ACT Research Report Series 94-2

A Quadratic Curve Equating Method to Equate the First Three Moments in Equipercentile Equating

Tian-You Wang
Michael J. Kolen

June 1994

ACT

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

A Quadratic Curve Equating Method to Equate the First Three Moments in Equipercentile Equating

Tianyou Wang
Michael J. Kolen

A Quadratic Curve Equating Method to Equate the First Three Moments in Equipercentile Equating

(Abstract)

In this paper, a quadratic curve equating method for equating different test forms under a random group data collection design is proposed. Procedures for implementing this method and related issues are described and discussed. The quadratic curve method was evaluated using real test data and simulated data in terms of model fit and equating error, and was compared to several other equating methods. It was found that the quadratic curve method fit many of the real test data examined and that when model fits the population, this method could perform better than other more sophisticated equating methods. Index terms: Equipercentile equating, smoothing procedures, quadratic curve equating, linear equating, random groups equating design.

A Quadratic Curve Equating Method to Equate the First Three Moments in Equipercentile Equating

In standardized testing, often multiple test forms are needed because examinees need to take the test at different occasions and one test form can be administered only once to ensure test security. In this situation, it is typically required that test scores derived from different forms are equivalent. Efforts can be made in the test construction process to make different forms as nearly equivalent as possible (e.g., forms can be built based on the same table of specifications; items can be selected to have approximately equal average difficulty level). But often these efforts are not enough to ensure test score equivalency for different forms. So, test equating based on test data is often performed to adjust test scores so that scores on different forms are more nearly equivalent. There are several designs for collecting test equating data. One of the designs is the random groups design, in which different test forms are administered to different but randomly equivalent groups of examinees.

Under the random groups equating design, the examinee groups that take different test forms (for simplicity, say, form X and form Y) are regarded as being sampled from the same population. The differences in score distributions for different test forms are attributed to form differences and sampling variations of the examinee groups. Equating form X to form Y involves transforming the X scores so that the transformed X scores have the same distribution as the Y scores. If an assumption can be made that the population distributions for X and Y scores have the same shape and only differ in mean and variance, then the linear equating method will be most appropriate. Linear equating takes the form

$$l_Y(x) = \sigma_Y \left[\frac{x - \mu_X}{\sigma_X} \right] + \mu_Y \ , \qquad (1)$$

where x is the score on form X, μ_X and μ_Y are means for form X and form Y, σ_X and σ_Y are standard deviations (*s.d.*) for form X and form Y, and $l_Y(x)$ is the equated form Y score for x.

If no assumptions can be made about the shape of the population score distributions, equipercentile equating method is the method of choice. Equipercentile equating for a discrete score distribution is given by

$$e_Y(x) = \frac{p^*(x) - \Pr[Y < u^*(x)]}{\Pr[Y = u^*(x)]} + u^*(x) - .5 \ , \qquad (2)$$

where Pr means probability, $p^*(x) = \Pr(X < x) + .5\Pr(X = x)$, and $u^*(x)$ is the smallest integer such that $p^*(x) = \Pr[Y < u^*(x)]$.

Equipercentile equating based on samples may have large sampling error because for any particular score, the equating relationship is based on local frequencies at that score point. Two types of smoothing techniques have been introduced to reduce random errors: pre-smoothing and post-smoothing. Pre-smoothing smoothes the score distributions for form X and form Y separately and equates the smoothed score distributions. Post-smoothing (Kolen, 1984) smoothes the equipercentile equating function directly.

Studies have been done to evaluate these methods (see Kolen, 1984, Fairbank, 1987, Cope & Kolen, 1990, Hanson, 1990, Hanson, Zeng, & Colton, 1991). Results from Hanson, Zeng, and Colton (1991) showed that smoothed equating was more accurate than unsmoothed equipercentile and linear methods in terms of mean squared errors. However, linear equating consistently had smaller random error, especially when sample sizes were small. This finding resulted because the linear method uses only means and standard deviations in computing the equating equation and these aggregate statistics typically have small sample variability. However, a fundamental limitation of linear methods is that if the shape of the distribution of X scores is different from that of Y scores in the population, it could be seriously biased. Although an increase in sample size could reduce standard errors of equating, it will not reduce bias. Angoff (1987) commented that equipercentile equating lacks a theoretical basis whereas linear equating makes strong statistical assumptions which are often violated. He suggested that consideration be given to equating methods which employ theoretical models that take into account higher moments. The purpose of this study is to propose a quadratic curve equating method, and to compare it with some other equating methods. If successful, the quadratic curve method would produce less random error than other pre- or post-smoothing equipercentile methods, and less bias than the linear method.

