree equation to those variance errors. Samples of 1,000 cases taking each form were then selected at random from specially defined subpopulations homogeneous with respect to sex, race, field of study, and level of performance (the last of these defined by scores on the GRE analytical test) and used as the basis of additional conversions between the two test forms.

Departures of the conversions based on each of these specially selected samples from the population conversion were calculated and evaluated in terms of the standard error of equating at five selected raw score points on

Form D3, adjusted for the disparity between the means, as well as the three higher moments--variance, skewness, and kurtosis--for these subpopulations and the corresponding moments of the total populations.

The results of this phase of the study showed that the conversions for the samples drawn from the Physical Science subpopulation differed significantly from that of the total population. When, in the second phase of the study, the appropriate adjustments for the moments were made, it was found that the Physical Sciences conversions fell clearly in line for the homogeneous GRE quantitative test, but not quite so clearly for the heterogeneous GRE verbal-plus-quantitative test. The conversions for all other subpopulation samples were acceptably within range of the population conversion. It was concluded that the assumption of population-independence for equating is supportable for homogeneous tests but, because of evidence of nonparallelism between the two forms, the assumption is not as clearly established by the data of this study for heterogeneous tests.

Acknowledgments

This study was supported jointly by the Graduate Record Examinations Board and by Educational Testing Service.

The authors are indebted to Henry I. Braun, Linda Leary, Frederic M. Lord, and William B. Schrader for their reviews of the manuscript and for the clarifications gained in discussions of this problem with them. The authors are especially indebted to Paul W. Holland for his indispensable advice throughout all phases of the statistical work.

Although the theory and practice of test equating have always been of great interest to test developers, they did not attract wide interest in psychometrics until about 1970. Since then, coincident with the need to compile data derived from the administration of the tests of different publishers (in order, for example, to evaluate the effectiveness of various compensatory education programs) and with the growing interest in and understanding of the many applications of item response theory, including equating, articles and books describing and evaluating the various methods of equating have appeared in the psychometric literature with increasing frequency. During these past 15 years, there has been an active interest in conducting studies to compare several methods with respect to the kinds of variations in design and application that affect their stability and possible bias, with respect to their underlying assumptions and their robustness in the face of violations of these assumptions, and with respect to their usefulness in practical applications. In addition to these studies, at least one new method of test equating was developed during this period (Holland & Wightman, 1982), and several modifications have been offered as improvements on methods that were already available and in frequent use.

In spite of this active and continuing interest, there still remains a long-standing assumption basic to all equating that, to this date, has not been examined in any detail. This is the assumption that the characteristics of the populations used for equating have no effect on the outcome of the equating. Indeed, reference to this assumption has been made in the past without question, as though it were established fact (see, for example, Angoff, 1966; Angoff, 1984, p. 86). Clearly, the casual attitudes so often observed in selecting samples for equating, and the conventional practice of applying equating results to the wide varieties of subpopulations suggests that there is no lack of confidence in the truth of the assumption.

The purpose of this study is to test the validity of this assumption explicitly and to understand a little more clearly the meaning of score equating. The question addressed here is this: Within the variation expected from random factors, is the conversion function that is developed to convert the scores on Form x of a test to the scale of Form y--assuming that Forms x and y have been designed to be parallel forms--the very same function, or does it vary, depending on whether it has been developed using samples coming from one or another subpopulation? To what extent is the conversion function population-independent, that is to say, person-free? And to state the obverse question: To what extent is it appropriate to apply a given conversion function to a particular individual or group, irrespective of the nature of the group? Once the two forms have been equated, to what extent is the equating function sufficiently general that we can use the equated scores on the two forms interchangeably for any and all types of individuals and groups? That is, to what extent is the equated measurement test-free?

It is by now generally accepted that it makes no logical sense to equate two tests that are known to measure different traits. The "equating" of, say, a verbal test to a math test is not expected to yield the same result for different populations. Men, for example, score at about the same level on verbal tests as women do, but substantially higher on math tests. As a consequence, an equating operation for tests as different as these will yield predictably different conversion equations for the two sexes. The question then remains: Will the conversion be the same across populations if the tests are parallel?

The statement of this question assumes that the matter of parallelism has been settled by means of some other criterion or some other set of procedures, for the similarity among the equating results for different subpoulations is itself sometimes taken as an operational test of parallelism for the two tests in question. The characteristic of parallelism and the data resulting from the several equating operations are so intimately related that the characteristic is sometimes assumed to coerce the operational result--that is, population-free equating; and conversely, the operational result is sometimes taken as a test of the characteristic. The present study is intended to test what the writers still regard as a highly reasonable, but as yet untested, hypothesis that the characteristic does indeed coerce the operational result, even for forms that were designed to be measuring the same set of traits and are therefore at least approximately parallel.

The study was designed with the exclusive use, to the extent possible, of empirical data, freeing it (again, to the extent possible) of any assumptions that might tend to limit the generalizations to be drawn from the study. Accordingly, measures of random variation were developed entirely empirically, and adjustments, derived from theory, were made in these measures only as necessary to reflect differences in the first four moments for the groups studied. Actually, the study was designed to investigate the assumption of population-free equating with both linear and equipercentile models. It was therefore understood that, for the study of equipercentile equating, an empirical measure of error had to be obtained; standard formulas do not exist for the varieties of procedures used to conduct equipercentile equating. One such formula does indeed exist (Lord, 1982), but expectedly, it applies only to a particular way of treating the data. Other ways of treating the data--and there are several--are not fully accounced for by Lord's formula.

By way of background for the procedures followed in the study, the following information will be useful. At the October 1981 administration of the GRE General Test, three forms of three tests were administered: a 76-item verbal test, a 60-item quantitative test, and a 50-item analytical ability test. The analytical ability test was introduced in the fall of 1977 for experimental purposes, pending a final decision by the GRE Board regarding its retention as a formal part of the testing program. During this experimental period, it was planned to conduct studies of reliability, validity, bias, and coachability for the analytical test, and its possible redundancy with the verbal and quantitative tests. These studies were to be undertaken to determine whether the analytical test would stand up under the kind of scrutiny it was expected to receive. For the purposes of the present study, however, only two of these three tests were planned for specific use. One of these was the quantitative test, which was chosen as the principal test of interest because performance on it was expected to show more variability among variously selected subpopulations than performance on the verbal test. On the other hand, it was expected that results based on the quantitative test alone--a highly homogeneous test--might not be considered an adequate test of the assumption that equating is population-independent. What was needed was a heterogeneous test as well, on which different types of subpopulations would show different patterns of response. Accordingly, a heterogeneous test was formed, constructed simply by adding the scores on the verbal and quantitative tests for each person tested. The study of the verbal-plus-quantitative test represented a second phase of the study, paralleling the study of the quantitative test in all respects.

Procedure

As already indicated, the data used for this study came from the regular October 1981 administration of the Graduate Record Examinations General Test (previously known as the GRE Aptitude Test), at which time a total of about 56,460 examinees were tested. In addition to other forms that were administered to special populations on or about the same testing date, the three principal test forms of the General Test were Forms 3DGR1, 3DGR2, and 3DGR3, which had all been developed at the same time with the same content and statistical specifications. These three test forms were distributed to the examinees at the administration in "spiralled" fashion. That is, the test books were packaged and distributed in such a way that every 1st, 4th, 7th, 10th, ... student in each testing room received the same form; every 2nd, 5th, 8th, 11th, ... student were also given the same form, but different from the form given the first group; and every 3rd, 6th, 9th, 12th, ... student were similarly given the same form, but different from the first two. For the purposes of the study, Forms 3DGR1 (hereafter referred to as Form D1) and Form 3DGR3 (to be referred to as Form D3) were selected, principally because they appeared to be the most similar of the three possible pairs of forms with respect to their statistical characteristics. From those students taking Forms D1 or D3, about 35,650 in total number, data for all students whose test supervisor (chief proctor) reported some testing irregularity were removed. Further, data for all students who failed to mark an answer on any one of the sections of the test were also removed. And, finally, data for all students who reported that English was not their preferred language were removed. As a result of these restrictions 13,470 cases (49.89%) taking Form D1 and

13,527 cases (50.11%) taking Form D3 were finally considered usable for the study.

From each of the two base populations, 100 samples of 1,000 cases each were selected at random. (Each individual in a sample was selected without replacement; each sample was selected with replacement.) The samples in each population were numbered as they were selected, from 1 to 100, and the samples in the two populations bearing the same number were paired, resulting in 100 random pairings.

Scores on Form D3 were then equated by both linear and equipercentile methods to scores on Form D1, once for each of the 100 pairings, by procedures appropriate to Design I (random groups—one test administered to each group; see Angoff, 1984, p. 94), in which, for linear equating, scores on two forms are defined as equivalent if their standard-score deviates $[(x - M_X)/s_X]$ in their respective, randomly equivalent groups are equal. For equipercentile equating, scores on the two forms are defined as equivalent if their percentile ranks for the two tests, in the two randomly equivalent groups, are equal.

As a result of these equating operations, a bundle of 100 linear conversions and a bundle of more complex functions resulting from the 100 equipercentile equatings were produced separately for the quantitative score and for the verbal-plus-quantitative score. Actually, the equipercentile functions were very nearly linear for the main body of the data; as mentioned earlier, the two forms were designed to be approximately parallel in their statistical and content characteristics.

In preparation for equipercentile equating, the observed frequencies in each of the 200 distributions of GRE quantitative score and of GRE verbal-plus-quantitative score (100 for Form D1 and 100 for Form D3) were first smoothed by a seven-point rolling average method developed by E. E. Cureton and J. W. Tukey (Cureton & Tukey, 1951; Angoff, 1984, page 12). Following this operation, the smoothed frequencies were smoothed a second time and by the same procedure. (As a possible alternative procedure, the smoothing was also carried out by fitting the negative hypergeometric function to the raw data (Keats & Lord, 1962). However, the negative hypergeometric, while preserving the mean and standard deviation, failed to give a satisfactory fit and was therefore not used in the study.) No smoothing was undertaken for the succession of equipercentile points relating Form D1 and Form D3; all points were connected by straight lines, and converted score points on Form D3 were read off by computer by linear interpolation.

For the sake of convenience and easier interpretation, raw scores on Form D1 (the test of reference; throughout this study, raw scores on Form D3 were equated to raw scores on Form D1) were converted, using all 13,470 cases available for that form, to a scale in which the mean for that group was redefined as 500 and the standard deviation as 100. As a result of this operation, it was possible to observe the random variation, empirically developed, of the 100 linear and 100 equipercentile equatings on a 500-100 scale. Specifically, this variation is the variation in standard scores, the scores converted to the standard (500-100) scale for Form D1, corresponding to each given raw score on Form D3. The equation converting raw scores on Form D1 to the standard reference scale is given as follows for the GRE quantitative score:

$$S_q = \left[\frac{100}{9.1426}\right]x_q + 500 - \left[\frac{100}{9.1426}\right][38.1861]$$

$$S_q = 10.9378\ x_q + 82.3281,$$

where x_q denotes raw GRE quantitative scores on Form D1 and S_q denotes the reference scale for those scores.

Similarly, the equation converting raw scores on Form D1 to the reference scale is given as follows for the GRE verbal-plus-quantitative scores:

$$S_c = \left[\frac{100}{19.0698}\right]x_c + 500 - \left[\frac{100}{19.0698}\right][84.1419]$$

$$S_c = 5.2439\ x_c + 58.7683,$$

where x_c denotes raw GRE verbal-plus-quantitative scores on Form D1 and S_c denotes the reference scale for those scores.

Once the 1,000-case samples referred to above were drawn from each of the two base populations, 11 special subpopulations were defined and formed, homogeneous with respect to sex, ethnic background, academic field, or score level, as follows: Men, Women; Whites, Minorities; Humanities Majors, Social Science Majors, Biological Science Majors, Physical Science Majors; and High-, Middle-, and Low-scoring examinees. Minorities were defined as including Blacks plus Hispanics (only).

The categories of High-, Middle-, and Low-scoring were constituted by selecting the highest-scoring 27 percent, the middle-scoring 46 percent, and the lowest-scoring 27 percent on the GRE analytical test. This, the role of defining the High-, Middle-, and Low-scoring groups, was the only role in the study played by the analytical test. When the cuts effected by the foregoing percentages intersected a score interval (which occurred in every instance), the appropriate number in the interval was assigned to one category or the other at random.

That the use of the analytical score was effective in making these assignments into the High-, Middle-, and Low-scoring groups may be inferred from Tables 1 and 2, which give the means, standard deviations, and intercorrelations among the verbal, quantitative, analytical, and verbal-plus-quantitative scores. The correlations of analytical with quantitative are .7149 in the Form D1 population and .7133 in the Form D3 population, high enough to use for selecting disparately scoring groups. The correlations of analytical with verbal-plus-quantitative are even higher: .7375

Table 1

Means, Standard Deviations, and Intercorrelations among Verbal, Quantitative, Analytical, and Verbal-Plus-Quantitative Scores for the Entire Population Taking Form D1

N = 13,470

	Verbal	Quantitative	Analytical	Verbal Plus Quantitative	Mean	Standard Deviation
Verbal		.5560	.6068	.9172	45.9558	12.4072
Quantative	.556		.7149	.8412	38.1861	9.1426
Analytical	.6068	.7149		.7375	28.0589	7.8178
Verbal Plus Quantitative	.9172	.8412	.7375		84.1419	19.0698

Table 2

Means, Standard Deviations, and Intercorrelations among Verbal, Quantitative, Analytical and Verbal-Plus-Quantitative Scores for the Entire Population Taking Form D3

N = 13,527

	Verbal	Quantitative	Analytical	Verbal Plus Quantitative	Mean	Standard Deviation
Verbal		.5900	.6560	.9074	46.0686	11.2228
Quantative	.5900		.7133	.8747	38.6139	9.7347
Analytical	.6560	.7133		.7651	26.3650	7.6019
Verbal Plus Quantitative	.9074	.8747	.7651		84.6825	18.6982

in the Form D1 population and .7651 in the Form D3 population. It may be seen in Tables 3 and 4, in the rows corresponding to High-Scoring, Middle-Scoring and Low-Scoring, that the means and standard deviations of the resulting subpopulations selected on the basis of the analytical score are in fact quite different.

Tables 1 and 2 also throw some light on the parallelism of Forms D1 and D3. Although the correlations of quantitative and analytical scores are remarkably similar (.7149 vs. .7133), the correlations of verbal scores with quantitative scores (.5560 vs. .5900) and verbal scores with analytical scores (.6068 vs. .6560) differ more than would be ideal for operationally parallel forms. This observation is supported by the fact that the standard deviations, especially for verbal scores (12.4072 vs. 11.2228), but also for quantitative scores (9.1426 vs. 9.7347), differ more than would be expected if the forms were closely parallel. Further discussion of the lack of parallelism of the two forms appears later in this report.

