Three instructors developed two forms of a final exam for an Intermediate Algebra course. After developing content specifications, the instructors used *TestGen* 3.3 (a computerized testbank developed by Addison-Wesley) to generate two forms of the test. The two 24-item test forms were spiraled to 428 students at the end of the Spring 2004 and Fall 2005 semesters ($N_X = 202$; $N_Y = 226$).

Figure 1: *Raw Score Relative Frequency Distributions for Form X & Form Y*

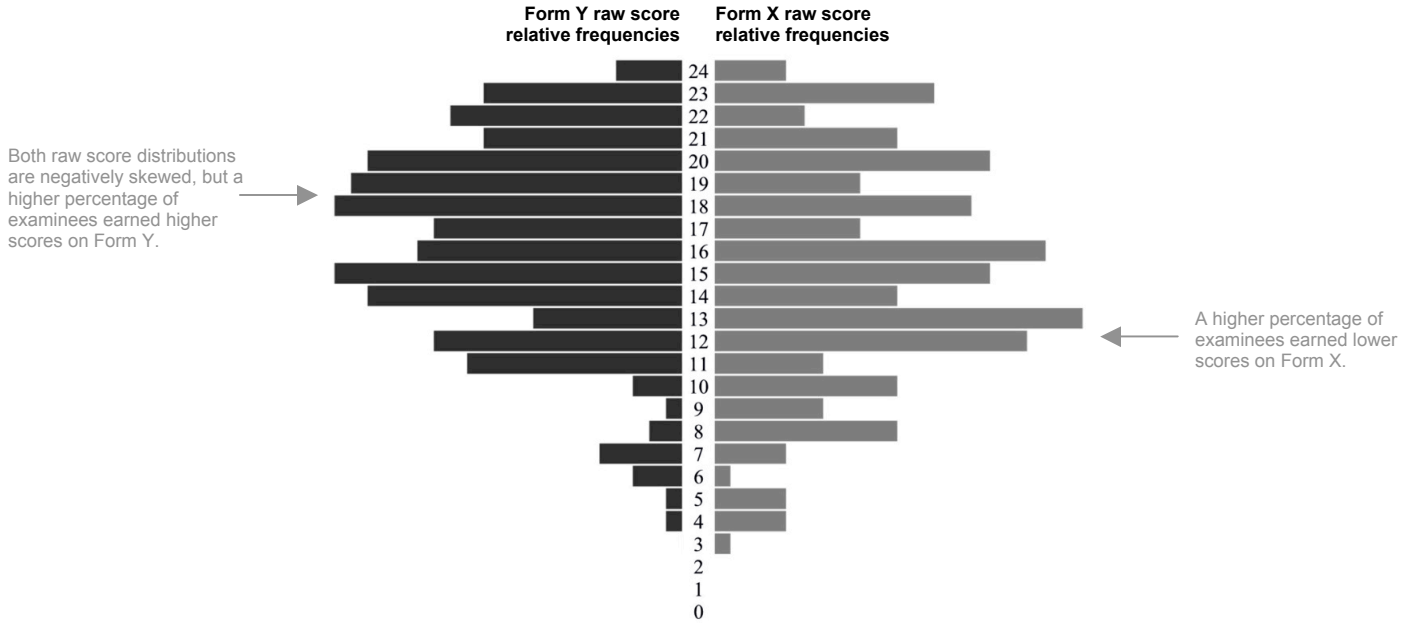
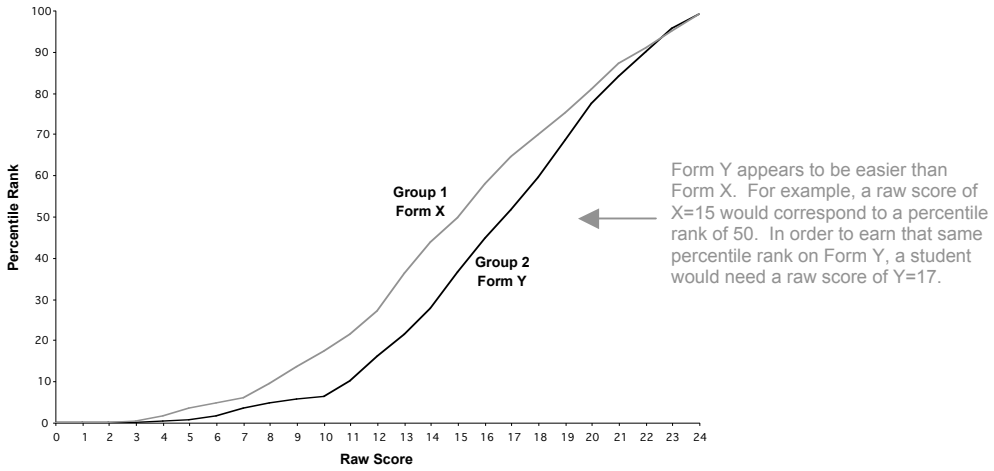