The Quadratic Curve Equating Method

In choosing such a nonlinear equating function, the following aspects were considered:

(1) The function should be more flexible than the linear function.

(2) The function should preserve beneficial properties of linear equating, such as using statistics with small random errors and being computationally simple.

(3) The performance of this method should be comparable to more complicated techniques like smoothed equipercentile equating in most, if not all, testing situations.

Based on the preceding considerations, a quadratic curve to relate scores on form X to form Y was proposed which takes the form

$$q(x) = ax^2 + bx + c \quad . \qquad (3)$$

The coefficients *a, b, c* are so determined such that the equated X scores will have the same mean, standard deviation (*s.d.*) and skewness as the form Y scores. The difference between this relationship and linear equating relationship is that it has one additional squared term and that skewness is taken into account in computing the equating function.

The assumption underlying this method is that if population distributions are used and the appropriate quadratic equating relationship is established to equate the first three moments, then all the moments of the equated X score distribution will be the same as those of the Y score distribution.

In order to determine coefficients *a, b,* and *c* using the method of moments, the following set of non-linear equations needs to be solved:

$$E[q(X)] - E(Y) = 0 \tag{4}$$
$$E\{[q(X)]^2\} - E(Y^2) = 0 \tag{5}$$
$$E\{[q(X)]^3\} - E(Y^3) = 0 \tag{6}$$

where E represents expectation. If $q(X)$ is substituted in these equations, we get:

$$E[aX^2 + bX + c] - E(Y) = 0 \tag{7}$$
$$E[(aX^2 + bX + c)^2] - E(Y^2) = 0 \tag{8}$$
$$E[(aX^2 + bX + c)^3] - E(Y^3) = 0 \tag{9}$$

The left hand side of these equations are functions of *a, b, c,* the first six moments of X scores, and the first three moments of Y scores. When population distributions are not known, sample moments are used. The Newton-Raphson method (Press, 1992) could be used to simultaneously solve this set of equations for *a, b* and *c* iteratively. Another easier way to find these coefficients is to utilize the property that linear transformation does not change the skewness of a score distribution. The procedure is as follows.

First, let us define skewness of Y as

$$Sk(Y) = \frac{E[Y - E(Y)]^3}{\{E[Y - E(Y)]^2\}^{3/2}} \;. \tag{10}$$

Find d so that $Z = X + dX^2$ will have same skewness as Y; i.e., Sk(Z)=Sk(Y). This can be done using an iterative numerical method. Second, Let b equal the ratio of *s.d.* of Y to *s.d.* of Z; i.e.,

$b = \frac{s.d.(Y)}{s.d.(Z)}$. We know that b Z will have same *s.d.* and skewness as Y because multiplication by the constant b does not change the skewness of Z. Then add a constant c so that $c+b(X+dX^2)$ has same mean, *s.d.* and skewness as Y. The three coefficients for equation (3) are then easily determined.

Some Technical Issues

Symmetry: One of the requirements for an equating method is symmetry. That is, the same equating relationship should result whether X is equated to Y or Y is equated to X. This quadratic function is clearly not symmetric because different orders of moments are used for X and Y scores. Kolen (1984) proposed an average of two equating relations obtained when X is equated to Y and Y is equated to X. However, this treatment still does not yield exactly symmetric results. For the quadratic method, a weighted average of two equating functions will be used. Suppose for a given score x, the equated score obtained from one direction is y_1, that from the other direction is y_2, and the two first derivatives at score point x are d_1 and d_2, then the weighted average is

$$y = w_1 y_1 + (1 - w_1) y_2 ,$$

where

$$w_1 = \frac{\tan[.5\arctan(d_1) + .5\arctan(d_2)] - d_2}{d_1 - d_2}, \quad \text{if } d_1 \neq d_2;$$

or

$$w_1 = 0.5, \qquad \text{if } d_1 = d_2.$$

This weighted average is guaranteed to be symmetric for the linear case. For the quadratic curves in this situation, the curvature can be expected to be very small. Thus, a good approximation to symmetry can be assumed. Note that generally the weights are different at different x scores.

