ACT Research Report Series 89-3

Application of Equipercentile Techniques to Test Scale Construction: Scaling and Equating of the ACT ASSET Placement Test Battery

Ronald T. Cope

June 1989

ACT

APPLICATION OF EQUIPERCENTILE TECHNIQUES TO TEST SCALE CONSTRUCTION: SCALING AND EQUATING OF THE ACT ASSET PLACEMENT TEST BATTERY

Ronald T. Cope

ABSTRACT

Six tests of the ASSET placement test program were scaled by adapting an equipercentile technique usually restricted to equating different forms of a single test. The purpose of applying this technique was to transform raw scores into scale scores having for all tests approximately the same percent of examinees of a representative weighted sample of ASSET user institutions scoring at or below given scale scores. Scale scores with such normative interpretations were intended to aid test users in changing from the use of the original Forms A to the new Forms B and C of the Basic Skills tests and Forms B of the Advanced Mathematics tests. Obtained percents of examinees at or below given scale scores showed close agreement across the six scaled tests of the ASSET battery. Scale score means and standard deviations were nearly equal across tests, with means about 40.5 and standard deviations about 6.2. Subsequent equating of basic skills Forms C to Forms B allowed conversion of raw scores on Forms C to scale scores.

APPLICATION OF EQUIPERCENTILE TECHNIQUES TO TEST SCALE CONSTRUCTION: SCALING AND EQUATING OF THE ACT ASSET PLACEMENT TEST BATTERY

Introduction

Raw Scores Versus Scale Scores and Equating

Equating different forms of a test and converting raw scores to scale scores have been increasingly common practices in educational testing. Equating and scaling now enjoy widespread use because they convey advantages over using only raw or number-right scores for reporting test results. If only one form of a test exists, reporting results in terms of raw scores may suffice. But after additional forms have been created, use of only raw scores creates problems of difficult interpretation and misinterpretation.

First, consider the advantages of equating. Creating multiple forms of a test with equal difficulty throughout the range of raw scores remains an extremely difficult, if not impossible, task. Without adjustment for these differences of difficulty, a student who happens to receive a hard form of a test is unfairly disadvantaged. Equating is a statistical procedure that adjusts for these inevitable differences. Equating allows scores from different forms of a test to be used interchangeably. One way to report the results of equating Forms A and B of a test would be to construct tables that convert Form B raw scores to Form A raw scores. Then both converted Form B raw scores and Form A raw scores could be treated as Form A raw scores. Placing all scores on the Form A raw score scale would allow direct comparison of groups taking the forms.

Converting raw scores to scale scores further enhances interpretation. With scaling, raw scores, which have no intrinsic normative meaning, can be transformed into a scale having easy-to-remember reference points with convenient normative interpretations. Scale scores avoid problems such as

having to explain to a student who took an easy form y of a test why points were subtracted from his or her raw score to convert them to raw scores of hard form x. Scale scores convey meaning abstracted from the particulars of individual forms' raw scores.

Test Equating: Equivalent Groups Design, Equipercentile Method

Procedures for equating different forms of a test consist of (1) a design for collecting data, and (2) a method of using the data to produce tables or formulas or charts giving raw-to-raw or raw-to-scale score conversions. (See Petersen, Kolen, and Hoover, 1989, and Angoff, 1984, for general discussions of test equating). The **equivalent-groups design** uses the test administration procedure known as **spiraling:** If, for example, two forms A and B are to be equated, they are packaged in the order ABABAB . . . so that the first examinee receives Form A; the second, Form B; the third, Form A; and so on. Within the limits of random sampling error, the groups taking the forms are equivalent. Consequently, statistical differences of scores obtained this way can be attributed to form differences.

One widely-used method for using the obtained data to complete the equating process associates raw scores of the same percentile rank. This **equipercentile method** is illustrated in Figure 1. Because of sampling error, the lines of the graph will be bumpy. To lessen the effect of sampling error and to better estimate the true relationship, the line of raw score equivalents can be smoothed. Alternatively, the score distributions can be smoothed before the graph is constructed.