The constitution of the Minority subpopulations is given in the following table:

	Numbers of Examinees Taking: Form D1	Form D3
Blacks	705	698
Puerto Ricans	33	30
Mexican Americans	195	163
Other Hispanic	111	109
	1,044	1,000

To make comparisons between the conversion functions developed from cases drawn from the 11 subpopulations with those developed from cases drawn from the total population, samples of 1,000, equal in size to each of the 100 samples drawn from the total population, were drawn from the 11 subpopulations. In all, 20 initial samples of size 1,000 were drawn from the subpopulations taking Form D1, two from each of the subpopulations except for the White and Minority subpopulations, from which only one sample each was drawn. These exceptions were occasioned by the fact that the two Minority groups--the group taking D1 and the group taking D3--were so small; the entire group of Minorities taking Form D3, for example, numbered only 1,000, indeed, exactly 1,000. To match that number, a sample of 1,000 was drawn from the total group of 1,044 taking Form D1. (For symmetry's sake, only one sample, also of 1,000 cases, was drawn from the White subpopulation.) Within each subpopulation the samples were numbered 1 or 2 in the order in which they were drawn.

Similarly, 20 samples of 1,000 cases each were drawn from the subpopulations taking Form D3--again, two from each of the subpopulations except the White and the Minority groups, from which only one sample each was drawn. (As mentioned above, the size of the Minority subpopulation taking Form D3 was almost as small as that taking Form D1, 1,044 cases, as compared with 1,000.) As in the sampling from the total population, each individual in a subpopulation sample was selected without replacement; each sample was selected with replacement. Here, too, within each subpopulation the samples were numbered 1 or 2 in the order in which they were drawn.

For reasons to be described below, two additional samples of 1,000 cases each, Samples 3 and 4, were drawn from each of the two Physical Science subpopulations to make a total of 22 pairs of samples drawn from particular subpopulations.

Finally, within each subpopulation the sample of 1,000 cases taking D1 was paired with the sample of 1,000 cases taking D3 bearing the same number, and Form D3 was equated to Form D1 by both linear and equipercentile methods as described above. All these processes were carried out for the GRE quantitative scores and again for the GRE verbal-plus-quantitative scores. The numbers of cases, means, standard deviations, and skewness and kurtosis values (normal kurtosis taken as zero) for GRE quantitative scores on Forms D1 and D3 are given in Table 3. Corresponding statistics for GRE verbal-plus-quantitative are given in Table 4.

In both Tables 3 and 4, the sums of the numbers of cases in the subpopulations defined by the categories of sex, ethnicity, and field of study fall short of the numbers of cases in the total populations taking Forms D1 and D3. This is so because in some instances examinees neglected to provide the information needed to classify them. In the case of the ethnic category, it is additionally so because in this study the Minority subpopulation was limited to Blacks, Puerto Ricans, Mexican Americans, and other Hispanics.

Table 3

Summary Statistics on GRE Quantitative Scores for the Total Populations and Subpopulations Taking Forms D1 and D3

	Form D1					Form D3				
	No. of Cases	Mean	Std Dev	Sk	Ku	No. of Cases	Mean	Std Dev	Sk	Ku
Total	13,470	38.19	9.14	-.2573	-.1339	13,527	38.61	9.73	-.1858	-.4194
Male	5,863	41.14	8.97	-.4468	.0303	5,700	42.00	9.55	-.4057	-.3038
Female	7,468	35.91	8.58	-.2405	-.0242	7,706	36.14	9.08	-.1521	-.3244
White	11,409	38.94	8.51	-.1506	-.1940	11,533	39.32	9.18	-.1151	-.4574
Minority	1,044	28.43	9.63	.2747	-.3837	1,000	28.76	9.96	.4325	-.2095
Humanities	1,923	36.07	8.76	-.2699	.0388	1,935	36.76	8.83	-.0385	-.4004
Social Science	5,840	35.56	8.53	-.2737	.0828	5,913	35.62	9.12	-.1357	-.2928
Biological Science	2,811	38.67	7.78	-.3576	.1624	2,892	39.36	8.26	-.3249	-.0953
Physical Science	2,264	46.65	7.16	-.7390	.6885	2,150	48.29	7.16	-.8855	.9950
High-Scoring	3,637	46.00	6.29	-.1549	-.5059	3,652	46.78	6.92	-.3513	-.3420
Mid-Scoring	6,196	38.48	6.77	-.0306	-.0269	6,223	39.10	7.32	.0004	-.2347
Low-Scoring	3,637	29.88	7.85	-.0727	-.0855	3,652	29.61	7.98	.1527	-.1498

Table 4

Summary Statistics on GRE Verbal-Plus-Quantitative Scores
for the Total Populations and Subpopulations
Taking Forms D1 and D3

	Form D1					Form D3				
	No. of Cases	Mean	Std Dev	Sk	Ku	No. of Cases	Mean	Std Dev	Sk	Ku
Total	13,470	84.14	19.07	-.3140	-.1648	13,527	84.68	18.70	-.2891	-.1814
Male	5,863	87.79	18.66	-.3903	-.0114	5,700	89.48	18.35	-.4105	-.0925
Female	7,468	81.36	18.85	-.2703	-.2205	7,706	81.18	18.14	-.2553	-.1447
White	11,409	86.15	17.46	-.1847	-.2783	11,533	86.45	17.21	-.1705	-.2649
Minority	1,044	61.32	19.73	.3557	-.3343	1,000	63.43	20.11	.4869	-.0870
Humanities	1,923	85.71	18.97	-.4833	-.0617	1,935	85.69	18.31	-.2076	-.3131
Social Science	5,840	79.99	19.09	-.2353	-.2840	5,913	80.02	18.62	-.2135	-.1977
Biological Science	2,811	84.40	16.98	-.3002	-.0327	2,892	85.19	16.20	-.3051	.0337
Physical Science	2,264	94.46	17.05	-.4697	.0979	2,150	97.40	15.65	-.6470	.6977
High-Scoring	3,637	100.81	12.42	-.1895	-.2534	3,652	101.23	12.22	-.1704	-.2153
Mid-Scoring	6,196	84.99	13.82	-.1026	-.1491	6,223	85.93	13.01	-.0145	-.1876
Low-Scoring	3,637	66.04	16.13	-.0808	-.1659	3,652	66.01	15.23	-.0426	-.1007

As expected, Table 3, based on the quantitative scores, shows that the highest-scoring subpopulation, even higher than that defined as "High-Scoring," is the Physical Science group, whose means on both Forms D1 and D3 are more than a standard deviation higher than those of any of the other curricular groups. Of these GRE subpopulations, the lowest, on both Forms D1 and D3, is the Minority group, whose means are more than a standard deviation lower than the White group.

Table 4, based on the verbal-plus-quantitative scores, fails to show Physical Science at quite the level seen in Table 3. This is to be expected, since their superiority in the verbal area is not nearly as pronounced as it is in the quantitative area, and the verbal component in the verbal-plus-quantitative score diminishes the superiority of the Physical Science group seen in Table 3. The Minority group, however, is still the lowest in Table 4 on both Form D1 and Form D3.

The differences among the subpopulations of interest in this study have a direct bearing on the methods used for evaluating the differences in the equating functions, since the standard error of equating is a function of the level and dispersion (as well as the skewness and kurtosis) of the groups used for equating. The matter of the choice of standard error in this context deserves some elaboration. In the first place, there is no single standard error of equating; it varies as a function of score level, following what appears to be second-degree function of the general form, $SE^2_{y*} = a + bx + cx^2$ (where SE^2_{y*} is the variance error of equated scores expressed on some scale, y, and x is the score that is converted, by means of the equating function, to the scale of y), showing a minimum in the general vicinity of the mean of the x-scores and becoming increasingly larger with increasing distance from that minimum point. Secondly, there were three possible variance error functions to choose from in evaluating the results of this study. One of these is a formula due to Lord (Lord, 1950; see also Angoff, 1984, page 97) that assumes normal distributions in the test forms and populations of interest. A second is ..mula due to Braun and Holland (1982, p. 33) that is more general, allowing consideration of degrees of skewness and kurtosis, but ignoring moments higher than the fourth. The third function was the system of empirical variance errors observed simply by calculating the variance of converted scores--developed from the 100 conversion functions described above--on Form D1 corresponding to each raw score on Form D3. This operation produces a function that is expected to be--and, in fact, was--perfectly smooth for linear equating, but that had to be fitted for equipercentile equating. After an examination and consideration of these three types of functions, it was decided to use the last, a smooth function of the empirically determined values, at least for this phase of the analysis. There were two reasons for coming to this decision: (1) the three functions did not appear to differ greatly or systematically; (2) since the study was designed from the outset to be as nearly empirical as possible, and since no analytical function for the variance error of equating could be developed for equipercentile equating, it was decided, for consistency's sake as well, to depend on empirically developed variance error throughout--at least, to the extent possible.

As indicated above, the empirical variance error function for equipercentile equating had to be fitted, and this was done by a second-degree equation. The plot of observed variance error points for equipercentile equating, corresponding to successive raw score values on Form D3, was found to be U-shaped, similar in form to that found for linear equating. However, these points were not only slightly erratic, as expected, they ceased to be monotonic in the regions near the extremes of the scale. Accordingly, the curves of equipercentile error variances, one for GRE quantitative scores and one for GRE verbal-plus-quantitative scores, were fitted only within the large central regions where the ascending bars of the U-curves displayed monotonicity.

Four second-degree equations, expressing the variance errors of equating on the standard reference scale as a function of raw score on Form D3, resulted from the foregoing operations and are given as follows for linear and for equipercentile equating of the GRE quantitative scores and the GRE verbal-plus-quantitative scores:

$$SE^2_{y*(\ell)} = 196.8437 - 8.8533x_q + 0.1102x_q^2 \quad (1)$$

$$SE^2_{y*(e)} = 665.5842 - 33.1299x_q + 0.4218x_q^2 \quad (2)$$

$$SE^2_{y*(\ell)} = 228.5855 - 4.6933x_c + 0.0259x_c^2 \quad (3)$$

$$SE^2_{y*(e)} = 718.0959 - 15.3412x_c + 0.0829x_c^2 \quad (4)$$

As may be seen by comparing in detail equation (1) with (2) and equation (3) with (4), the variance errors of equipercentile equating for these data are, for the most part, considerably larger than the variance errors of linear equating. The largest differences occur at the extremes of the Form D3 raw score scale and decline in size as one approaches the general vicinity of the mean. At their minimums the differences in variance error between the two equating methods are very small. For a short interval in the region of the minimum, the variance error of equipercentile equating is actually slightly smaller than the variance error of linear equating.

Equations (1)-(4) are repeated in equations (5)-(8), this time expressed in terms of standardized deviations, z(x), from the means on Form D3 (x) for the total population (M_{x_q} = 38.6139, s_{x_q} = 9.7347; M_{x_c} = 84.6825, s_{x_c} = 18.6982).

$$SE^2_{y*(\ell)} = 19.3386 - 3.3148z(x)_q + 10.4459z^2(x)_q \quad (5)$$

$$SE^2_{y*(e)} = 15.1670 - 5.4351z(x)_q + 39.9678z^2(x)_q \quad (6)$$

$$SE^2_{y*(\ell)} = 17.0949 - 5.6400z(x)_c + 9.0657z^2(x)_c \quad (7)$$

$$SE^2_{y*(e)} = 13.2385 - 24.4178z(x)_c + 28.9732z^2(x)_c \quad (8)$$

Here, the observation made in the preceding paragraph, comparing linear and equipercentile equating errors, may be verified directly. At the mean, where z(x)= 0, the variance errors of equipercentile equating are clearly smaller, but only by a small amount, than the variance errors of linear equating.

There are additional matters to consider in the choice of the appropriate standard error. As alluded to above, the error of determining an equated score on Form y corresponding to a given score on Form x of a test is in large part a function of the standardized distance of that x-score from the mean of the distribution of x-scores. For example, the formula for the variance error of equating for Design I equating (Angoff, 1984, p. 94), assuming normal distributions for both Forms x and y, is shown by Lord (1950; see also Angoff, 1984, p. 97) to be

$$SE^2_{y*} = \frac{2s^2_y}{n_t}(z^2_x + 2), \quad (9)$$

where SE^2_{y*} = the variance error of equated y-scores,

n_t = the sum of the numbers of cases in the two groups used for equating, and

z_x = $(x - M_x)/s_x$.

It can be determined from equation (9) that the standard error of equating is 1.22 times as large at one standard deviation from the mean as at the mean, 1.73 times as large at two standard deviations from the mean as at the mean, and 2.35 times as large at three standard deviations from the mean as at the mean. This being the case, any equating based on a distribution of x scores with a mean other than that characterizing the total group taking Form x would yield converted scores with different standard errors from those appropriate to the total population. A low-scoring group, for example, would have smaller standard errors than the total group for scores near the lower end of the scale and larger errors near the upper end of the scale. A high-scoring group, on the other hand, would have larger standard errors than the total group for scores near the lower end of the scale and smaller errors near the upper end of the scale. Accordingly, to convert an equation for variance error that was based on the mean and standard deviation of one group, say Group 1, to an equation for variance error appropriate to another group, say Group 2, with a different mean and standard deviation, the mean and standard deviation implicit in the equation appropriate to Group 1, say $SE^2_{y*} = a + bx_1 + cx_1^2$, needs to be replaced by the mean and standard deviation of Group 2. In effect, this change calls for the substitutionion of the value, $(s_{x_1}/s_{x_2})(x_2 - M_{x_2}) + M_{x_1}$, for x_1. If the more general equation is used, of the form, $SE^2_{y*} = A + Bz(x) + Cz^2(x)$, as in equations (5)-(8), then the expression $(x - M_x)/s_x$, containing the values of mean and standard deviation for the appropriate group, needs to be used in place of $z(x)$.

In addition to the specification of the first and second moments (mean and standard deviation), it is appropriate also to consider specifying the third and fourth moments if they differ from the third and fourth moments that characterize the total group from which the 100 samples of 1,000 cases were drawn. As mentioned earlier, the formula for the variance error of equating given in equation (9) assumes normal distributions in x and y, in which skewness and kurtosis are taken as zero. A more general formula for the variance error of equating, which permits the specification of the third and fourth moments (but ignoring higher order terms), is given by Braun and Holland (1982, page 33) and reproduced here, with minor changes in notation, as equation (10):

$$SE^2_{y*} = \frac{s^2_y}{n_x}\{1 + Sk(x)z(x) + \frac{[2 + Ku(x)]}{4}z^2(x)\} + \frac{s^2_y}{n_y}\{1 + Sk(y)z(x) + \frac{[2 + Ku(y)]}{4}z^2(x)\}, \qquad (10)$$

where s^2_{y*} = the variance of scores on Form D1 expressed on the 500-100 scale,

n_x and n_y = the number of cases taking Forms D3 and D1, respectively,

$Sk(x)$ and $Sk(y)$ = the values of skewness in the D3 and D1 distributions, respectively,

$Ku(x)$ and $Ku(y)$ = the corresponding values of kurtosis,

and $z(x)$ = the standarized distance from the score of interest to its own mean.