Equating at Extreme Scores: Equating at both ends of score range is problematic for nearly all equating methods. This issue also concerns the quadratic method. In implementing the post-smoothing method, Kolen (1984) excluded the upper half percent and the lower half percent of the data in computing the spline function and use two straight lines to link the ends of the spline to the two unequated end scores. This practice was also adopted in the present study.

Methodology and Data

The proposed quadratic curve equating was evaluated in two aspects: model fit and sampling error.

Model Fit

Like the linear equating method, this new method makes an assumption about the true population equating relationship. The assumption underlying the new method basically states that the true population equating relationship is quadratic in form. How well this method performs logically depends on how close the true equating is to the quadratic form in real testing situations. Although population score distributions are almost never available in practice, sample distributions, especially large sample distributions can provide valuable information about this issue. In this study, real test data were used to assess model fit for this method. For each pair of test forms, five different equating methods were applied and plotted for visual examination: unsmoothed equipercentile equating, spline post-smoothing with smoothing parameter s=0.2 and s=0.5, quadratic curve equating, and linear equating. The first four central moments of the equated X scores with the quadratic curve method were also computed. If the assumptions of this method are met, the kurtosis of the equated X scores would be expected to be close to that of the Y scores. Under normality, the sampling variance of kurtosis equals 24/N where N is the sample size (see Kendall & Stuart, 1977, pp. 258). The extent to which the model fits the data can be partially assessed by comparing the difference of the kurtosis after equating and the standard deviation of absolute kurtosis differences.

The first two pairs of data are the same as the first two pairs used in Hanson, Zeng, and Colton (1991). The first pair consists of two 30-item subsets from a professional licensure exam. The second pair consists two 20-item subsets of two forms of Reading subtests of ACT Assessment. Each of these data sets have very large sample sizes (over 38,000 and 82,000 respectively). So the unsmoothed equipercentile equating relationship can be taken as the population equating.

Data from an operational equating of the ACT Assessment were also used. These data contain seven forms (form A to form G) for each of the four tests: English with 75 items, Mathematics with 60 items, Reading with 40 items and Science with 40 items. For each test, seven pairs of distributions were used for equating (form A to B, B to C ... G to A).

Sampling Error

Because the quadratic curve method uses aggregate statistics, like the linear method, it would be expected to have less sampling variance than the unsmoothed equipercentile method or even the smoothed equipercentile methods when the model assumptions are met. Parametric bootstrapping methods (Efron, 1982) were used to assess the sampling error of this equating method and compare it to the unsmoothed equipercentile method, the linear method, and the spline post-smoothing method. First, a pair of population distributions was defined using either smoothed sample distributions or very large sample distributions. Second, sample distributions with sample size N (three different sample sizes were used in this study. For the long test of 75 items N=500, 2000, and 3000; for short tests of 20, 30 and 40 items, N=250, 500, and 2000) were generated from the population distributions by computer and equatings with various methods were performed. Third, the second step was repeated n (in this study, n=200) times and evaluative indices were computed for each score point.

The study by Hanson, Zeng, and Colton (1991) showed that pre-smoothing and post-smoothing yielded comparable results in terms of mean squared error. So it is sufficient to just use post-smoothing to represent smoothed equipercentile methods.

Three types of population distributions were used. The first type was two pairs of observed distributions with very large sample sizes. These observed distributions were taken directly as the population distributions. These were the 30-item licensure exam subtests and the 20-item Reading subtests described previously.

The second and third types were the results of smoothing ACT Assessment score distributions using log-linear smoothing method (Hanson, 1992). The second type was intended to represent situations where the equipercentile equatings with smoothed score distributions were close to the quadratic function. The third type were to represent situations where the equipercentile equatings with smoothed score distributions were not close to the quadratic function. The third type was used also to assess the robustness of the quadratic curve method to model violation. From initial examination of the equating functions from different methods, English forms A and B, Science forms G and A were selected to represent the second type and Reading forms A and B were selected to represent the third type. Pearson χ^2 statistics for model fit were examined to determine the degrees of the log-linear model. Figure 7 gave plots of the equipercentile equating and other equatings based on three pairs of smoothed score distributions.