One Limitation of Equating

Forms of tests admit of equating when they are built to identical content specifications and have about the same level of difficulty. Tests that

measure different skills, however, cannot be equated. Accordingly, scores on tests measuring different skills can be compared only on a normative basis--that is, in relation to particular groups of examinees. For example, a student might have scores on tests of English skills and algebra with percentile ranks of 50 among all U.S. twelfth graders. But the student could not meaningfully be described as "equally good at English and algebra." This distinction is important for the application of equating techniques to scaling.

Adaptation of the Equipercentile Method to Scaling Different Tests

Although equating must be restricted to alternate forms of a single test, the statistical technique of relating raw scores of equal percentile rank was applied in this study to the creation of the scale of a battery of tests. The purpose of such an application is to impart similar normative meanings to scale scores of the different tests. Scale scores with this property facilitate user interpretation. Here the technique was applied to both the scaling and the subsequent equating of the Basic Skills and Advanced Mathematics tests of the ACT ASSET battery of collegiate placement tests: Language Skills, Reading Skills, Numerical Skills, Elementary Algebra, Intermediate Algebra, and College Algebra. (The Advanced Mathematics tests include a Geometry test, but its scaling awaits procurement of a sample of students in collegiate geometry classes.)

Method

Subjects

The samples for Writing Skills, Reading Skills, and Numerical Skills tests consisted of examinees in a weighted representative sample of ASSET user institutions. Subjects taking Forms A and B of the Writing Skills test

numbered respectively 2,469 and 2,463; Reading Skills test, 2,464 and 2,454; Numerical Skills test, 2,445 and 2,402. Groups of students from appropriate college classes served as subjects for the Advanced Mathematics tests. Respective numbers of subjects taking Elementary Algebra, Intermediate Algebra, and College Algebra were 773, 413, and 682.

Measures

The ASSET battery consists of the following tests, all of which have a time limit of 25 minutes:

Test	No. of Items	New Form(s)
Basic Skills:		
Writing Skills	36	B and C
Reading Skills	24	B and C
Numerical Skills	32	B and C
Advanced Mathematics:		
Elementary Algebra	25	B
Intermediate Algebra	25	B
College Algebra	25	B
Geometry	25	B

Procedure

The score scale for new Forms B and C was constructed to meet the following standards: The scale should be unique and not easily confused with other score scales, it should have an inherent normative interpretation, it should contain an appropriate number of score points, and it should have a suitable central reference point. A scale of 40 was chosen as the central reference point such that 50 percent of the smoothed weighted distributions of examinees would fall at or below this score. Smoothing was accomplished by

fitting the four-parameter-beta compound binomial distribution (Lord, 1965) to the weighted distributions. Weighting of the raw score distributions was intended to correct for the underrepresentation of schools in the stratified (region) sample. The sample had been stratified by region, and a given examinee's probability of selection within region was proportional to school size.

Form B of the ASSET Numerical Skills test was selected as the starting point for the score scale. Distributions of scores on Forms B and C of Numerical Skills were more nearly symmetrical than those of scores on the Writing Skills and Reading Skills tests. The latter two tests were easier for the norming sample than Numerical Skills and had negatively skewed distributions. Distributions of scores on Elementary Algebra, Intermediate Algebra, and College Algebra were closer in shape to the Numerical Skills distributions. However, none of the algebra tests could be considered as starting points for the scaling because scaling and equating of the Basic Skills tests had to be completed before scores on the Advanced Mathematics tests became available.

Two considerations motivated the choice of the most nearly symmetrical usable distribution as the starting point: (1) The negative skewness of the Writing Skills and Reading Skills score distributions appeared to result from the relatively low ceilings of these tests; in light of the nearly symmetrical score distributions of the more difficult Numerical Skills test, there was no strong indication that the skills measured by the other two tests were inherently skewed in the samples studied, (2) Raw and scale scores of complementary percents of examinees scoring at or below would be approximately the same distance from the central reference score; for example, the raw and scale scores having 75% and 25% at or below would be approximately the same

distance, respectively, above and below the 50% at or below central reference raw and scale scores. Imparting this property to the score scale was intended to further enhance interpretability.