When $n_x = n_y = n$, equation (10) reduces to equation (11):

$$SE^2_{y*} = \frac{s^2_y}{n}\{2 + [Sk(x) + Sk(y)]z(x) + \frac{[4 + Ku(x) + Ku(y)]}{4}z^2(x)\}, \qquad (11)$$

in which normal kurtosis is taken as zero. It may be seen that when the distributions of x and y are both normal, equation (11) reduces to equation (9).

With the foregoing types of data and formulas available, it is possible to make several types of comparisons. The present study was restricted, however, to answering the following questions: (1) To what extent do the conversions based on the samples drawn from each of the several subpopulations agree with the overall conversion based on the total population of all cases (N = 13,470 taking Form D1; N = 13,527 taking Form D3)? To evaluate these differences, the variance errors of equating were taken from the empirical variances of converted scores on Form D1 observed at selected Form D3 raw scores--scores 10, 20, 30, 40, and 50 on the GRE quantitative test, and scores 25, 50, 75, 100, and 125 on the GRE verbal-plus-quantitative test. These variance errors were taken as observed in the equatings based on the 100 pairs of samples drawn from the overall population shown in equations (1)-(4), unadjusted for the mean, standard deviation, skewness, or kurtosis of the particular subpopula-tions. To repeat, then: the question addressed in this series of tests is: How well would an equating based on a general population serve members of special subpopulations whose means and standard deviations differed from the overall?

On the other hand, the foregoing question (1) may not be the most appropriate one to ask. One can argue that the standard errors based on the 100 pairs of samples drawn from the total population are not appropriate for groups whose distribution moments, especially the mean and standard deviation, differ from those of the total population. Accordingly, we may ask: (2) How well does an equating for a subpopulation agree with the equating for the overall population if the variance errors for the latter were adjusted to reflect the fact that the subpopulation's first four moments differed from those of the overall group?

Ordinarily, this second question might have been answered simply by substituting the observed values for the first four moments in equation (10) using the resulting variance errors of equating to evaluate the disparity between the subpopulation conversions and the overall conversion at various score points of interest. However, this procedure was judged not to be entirely satisfactory. In the first place, it was intended that the study should depend more fully than this procedure would call for on the empirically observed variation of the 100 conversions based on samples drawn from the general population. Secondly, there was some concern that the observed values of skewness and kurtosis for the subpopulations were too unstable to use, based as they were in some instances on relatively small groups, without some attempt to improve them. Now it was observed that, as expected, there was a substantial negative correlation between the mean of a subpopulation and the value of the subpopulation's skewness, a relationship that may generally be sensed from an observation of the values of mean and skewness shown in Tables 3 and 4. For GRE quantitative scores, Form D1, this correlation was -.71; for GRE quantative scores, Form D3, the correlation was -.90; for GRE verbal-plus-quantitative scores, Form D1, -70; and GRE verbal-plus-quantitative scores, Form D3, -.72. (These correlations were considerably affected, it should be noted, by the extreme bivariate points for the Minority group and the Physical Science groups. The former yielded consistently low means and relatively high positive values of skewness; the latter had very high means and relatively low skewness values.) It was therefore possible to use an entire data set of bivariate points for the 12 populations (11 subpopulations plus the total) to derive more reliable estimates of skewness (a relatively unreliable statistic) from knowledge of the subpopulation mean (a relatively stable statistic) by regressing the former on the latter. Accordingly, new regressed estimates of skewness were calculated and the factor, $[\hat{Sk}(x) + \hat{Sk}(y)]_i/[\hat{Sk}(x) + \hat{Sk}(y)]_p$, was formed (in which each of the values of skewness is an estimated value, i represents one of the 11 subpopulations and p represents the total population) and used to correct, that is, moderate, the coefficient of $z(x)$ in the equations corresponding to equations (5) and (7). These equations, it is recalled, were empirically derived from samples drawn at random from the overall population.

The observed and estimated skewnesses, and the factors used to moderate the coefficients of $z(x)$ in equations (5) and (7) are given in Tables 5 and 6. The foregoing describes the process of correcting for the fact that the skewnesses of each of the subpopulations differed from those of the total population and doing so by taking advantage of their predictability from knowledge of the means. Efforts to make stable estimates of kurtosis, however, were not as successful, and the judgment was made that, in view of the instability of the kurtosis values, no special efforts would be made to adjust them. Instead, the actual values of kurtosis that were observed for the entire populations of 13,470 and 13,527 cases (taking Forms D1 and D3, respectively) were, with one exception, used as they were found. In effect, this meant using the observed coefficients of $z^2(x)$ in equations (5) and (7)--as well as the constant values (A)--for each of the various subpopulations. The exception noted above was the case of the Physical Science subpopulation, whose kurtosis values were generally so extreme that, it was judged, they could not be ignored. Therefore, for the Physical Science group, a correction factor was developed, similar in construction to that used for adjusting the empirical skewness coefficients: $[10 + 2.5Ku(x) + 2.5Ku(y)]_{ps}/[10 + 2.5Ku(x) + 2.5Ku(y)]_p$, where the subscript ps represents the Physical Science subpopulation and p represents the total population. In a fashion similar to the adjustments for skewness, this factor was used to moderate the coefficient

Table 5

Observed and Estimated Values of Skewness,
and Correcti n Factors, by Subpopulation

GRE Quantitative Test

Subpopulation	Values of Skewness-Form D1		Values of Skewness-Form D3		Correction Factor
	Observed	Regressed	Observed	Regressed	
Total	-.2573	-.2381	-.1858	-.1799	---
Male	-.4468	-.3327	-.4057	-.3488	1.6303
Female	-.2405	-.1652	-.1521	-.0561	0.5294
White	-.1506	-.2621	-.1151	-.2151	1.1415
Minority	.2747	.0745	.4325	.3119	-0.9243
Humanities	-.2699	-.1705	-.0385	-.0875	0.6172
Social Science	-.2737	-.1541	-.1357	-.0305	0.4416
Biological Science	-.3576	-.2538	-.3249	-.2173	1.1270
Physical Science	-.7390	-.5091	-.8855	-.6631	2.8042
High-Scoring	-.1549	-.4884	-.3513	-.5877	2.5742
Middle-Scoring	-.0306	-.2475	.0004	-.2043	1.0808
Low-Scoring	-.0727	.0281	.1527	.2695	-0.7118

Table 6

Observed and Estimated Values of Skewness,
and Correction Factors, by Subpopulation

GRE Verbal-Plus-Quantitative Test

Subpopulation	Values of Skewness-Form D1		Values of Skewness-Form D3		Correction Factor
	Observed	Regressed	Observed	Regressed	
Total	-.3140	-.2373	-.2891	-.2008	---
Male	-.3903	-.2905	-.4105	-.2867	1.3174
Female	-.2703	-.1969	-.2553	-.1381	0.7645
White	-.1847	-.2666	-.1705	-.2325	1.1392
Minority	.3557	.0952	.4869	.1799	-0.6279
Humanities	-.4833	-.2601	-.2076	-.2188	1.0931
Social Science	-.2353	-.1768	-.2135	-.1173	0.6713
Biological Science	-.3002	-.2411	-.3051	-.2098	1.0293
Physical Science	-.4697	-.3876	-.6470	-.4287	1.8632
High-Scoring	-.1895	-.4801	-.1704	-.4972	2.2308
Middle-Scoring	-.1026	-.2496	-.0145	-.2232	1.0793
Low-Scoring	-.0808	.0265	-.0426	.1338	-0.3659

of $z^2(x)$ in equations (5) and (7). As implied by the foregoing, these adjustments were made in the linear equations (only) for both GRE quantitative and GRE verbal-plus-quantitative scores. Because so little is known about the parameters for the variance error of equipercentile equating, no adjustments were made in equations (6) and (8) except to relocate the scores of interest with respect to their means and standard deviations.

Table 7 gives the equations for the variance errors of linear equating for GRE quantitative scores for each of the several subpopulations, in both standard score and raw score form. As described above, these equations have been adjusted for the first four moments of the subpopulation distributions. Table 8 gives the equations for equipercentile equating, which are, also as described above, adjusted only for mean and standard deviation. Tables 9 and 10 give the corresponding equations for GRE verbal-plus-quantitative scores.

Table 7

Equations for Empirical Variance Errors of Linear Equating, Adjusted for the First Four Moments of the Distributions for the Subpopulations of Interest

GRE Quantitative

	$SE^2_{y*} = A + Bz(x)_q + Cz^2(x)_q$			$SE^2_{y*} = a + bx_q + cx^2_q$		
Subpopulation	A	B	C	a	b	c
Total	19.3386	-3.3148	10.4459	196.8437	-8.8533	0.1102
Male	Ditto	-[illegible] 4040	Ditto	245.0408	-10.1826	0.1145
Female	Ditto	- [illegible].7549	Ditto	191.6012	-9.3411	0.1266
White	Ditto	-3.7840	Ditto	227.3162	-10.1668	0.1240
Minority	Ditto	3.0637	Ditto	97.6056	-5.7496	0.1053
Humanities	Ditto	-2.0458	Ditto	208.8877	-10.0801	0.1339
Social Science	Ditto	-1.4638	Ditto	184.4776	-9.1111	0.1256
Biological Science	Ditto	-3.7356	Ditto	274.2450	-12.4992	0.1530
Physical Science	Ditto	-9.2952	17.2249	865.8115	-33.7576	0.3361
High-Scoring	Ditto	-8.5327	10.4459	554.6409	-21.6519	0.2182
Middle-Scoring	Ditto	-3.5826	Ditto	336.4294	-15.7289	0.1949
Low-Scoring	Ditto	-2.3596	Ditto	154.4809	-9.4230	0.1641

Table 8

Equations for Empirical Variance Errors of Equipercentile Equating, Adjusted for the Means and Standard Deviations of the Distributions for the Subpopulations of Interest

GRE Quantitative

	$SE^2_{y*} = a + bx_q + cx^2_q$		
Subpopulation	a	b	c
Total	665.5842	-33.1299	0.4218
Male	811.7291	-37.3646	0.4381
Female	669.1853	-35.6000	0.4843
White	772.1808	-37.9148	0.4746
Minority	364.1782	-23.7216	0.4029
Humanities	730.4559	-38.2972	0.5125
Social Science	646.3721	-34.8426	0.4807
Biological Science	948.2828	-46.7520	0.5855
Physical Science	1870.4530	-76.0760	0.7798
High-Scoring	1879.3237	-78.9107	0.8350
Middle-Scoring	1184.2311	-59.0518	0.7456
Low-Scoring	585.9281	-37.8670	0.6279

Table 9

Equations for Empirical Variance Errors of Linear Equating, Adjusted for the First Four Moments of the Distributions for the Subpopulations of Interest

GRE Verbal-Plus-Quantitative

	$SE^2_{y*} = A + Bz(x)_c + Cz^2(x)_c$			$SE^2_{y*} = a + bx_c + cx^2_c$		
Subpopulation	A	B	C	a	b	c
Total	17.0949	-5.6400	9.0657	228.5855	-4.6933	0.0259
Male	Ditto	-7.4302	Ditto	268.7716	-5.2206	0.0269
Female	Ditto	-4.3117	Ditto	218.0270	-4.7124	0.0276
White	Ditto	-6.4249	Ditto	278.0194	-5.6631	0.0306
Minority	Ditto	3.5416	Ditto	96.1123	-2.6675	0.0224
Humanities	Ditto	-6.1651	Ditto	244.5884	-4.9732	0.0270
Social Science	Ditto	-3.7860	Ditto	200.8596	-4.3395	0.0262
Biogical Science	Ditto	-5.8052	Ditto	298.1843	-6.2411	0.0345
Physical Science	Ditto	-10.5085	11.8987	543.3550	-10.1344	0.0486
High-Scoring	Ditto	-12.5817	9.0657	742.9482	-13.3119	0.0607
Middle-Scoring	Ditto	- 6.0872	Ditto	452.9112	-9.6751	0.0536
Low-Scoring	Ditto	2.0635	Ditto	178.3997	-5.0231	0.0391

Table 10

Equations for Empirical Variance Errors of Equipercentile Equating, Adjusted for the Means and Standard Deviations of the Distributions for the Subpopulations of Interest

GRE Verbal-Plus-Quantitative

$$SE^2_{y*} = a + bx_c + cx_c^2$$

Subpopulation	a	b	c
Total	718.0959	-15.3412	0.0829
Male	820.8492	-16.7211	0.0860
Female	703.0153	-15.6471	0.0881
White	866.6386	-18.3243	0.0978
Minority	378.4867	-10.3019	0.0716
Humanities	762.3566	-16.1513	0.0865
Social Science	653.4806	-14.6901	0.0836
Biological Science	942.4082	-20.3080	0.1104
Physical Science	1287.3908	-24.6022	0.1183
High-Scoring	2202.2282	-41.2518	0.1939
Middle-Scoring	1438.8581	-31.3022	0.1712
Low-Scoring	663.1466	-18.0895	0.1249

Results

Tables 11, 12, 13, and 14 give the results for Question 1: To what extent do the conversions based on the subpopulation samples agree with the total-population conversion if we take as the standard errors of equating those that were obvserved in the variation of the 100 conversions based on samples drawn at random from the total? With one notable exception, relating to the samples drawn from the Physical Science subpopulation, the results given in Table 11, based on linear equating for GRE quantitative scores, show that the samples drawn from their respective subpopulations yield conversions that are very much in line with the conversion based on the total population. There are occasional departures from expectation, but these are few in number and they tend to occur in the regions of the scale far removed from the vicinity of the mean for the particular subpopulation in question. For example, Sample 1 of the High-Scoring group shows significant departures (at the 5% level) at the lower end of the scale; one of the Middle-Scoring and one of the Low-Scoring samples show significant departures at the upper end of the scale. One clear explanation for these departures is that when the mean of the particular group differs from that of the total population, the standard errors used in the Question 1 phase of the study, taken from equations (1)-(4), are inappropriately small for scores at great distances from the mean of the group. (Corresponding, they are too large for scores close to the mean of the group.) As implied earlier, conversion lines based on samples drawn at random from a population tend to intersect in the region of the mean of the population and fan out at the ends of the scale, with greater fanning at greater distances from the mean.