The evaluative indices are bias, standard error (s.e.), and root mean squared error (*RMSE*). For any score x on form X, denote $e(x)$ as the true (or population) equated score and $\hat{e}_s(x)$ as equated score based on sample s with any particular equating method. The mean equated score based on n samples is

$$\bar{e}(x) = \frac{1}{n}\sum_{s=1}^{n}\hat{e}_s(x).$$

The estimated bias is

$$\bar{\hat{e}}(x) - e(x).$$

The estimated standard error is

$$\sqrt{\frac{1}{n}\sum_{s=1}^{n}\left(\hat{e}_s(x) - \bar{\hat{e}}(x)\right)^2}.$$

The estimated root mean squared error is

$$\sqrt{\frac{1}{n}\sum_{s=1}^{n}\left(\hat{e}_s(x) - e(x)\right)^2}.$$

These indices are computed for all the raw score points and estimated root mean squared errors are plotted for all the methods being compared. Weighted averages of these indices weighted by the X score population distribution are also computed by the formula:

$$\sum_{x=1}^{k}(index)Pr(X = x),$$

where k is the number of items.

The root mean squared errors for five above mentioned equating methods were plotted along the score scale for visual comparison. Weighted averages of absolute bias, standard error, root mean squared error(*RMSE*) and estimated standard error of *RMSE* were tabulated.

Results

Descriptive statistics for all real data equatings are summarized in Table 1 (see Table 1 at the end of the report). Because of the adjustment to achieve symmetry, the first three central moments of form X after equating were not exactly the same of those of form Y. But this adjustment had little effect on means and standard deviations except for one case. (ACT Mathematics form C and form D where the adjustment resulted in a 0.019 difference in s. d.); skewness was affected only at the third decimal point except for one case. The equating functions for the real test data sets are plotted in Figure 1 through Figure 6. The plots showed that the quadratic equating is more flexible and provides better fit in most cases than the linear equating method.

Figures 1 and 2 plot the equating functions for the two sets of large sample data. Here, the unsmoothed equipercentile equatings are regarded as population equatings. The post-smoothing functions were closer to the population equating because that the unsmoothed equipercentile equating is very smooth. The quadratic curve equating appears to fit the two population equatings quite well. The maximum biases are within 0.2 score points. The kurtosis difference were reduced by about half after equating in both cases.

The plots of equating functions based on ACT Assessment operational equating data sets (Figure 3 through 6) showed that the quadratic curve method performed quite well in smoothing the equipercentile equating function in most cases. In many cases, the quadratic equating function was between post-smoothings with s=0.2 and s=0.5. The quadratic method did not perform well in 3 to 4 cases (Math C to D, Math D to E, Reading A to B, and perhaps Reading B to C), where the unsmoothed equipercentile equatings displayed an "S" shape. Examination of the kurtosis revealed that these cases correspond to the highest after-equating kurtosis differences among all the 28 quadratic equatings. The standard error of sample kurtosis for a sample size of 2900 is 0.091. The standard error for the kurtosis difference for two independent samples of size of 2900 is thus about 0.129. The average absolute kurtosis difference before equating is 0.163. The average (over all equatings conducted) absolute kurtosis difference after equating is 0.097 which is smaller than the standard deviation of the sample difference. The number of absolute kurtosis differences that are smaller than 0.129 is 20 out of 28. These results suggested that the quadratic curve method fits this set of operational equating data sets reasonably well.

Population equatings for three pairs of distributions used in the simulation are shown in Figure 7. Root mean squared errors for five different equating methods based on five pairs of population distributions were plotted in Figure 8 through Figure 12. Figures 8 and 9 contain plots of *RMSE* of equatings for samples drawn from two large sample data sets: the 30-item Licensure test and the 20-item ACT Reading test.

For the Licensure test, the quadratic method performed about the same as the postsmoothing methods for the small sample size and performed better than these methods for the large sample size. For the 20-item Reading test, the quadratic method performed better than the smoothing methods at some score ranges but not at others. Note that in these two cases, linear equating had remarkably small *RMSE*, especially when sample sizes were small. This finding is probably due to the small form difference in both cases.