Fifty percent of scores of the Numerical Skills Form B smoothed weighted distribution were found to equal or exceed a raw score of 17; therefore, this raw score was set to yield a scale score of 40. Raw scores above and below 17 were tentatively mapped onto scale scores on a point-for-point basis: a raw score of 16 yielded a scale score of 39, a raw score of 18 yielded a scale score of 41, and so on. Consequently, the 0-32 raw score scale yielded a score scale of 23-55, and the conversion formula was $S_{NB} = N_B + 23$, where S_{NB} is scaled score and N_B is Numerical Skills Form B raw score.

The next step consisted of seeing whether score values such as 30 and 50 had suitable normative values. Approximately 95% of Numerical Skills Form B scale scores fell at or below 50, 75% fell at or below 45, 25% fell at or below 35, and 5% fell at or below 30. These numbers being convenient and easily-remembered, the conversions were allowed to stand.

Next, raw scores on Form B of the remaining tests were converted to scale scores. First raw scores of these tests were related by the equipercentile method to Numerical Skills Form B raw scores. Then the raw-to-scale score conversions for Numerical Skills Form B were used to derive the raw-to-scale score conversions for Form B of the remaining tests. Figure 2 gives a schematic representation of the scaling process.

Finally, for each of the three Basic Skills tests, Form C was equated to Form B by the equipercentile method. The equating produced conversions of Form C raw scores to scale scores. Figure 3 depicts the equating.

Results and Discussion

Table 1 gives number of examinees, moments of raw and scale scores, KR20 internal consistency estimates, and scale-score standard errors of measurement for each of the scaled ASSET tests. Although the raw score distributions of the Writing Skills and Reading Skills tests have marked skewness, scale scores have, as a result of the scaling technique, little skewness and nearly equal moments.

Tables 2 and 3 give conversions of Forms B and C Basic Skills tests raw-to-scale score. The marked skewness of raw score distributions of Writing Skills and Reading Skills caused squeezing and stretching of the scale scores: Low scores are compressed in the sense that often two consecutive raw scores convert to the same scale score and stretched in the sense that consecutive high scores sometimes convert to nonconsecutive scale scores (for example, Writing Skills Form B raw scores of 9 and 10 both convert to a scale score of 28, and raw scores of 34 and 35 convert to respective scale scores of 50 and 52). The easiness of these two tests in relation to Numerical Skills shows in perfect raw scores' scaling to only 53 or 54 instead of the maximum possible 55. In contrast, Numerical Skills Form C scale scores have only one duplication and one skip and, like Numerical Skills Form B, scale to 55.

Table 4 presents raw-to-scale conversions for the scaled Advanced Mathematics tests. The greater difficulty of these tests than that of the Basic Skills tests is most clearly evident at the top of the scales: For example, the top three raw scores of Elementary Algebra all convert to the maximum possible scale score 55. Little if any squeezing or stretching is evident, but some skips result from the relatively low number of items--25 for all Advanced Mathematics tests versus 32 for Numerical Skills.

Table 5 (for Basic Skills tests) and Table 6 (for Advanced Mathematics tests) present percentages of examinees at or below given scale scores. Percentages are similar across tests for given scale scores, and vary by no more than 5 points. Note that these tables are based upon actual--not smoothed--weighted score distributions. Percentages for skipped scale scores were obtained by interpolation.

As time passes, representative groups of ASSET users may change in overall performance and in their profiles of relative performance on the different tests; if this happens, the percents at or below given scale scores will change. By then, however, users will be more familiar with the score scale and the performance of students in their own institutions. They will accordingly be less dependent upon percents of the scaling sample scoring at or below for interpretation and will have acquired a sense of the levels of performance represented by different scale scores.