Figure 2: *Percentile Ranks for equating Form X and Form Y of the Intermediate Algebra Exam*



This project:
1. Mean, linear, unsmoothed equipercentile equating results
2. Pre- and post-smoothed equipercentile equating results
3. Selecting the "best" equating method (focus on equating error)
4. Appropriateness of equating in this situation

1. Mean, linear, unsmoothed equipercentile equating results (RAGE-RGEQUATE) $m_y(x) = x + 1.4954$

$l_y(x) = 0.8616x + 3.5665$

Table 2: *Raw-to-raw score conversion tables (Random Groups Design)*

	Target Form	New Form	Mean		Linear		Equipercentile (No Smoothing)	
	Y Raw	**X Raw**	**Unrounded**	**Rounded**	**Unrounded**	**Rounded**	**Unrounded**	**Rounded**
	0	0	1.4954	1	3.5665	4	0*	0
	1	1	2.4954	2	4.4280	4	1*	1
	2	2	3.4954	3	5.2896	5	2*	2
	3	3	4.4954	4	6.1512	6	4.059	4
	4	4	5.4954	5	7.0127	7	5.952	6
	5	5	6.4954	6	7.8743	8	7.066	7
	6	6	7.4954	7	8.7358	9	7.814	8
	7	7	8.4954	8	9.5974	10	9.642	10
	8	8	9.4954	9	10.4590	10	10.904	11
	9	9	10.4954	10	11.3205	11	11.581	12
	10	10	11.4954	11	12.1821	12	12.177	12
	11	11	12.4954	12	13.0437	13	12.957	13
	12	12	13.4954	13	13.9052	14	13.920	14
	13	13	14.4954	14	14.7668	15	14.961	15
	14	14	15.4954	15	15.6284	16	15.841	16
	15	15	16.4954	16	16.4899	16	16.729	17
	16	16	17.4954	17	17.3515	17	17.829	18
	17	17	18.4954	18	18.2130	18	18.552	19
	18	18	19.4954	19	19.0746	19	19.138	19
	19	19	20.4954	20	19.9362	20	19.766	20
	20	20	21.4954	21	20.7977	21	20.443	20
	21	21	22.4954	22	21.6593	22	21.565	22
	22	22	23.4954	23	22.5209	23	22.164	22
	23	23	24.4954	24	23.3824	23	22.901	23
	24	24	25.4954	24	24.2440	24	23.941	24
Mean	**16.4558**	14.9604	**16.4558**	15.9406	**16.4558**	16.4109	16.4584	16.5396
SD	**4.3017**	4.9929	4.9929	4.9589	**4.3017**	4.3252	4.2890	4.2811
Skewness	**-0.4354**	-0.1765	-0.1765	-0.2037	-0.1765	-0.1729	-0.4268	-0.4422
Kurtosis	**2.7448**	2.3566	2.3566	2.3367	2.3566	2.3190	2.7311	2.7800