Figures 10 and 11 present *RMSE* plots for situations in which the quadratic function apparently fit the population equating relationship well. For these two cases, both the smoothing methods and the quadratic method improved over the unsmoothed equipercentile methods. The amount of improvement of post-smoothing methods is consistent with the results in Hanson, Zeng, and Colton (1991). Clearly, in these two cases, the quadratic method performed better than all other methods regardless of the sample sizes. But the better performance is more consistent along the score scale for small samples than for large samples.

Figure 12 contains plots of *RMSE* for a situation where the population equating relationship does not fit a quadratic function. Apparently, there is no advantage to using this method over using the unsmoothed equipercentile method. Interestingly, when the sample size is small, the linear method had the smallest *RMSE* in the middle score range; when sample size is large, virtually no method shows real improvement over unsmoothed equipercentile method in this case.

Tables 2 through 6 (see Table 2 through 6 at the end of the report) contain the average values of absolute bias, standard error, and *RMSE*, weighted by the X score population frequencies. For the first and second type of populations, all the standard errors were much larger than the absolute bias except for the linear method. So the *RMSE* values were mainly attributed to standard errors. For the first type populations (Table 2 and 3), the quadratic method had slightly better average performance than the smoothing methods in one case and had slightly worse average performance in the other case. For the second type populations (Table 4 and 5), the quadratic method generally had better average performance than the smoothing methods. This is more evident for small sample sizes than for large sample sizes. As we already knew, unsmoothed equipercentile equating is unbiased. Increasing sample size (thus reducing random error) to certain level would surely make this method the best choice. Post smoothing with small smoothing parameter is close to unsmoothed equipercential method. It is observed from the results of the present study that when sample size is large, post-smoothing with small smoothing parameter is usually a favorable choice. For the third type populations (Table 6), the quadratic method had smaller average standard error but not absolute bias. The larger average *RMSE* were attributed to larger bias. Post-smoothing with larger smoothing parameters produced larger bias, but a smaller standard error than that with smaller smoothing parameters. In almost all the cases, linear methods always had smaller standard errors.

Discussion and Conclusion

In searching for an appropriate polynomial function to model the equating relationship, we considered adding one cubic term to the quadratic function so that kurtosis could also be equated. But doing so was found undesirable for two reasons. First, it makes computation much more complicated. Second, sample kurtosis has much more random error than skewness. The variance of sample kurtosis is four times that of sample skewness (see Kendall & Stuart, 1977, pp. 258). Higher order polynomial functions might be investigated in the future if these issues can be properly resolved. The quadratic function could be studied as the first step in this direction.

Linear and equipercentile equating both have advantages and limitations. Smoothing methods are aimed at reducing the random error of the equipercentile methods but they usually involve complicated mathematical manipulation and computer programming. Also, they often require subjective judgment about model parameters. The quadratic equating method proposed in this paper provides another approach to reduce random error as well as bias. Both the idea and computation are simple, and implementation of the quadratic method does not require subjective judgment.

The results based on the real test data in this study showed that the quadratic method worked well for most but not all of the test data. When the population equating relationship was close to a quadratic in form, this method clearly displayed smaller random error and bias than other sophisticated methods for both small and large sample sizes. However, procedures need to be derived to judge whether or not the quadratic method adequately fits the population based on sample data. An examination of the equipercentile equating relationship and the kurtosis difference before and after the quadratic equating might be helpful if this procedure were to be used in practice.

References

Angoff, W. H. (1987). Technical and practical issues in equating: A discussion of four papers. *Applied Psychological Measurement, 11*, 291-300.

Cope, R. T., & Kolen, M. J. (1990). *A study of methods for estimating distributions of test scores*. American College Testing Research Report 90-5. Iowa City, IA: American College Testing.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Fairbank, B. A. (1987). The use of presmoothing and postsmoothing to increase the precision of equipercentile equating. *Applied Psychological Measurement, 11*, 245-262.

Hanson, B. A. (1990). *An investigation of methods for improving estimation of test score distributions*. American College Testing Research Report 90-4. Iowa City, IA: American College Testing.

Hanson, B. A. (1992). *Description of a program for smoothing univariate test score distributions*. Iowa City, IA: American College Testing.

Hanson, B. A., Zeng, L., & Colton, D. (1991). *A comparison of presmoothing and postsmoothing methods in equipercentile equating*. Paper presented to the annual meeting of the National Council on Measurement in Education, Chicago.

Kolen, M. J. & Brennan, R. L. (in press). *Test equating methods and practices* . Springer-Verlag.