REFERENCES

Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service. (Reprint of chapter in R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, DC: American Council on Education, 1971.)

Lord, F. M. (1965). A strong true-score theory, with applications. Psychometrika, 30, 239-270.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In Linn, R. L. (Ed.), Educational measurement (3rd ed.). (pp. 221-262). Washington, DC: American Council on Education.

TABLE 1

Moments, Reliabilities, and Standard Errors of Measurement of ASSET Tests

Test/Form	N	Raw Scores				Scale Scores				KR20	Scale Score Standard Error of Measurement
		Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis		
Writ. Sk. Form B	2,469	25.19	6.54	-0.59	2.77	40.49	6.21	-0.03	2.32	.87	2.24
Writ. Sk. Form C	2,463	24.65	7.12	-0.59	2.60	40.49	6.21	-0.03	2.31	.87	2.24
Rdng. Sk. Form B	2,464	16.32	4.25	-0.40	2.61	40.49	6.20	-0.03	2.30	.78	2.91
Rdng. Sk. Form C	2,454	16.80	4.42	-0.56	2.69	40.49	6.18	-0.04	2.30	.78	2.90
Num. Sk. Form B	2,445	17.49	6.21	-0.02	2.32	40.49	6.21	-0.02	2.32	.86	2.32
Num. Sk. Form C	2,402	18.05	6.33	-0.10	2.27	40.48	6.21	-0.01	2.32	.85	2.41
Elementary Alg.	773	11.51	3.77	0.43	2.94	40.49	6.19	-0.01	2.33	.66	3.61
Intermediate Alg.	827	12.25	4.29	0.04	2.55	40.49	6.20	-0.02	2.33	.72	3.28
College Algebra	682	12.73	3.92	0.16	2.56	40.49	6.19	-0.01	2.32	.68	3.50

TABLE 2

Raw Scores of ASSET Basic Skills Tests Converted to Scale Scores

Form B

Writing Skills				Reading Skills		Numerical Skills			
Raw Score	Scale Score	Raw Score	Scale Score	Raw Score	Scale Score	Raw Score	Scale Score	Raw Score	Scale Score
0	23	20	35	0	23	0	23	20	43
1	23	21	36	1	23	1	24	21	44
2	24	22	37	2	24	2	25	22	45
3	24	23	38	3	25	3	26	23	46
4	25	24	38	4	26	4	27	24	47
5	25	25	39	5	27	5	28	25	48
6	26	26	40	6	28	6	29	26	49
7	27	27	41	7	29	7	30	27	50
8	27	28	42	8	30	8	31	28	51
9	28	29	44	9	31	9	32	29	52
10	28	30	45	10	32	10	33	30	53
11	29	31	46	11	33	11	34	31	54
12	30	32	47	12	34	12	35	32	55
13	30	33	49	13	35	13	36		
14	31	34	50	14	37	14	37		
15	32	35	52	15	38	15	38		
16	32	36	54	16	39	16	39		
17	33			17	41	17	40		
18	34			18	43	18	41		
19	34			19	44	19	42		
				20	46				
				21	48				
				22	49				
				23	51				
				24	53				

TABLE 3

Raw Scores of ASSET Basic Skills Tests Converted to Scale Scores

Form C

Writing Skills				Reading Skills		Numerical Skills			
Raw Score	Scale Score	Raw Score	Scale Score	Raw Score	Scale Score	Raw Score	Scale Score	Raw Score	Scale Score
0	23	20	36	0	23	0	23	20	42
1	23	21	37	1	23	1	24	21	43
2	24	22	37	2	24	2	25	22	44
3	25	23	38	3	25	3	26	23	45
4	25	24	39	4	26	4	27	24	46
5	26	25	40	5	27	5	28	25	47
6	27	26	41	6	28	6	29	26	48
7	27	27	42	7	29	7	30	27	49
8	28	28	43	8	30	8	31	28	50
9	29	29	44	9	31	9	32	29	52
10	29	30	45	10	32	10	33	30	53
11	30	31	46	11	33	11	34	31	54
12	31	32	47	12	34	12	35	32	55
13	31	33	49	13	35	13	36		
14	32	34	50	14	36	14	37		
15	33	35	52	15	37	15	37		
16	33	36	54	16	39	16	38		
17	34			17	40	17	39		
18	35			18	41	18	40		
19	35			19	43	19	41		
				20	45				
				21	46				
				22	48				
				23	51				
				24	53				