Figure 3: *Results from mean, linear, and equipercentile equating methods*

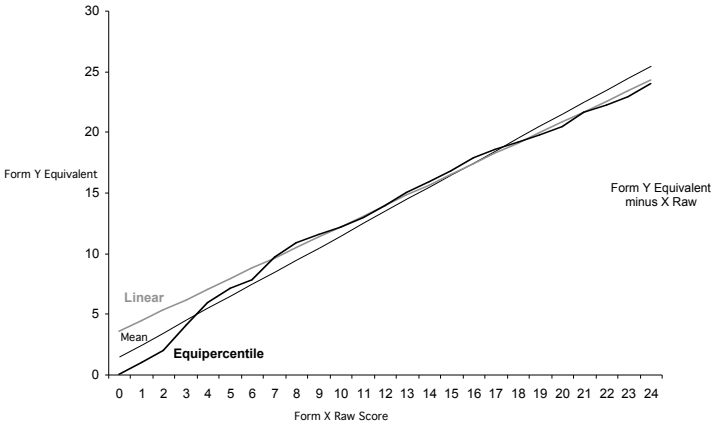
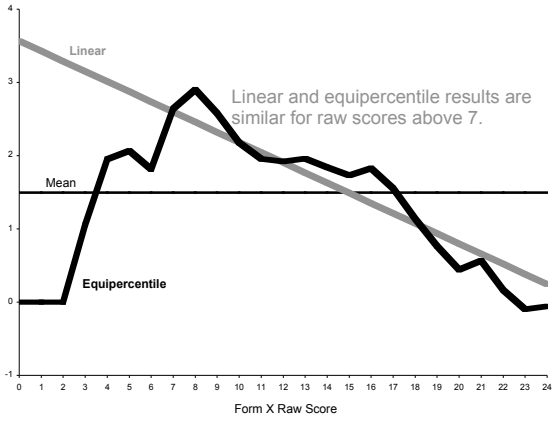


Figure 4: *Deviations between equating methods and identity equating*



2. Pre- and post-smoothed equipercentile equating results

Cubic Splines Postsmoothing

Figure 5: *Raw-to-raw score equivalents for cubic spline postsmoothing (S = 0.05, 0.10, 0.20, and 0.30)*