Kolen, M. J. (1984). Effectiveness of Analytic Smoothing in Equipercentile Equating. *Journal of Educational Statistics, 9*, 25-44.

Press, W. H. (1992). Numerical recipes in FORTRAN: the art of scientific computing. Cambridge: Cambridge University Press.

Table 1. Descriptive Statistics for Observed Data before and after Quadratic Curve Equating

	Mean	S.d.	Skewness	Kurtosis	Kurt. diff. before equat.	Kurt. diff. after equat.	Sample size
Licensure Subtest (30 items)							
New form before equating	18.880	3.680	-0.130	2.786			38765
New form after equating	19.157	3.430	-0.304	2.934			
old form	19.157	3.430	-0.308	3.051	0.265	0.117	38765
ACT Reading Subtest (20 items)							
New form before equating	12.300	3.757	-0.205	2.391			82062
New form after equating	12.688	3.580	-0.278	2.449			
old form	12.688	3.580	-0.280	2.522	0.131	0.073	83693
ACT English (75 items)							
Form A before equating	48.482	13.088	-0.089	2.187	0.185	0.030	2968
Form A after equating	51.325	12.753	-0.320	2.373			
Form B before equating	51.325	12.755	-0.322	2.423	0.236	0.050	2748
Form B after equating	48.571	12.207	-0.168	2.286			
Form C before equating	48.571	12.207	-0.167	2.302	0.121	0.016	2921
Form C after equating	51.156	12.206	-0.409	2.553			
Form D before equating	51.156	12.205	-0.406	2.532	0.230	0.021	2903
Form D after equating	50.741	12.204	-0.312	2.436			
Form E before equating	50.741	12.204	-0.313	2.395	0.137	0.041	2880
Form E after equating	51.273	12.770	-0.380	2.465			
Form F before equating	51.273	12.770	-0.381	2.521	0.126	0.056	2853
Form F after equating	50.070	12.876	-0.306	2.428			
Form G before equating	50.070	12.876	-0.308	2.372	0.149	0.056	2800
Form G after equating	48.482	13.091	-0.085	2.217			
ACT Mathematics (60) items							
Form A before equating	28.463	10.569	0.481	2.535	0.282	0.143	2968
Form A after equating	30.300	12.198	0.256	2.327			
Form B before equating	30.301	12.187	0.276	2.148	0.287	0.179	2748
Form B after equating	29.758	11.563	0.415	2.281			
Form C before equating	29.758	11.567	0.424	2.463	0.315	0.182	2921
Form C after equating	31.080	12.741	0.179	2.298			
Form D before equating	31.082	12.722	0.208	2.060	0.403	0.238	2903
Form D after equating	28.937	11.448	0.296	2.116			
Form E before equating	28.937	11.450	0.305	2.367	0.307	0.251	2880
Form E after equating	29.819	11.358	0.380	2.444			
Form F before equating	29.819	11.358	0.380	2.424	0.057	0.020	2853
Form F after equating	30.389	11.109	0.321	2.375			
Form G before equating	30.389	11.108	0.324	2.253	0.171	0.122	2800
Form G after equating	28.463	10.566	0.474	2.392			

Table 1 (continued). Descriptive Statistics for Observed Data before and after Quadratic Curve Equating