It is noteworthy in this connection that the conversion for the Minority sample showed no significant values at any of the five points. Because of

Table 11

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Linear Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores 10	20	30	40	50
Male					
Sample 1					
Difference	1.15	-0.45	-2.06	-3.66	-5.26
SE Equating (Unadj)	10.92	7.99	5 52	4.37	5.45
t-value	0.11	-0.06	-0.37	-0.84	-0.96
Sample 2					
Difference	-2.31	-3.20	-4.09	-4.99	-5.88
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-0.21	-0.40	-0.74	-1.14	-1.08
Female					
Sample 1					
Difference	6.78	4.99	3.20	1.41	-0.38
SE Equating (Unadj)	10 92	7.99	5.52	4.37	5.45
t-value	0.62	0.62	0.58	0.32	-0.07
Sample 2					
Difference	-15.54	-13.45	-11.36	-9.27	-7.18
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-1.42	-1.68	-2.06*	-2.12*	-1.32
White					
Sample 1					
Difference	12.20	6.00	-0.19	-6.38	-12.58
SE Equating (Unadj)	10 92	7.99	5.52	4.37	5 45
t-value	1.12	0.75	-0.03	-1.46	-2.31*
Minority					
Sample 1					
Difference	-9.81	-6.84	-3.87	-0.90	2.07
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5 45
t-value	-0.90	-0.86	-0.70	-0.21	0.38

Table 11 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Linear Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores 10	20	30	40	50
Humanities					
Sample 1					
Difference	-11.56	-8.23	-4.90	-1.57	1.75
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-1.06	-1.03	-0.89	-0.36	0.32
Sample 2					
Difference	-14.77	-11.08	-7.38	-3.69	0.00
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-1.35	-1.39	-1.34	-0.84	0.00
Social Science					
Sample 1					
Difference	9.06	5.58	2.09	-1.39	-4.87
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	0.83	0.70	0.38	-0.32	-0.89
Sample 2					
Difference	0.43	0.58	0.74	0.89	1.04
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	0.04	0.07	0.13	0.20	0.19
Biological Science					
Sample 1					
Difference	10.31	6.27	2.23	-1.81	-5.85
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	0.94	0.78	0.40	-0.41	-1.07
Sample 2					
Difference	-17.04	-12.27	-7.50	-2.73	2.04
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-1.56	-1.54	-1.36	-0.63	0.37

Table 11 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Linear Equating

Standard Errors Unadjusted

	Form D3 Raw Scores				
Subpopulation Sample	10	20	30	40	50
Physical Science					
Sample 1					
Difference	-7.85	-6.87	-5.89	-4.92	-3.94
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-0.72	-0.86	-1.07	-1.13	-0.72
Sample 2					
Difference	-60.39	-46.58	-32.77	-18.96	-5.15
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-5.53**	-5.83**	-5.94**	-4.34**	-0.94
Sample 3					
Difference	-41.10	-31.67	-22.24	-12.82	-3.39
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-3.76**	3.96**	-4.03**	-2.93**	-0.62
Sample 4					
Difference	-25.50	-20.15	-14.80	-9.46	-4.11
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-2.33*	-2.52*	-2.68**	-2.16*	-0.75
High-Scoring					
Sample 1					
Difference	23.56	17.94	12.33	6.71	1.09
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	2.16*	2.25*	2.23*	1.54	0.20
Sample 2					
Difference	7.39	4.39	1.40	-1.60	-4.60
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	0.68	0.55	0.25	-0.37	-0.84

Table 11 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Linear Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores 10	20	30	40	50
Middle-Scoring					
Sample 1					
Difference	-7.57	-4.76	-1.94	0.87	3.68
SE Equating (Unadj)	10.92	7.99	5 52	4.37	5.45
t-value	-0.69	-0.60	-0.35	0.20	0.67
Sample 2					
Difference	1.28	-1.99	-5.26	-8.52	-11.79
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	0.12	-0.25	-0.95	-1.95	-2.16*
Low-Scoring					
Sample 1					
Difference	-7.81	-3.77	0.27	4.31	8
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-0.72	-0.47	0.05	0.99	1.53
Sample 2					
Difference	-4.41	0.32	5.05	9.78	14.50
SE Equating (Unadj)	10.92	7.99	5.52	4.37	5.45
t-value	-0.40	0.04	0.91	2.24*	2.66**

*$p < 0.05$
**$p < 0.01$

Note: Negative values in this table indicate that the sample conversion is lower than the total-population conversion.

Table 12

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Equipercentile Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores 10	20	30	40	50
Male					
Sample 1					
Difference	10.45	-23.62	5.02	-2.65	-7.53
SE Equating (Unadj)	19.40	13.10	7.16	3.90	7.97
t-value	0.54	-1.80	0.70	-0.68	-0.94
Sample 2					
Difference	40.49	-1.94	-6.65	-4.58	-2.26
SE Equating (Unadj)	19.40	13.10	7.16	3.90	7.97
t-value	2.09*	-0.15	-0.93	-1.17	-0.28
Female					
Sample 1					
Difference	14.81	-2.47	5.20	-1.49	3.78
SE Equating (Unadj)	19.40	13.10	7.16	3.90	7.97
t-value	0.76	-0.19	0.73	-0.38	0.47
Sample 2					
Difference	-25.85	-14.17	-8.16	-9.38	-10.78
SE Equating (Unadj)	19.40	13.10	7.16	3.90	7.97
t-value	-1.33	-1.08	-1.14	-2.40*	-1.35
White					
Sample 1					
Difference	a	0.44	0.91	-6.94	-12.74
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	0.03	0.13	-1.78	-1.60
Minority					
Sample 1					
Difference	4.41	-3.13	-1.59	-9.82	-1.71
SE Equating (Unadj)	19.40	13.10	7.16	3.90	7.97
t-value	0.23	-0.24	-0.22	-2.52*	-0.22

Table 12 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Equipercentile Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores 10	20	30	40	50
Humanities					
Sample 1					
Difference	a	-0.19	-2.53	-4.19	9.24
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	-0.01	-0.35	-1.07	1.16
Sample 2					
Difference	a	-5.56	-4.65	-4.05	-0.32
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	-0.42	-0.65	-1.04	-0.04
Social Science					
Sample 1					
Difference	a	7.23	5.52	-5.47	-4.01
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	0.55	0.77	-1.40	-0.50
Sample 2					
Difference	11.73	-5.77	-1.29	3.91	-4.11
SE Equating (Unadj)	19.40	13.10	7.16	3.90	7.97
t-value	0.60	-0.44	-0.18	1.00	-0.52
Biological Science					
Sample 1					
Difference	a	22.92	-7.02	-1.04	-4.80
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	1.75	-0.98	-0.27	-0.60
Sample 2					
Difference	a	-25.12	-5.84	-3.60	0.46
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	-1.92	-0.82	-0.92	0.06

Table 12 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Equipercentile Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores 10	20	30	40	50
Physical Science					
Sample 1					
Difference	a	20.83	-2.17	-10.18	-2.25
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	1.59	-0.30	-2.61**	-0.28
Sample 2					
Difference	a	-45.06	-36.49	-21.86	-2.48
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	-3.44**	-5.10**	-5.61**	-0.31
Sample 3					
Difference	a	-28.86	-26.99	-15.43	-1.38
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	-2.20*	-3.77**	-3.96**	-0.17
Sample 4					
Difference	a	-51.50	-24.98	-7.51	-2.72
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	-3.93**	-3.49**	-1.93	-0.34
High-Scoring					
Sample 1					
Difference	a	a	19.77	8.17	2.79
SE Equating (Unadj)	-	-	7.16	3.90	7.97
t-value	-	-	2.76**	2.10*	0.35
Sample 2					
Difference	a	a	5.24	0.25	-7.33
SE Equating (Unadj)	-	-	7.16	3.90	7.97
t-value	-	-	0.73	0.06	-0.92

Table 12 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Equipercentile Equating

Standard Errors Unadjusted

	Form D3 Raw Scores				
Subpopulation Sample	10	20	30	40	50
Middle-Scoring					
Sample 1					
Difference	a	-28.72	-9.94	2.94	3.38
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	-2.19*	-1.39	0.75	0.42
Sample 2					
Difference	a	0.25	-12.17	-8.04	-15.26
SE Equating (Unadj)	-	13.10	7.16	3.90	7.97
t-value	-	0.02	-1.70	-2.06*	-1.92
Low-Scoring					
Sample 1					
Difference	11.10	1.12	-2.84	3.10	8.02
SE Equating (Unadj)	19.40	13.10	7.16	3.90	7.97
t-value	0.57	0.09	-0.40	0.79	1.01
Sample 2					
Difference	16.57	6.87	3.19	8.79	3.09
SE Equating (Unadj)	19.40	13.10	7.16	3.90	7.97
t-value	0.85	0.52	0.45	2.26*	0.39

[a]No data are available to establish conversions at these points.

*$p < 0.05$
**$p < 0.01$

Note: Negative values in this table indicate that the sample conversion is lower than the total-population conversion.

Table 13

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores: 25	50	75	100	125
Male					
Sample 1					
Difference	1.75	-2.65	-7.06	-11.46	-15.86
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.15	-0.35	-1.49	-2.66**	-2.31*
Sample 2					
Difference	-10.08	-9.29	-8.50	-7.71	-6.92
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-0.89	-1.21	-1.79	-1.79	-1.01
Sample 1					
Difference	1.98	3.91	5.84	7.77	9.70
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.18	0.51	1.23	1.80	1.41
Sample 2					
Difference	-15.78	-11.42	-7.06	-2.70	1.65
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-1.40	-1.49	-1.49	-0.63	0.24
White					
Sample 1					
Difference	3.96	-1.42	-6.80	-12.18	-17.56
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.35	-0.18	-1.43	-2.83**	-2.56*
Minority					
Sample 1					
Difference	3.80	-1.71	-7.23	-12.74	-18.26
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.34	-0.22	-1.53	-2.96**	-2.66**

Table 13 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Standard Errors Unadjusted

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Humanities					
Sample 1					
Difference	3.23	3.39	3.54	3.70	3.85
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.29	0.44	0.75	0.86	0.56
Sample 2					
Difference	1.28	1.70	2.11	2.52	2.93
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.11	0.22	0.45	0.59	0.43
Social Science					
Sample 1					
Difference	7.16	4.76	2.36	-0.04	-2.44
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.63	0.62	0.50	-0.01	-0.36
Sample 2					
Difference	-3.98	-1.21	1.55	4.31	7.07
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-0.35	-0.16	0.33	1.00	1.03
Biological Science					
Sample 1					
Difference	-0.39	-2.05	-3.72	-5.38	-7.05
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-0.03	-0.27	-0.78	-1.25	-1.03
Sample 2					
Difference	-24.09	-14.98	-5.87	3.24	12.35
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-2.13*	-1.95	-1.24	0.75	1.80

Table 13 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores				
	25	50	75	100	125
Physical Science					
Sample 1					
Difference	-24.40	-20.08	-15.76	-11.44	-7.12
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-2.16*	-2.62**	-3.33**	-2.65**	-1.04
Sample 2					
Difference	-59.22	-44.06	-28.91	-13.75	1.40
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-5.25**	-5.75**	-6.10**	-3.19**	0.20
Sample 3					
Difference	-40.63	-29.70	-18.77	-7.84	3.10
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-3.60**	-3.88**	-3.96**	-1.82	0.45
Sample 4					
Difference	-25.48	-20.90	-16.33	-11.75	-7.18
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-2.26*	-2.73**	-3.45**	-2.73**	-1.05
High-Scoring					
Sample 1					
Difference	6.99	5.12	3.26	1.39	-0.48
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.62	0.67	0.69	0.32	-0.07
Sample 2					
Difference	3.99	1.14	-1.70	-4.55	-7.39
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.35	0.15	-0.36	-1.06	-1.08

Table 13 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Standard Errors Unadjusted

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Middle-Scoring					
Sample 1					
Difference	-21.95	-12.53	-3.11	6.32	15.74
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-1.94	-1.63	-0.66	1.47	2.29*
Sample 2					
Difference	-35.08	-23.73	-12.38	-1.03	10.32
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	-3.11**	-3.10**	-2.61**	-0.24	1.50
Low-Scoring					
Sample 1					
Difference	7.26	4.50	1.74	-1.02	-3.78
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.64	0.59	0.37	-0.24	-0.55
Sample 2					
Difference	4.92	7.50	10.08	12.67	15.25
SE Equating (Unadj)	11.29	7.66	4.74	4.31	6.86
t-value	0.44	0.98	2.13*	2.94**	2.22*

*$p < 0.05$
**$p < 0.01$

Note: Negative values in this table indicate that the sample conversion is lower than the total-population conversion.

Table 14

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Equipercentile Equating

Standard Errors Unadjusted

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Male					
Sample 1					
Difference	a	-16.21	-3.02	-14.49	-17.33
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	-1.29	-0.52	-4.07**	-1.77
Sample 2					
Difference	a	7.33	-8.91	-6.81	5.58
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	0.58	-1.54	-1.91	0.57
Female					
Sample 1					
Different	23.20	-0.39	10.00	8.62	-3.97
SE Equating (Unadj)	19.66	12.58	5 80	3.56	9.76
t-value	1.18	-0.03	1.72	2.42*	-0.41
Sample 2					
Difference	a	-12.09	-4.35	0.07	-2.80
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	-0.96	-0.75	0.02	-0.29
White					
Sample 1					
Difference	a	4.62	-8.69	-13.25	-8.00
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	0.37	-1.50	-3.72**	-0.82
Minority					
Sample 1					
Difference	1.79	-1.04	-1.65	-18.40	-38.50
SE Equating (Unadj)	19.66	12.58	5.80	3.56	9.76
t-value	0.09	-0.08	-0.28	-5.17**	-3.94**

Table 14 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Equipercentile Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores 25	50	75	100	125
Humanities					
Sample 1					
Difference	a	2.26	3.26	3.45	-13.33
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	0.18	0.56	0.97	-1.37
Sample 2					
Difference	a	0.54	2.28	2.00	0.76
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	0.04	0.39	0.56	0.08
Social Science					
Sample 1					
Difference	15.31	15.06	0.74	1.27	22.50
SE Equating (Unadj)	19.66	12.58	5.80	3.56	9 76
t-value	0.78	1.20	0.13	0.36	2.30*
Sample 2					
Difference	16.35	4.50	2.91	5.84	1.93
SE Equating (Unadj)	19.66	12.58	5.80	3.56	9.76
t-value	0.83	0.36	0.50	1.64	0.20
Biological Science					
Sample 1					
Difference	a	3.27	-4.15	-5.23	-23.28
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	0.26	-0.72	-1.47	-2.38*
Sample 2					
Difference	a	-26.55	-9.84	6.49	2.90
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	-2.11*	-1.70	1.82	0.30

Table 14 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Equipercentile Equating

Standard Errors Unadjusted

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Physical Science					
Sample 1					
Difference	a	-9.11	-20.55	-11.33	2.80
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	-0.72	-3.54**	-3.18**	0.29
Sample 2					
Difference	a	-23.24	-31.73	-15.52	5.43
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	-1.85	-5.47**	-4.36**	0.56
Sample 3					
Difference	a	-18.09	-24.81	-9.25	1.22
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	-1.44	-4.28**	-2.60**	0.12
Sample 4					
Difference	a	-15.78	-23.77	-12.03	0.03
SE Equating (Unadj)	-	12.58	5.80	3 56	9.76
t-value	-	-1.25	-4.10**	-3.38**	0.00
High-Scoring					
Sample 1					
Difference	a	a	8.78	1.77	0.22
SE Equating (Unadj)	-	-	5.80	3.56	9.76
t-value	-	-	1.51	0.50	0.02
Sample 2					
Difference	a	a	9.39	-6.04	-1.85
SE Equating (Unadj)	-	-	5.80	3.56	9.76
t-value	-	-	1.62	-1.70	-0.19

Table 14 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Equipercentile Equating

Standard Errors Unadjusted

Subpopulation Sample	Form D3 Raw Scores 25	50	75	100	125
Middle-Scoring					
Sample 1					
Difference	a	-55.91	-1.64	7.29	-14.66
SE Equating (Unadj)	-	12.58	5.80	3.56	9.76
t-value	-	-4.44**	-0.28	2.05*	-1.50
Sample 2					
Difference	a	-24.76	-11.75	-2.89	a
SE Equating (Unadj)	-	12.58	5.80	3.56	-
t-value	-	-1.97*	-2.03*	-0.81	-
Low-Scoring					
Sample 1					
Difference	16.43	5.80	4.82	-8.08	a
SE Equating (Unadj)	19.66	12.58	5.80	3.56	-
t-value	0.84	0.46	0.83	-2.27*	-
Sample 2					
Difference	22.17	4.89	8.83	21.90	a
SE Equating (Unadj)	19.66	12.58	5.80	3.56	-
t-value	1.13	0.39	1.52	6.15**	-

[a]No data are available to establish conversions at these points.