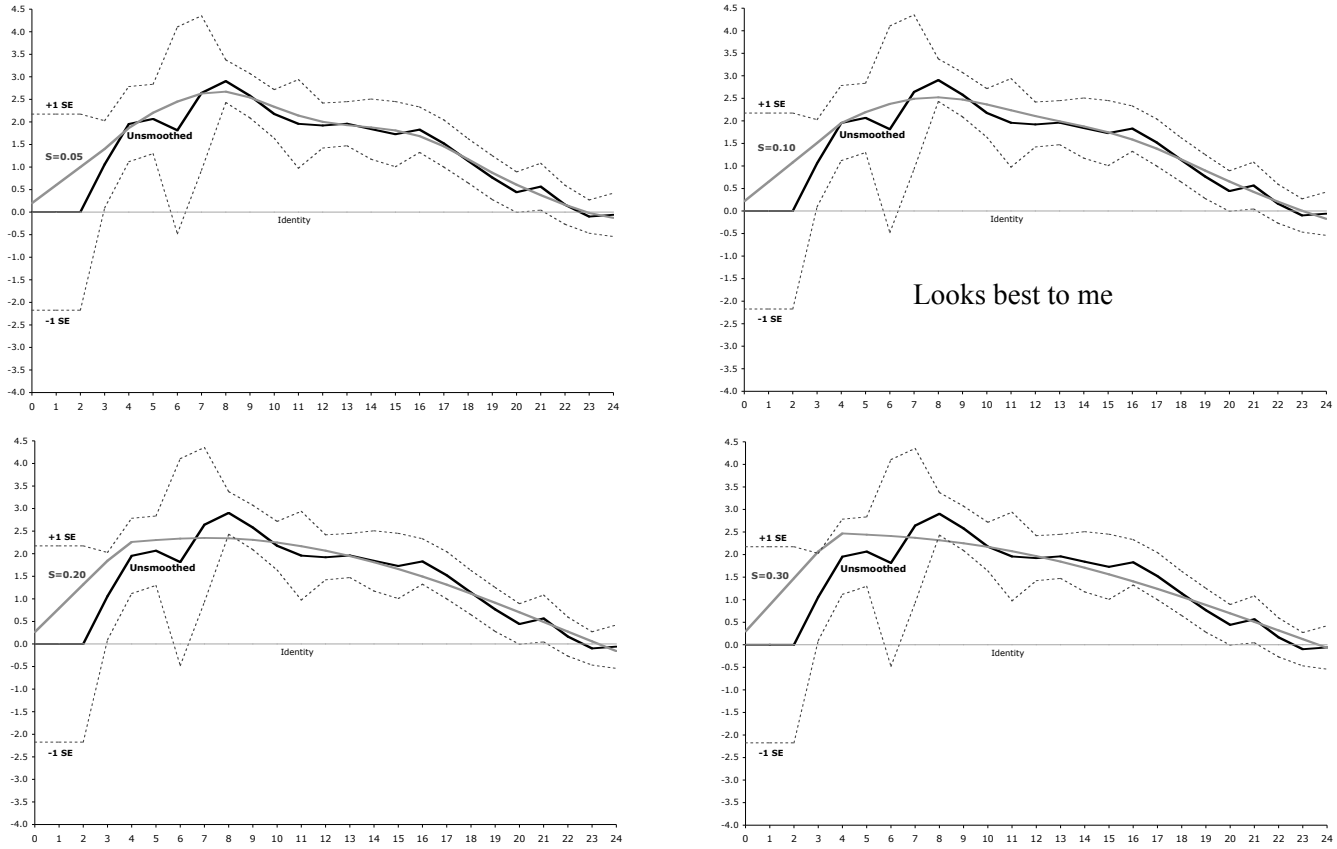


Table 4: *Rounded score conversions for postsmoothing*

	Target	New Form	No	Rounded Scores with Smoothing level (S = 0.__)								
	Y Raw	**X Raw**	**Smooth**	**.01**	**.05**	**.10**	**.20**	**.30**	**.40**	**.50**	**.75**	**1.00**
Mean	16.4558	14.9604	16.5396	16.5644	16.5248	16.4802*	16.3416	16.3911	16.3564	13.3564	13.3564	13.3069
SD	4.3017	4.9929	4.2811	4.2923*	4.2742	4.3272	4.3722	4.4299	4.3432	4.3432	4.3432	4.3526
Sk	-0.4354	-0.1765	-0.4422	-0.4579	-0.4418	-0.4325*	-0.3914	-0.3642	-0.2344	-0.2344	-0.2344	-0.2003
Ku	2.7448	2.3566	2.7801	2.7447*	2.7709	2.6430	2.5957	2.5497	2.3789	2.3789	2.3789	2.3488