	Mean	S.d.	Skewness	Kurtosis	Kurt. diff. before equat.	Kurt. diff. after equat.	Sample size
	ACT Reading (40) items						
Form A before equating	24.804	6.584	-0.024	2.375	0.108	0.059	2968
Form A after equating	25.350	7.581	-0.065	2.386			
Form B before equating	25.350	7.581	-0.061	2.117	0.258	0.269	2748
Form B after equating	25.669	6.577	-0.141	2.158			
Form C before equating	25.669	6.578	-0.150	2.466	0.349	0.308	2921
Form C after equating	25.837	6.896	-0.185	2.492			
Form D before equating	25.837	6.896	-0.185	2.459	0.007	0.033	2903
Form D after equating	25.314	6.955	-0.099	2.408			
Form E before equating	25.314	6.954	-0.102	2.312	0.147	0.096	2880
Form E after equating	24.731	6.821	0.026	2.275			
Form F before equating	24.731	6.822	0.031	2.385	0.073	0.110	2853
Form F after equating	25.452	6.511	-0.139	2.458			
Form G before equating	25.452	6.512	-0.140	2.483	0.098	0.025	2800
Form G after equating	24.804	6.585	-0.022	2.434			
	ACT Science (40) items						
Form A before equating	24.153	6.439	-0.192	2.553	0.148	0.042	2968
Form A after equating	22.661	7.077	0.200	2.543			
Form B before equating	22.659	7.064	0.170	2.373	0.180	0.170	2748
Form B after equating	22.227	6.964	0.231	2.400			
Form C before equating	22.330	6.964	0.232	2.431	0.058	0.031	2921
Form C after equating	24.122	6.640	-0.044	2.415			
Form D before equating	24.122	6.642	-0.048	2.496	0.065	0.081	2903
Form D after equating	22.965	6.515	0.061	2.477			
Form E before equating	22.965	6.515	0.060	2.463	0.033	0.014	2880
Form E after equating	22.374	6.334	0.175	2.495			
Form F before equating	22.374	6.334	0.173	2.443	0.020	0.052	2853
Form F after equating	22.439	7.073	0.110	2.426			
Form G before equating	22.439	7.072	0.111	2.405	0.038	0.021	2800
Form G after equating	24.153	6.438	-0.191	2.511			

Table 2. Average Absolute Bias, Standard Error and Root Mean Squared Error for the Licensure Subtest.

Sample Size = 250

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.041	0.457	0.462	0.027
Linear	0.113	0.388	0.410	0.020
Quad. Curve	0.051	0.403	0.410	0.021
Post Smooth 0.2	0.024	0.405	0.409	0.021
Post Smooth 0.5	0.069	0.387	0.399	0.019

Sample Size = 500

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.026	0.298	0.303	0.010
Linear	0.117	0.242	0.276	0.008
Quad. Curve	0.049	0.257	0.265	0.008
Post Smooth 0.2	0.049	0.257	0.267	0.008
Post Smooth 0.5	0.088	0.247	0.269	0.007

Sample Size = 2000

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.019	0.159	0.164	0.003
Linear	0.111	0.132	0.180	0.003
Quad. Curve	0.034	0.140	0.148	0.002
Post Smooth 0.2	0.024	0.143	0.150	0.002
Post Smooth 0.5	0.059	0.140	0.158	0.003

Table 3. Average Absolute Bias, Standard Error and Root Mean Squared Error for the ACT Reading Subtest.

Sample Size = 250

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.019	0.436	0.436	0.021
Linear	0.062	0.365	0.373	0.016
Quad. Curve	0.028	0.381	0.383	0.017
Post Smooth 0.2	0.034	0.380	0.382	0.017
Post Smooth 0.5	0.051	0.361	0.366	0.016

Sample Size = 500

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.013	0.301	0.301	0.009
Linear	0.063	0.249	0.259	0.007
Quad. Curve	0.031	0.264	0.266	0.008
Post Smooth 0.2	0.026	0.258	0.259	0.007
Post Smooth 0.5	0.047	0.244	0.250	0.007

Sample Size = 2000

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.010	0.151	0.152	0.002
Linear	0.068	0.124	0.145	0.002
Quad. Curve	0.039	0.132	0.139	0.002
Post Smooth 0.2	0.023	0.130	0.133	0.002
Post Smooth 0.5	0.046	0.122	0.133	0.002

Table 4. Average Absolute Bias, Standard Error and Root Mean Squared Error for the ACT English Test (A to B).

Sample Size = 500

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.081	1.146	1.154	0.139
Linear	0.750	0.919	1.230	0.138
Quad. Curve	0.150	0.978	1.001	0.102
Post Smooth 0.2	0.102	1.029	1.039	0.113
Post Smooth 0.5	0.334	0.984	1.052	0.110

Sample Size = 2000

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.072	0.592	0.598	0.037
Linear	0.685	0.471	0.904	0.057
Quad. Curve	0.123	0.504	0.538	0.028
Post Smooth 0.2	0.073	0.530	0.541	0.029
Post Smooth 0.5	0.218	0.513	0.566	0.031

Sample Size = 3000

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.038	0.470	0.476	0.024
Linear	0.733	0.370	0.849	0.044
Quad. Curve	0.121	0.396	0.434	0.019
Post Smooth 0.2	0.066	0.419	0.433	0.020
Post Smooth 0.5	0.193	0.406	0.462	0.023

Table 5. Average Absolute Bias, Standard Error and Root Mean Squared Error for the ACT Science Test (G to A).