TABLE 4

Raw Scores of Advanced Mathematics Tests
Converted to Scale Scores

Elementary Algebra		Intermediate Algebra		College Algebra	
Raw Score	Scale Score	Raw Score	Scale Score	Raw Score	Scale Score
0	23	0	23	0	23
1	23	1	25	1	24
2	24	2	26	2	25
3	26	3	27	3	26
4	27	4	29	4	27
5	29	5	30	5	28
6	31	6	31	6	30
7	32	7	33	7	31
8	34	8	34	8	33
9	36	9	36	9	34
10	38	10	37	10	36
11	40	11	39	11	38
12	42	12	40	12	40
13	44	13	42	13	41
14	45	14	43	14	43
15	47	15	45	15	44
16	48	16	46	16	46
17	49	17	48	17	47
18	50	18	49	18	49
19	51	19	50	19	50
20	52	20	51	20	51
21	53	21	52	21	52
22	54	22	53	22	53
23	55	23	54	23	54
24	55	24	55	24	55
25	55	25	55	25	55

TABLE 5

1989 ASSET Basic Skills Scale Score User Norms (Forms B and C)
Weighted Percent of Students Scoring at or Below Scale Score
(*Interpolated)

Writing Skills		Reading Skills		Numerical Skills	
Scale Score	Cumulative Percent	Scale Score	Cumulative Percent	Scale Score	Cumulative Percent
23	0	23	0	23	0
24	0	24	0	24	0
25	0	25	0	25	0
26	0	26	1	26	0
27	1	27	1	27	1
28	2	28	2	28	2
29	4	29	3	29	3
30	6	30	5	30	6
31	8	31	7	31	8
32	11	32	10	32	11
33	14	33	14	33	15
34	19	34	19	34	18
35	24	35	25	35	22
36	27	36	27	36	26
37	33	37	34	37	35
38	40	38	38	38	40
39	46	39	45	39	45
40	50	40	49	40	50
41	55	41	58	41	56
42	61	42	63*	42	62
43	64	43	67	43	67
44	70	44	71	44	73
45	76	45	75	45	78
46	82	46	84	46	82
47	87	47	87*	47	86
48	90*	48	91	48	90
49	92	49	94	49	93
50	96	50	96*	50	95
51	98*	51	98	51	96
52	99	52	99*	52	98
53	99*	53	100	53	99
54	100			54	100
				55	100

TABLE 6

1989 ASSET Advanced Mathematics Tests Scale Score User Norms
Percent of Students Scoring at or Below Scale Score
(*Interpolated)

Elementary Algebra		Intermediate Algebra		College Algebra	
Scale Score	Percent at or below	Scale Score	Percent at or below	Scale Score	Percent at or below
23	0	23	1	23	0
24	0	24	1	24	0
25	0	25	1	25	0
26	1	26	1	26	0
27	2	27	2	27	1
28	4*	28	3*	28	2
29	5	29	3	29	4*
30	6*	30	5	30	5
31	7	31	9	31	10
32	14	32	12*	32	13*
33	18*	33	14	33	16
34	21	34	21	34	22
35	26*	35	25*	35	27*
36	31	36	29	36	31
37	37*	37	36	37	35*
38	42	38	40*	38	38
39	48*	39	43	39	43*
40	53	40	51	40	48
41	58*	41	56*	41	58
42	62	42	61	42	63*
43	67*	43	70	43	67
44	72	44	73*	44	75
45	80	45	76	45	79*
46	84*	46	82	46	83
47	87	47	85*	47	89
48	91	48	88	48	91*
49	93	49	93	49	93
50	95	50	96	50	95
51	97	51	98	51	97
52	98	52	99	52	98
53	99	53	99	53	99
54	99	54	100	54	100
55	100	55	100	55	100