S = 0.10 smoothing level appears to yield the best results

Polynomial Log-Linear Presmoothing

Figure 6: *Presmoothing Form Y distribution (Using C = 1, 2, 3, 4, 5, 6)*

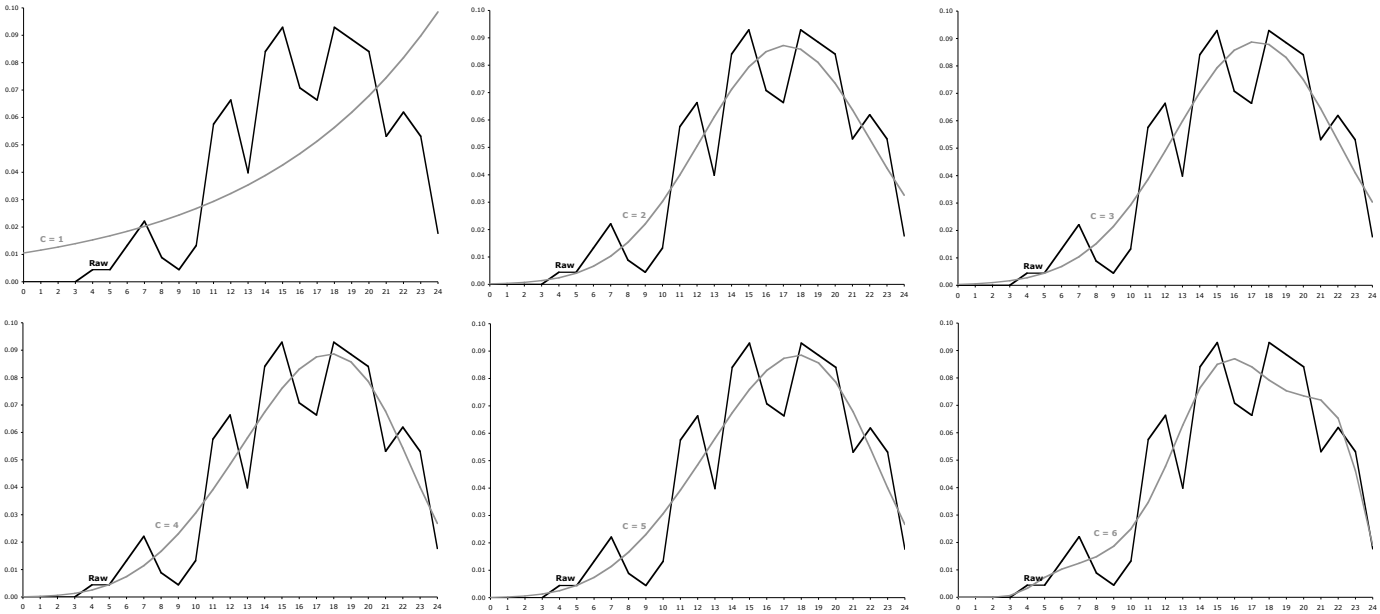


Figure 6: *Presmoothing Form X distribution (Using C = 1, 2, 3, 4, 5, 6)*

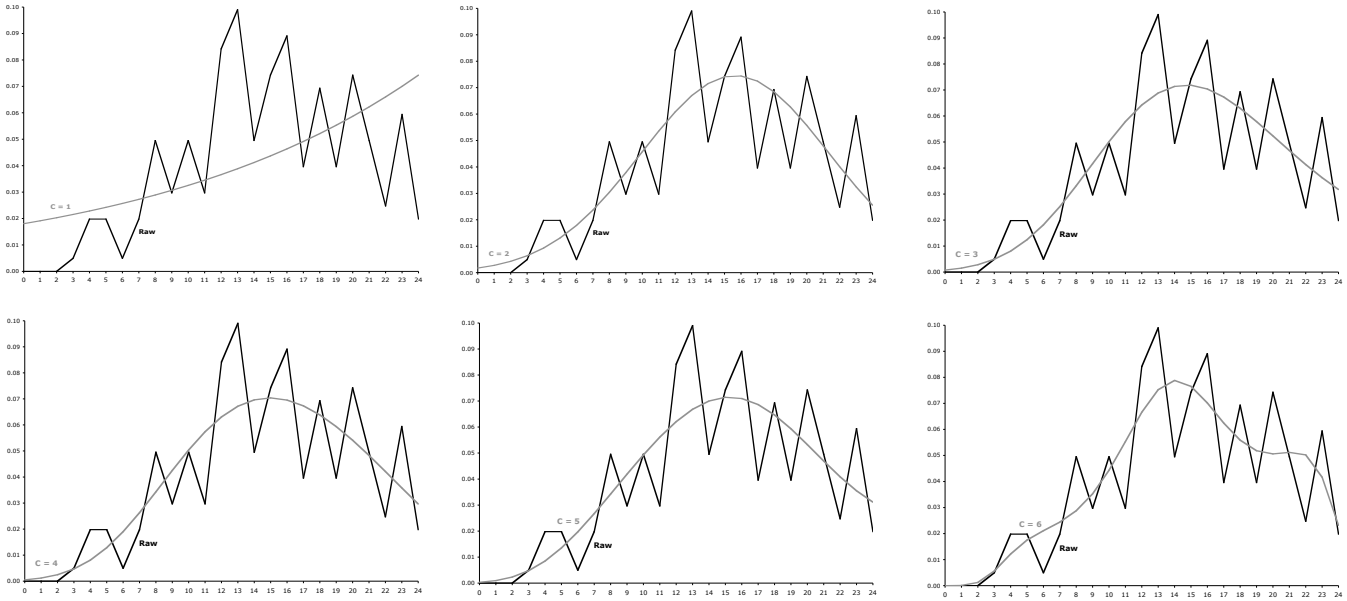


Table 7: *Rounded presmoothing equivalents*

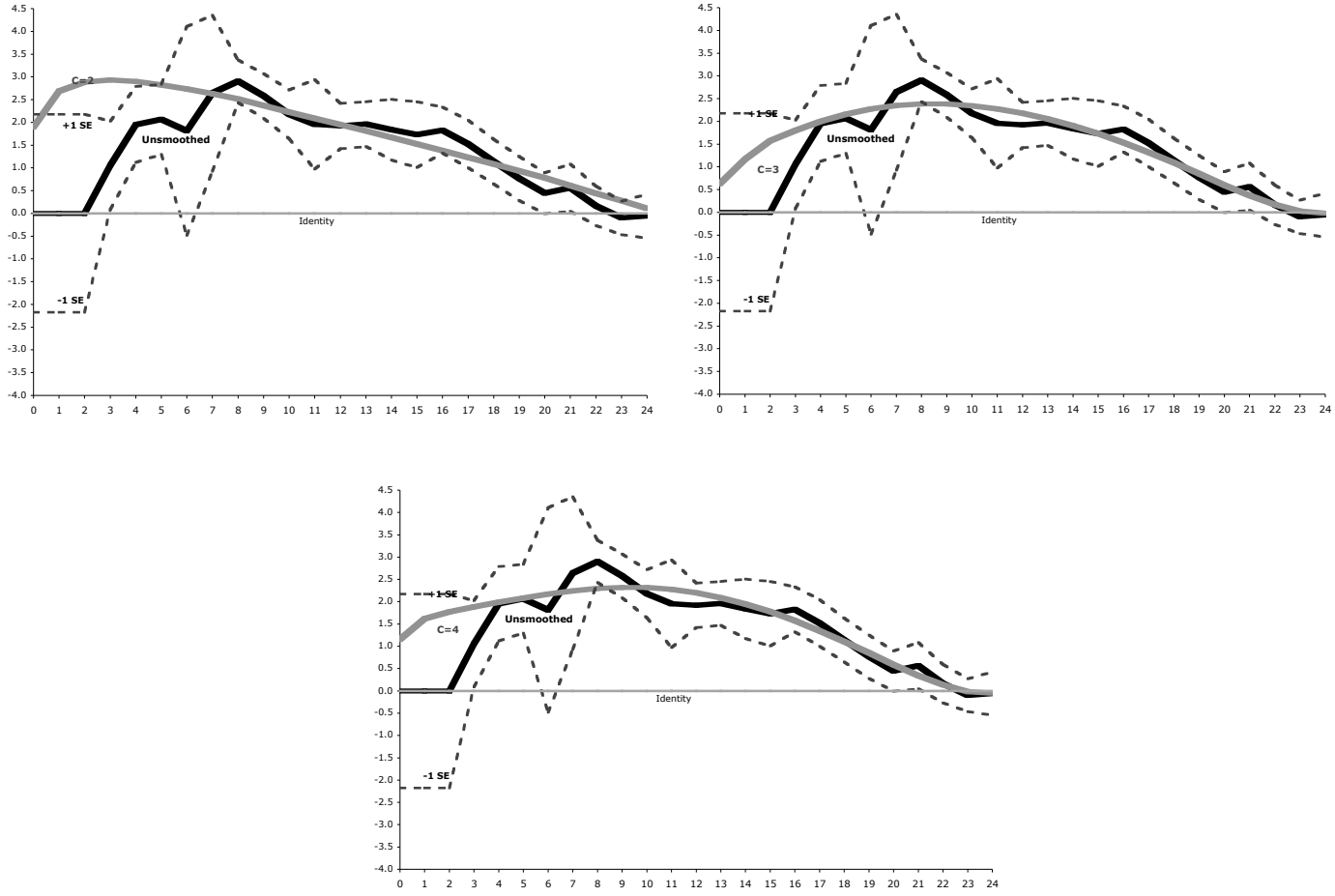
	Target	New Form	Rounded Scores with Presmoothing level (C = __)									
	Y Raw	X Raw	1	2	3	4	5	6	7	8	9	10
Mean	16.4558	14.9604	16.7426	16.5099	16.4307*	16.4307	16.4307	1.65297	16.5347	16.5050	16.5347	16.5248
SD	4.3017	4.9929	4.3977	4.2219	4.3948	4.3948	4.3948	4.2603	4.2510	4.2859*	4.2510	4.2742
Sk	-0.4354	-0.1765	-0.4332*	-0.2235	-0.4417	-0.4417	-0.4417	-0.4219	-0.4157	-0.4136	-0.4517	-0.4418
Ku	2.7448	2.3566	2.6365	2.3717	2.5777	2.5777	2.5777	2.7046	2.6989	2.6456	2.6989	2.7709*