*$p < 0.05$
**$p < 0.01$

Note: Negative values in this table indicate that the sample conversion is lower than the total-population conversion.

their low means on these two forms, one would ordinarily have expected significant departures for their conversion line at the upper end of the scale. These departures did not occur in Table 11; they did, however, occur in Table 12, which is based on the results of equipercentile equating.

Some mention should also be made here of the fact that, with the exception of the White and Minority groups, more than one sample was drawn from each of the subpopulations, even though some of the subpopulations were relatively small. Clearly, with small populations, there is likely to be far too much data common to the samples to yield experimentally independent observations. But having drawn the samples and calculated the results, the authors judged that the correct course of action would be to report all the data. As it stands, then, it should be kept in mind that the two or more samples drawn from the smaller subpoulations yield somewhat redundant information.

This discussion brings us to the data for the Physical Science group. Initially, only two pairs of samples were drawn from this subpopulation, as for all but two of the other subpopulations, and correspondingly, only two linear conversions were carried out for this group. When the results of these operations were reviewed (see the results for Samples 1 and 2), they were thought to be too different to accept without question, especially because of the large proportion of data likely to be common to them. Accordingly, the sampling process was reviewed, and when that was found to be free of error, two more samples, Samples 3 and 4, were also drawn to determine the direction in which the data were leaning. The results in Table 11 seem to indicate that, in spite of its consistency with the findings for all the other samples drawn from the 10 other subpopulations, the conversion for Sample 1 was as divergent from the Physical Science subpopulation conversion in the direction as was Sample 2 in the other direction. The conversions for Samples 3 and 4 were more intermediate and closer to their subpopulation conversion.

In general, these results show that the conversions based on the samples drawn from the Physical Science subpopulation are clearly at odds with the Total-Population conversion, more so, as expected from the high level of its mean, at the lower end of the scale. Indeed, the conversion lines for all four samples fall below the overall conversion line, as does the line based on the entire Physical Science subpopulation. Obviously, one explanation for the behavior of the Physical Science group is the substantial difference between its mean and the population mean. However, other groups, whose means are similarly different from the population mean, do not shows a degree of departure of this magnitude. The Minority group, whose means are even more distant from the overall population means than the Physical Science group, shows no significant t-values at all; and the results for the High-Scoring group, whose means are also much higher than those of the total population, do not show significance at any point beyond the 5 percent level. This is all the more perplexing because of the degree of overlap between the Physical Science and the High-Scoring groups. There were 997 cases common to these two groups in the D1 population, representing 44 percent of the Physical Science group and 27 percent of the High-Scoring group; and there were 982 cases common to the two groups in the D3 population, representing 46 percent of the Physical Science group and 27 percent of the High-Scoring group. With this degree of overlap, one might, perhaps, have expected greater similarity in their conversion equations.

Examination of Table 12, which gives corresponding results for equipercentile equating, reveals that it is essentially consistent with Table 11 in all respects. As in Table 11, the only subpopulation that shows extensive departures from the population conversion is the Physical Science group. It is noted parenthetically that, in Table 12 as in other tables that report the results of equipercentile conversions, there are some values that cannot be determined, simply because data are not available at those points. Unlike linear equating that, by definition, is based on the first two moments and yields a line capable of extrapolation without limit, equipercentile equating is largely a local process, depending on the availability of data at the score points of interest. If there are no data, there cannot be a conversion except, of course, by arbitrary extrapolation without the benefit of actual data.

Table 13, based on linear equating data taken from the heterogeneous verbal-plus-quantitative test, yields information generally consistent with its counterpart, Table 11, which is also based on linear equating but takes its data from the homogeneous quantitative test. Indeed, Table 13 shows slightly larger absolute t-values, on average, and a somewhat greater number of statistically significant values than are found in Table 11. As expected, the same results, generally, are found for equipercentile equating, as will be seen in comparing Table 14 with Table 12. No satisfactory method of evaluating the statistical significance of the differences between Tables 11 and 12 on the one hand and Tables 13 and 14 on the other is available. The interpretation is further complicated by the fact, as observed earlier in reviewing Tables 1 and 2, that the forms of the heterogeneous tests are not strictly parallel, an observation supported by an examination of the distribution of item content over the 21 or 22 reading comprehension items in the verbal component of the GRE verbal-plus-quantitative test. It is customary in the construction of items of this type to represent each of the four curricular areas (humanities, social science, biological science, and physical science) with the use of either a long passage followed by seven or eight items, or a short passage followed by three or four items, and to allow the association of passage length (and number of items) with curricular area to vary from form to form. In the construction of Forms D1 and D3, the lengths of the passages and the numbers of items based on them were distributed as shown in the table below. From the information given in this table, it appears that Form D1 gave more weight to humanities and biological science and less to social science and physical science; Form D3

	Humanities		Social Science		Biological Science		Physical Science	
	Passage Length	No. of Items	Passage Length	No. of Items	Passage Length	No. of Items	Passage Length	No. of Items
Form D1	Long	8	Short	3	Long	7	Short	3
Form D3	Short	3	Long	7	Long	8	Short	4

gave more weight to social and biological science and less to humanities and physical science. It is significant, however, that while physical science received relatively little weight in either form, it received more weight in D3 than in D1. Although it is difficult to speculate on the effects these differences may have exerted on the study, it seems clear that the two forms were not in fact strictly parallel and that their differences may well have interacted with the performances of the four curriculum groups in such a way as to affect their individual conversions as observed in this study.

In reviewing the results summarized in Tables 11-14, it must be borne in mind that the departures of the conversion lines for the subpopulation samples from the overall conversion line were evaluated statistically by using standard errors appropriate to a population with its own particular mean and standard deviation, not the mean and standard deviation of the subpopulation. As already pointed out, some of these subpopulations are characterized by mean scores at levels quite different from that of the general population and by degrees of skewness that are consistent with their means. The standard errors appropriate to these subpopulations are different from those appropriate to the general population. Accordingly, Tables 15-18 are presented, which correspond respectively to Tables 11-14. They differ from the preceding tables in the respect that the empirical standard error formulas used to calculate the observed differences in the conversion lines have been adjusted with respect to the position of the mean and its standardized distance from the score points of interest. In addition, the standard error formulas for linear equating, in Tables 15 and 17, have also been adjusted for the third moment (skewness) and, where possible, the fourth moment (kurtosis).

Figures 1 and 2 illustrate the discrepancies between sample conversions and the Total-Population conversion. Both figures have been oriented in such a way as to show the overall conversion as a base line and the sample conversion in relation to it. Figure 1 shows the discrepancy between the Female Sample 1 conversion and the base line conversion in relation to 1.96 standard errors of equating (the "box") and in relation to 2.58 standard errors of equating (the "whisker"), when appropriate adjustments have been made in the standard errors for the first four moments of the distributions. The close agreement between the Female Sample 1 conversion and the Total-Population conversion is evident in Figure 1. Figure 2 gives a corresponding picture, but for a case of a large discrepancy, that between the Physical

Table 15

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Linear Equating

Standard Errors Adjusted for the First Four Moments

Subpopulation Sample	Form D3 Raw Scores				
	10	20	30	40	50
Male					
Sample 1					
Difference	1.15	-0.45	-2.06	-3.66	-5.26
SE Equating (Adj)	12.44	9.34	6.53	4.57	4.71
t-value	0.09	-0.05	-0.31	-0.80	-1.12
Sample 2					
Difference	-2.31	-3.20	-4.09	-4.99	-5.88
SE Equating (Adj)	12.44	9.34	6.53	4.57	4.71
t-value	-0.19	-0.34	-0.63	-1.09	-1.25
Female					
Sample 1					
Difference	6.78	4.99	3.20	1.41	-0.38
SE Equating (Adj)	10.53	7.44	5.03	4.53	6.40
t-value	0.64	0.67	0.64	0.31	-0.06
Sample 2					
Difference	-15.54	-13.45	-11.36	[illegible]	-7.18
SE Equating (Adj)	10.53	7.44	5.03	4.53	6.40
t-value	-1.48	-1.81	-2.26*	-2.05*	-1.12
White					
Sample 1					
Difference	12.20	6.00	-0.19	-6.38	-12.58
SE Equating (Adj)	11.75	8.58	5.83	4.37	5.39
t-value	1.04	0.70	-0.03	-1.46	-2.33*
Minority					
Sample 1					
Difference	-9.81	-6.84	-3.87	-0.90	2.07
SE Equating (Adj)	7.12	4.97	4.46	6.01	8.56
t-value	-1.38	-1.38	-0.87	-0.15	0.24

Table 15 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Linear Equating

Standard Errors Adjusted for the First Four Moments

Subpopulation Sample	Form D3 Raw Scores 10	20	30	40	50
Humanities					
Sample 1					
Difference	-11.56	-8.23	-4.90	-1.57	1.75
SE Equating (Adj)	11.02	7.80	5.20	4.47	6.30
t-value	-1.05	-1.05	-0.94	-0.35	0.28
Sample 2					
Difference	-14.77	-11.08	-7.38	-3.69	0.00
SE Equating (Adj)	11.02	7.80	5.20	4.47	6.30
t-value	-1.34	-1.42	-1.42	-0.83	0.00
Social Science					
Sample 1					
Difference	9.06	5.58	2.09	-1.39	-4.87
SE Equating (Adj)	10.29	7.25	4.92	4.59	6.56
t-value	0.88	0.77	0.43	-0.30	-0.74
Sample 2					
Difference	0.43	0.58	0.74	0.89	1.04
SE Equating (Adj)	10.29	7.25	4.92	4.59	6.56
t-value	0.04	0.08	0.15	0.19	0.16
Biological Science					
Sample 1					
Difference	10.31	6.27	2.23	-1.81	-5.85
SE Equating (Adj)	12.83	9.25	6.08	4.37	5.64
t-value	0.80	0.68	0.37	-0.41	-1.04
Sample 2					
Difference	-17.04	-12.27	-7.50	-2.73	2.04
SE Equating (Adj)	12.83	9.25	6.08	4.37	5.64
t-value	-1.33	-1.33	-1.23	-0.62	0.36

Table 15 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Linear Equating

Standard Errors Adjusted for the First Four Moments

	Form D3 Raw Scores				
Subpopulation Sample	10	20	30	40	50
Physical Science					
Sample 1					
Difference	-7.85	-6.87	-5.90	-4.92	-3.94
SE Equating (Adj)	23.70	18.03	12.47	7.29	4.25
t-value	-0.33	-0.38	-0.47	-0.67	-0.93
Sample 2					
Difference	-60.39	-46.58	-32.77	-18.96	-5.15
SE Equating (Adj)	23.70	18 03	12 47	7 29	4.25
t-value	-2.55*	-2.58**	-2.63**	-2.60**	-1.21
Sample 3					
Difference	-41.10	-31.67	-22.24	-12.82	-3.39
SE Equating (Adj)	23.70	18.03	12.47	7.29	4.25
t-value	-1.73	-1.76	-1.78	-1.76	-0.80
Sample 4					
Difference	-25.50	-20.15	-14.80	-9.46	-4.11
SE Equating (Adj)	23.70	18.03	12.47	7.29	4.25
t-value	-1.08	-1.12	-1.19	-1.30	-0.97
High-Scoring					
Sample 1					
Difference	23.56	17.94	12.33	6.71	1.09
SE Equating (Adj)	18.97	14.45	10.07	6.14	4.20
t-value	1.24	1.24	1.22	1.09	0.26
Sample 2					
Difference	7.39	4.39	1.40	-1.60	-4.60
SE Equating (Adj)	18.97	14.45	10.07	6.14	4.20
t-value	0.39	0.30	0.14	-0.26	-1.09

Table 15 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Linear Equating

Standard Errors Adjusted for the First Four Moments

	Form D3 Raw Scores				
Subpopulation Sample	10	20	30	40	50
Middle-Scoring					
Sample 1					
Difference	-7.57	-4.76	-1.94	0.87	3.68
SE Equating (Adj)	14.09	9.99	6.32	4.37	6.09
t-value	-0.54	-0.48	-0.31	0.20	0.60
Sample 2					
Difference	1.28	-1.99	-5.26	-8.52	-11.79
SE Equating (Adj)	14.09	9.99	6.32	4.37	6.09
t-value	0.09	-0.20	-0.83	-1.95	-1.93
Low-Scoring					
Sample 1					
Difference	-7.81	-3.77	0.27	4.31	8.34
SE Equating (Adj)	8.76	5.63	4.41	6.33	9.67
t-value	-0.89	-0.67	0.06	0.68	0.86
Sample 2					
Difference	-4.41	0.32	5.05	9.78	14.50
SE Equating (Adj)	8.76	5.63	4.41	6.33	9.67
t-value	-0.50	0.06	1.14	1.54	1.50

*$p < 0.05$
**$p < 0.01$

Note: Negative values in this table indicate that the sample conversion is lower than the total-population conversion.