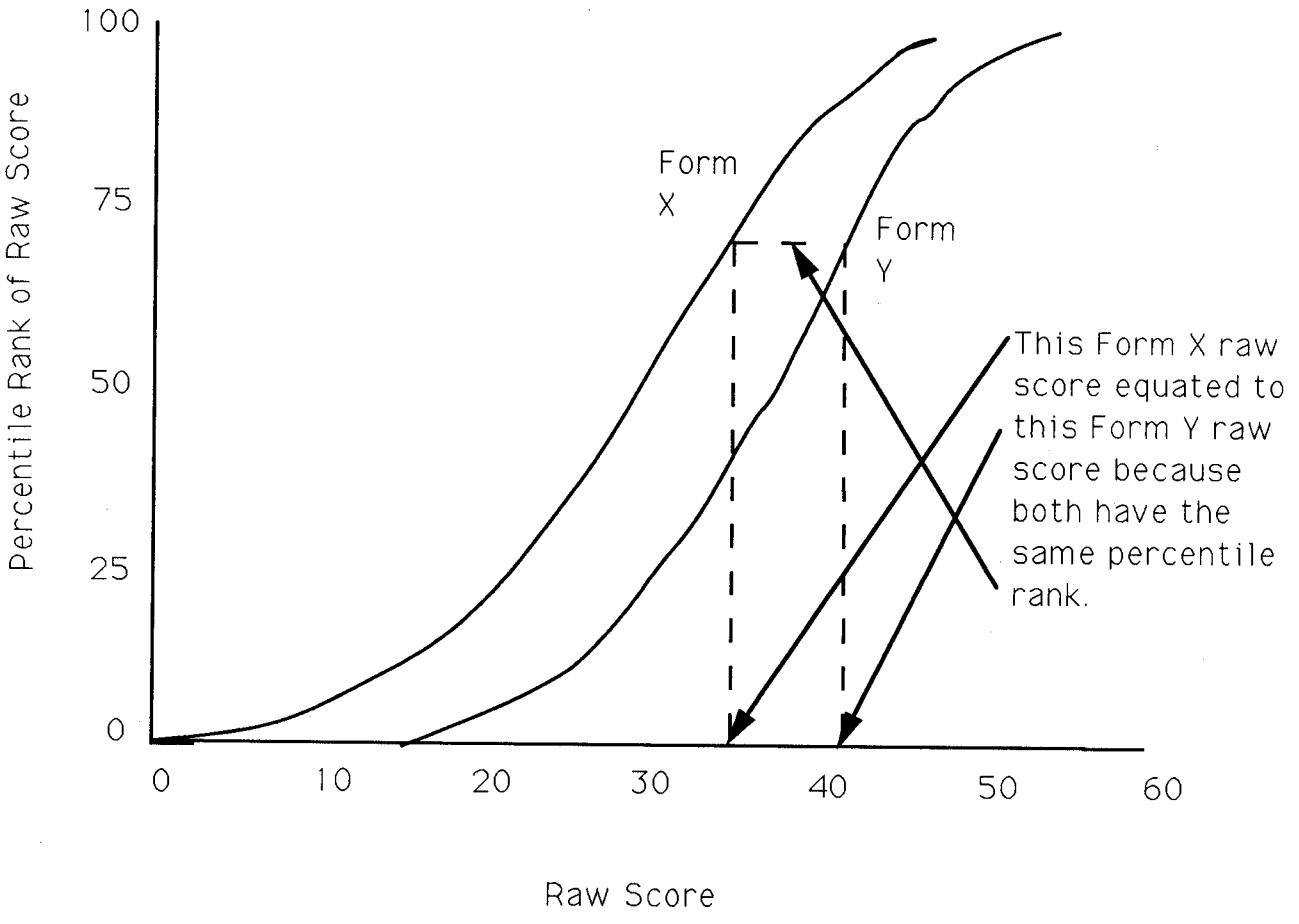

Figure 1. Equipercentile Equating.