Table 5: *Moments and fit statistics for presmoothing*

Form Y:							
Degree	Mean	SD	sk	ku	Chi-Squared	df	Difference
C= 6	16.455752	4.301707	-0.435372	2.744803	19.211	18	1.176
C= 5	16.455752	4.301707	-0.435373	2.744804	23.487	19	4.276*
C= 4	16.455742	4.301725	-0.435389	2.744853	23.488	20	0.001
C= 3	16.455752	4.301707	-0.435372	2.877737*	24.059	21	0.571
C= 2	16.455742	4.301718	-0.381072*	2.756566*	24.262	22	0.202
C= 1	16.455734	6.357974*	-0.788613*	2.673953*	108.063*	23	83.802*

Form X:							
Degree	Mean	SD	sk	ku	Chi-Squared	df	Difference
C= 6	14.960381	4.992923	-0.176508	2.356626	26.446	18	0.438
C= 5	14.960394	4.992910	-0.176501	2.356614	30.389	19	3.943*
C= 4	14.960396	4.992907	-0.176500	2.356611	30.498	20	0.109
C= 3	14.960391	4.992913	-0.176503	2.410096*	30.674	21	0.177
C= 2	14.960396	4.992908	-0.311584*	2.592827*	32.255	22	1.581
C= 1	14.960387	6.840647*	-0.506983*	2.153671*	91.677*	23	59.422*