Table 16

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Equipercentile Equating

Standard Errors Adjusted for Mean and Standard Deviation Only

	Form D3 Raw Scores				
Subpopulation Sample	10	20	30	40	50
Male					
Sample 1					
Differences	10.45	-23.62	5.02	-2.65	-7.53
SE Equating (Adj)	21.95	15.48	9.22	4.25	6 22
t-value	0.48	-1.53	0.54	-0.62	-1.21
Sample 2					
Differences	40.49	-1.94	-6.65	-4.58	-2.26
SE Equating (Adj)	21.95	15.48	9.22	4.25	6.22
t-value	1.84	-0.13	-0.72	-1.08	-0.36
Female					
Sample 1					
Differences	14.81	-2.47	5.20	-1.49	3.78
SE Equating (Adj)	19.02	12.28	6.09	4.48	10.00
t-value	0.78	-0.20	0.85	-0.33	0.38
Sample 2					
Differences	-25.85	-14.17	-8.16	-9.38	-10.78
SE Equating (Adj)	19.02	12.28	6 09	4.48	10 00
t-value	-1.36	-1.15	-1.34	-2.09*	-1.08
White					
Sample 1					
Difference	a	0.44	0.91	-6.94	-12.74
SE Equating (Adj)	-	14.27	7.87	3.87	7.94
t-value	-	0.03	0.12	-1.79	-1.61
Minority					
Sample 1					
Difference	4.41	-3.13	-1.59	-9.82	-1.71
SE Equating (Adj)	12.93	7.13	3.89	7.74	13.61
t-value	0.34	-0.44	-0.41	-1.27	-0.13

Table 16 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Equipercentile Equating

Standard Errors Adjusted for Mean and Standard Deviation Only

Subpopulation Sample	Form D3 Raw Scores				
	10	20	30	40	50
Humanities					
Sample 1					
Difference	a	-0.19	-2.53	-4.19	9.24
SE Equating (Adj)	-	13.02	6.54	4.31	9.84
t-value	-	-0.01	-0.39	-0.97	0.94
Sample 2					
Difference	a	-5.56	-4.65	-4.05	-0.32
SE Equating (Adj)	-	13.02	6.54	4.31	9.84
t-value	-	-0.43	-0.71	-0.94	-0.03
Social Science					
Sample 1					
Difference	a	7.23	5.52	-5.47	-4.01
SE Equating (Adj)	-	11.91	5.81	4.67	10.29
t-value	-	0.61	0.95	-1.17	-0.39
Sample 2					
Difference	11.73	-5.77	-1.29	3.92	-4.11
SE Equating (Adj)	18.60	11.91	5.81	4.67	10.29
t-value	0.63	-0.48	-0.22	0.84	-0.40
Biological Science					
Sample 1					
Difference	a	22.92	-7.02	-1.04	-4.80
SE Equating (Adj)	-	15.73	8.52	3.87	8.63
t-value	-	1.46	-0.82	-0.27	-0.56
Sample 2					
Difference	a	-25.12	-5.84	-3.60	0.46
SE Equating (Adj)	-	15.73	8.52	3.87	8.63
t-value	-	-1.60	-0.69	-0.93	0.05

Table 16 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Equipercentile Equating

Standard Errors Adjusted for Mean and Standard Deviation Only

	Form D3 Raw Scores				
Subpopulation Sample	10	20	30	40	50
Physical Science					
Sample 1					
Difference	a	20.83	-2.17	-10.18	-2.25
SE Equating (Adj)	-	25.71	17.03	8 67	4.02
t-value	-	0.81	-0.13	-1.17	-0.56
Sample 2					
Difference	a	-45.06	-36.49	-21.86	-2.48
SE Equating (Adj)	-	25.71	17.03	8.67	4.02
t-value	-	-1.75	-2.14*	-2.52*	-0.62
Sample 3					
Difference	a	-28.86	-26.99	-15.43	-1.38
SE Equating (Adj)	-	25.71	17.03	8 67	4.02
t-value	-	-1.12	-1.58	-1.78	-0.34
Sample 4					
Difference	a	-51.50	-24.98	-7.51	-2.72
SE Equating (Adj)	-	25.71	17 03	8.67	4.02
t-value	-	-2.00*	-1.47	-0.87	-0.68
High-Scoring					
Sample 1					
Difference	a	a	19.77	8.17	2.79
SE Equating (Adj)	-	-	16.23	7 67	4.61
t-value	-	-	1.22	1.07	0.60
Sample 2					
Difference	a	a	5.24	0.25	-7.33
SE Equating (Adj)	-	-	16.23	7.67	4.61
t-value	-	-	0.32	0.03	-1.59

Table 16 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Quantitative Scores; Equipercentile Equating

Standard Errors Adjusted for Mean and Standard Deviation Only

	Form D3 Raw Scores				
Subpopulation Sample	10	20	30	40	50
Middle-Scoring					
Sample 1					
Difference	a	-28.72	-9.94	2.94	3.38
SE Equating (Adj)	-	17.36	9.15	3.89	9.78
t-value	-	-1.65	-1.09	0.76	0.35
Sample 2					
Difference	a	0.25	-12.17	-8.04	-15.26
SE Equating (Adj)	-	17.36	9.15	3.89	9.78
t-value	-	0.01	-1.33	-2.07*	-1.56
Low-Scoring					
Sample 1					
Difference	11.10	1.12	-2.84	3.10	8.02
SE Equating (Adj)	16.42	8.93	3.87	8.71	16.19
t-value	0.68	0.13	-0.73	0.36	0.50
Sample 2					
Difference	16.57	6.87	3.19	8.79	3.09
SE Equating (Adj)	16.43	8.93	3.87	8.71	16.19
t-value	1.01	0.77	0.82	1.01	0.19

[a]No data are available to establish conversions at these points.

*$p < 0.05$
**$p < 0.01$

Note: Negative values in this table indicate that the sample conversion is lower than the total-population conversion.

Table 17

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Standard Errors Adjusted for the First Four Moments

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Male					
Sample 1					
Difference	1.75	-2.65	-7.06	-11.46	-15.86
SE Equating (Adj)	12.45	8.66	5.35	3.98	6.06
t-value	0.14	-0.31	-1.32	-2.88**	-2.62**
Sample 2					
Difference	-10.08	-9.29	-8.50	-7.71	-6.92
SE Equating (Adj)	12.45	8.66	5.35	3.98	6 06
t-value	-0.81	-1.07	-1.59	-1.94	-1.14
Female					
Sample 1					
Difference	1.98	3.91	5.84	7.77	9.70
SE Equating (Adj)	10.84	7.16	4.43	4.73	7.72
t-value	0.18	0.55	1.32	1.64	1.26
Sample 2					
Difference	-13.78	-11.42	-7.06	-2.70	1.65
SE Equating (Adj)	10.84	7.16	[illegible]43	4 73	7 72
t-value	-1.46	-1.59	-1.59	-0.57	0.21
White					
Sample 1					
Difference	3.96	-1.42	-6.80	-12.18	-17.56
SE Equating (Adj)	12 47	8.45	5.04	4.20	6.94
t-value	0.32	-0.17	-1.35	-2.90**	-2.53*
Minority					
Sample 1					
Difference	3.80	-1.71	-7.23	-12.74	-18.26
SE Equating (Adj)	6 59	4.33	4.70	7.31	10.63
t-value	0.58	-0.40	-1.54	-1.74	-1.72

Table 17 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Standard Errors Adjusted for the First Four Moments

Subpopulation Sample	Form D3 Raw Scores 25	50	75	100	125
Humanities					
Sample 1					
Difference	3.23	3.39	3.54	3.70	3.85
SE Equating (Adj)	11.71	7.97	4.88	4.22	6.76
t-value	0.28	0.42	0.73	0.88	0.57
Sample 2					
Difference	1.28	1.70	2.11	2.52	2.93
SE Equating (Adj)	11.71	7.97	4.88	4.22	6.76
t-value	0.11	0.21	0.43	0.60	0.43
Social Science					
Sample 1					
Difference	7.16	4.76	2.36	-0.04	-2.44
SE Equating (Adj)	10.37	6.84	4.33	4.84	7.80
t-value	0.69	0.70	0.54	-0.01	-0.31
Sample 2					
Difference	-3.98	-1.21	1.55	4.31	7.07
SE Equating (Adj)	10.37	6.84	4.33	4.84	7.80
t-value	-0.38	-0.18	0.36	0.89	0.91
Biological Science					
Sample 1					
Difference	-0.39	-2.05	-3.72	-5.38	-7.05
SE Equating (Adj)	12.80	8.51	4.93	4.40	7.59
t-value	-0.03	-0.24	-0.75	-1.22	-0.93
Sample 2					
Difference	-24.09	-14.98	-5.87	3.24	12.35
SE Equating (Adj)	12.80	8.51	4.93	4.40	7.59
t-value	-1.88	-1.76	-1.19	0.74	1.63

Table 17 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Standard Errors Adjusted for the First Four Moments

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Physical Science					
Sample 1					
Difference	-24.40	-20.08	-15.76	-11.44	-7.12
SE Equating (Adj)	17 90	12.57	7.52	3.96	5.96
t-value	-1.36	-1.60	-2.10*	-2.89**	-1.19
Sample 2					
Difference	-59.22	-44.06	-28.91	-1[illegible].75	1.40
SE Equating (Adj)	17.90	12 57	7.52	3.96	5.96
t-value	-3.31**	-3.50**	-3.84**	-3.47**	0.24
Sample 3					
Difference	-40.63	-29.70	-18.77	-7.84	3.10
SE Equating (Adj)	17.90	12.57	7.52	3.96	5.96
t-value	-2.27*	-2.36*	-2.50*	-1.98*	0.52
Sample 4					
Difference	-25.48	-20.90	-16.33	-11.75	-7.18
SE Equating (Adj)	17.90	12.57	7.52	3.96	5.96
t-value	-1.42	-1.66	-2.17*	-2.97**	-1.20
High-Scoring					
Sample 1					
Difference	6.99	5.12	3.26	1.39	-0.48
SE Equating (Adj)	21 17	5.13	9.26	4.30	5.19
t-value	0.33	0.34	0.35	0.32	-0.09
Sample 2					
Difference	3.99	1.14	-1.70	-4.55	-7.39
SE Equating (Adj)	21.17	15 13	9 26	4.30	5 19
t-value	0.19	0.08	-0.18	-1.06	-1.42

Table 17 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Standard Errors Adjusted for the First Four Moments

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Middle-Scoring					
Sample 1					
Difference	-21.95	-12.53	-3.11	6.32	15.74
SE Equating (Adj)	15.64	10.15	5.35	4.59	8.98
t-value	-1.40	-1.23	-0.58	1.37	1.75
Sample 2					
Difference	-35.08	-23.73	-12.38	-1.03	10.32
SE Equating (Adj)	15.64	10.15	5.35	4.59	8.98
t-value	-2.24*	-2.34*	-2.31*	-0.22	1.15
Low-Scoring					
Sample 1					
Difference	7.26	4.50	1.74	-1.02	-3.78
SE Equating (Adj)	8.79	4.99	4.63	8.18	12.69
t-value	0.83	0.90	0.37	-0.13	-0.30
Sample 2					
Difference	4.92	7.50	10.08	12.67	15.25
SE Equating (Adj)	8.79	4.99	4.63	8.18	12.69
t-value	0.56	1.50	2.18*	1.55	1.20

*$p < 0.05$
**$p < 0.01$

Note: Negative values in this table indicate that the sample conversion is lower than the total-population conversion.

Table 18

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Equipercentile Equating

Standard Errors Adjusted for Mean and Standard Deviation Only

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Male					
Sample 1					
Difference	a	-16.21	-3.02	-14.49	-17.33
SE Equating (Adj)	-	14.14	7.11	2.96	8.63
t-value	-	-1.15	-0.42	-4.89**	-2.01*
Sample 2					
Difference	a	7.33	-8.91	-6.81	5.58
SE Equating (Adj)	-	14.14	7 11	2.96	8.63
t-value	-	0.52	-1.25	-2.30*	0.65
Female					
Sample 1					
Difference	23.20	-0.39	10.00	8.62	-3.97
SE Equating (Adj)	19.15	11.87	4.99	4.37	11.11
t-value	1.21	-0.03	2.00*	1.97*	-0.36
Sample 2					
Difference	a	-12.09	-4.35	0.07	-2.80
SE Equating (Adj)	-	11.87	4 99	4.37	11.11
t-value	-	-1.02	-0.87	0.02	-0.25
White					
Sample 1					
Difference	a	4.62	-8.69	-13.25	-8.00
SE Equating (Adj)	-	13.96	6 50	3.46	10.19
t-value	-	0.33	-1.34	-3.83**	-0.79
Minority					
Sample 1					
Difference	1.79	-1.04	-1.65	-18.40	-38.50
SE Equating (Adj)	12.87	6.52	2.96	8.04	14.49
t-value	0.14	-0.16	-0.56	-2.29*	-2.66**

Table 18 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Equipercentile Equating

Standard Errors Adjusted for Mean and Standard Deviation Only

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Humanities					
Sample 1					
Difference	a	2.26	3.26	3.45	-13.33
SE Equating (Adj)	-	13 07	6.11	3 44	9.72
t-value	-	0.17	0.53	1.00	-1.37
Sample 2					
Difference	a	0.54	2.28	2.00	0.76
SE Equating (Adj)	-	13.07	6.11	3 44	9.72
t-value	-	0.04	0.37	0.58	0.08
Social Science					
Sample 1					
Difference	15.31	15.06	0.74	1.27	22.50
SE Equating (Adj)	18.40	11 31	4.68	4.52	11.11
t-value	0.83	1.33	0.16	0.28	2.03*
Sample 2					
Difference	16.35	4.50	2.91	5.84	1.93
SE Equating (Adj)	18 40	11.31	4.68	4.52	11.11
t-value	0.89	0.40	0.62	1.29	0.17
Biological Science					
Sample 1					
Difference	a	3.27	-4.15	-5.23	-23.28
SE Equating (Adj)	-	14.24	6.33	3.89	11.32
t-value	-	0.23	-0.66	-1.34	-2.06*
Sample 2					
Difference	a	-26.55	-9.84	6.49	2.90
SE Equating (Adj)	-	14.24	6.33	3.89	11.32
t-value	-	-1.86	-1.56	1.67	0.26

Table 18 (continued)

Differences between Sample Conversions and Total-Population Conversions, Standard Errors of Equating, and Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Equipercentile Equating