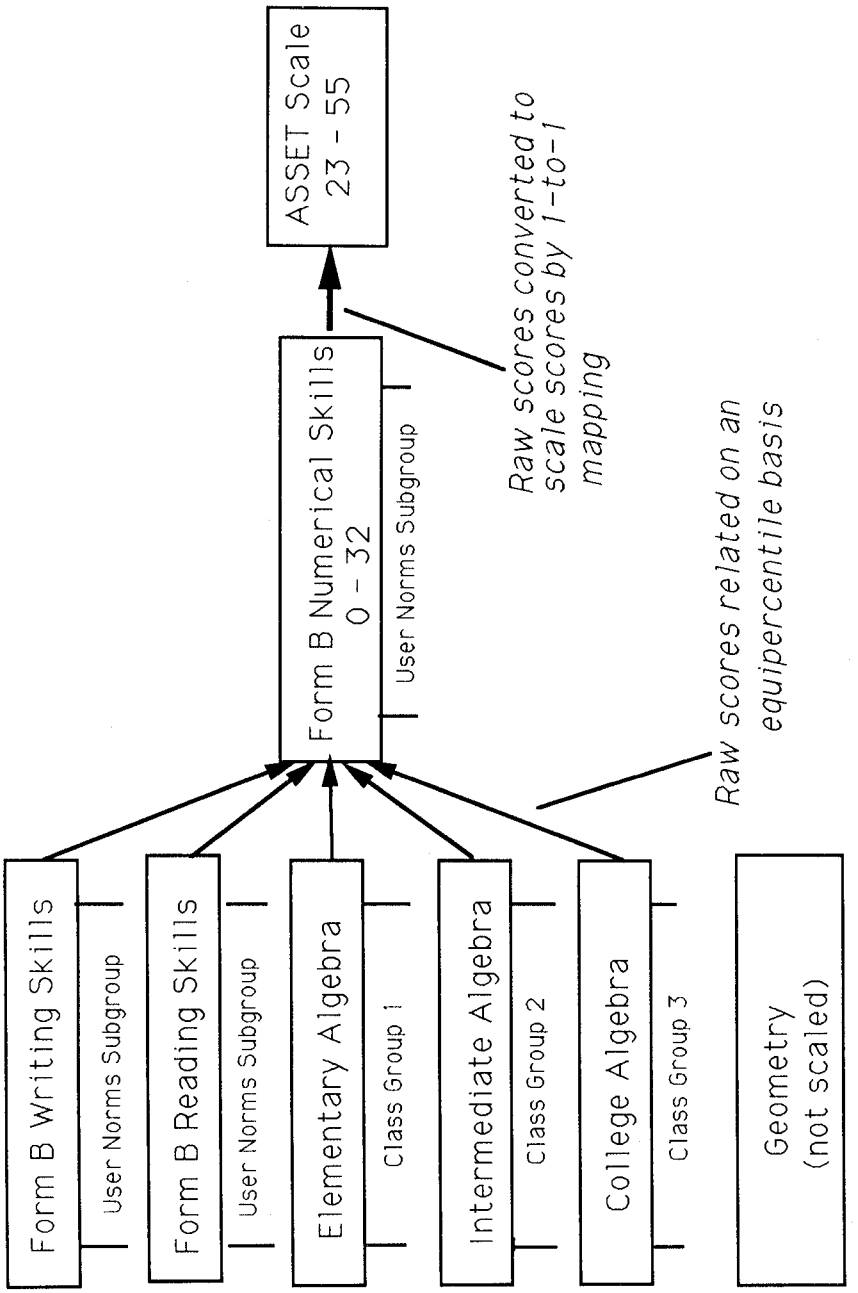

Figure 2. Setting the ASSET Score Scale and Converting Raw Scores to Scale Scores.