Figure 7: *Raw-to-raw score equivalents (Presmoothing with C=2, 3, and 4)*



C = 3 appears to yield the best results

3. Selecting the "best" equating method (focus on equating error)

Table 8: *Rounded equipercentile equating equivalents*

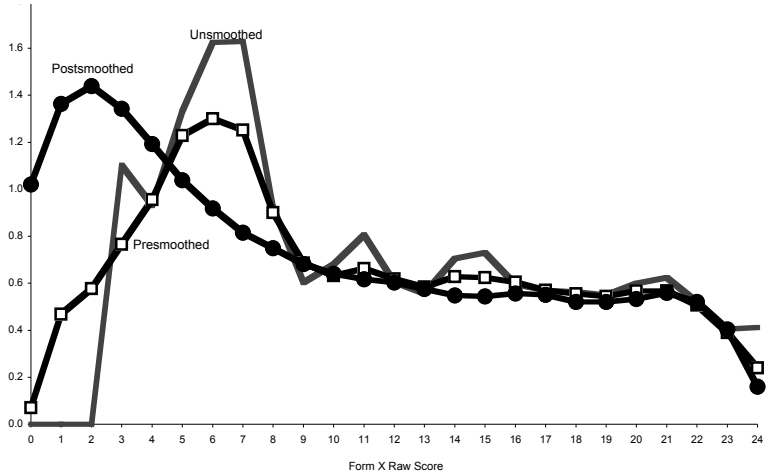
Target Y Raw	New Form X Raw	Unsmoothed	Presmoothed (C = 3)	Postsmoothed (S = 0.10)
0	0	1	1	1
1	1	1***	2	2
2	2	2***	4***	3
3	3	4***	5	5
4	4	6	6	6
5	5	7	7	7
6	6	8	8	8
7	7	10***	9	9
8	8	11***	10***	11
9	9	12***	11	11
10	10	12	12	12
11	11	13	13	13
12	12	14	14	14
13	13	15	15	15
14	14	16	16	16
15	15	17	17	17
16	16	18	18	18
17	17	19***	18	18
18	18	19	19	19
19	19	20	20	20
20	20	20***	21	21
21	21	22***	21	21
22	22	22	22	22
23	23	23	23	23
24	24	24	24	24
Mean	16.4558	14.9604	16.5396	16.4307	16.4802
SD	4.3017	4.9929	4.2811	4.3948	4.3272
Sk	-0.4354	-0.1765	-0.4422	-0.4417	-0.4325
Ku	2.7448	2.3566	2.7801	2.5777	2.643

Standard Errors of Equating

Table 8: ***Average*** *Bootstrap Standard Errors of Equating (Unrounded and Roundde Equivalents)*

	Mean	Linear	Unsmoothed	Presmoothed (C=3)	Postsmoothed (S=0.10)
Unrounded	0.44266	0.47298	0.59144	0.56562	0.51619
Rounded	0.50177	0.54704	0.66376	0.62868	0.58937

Figure 9: *Bootstrap standard errors for unsmoothed and smoothed equipercentile methods (rounded)*



4. Appropriateness of equating in this situation

A. Based on visual examination of score distributions, equating relationships, and standard errors, it appears as though postsmoothed equipercentile equating with S = 0.10 provides the best results. But are these results any good?

B. Compare to the "Conditions Conducive to a Satisfactory Equating" by Kolen & Brennan

Condition: "Equated scores on alternate forms can be used interchangeably only if the alternate forms are built to carefully designed *content and statistical specifications*" (p.269).

Problem: Even though a computer generated the test forms, the forms do not appear to be parallel.
Form X introduced decimals and negative numbers; increasing the difficulty and changing the nature of items
Because of this difference in content, the test forms should not be equated

Condition: Test forms should be at least 30-items in length – using items that have been pretested

Problem: These test forms contain 24 items. The items were not pretested.

Condition: Random groups equipercentile equating methods require large representative samples of examinees

Problem: Relatively small sample sizes. With continuing growth in enrollment, these samples may not be representative.

C. No equating method would be appropriate for this data. I would recommend instructors develop their own tests based on the content they cover in class and set their own performance standards.