Standard Errors Adjusted for Mean and Standard Deviation Only

Subpopulation Sample	Form D3 Raw Scores				
	25	50	75	100	125
Physical Science					
Sample 1					
Difference	a	-9.11	-20.55	-11.33	2.80
SE Equating (Adj)	-	18.79	10 37	3.16	7.76
t-value	-	-0.49	-1.98*	-3.58**	0.36
Sample 2					
Difference	a	-23.24	-31.73	-15.52	5.43
SE Equating (Adj)	-	18.79	10.37	3.16	7.76
t-value	-	-1.24	-3.06**	-4.91**	0.70
Sample 3					
Difference	a	-18.09	-24.81	-9.25	1.22
SE Equating (Adj)	-	18.79	10.37	3 16	7.76
t-value	-	-0.96	-2.39*	-2.93**	0.16
Sample 4					
Difference	a	-15.78	-23.77	-12.03	0.03
SE Equating (Adj)	-	18 79	10 37	3 16	7.76
t-value	-	-0.84	-2.29*	-3.81**	0.00
High-Scoring					
Sample 1					
Difference	a	a	8.78	1.77	0.22
SE Equating (Adj)	-	-	14.11	4.00	8.68
t-value	-	-	0.62	0.44	0.03
Sample 2					
Difference	a	a	9.39	-6.04	-1.85
SE Equating (Adj)	-	-	14.11	4.00	8 68
t-value	-	-	0.67	-1.51	-0.21

Table 18 (continued)

Differences between Sample Conversions and Total-Population Conversions,
Standard Errors of Equating, and
Standardized Differences (t-values) at Selected Raw Scores on Form D3

GRE Verbal-Plus-Quantitative Scores; Equipercentile Equating

Standard Errors Adjusted for Mean and Standard Deviation Only

	Form D3 Raw Scores				
Subpopulation Sample	25	50	75	100	125
Middle-Scoring					
Sample 1					
Difference	a	-55.91	-1.64	7.29	-14.66
SE Equating (Adj)	-	17.37	7.36	4.55	14.18
t-value	-	-3.22**	-0.22	1.60	-1.03
Sample 2					
Difference	a	-24.76	-11.75	-2.89	a
SE Equating (Adj)	-	17.37	7.36	4.55	-
t-value	-	-1.43	-1.60	-0.64	-
Low-Scoring					
Sample 1					
Difference	16.43	5.80	4.82	-8.08	a
SE Equating (Adj)	17.00	8.42	2.99	10.15	-
t-value	0.97	0.69	1.61	-0.80	-
Sample 2					
Difference	22.17	4.89	8.83	21.90	a
SE Equating (Adj)	17.00	8.42	2.99	10.15	-
t-value	1.30	0.58	2.96**	2.16*	-

[a]No data are available to establish conversions at these points.

*$p < 0.05$
**$p < 0.01$

Note: Negative values in this table indicate that the sample conversion is lower than the total-population conversion.

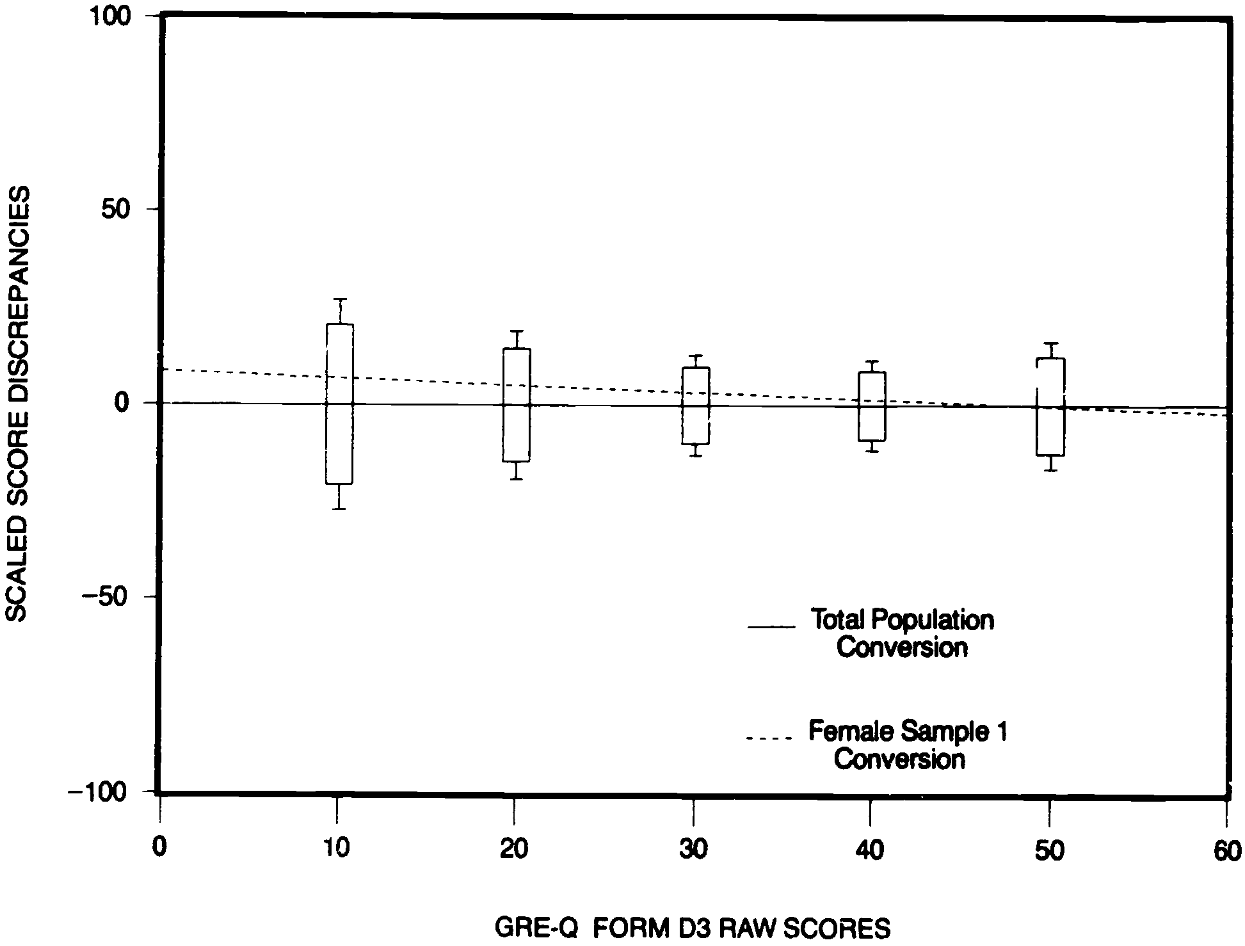

Figure 1. Illustration of Good Agreement between Subpopulation Sample Conversion and Total Population Conversion

GRE-Quantitative; Linear Equating
Standard Errors Adjusted for First Four Moments

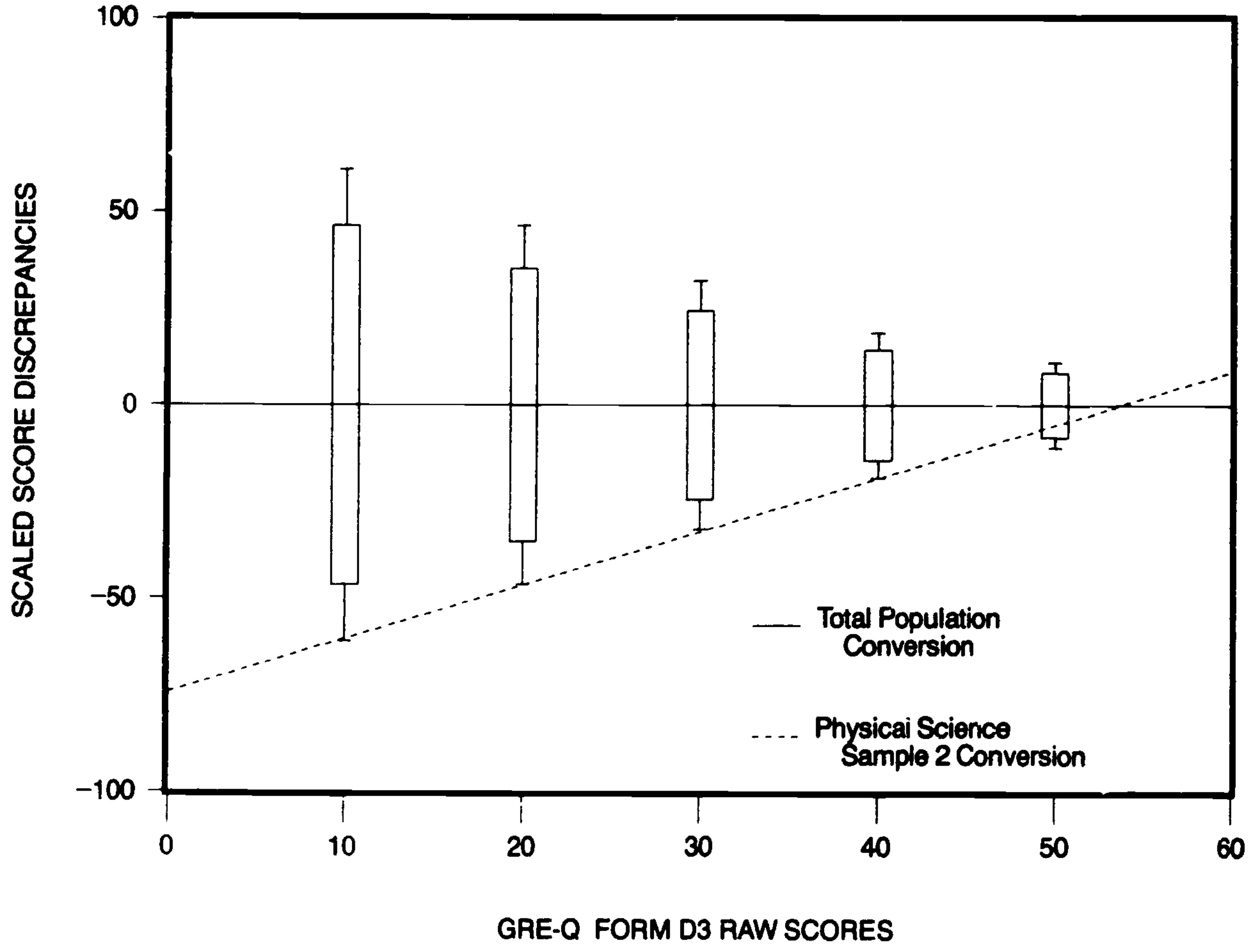

Figure 2. Illustration of Relatively Poor Agreement between Subpopulation Sample Conversion and Total Population Conversion

GRE-Quantitative; Linear Equating
Standard Errors Adjusted for First Four Moments

Science Sample 2 conversion and the base line conversion, also in terms of 1.96 standard errors and 2.58 standard errors, and also where adjustments have been made in the moments of the distribution.

Returning to Table 15 and its equipercentile counterpart, Table 16, the values of t convey a much more favorable picture of population-independency for for equating than was found in Tables 11 and 12, where no adjustments were made in the standard error formulas for the moments of the distributions. The Physical Science group is especially noteworthy here because in this table Sample 2, which yielded the steepest slope and lowest intercept of all the conversions calculated in this study, was the only one that showed consistently significant t-values. Moreover, although the discrepancies between the sample conversions for Physical Science and the conversion for the Total-Population were large, they have virtually no practical significance; the proportions of cases affected by these discrepancies were very small. For example, only 2.0 percent of the Physical Science subpopulation earned raw scores lower than 30 on Form D1; 1.6 percent earned raw scores lower than 30 on Form D3. The conclusion to be drawn from Tables 15 and 16 is that, at least for this homogeneously constructed test--and presumably for other homogeneous tests--the assumption of population-independency is unchallenged.

A slightly different picture emerges when we examine the results for the heterogeneous verbal-plus-quantitative test in Tables 17 and 18. These tables show a few scattered significant t-values in various places throughout the tables, most of them without clear pattern. The exception is, again, the Physical Science subpopulation. Sample 1 shows the smallest scaled score differences and t-values and Sample 2, the largest. In general, the values are somewhat larger than those in Tables 15 and 16, which were based on the homogeneous quantitative test. What is of some interest here is that the large significant differences are found consistently only for the Physical Science group, not for the High-Scoring group, whose mean GRE verbal-plus quantitative score is even higher than the Physical Science group's mean. This is especially puzzling in view of the overlap, noted in connection with Tables 11 and 12, between these two high-performing subpopulations. It can only be concluded that because, as observed earlier, Forms D1 and D3 are not as closely parallel as would have been preferred fo a test of the assumption, the findings for the Physical Science samples with respect to their use in the equating of this heterogeneous verbal-plus-quantitative it may well be a consequence of the lack of parallelism between the two forms of the test.

Incidental findings, ancillary to the central purpose of this study, have led to the following conclusions:

1. Linear equating of two parallel forms of a test, using random groups and no anchor test, yields smaller standard errors than equipercentile equating. This is true at all points along the scale of scores except in the vicinity of the mean, where the errors are very nearly the same. The disparity between the two methods of treating the data increases sharply with increasing distances from the mean.

2. The method of smoothing data derived from the negative hypergeometric distribution (Keats & Lord, 1962) yielded smoothed frequencies that departed considerably from the observed frequencies for long runs of data. It was therefore rejected as a method of choice for smoothing these data in favor of a method developed by Cureton and Tukey (1951). (See also Angoff, 1984, page 12.) The Cureton-Tukey method makes use of a weighted rolling average of frequencies that neighbor the frequency of interest, with

smaller weights for more distant frequencies. From their experience in this study, the authors recommend that the results of any smoothing method be examined closely before applying it. It is entirely likely that procedures like the Cureton-Tukey smoothing will be adequate for most equipercentile equating.

3. Samples should be chosen for equating that yield as nearly as possible rectangular distributions of scores, that is, with approximately equal numbers of cases at each point along the scale. This recommendation runs counter to the way in which equating samples are typically drawn, samples that yield bell-shaped distributions in which the mean of the distribution and the preponderance of cases fall at about the same point, and the frequencies decline in number in the direction of the extremes. Because the type of sampling recommended here (to achieve rectangular distributions) will reduce kurtosis, it may be inferred from equations (10) and (11) that it will also reduce the size of the coefficient of $z^2(x)$. If, in addition, efforts are made to select samples that yield symmetrical distributions, the size of the coefficient of $z(x)$ will also be reduced. Both of these actions will have the effect of enhancing the precision of equating at the extremes of the scale where equating errors are likely to be very large.