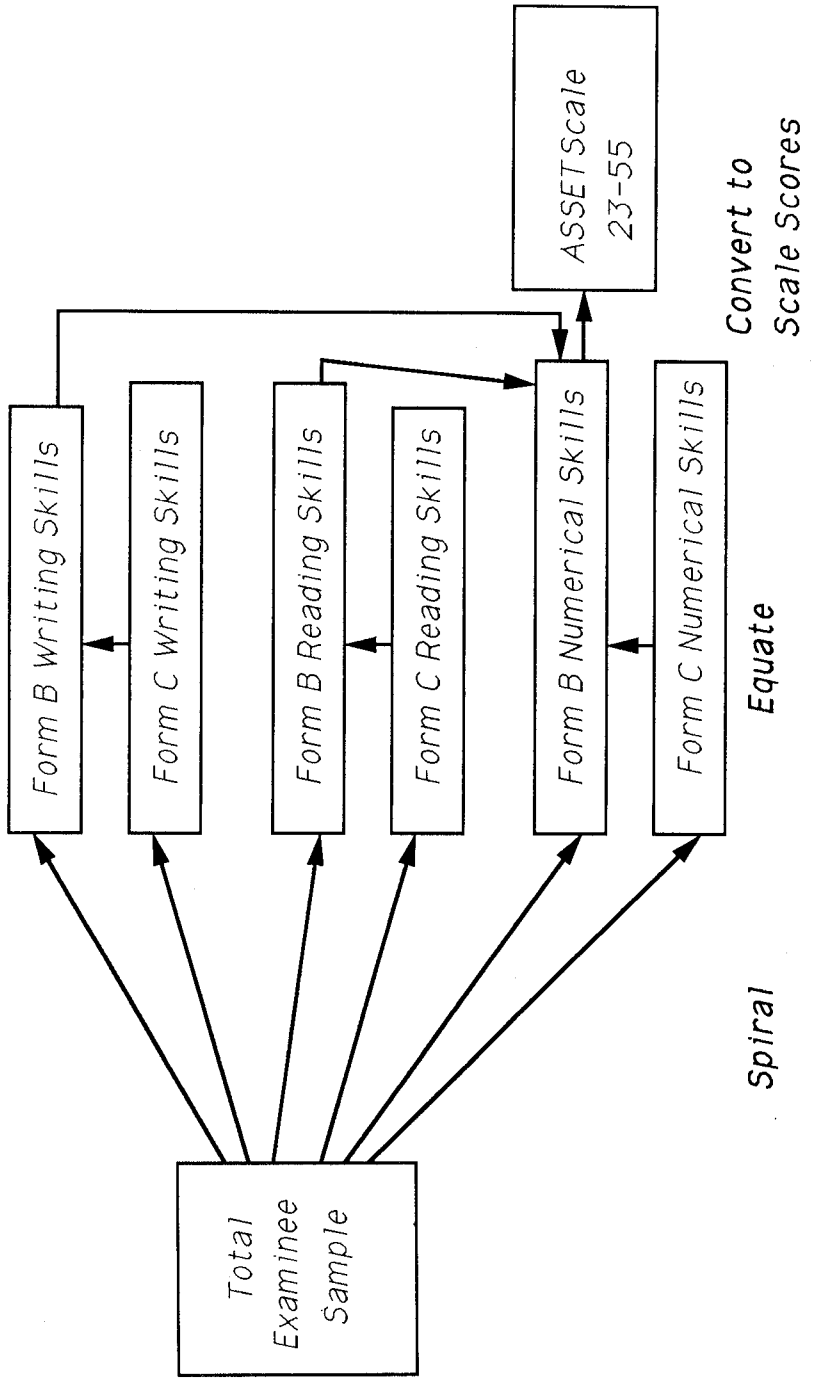

Figure 3. Equating Forms C to Forms B.