Summary and Conclusions

In this study, two general populations of students, one taking Form D1, the other taking Form D3 of the GRE General Test at the regular October 1981 administration, were used to equate the two forms of the homogeneous GRE quantitative test and also the specially constituted heterogenous GRE verbal-plus-quantitative test. Conversions developed from samples drawn from homogeneously defined subgroups of the total population revealed that the conversion for GRE quantitative scores, derived from the total population, applied reasonably well to its various component subgroups. However, the conversion for GRE verbal-plus-quantitative scores, derived on the basis of total-population data, is in some doubt; it applies well for every one of the subgroups except for the Physical Science group. This being the case, it can only be concluded that the assumption of population-independence for homogeneous parallel tests is clearly supported by these results, but because the forms used in this study were not very closely parallel, support for the corresponding assumption for heterogeneous tests is less clearly established.

It is difficult on logical grounds for the authors to accept the conclusion that the conversion line for parallel heterogeneous forms is dependent on the nature of the population on which the equating is based when the conversions for its component homogeneous tests are independent of the nature of that population. It would seem that a score of 100, say, on one form of a heterogeneous test, however that score was reached--by scoring 80 on verbal and 20 on quantitative for example, or 20 on verbal and 80 on quantitative--would have to convert to the same score on a parallel form of that heterogeneous test, whether the members of the population on which the equating is based are verbally inclined or whether they are quantitatively inclined. The key assumptions in coming to this conclusion are (1) that the two heterogeneous forms contain homogeneous component subtests that are parallel, at least approximately so, and (2) that the group taking one form and the group taking the other form are randomly drawn from the same population. It would seem to the authors then that, unlike the "equating" of nonparallel tests and unlike the use of nonrandom groups for equating, there would be no opportunity in properly designed and executed equating studies for different patterns of response made by different populations to affect the equating adversely. It would follow that the results of equating, even the equating of scores on heterogeneous, but parallel, forms of a test, should therefore be unique and generalizable to all populations.

In this connection it is noted that in Lord's (1982) discussion of equating, he refers to Braun and Holland's (1982, page 15) definition, "Form X and Form Y are equated on [population] P if the distribution of the transformed [x] scores in population P is the same as the distribution of the (untransformed) [y] scores," as "beyond reproach." For the general case, where, for example, Form x and Form y are not necessarily parallel, this definition is fully supported; and Lord provides an example of two tests of sharply different difficulty for whom an identity transformation, clearly misrepresenting the difference in difficulty between the forms, will result if the population used for equating guesses at random on almost all the items of the two forms. The question addressed by this study is whether the Braun-Holland definition is not too general, whether the equating of two parallel forms--again, even forms that are only approximately parallel and reasonably appropriate for the samples used to equate them--is not, allowing for random variation, sufficiently unique and general and independent, for most purposes, of the type of population used for equating. The results of this study indicate that at least for homogeneous parallel tests an equating function, properly developed, is in fact population-independent. As indicated elsewhere in this report, the assumption of population-independence is less clearly established here for heterogeneous tests, very likely because of the lack of complete parallelism of the two heterogeneous forms used in this study.

APPENDIX

Table I

Conversion Parameters and Conversions
from Selected Form D3 Raw Scores to the Form D1 Reference Scale,
Based on the Total Population, Subpopulations, and Subpopulation Samples

GRE Quantitative Scores; Linear Equating

	Parameters		Converted Scores Corresponding to Form D3 Raw Score				
Group	Slope	Intercept	10	20	30	40	50
Total Population	10.2726	103.3360	206	309	412	514	617
Male							
Total Subpop'n	10.2681	101.0457	204	306	409	512	614
Sample 1	10.1123	106.0891	207	308	409	511	612
Sample 2	10.1832	101.9240	204	306	407	509	611
Female							
Total Subpop'n	10.3356	101.6045	205	308	412	515	618
Sample 1	10.0937	111.9021	213	314	415	516	617
Sample 2	10.4814	85.7127	191	295	400	505	610
White							
Total Subpop'n	10.1384	109 5728	211	312	414	515	616
Sample 1	9.6533	121.7260	218	315	411	508	604
Minority							
Total Subpop'n	10.5720	89.1560	195	301	406	512	618
Sample 1	10.5695	90.5590	196	302	408	513	619
Humanities							
Total Subpop'n	10.8483	78.0707	187	295	404	512	620
Sample 1	10.6053	88.4537	195	301	407	513	619
Sample 2	10.6419	84.8739	191	298	404	511	617
Social Science							
Total Subpop'n	10.2348	106.7015	209	311	414	516	618
Sample 1	9.9245	115.8739	215	314	414	513	612
sample 2	10.2879	103.6146	206	309	412	515	618
Biological Science							
Total Subpop'n	10.3010	99.8530	203	306	409	512	615
Sample	9.8687	117.6855	216	315	414	512	611
Sample 2	10.7496	81.5246	189	297	404	512	619

Table I (continued)

Conversion Parameters and Conversions
from Selected Form D3 Raw Scores to the Form D1 Reference Scale,
Based on the Total Population, Subpopulations, and Subpopulation Samples

GRE Quantitative Scores; Linear Equating

Group	Parameters		Converted Scores Corresponding to Form D3 Raw Score				
	Slope	Intercept	10	20	30	40	50
Physical Science							
Total Subpop'n	10.9418	64.1202	174	283	392	502	611
Sample 1	10.3703	94.5100	198	302	406	509	613
Sample 2	11.6538	29.1306	146	262	379	495	612
Sample 3	11.2153	52.8106	165	277	389	501	614
Sample 4	10.8073	72.4920	181	289	397	505	613
High-Scoring							
Total Subpop'n	9.9475	120.0987	220	319	419	518	617
Sample 1	9.7110	132.5108	230	327	424	521	618
Sample 2	9.9730	113.7201	213	313	413	513	612
Middle-Scoring							
Total Subpop'n	10.1121	107.7814	209	310	411	512	613
Sample 1	10.5538	92.9565	198	304	410	515	621
Sample 2	9.9459	107.8810	207	307	406	506	605
Low-Scoring							
Total Subpop'n	10.7956	90.2983	198	306	413	521	629
Sample 1	10.6765	91.4856	198	305	412	519	625
Sample 2	10.7455	94.1957	202	309	417	524	631

Table II

Conversions from Selecte Form D3 Raw Scores
to the Form D1 Reference Scale,
Based on the Total Population, Subpopulations, and Subpopulation Samples

GRE Quantitative Scores; Equipercentile Equating

Group	Converted Scores Corresponding to Form D3 Raw Score				
	10	20	30	40	50
Total Population	189	300	416	514	615
Male					
Total Subpop'n	159	297	415	51[illegible]	612
Sample 1	200	277	421	511	608
Sample 2	230	298	410	509	613
Female					
Total Subpop'n	193	301	416	514	618
Sample 1	204	298	421	512	619
Sample 2	163	286	408	504	605
White					
Total Subpop'n	166	304	417	515	615
Sample 1	a	301	417	507	603
Minority					
Total Subpop'n	193	295	413	503	615
Sample 1	194	297	415	504	614
Humanities					
Total Subpop'n	a	282	410	511	619
Sample 1	a	300	414	510	625
Sample 2	a	295	412	510	615
Social Science					
Total Subpop'n	200	302	418	516	616
Sample 1	a	308	422	508	611
Sample 2	201	295	415	518	611
Biological Science					
Total Subpop'n	a	302	412	512	614
Sample 1	a	323	409	513	610
Sample 2	a	275	410	510	616

Table II (continued)

Conversions from Selected Form D3 Raw Scores
to the Form D1 Reference Scale,
Based on the Total Population, Subpopulations, and Subpopulation Samples

GRE Quantitative Scores; Equipercentile Equating

Group	Converted Scores Corresponding to Form D3 Raw Score				
	10	20	30	40	50
Physical Science					
Total Subpop'n	a	293	400	498	610
Sample 1	a	321	414	504	613
Sample 2	a	255	380	492	613
Sample 3	a	271	389	498	614
Sample 4	a	249	391	506	613
High-Scoring					
Total Subpop'n	a	a	425	518	617
Sample 1	a	a	436	522	618
Sample 2	a	a	421	514	608
Middle-Scoring					
Total Subpop'n	a	297	413	512	612
Sample 1	a	272	406	517	619
Sample 2	a	301	404	506	600
Low-Scoring					
Total Subpop'n	189	301	418	514	621
Sample 1	200	301	413	517	623
Sample 2	206	307	419	523	618

[a] No data are available at these points to establish conversions.

Table III

Conversion Parameters and Conversions from Selected Form D3 Raw Scores to the Form D1 Reference Scale, Based on the Total Population, Subpopulations, and Subpopulation Samples

GRE Verbal-Plus-Quantitative Scores; Linear Equating

	Parameters		Converted Scores Corresponding to Form D3 Raw Score				
Group	Slope	Intercept	25	50	75	100	125
Total Population	5.3481	47.1093	181	315	448	582	716
Male							
Total Subpop'n	5.3306	42.1521	175	309	442	575	708
Sample 1	5.1720	53.2613	183	312	441	570	700
Sample 2	5.3797	36.2442	171	305	440	574	709
Female							
Total Subpop'n	5.4489	43.0848	179	316	452	588	724
Sample 1	5.4253	47.1570	183	318	454	590	725
Sample 2	5.5224	26.9756	165	303	441	579	717
White							
Total Subpop'n	5.3197	50.6382	184	317	450	583	716
Sample 1	5.1329	56.4541	185	313	441	570	698
Minority							
Total Subpop'n	5.1444	53.9863	183	311	440	568	697
Sample 1	5.1275	56.4248	185	313	441	569	697
Humanities							
Total Subpop'n	5.4335	42.6379	178	314	450	586	722
Sample 1	5.3543	50.1868	184	318	452	586	719
Sample 2	5.3646	47.9815	182	316	450	584	719
Social Science							
Total Subpop'n	5.3773	47.9108	182	317	451	586	720
Sample 1	5.2521	56.6652	188	319	451	582	713
Sample 2	5.4586	40.3703	177	313	450	586	723
Biological Science							
Total Subpop'n	5.4964	33.1468	171	308	445	583	720
Sample 1	5.2815	48.3868	180	312	444	577	709
Sample 2	5.7125	13.9108	157	300	442	585	728

Table III (continued)

Conversion Parameters and Conversions
from Selected Form D3 Raw Scores to the Form D1 Reference Scale,
Based on the Total Population, Subpopulations, and Subpopulation Samples

GRE Verbal-Plus-Quantitative Scores; Linear Equating

Group	Parameters		Converted Scores Corresponding to Form D3 Raw Score				
	Slope	Intercept	25	50	75	100	125
Physical Science							
Total Subpop'n	5.7117	-2.2502	141	283	426	569	712
Sample 1	5.5209	18.3923	156	294	432	570	709
Sample 2	5.9543	-27.2611	122	270	419	568	717
Sample 3	5.7854	-4.4566	140	285	429	574	719
Sample 4	5.5311	17.0565	155		432	570	708
High-Scoring							
Total Subpop'n	5.3295	47.9151	181	314	448	581	714
Sample 1	5.2734	55.9681	188	320	451	583	715
Sample 2	5.2343	53.9418	185	316	447	577	708
Middle-Scoring							
Total Subpop'n	5.5707	25.7150	165	304	444	583	722
Sample 1	5.7250	15.7345	159	302	445	588	731
Sample 2	5.8021	0.6808	146	291	436	581	726
Low-Scoring							
Total Subpop'n	5.5541	38.4494	177	316	455	594	733
Sample 1	5.2377	57.1245	188	319	450	581	712
Sample 2	5.4514	49.4465	186	322	458	595	731

Table IV

Conversions from Selected Form D3 Raw Scores
to the Form D1 Reference Scale,
Based on the Total Population, Subpopulations, and Subpopulation Samples

GRE Verbal-Plus-Quantitative Scores; Equipercentile equating

	Converted Scores Corresponding to Form D3 Raw Score				
Group	25	50	75	100	125
Total Population	181	314	448	582	712
Male					
Total Subpop'n	163	310	442	574	712
Sample 1	a	298	445	568	695
Sample 2	a	321	439	575	718
Female					
Total Subpop'n	188	316	451	589	713
Sample 1	204	313	458	591	708
Sample 2	a	302	444	582	709
White					
Total Subpop'n	a	317	449	583	713
Sample 1	a	318	439	569	704
Minority					
Total Subpop'n	183	311	445	563	673
Sample 1	183	313	446	564	674
Humanities					
Total Subpop'n	a	304	450	586	705
Sample 1	a	316	451	586	699
Sample 2	a	314	450	584	713
Social Science					
Total Subpop'n	191	317	452	586	714
Sample 1	197	329	449	583	735
Sample 2	198	318	451	588	714
Biological Science					
Total Subpop'n	a	305	444	584	711
Sample 1	a	317	444	577	689
Sample 2	a	287	438	589	715

Table IV (continued)

Conversions from Selected Form D3 Raw Scores
to the Form D1 Reference Scale,
Based on the Total Population, Subpopulations, and Subpopulation Samples

GRE Verbal-Plus-Quantitative Scores; Equipercentile equating

	Converted Scores Corresponding to Form D3 Raw Score				
Group	25	50	75	100	125
Physical Science					
Total Subpop'n	170	301	419	568	713
Sample 1	a	305	427	571	715
Sample 2	a	291	416	567	718
Sample 3	a	296	423	573	713
Sample 4	a	298	424	570	712
High-Scoring					
Total Subpop'n	a	a	450	581	711
Sample 1	a	a	457	584	712
Sample 2	a	a	457	576	710
Middle-Scoring					
Total Subpop'n	a	299	443	583	734
Sample 1	a	258	446	589	697
Sample 2	a	289	436	579	a
Low-Scoring					
Total Subpop'n	181	315	454	589	a
Sample 1	198	320	453	574	a
Sample 2	203	319	457	604	a

[a]No data are available at these points to establish conversions.

References

Angoff, W. H. (1966). Can useful general-purpose equivalency tables be prepared for college admissions tests? In A. Anastasi (Ed.), Testing problems in perspective (pp. 251-264). Washington, DC: American Council on Education.

Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service. Also in R. L. Thorndike (Ed.), (1971), Educational Measurement (pp. 508-600). Washington, DC: American Council on Education.

Braun, H. I., and Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland and D. B. Rubin (Eds.), Test equating (pp. 9-49). New York: Academic Press.

Cureton, E. E., and Tukey, J. W. (1951). Smoothing frequency distributions, equating tests, and preparing norms. American Psycholcgist, 6, 404. (Abstract)

Holland, P. W., and Wightman, L. E. (1982). Section pre-equating: A preliminary investigation. In P. W. Holland and D. B. Rubin (Eds.), Test equating (pp. 271-297). New York: Academic Press.

Keats, J. A., and Lord, F. M. (1962). A theoretical distribution for mental test scores. Psychometrika, 7, 59-72.

Lord, F. M. (1950). Notes on comparable scales for test scores (Research Bulletin 50-48). Princeton, NJ: Educational Testing Service.

Lord, F. M. (1982). The standard error of equipercentile equating. Journal of Educational Statistics, 7, 165-174.