Sample Size = 250

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.078	0.809	0.816	0.068
Linear	0.449	0.642	0.813	0.064
Quad. Curve	0.101	0.680	0.694	0.048
Post Smooth 0.2	0.056	0.721	0.728	0.052
Post Smooth 0.5	0.201	0.688	0.726	0.048

Sample Size = 500

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.024	0.600	0.603	0.037
Linear	0.461	0.483	0.695	0.042
Quad. Curve	0.101	0.516	0.530	0.028
Post Smooth 0.2	0.072	0.540	0.547	0.030
Post Smooth 0.5	0.215	0.527	0.579	0.033

Sample Size = 2000

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.020	0.298	0.302	0.009
Linear	0.463	0.238	0.539	0.019
Quad. Curve	0.103	0.254	0.281	0.008
Post Smooth 0.2	0.053	0.267	0.276	0.008
Post Smooth 0.5	0.132	0.261	0.299	0.009

Table 6. Average Absolute Bias, Standard Error and Root Mean Squared Error for the ACT Reading Test (A to B).

Sample Size = 250

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.056	0.969	0.973	0.103
Linear	0.351	0.787	0.883	0.081
Quad. Curve	0.339	0.854	0.932	0.097
Post Smooth 0.2	0.119	0.866	0.877	0.083
Post Smooth 0.5	0.258	0.787	0.839	0.076

Sample Size = 500

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.042	0.685	0.690	0.047
Linear	0.355	0.564	0.689	0.045
Quad. Curve	0.336	0.620	0.722	0.057
Post Smooth 0.2	0.101	0.623	0.635	0.039
Post Smooth 0.5	0.244	0.587	0.647	0.040

Sample Size = 2000

	Abs. Bias	S.E.	RMSE	s.e.(RMSE)
Unsmoothed	0.027	0.341	0.346	0.012
Linear	0.359	0.283	0.476	0.018
Quad. Curve	0.339	0.301	0.472	0.019
Post Smooth 0.2	0.102	0.314	0.335	0.011
Post Smooth 0.5	0.192	0.308	0.374	0.013

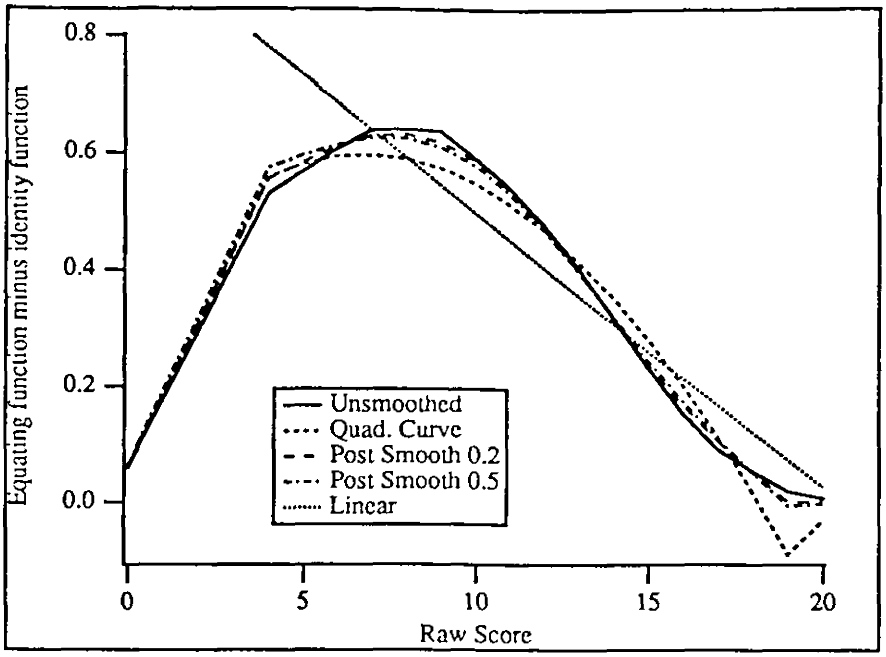

Figure 1. Equating functions for ACT Reading subscore using observed data.

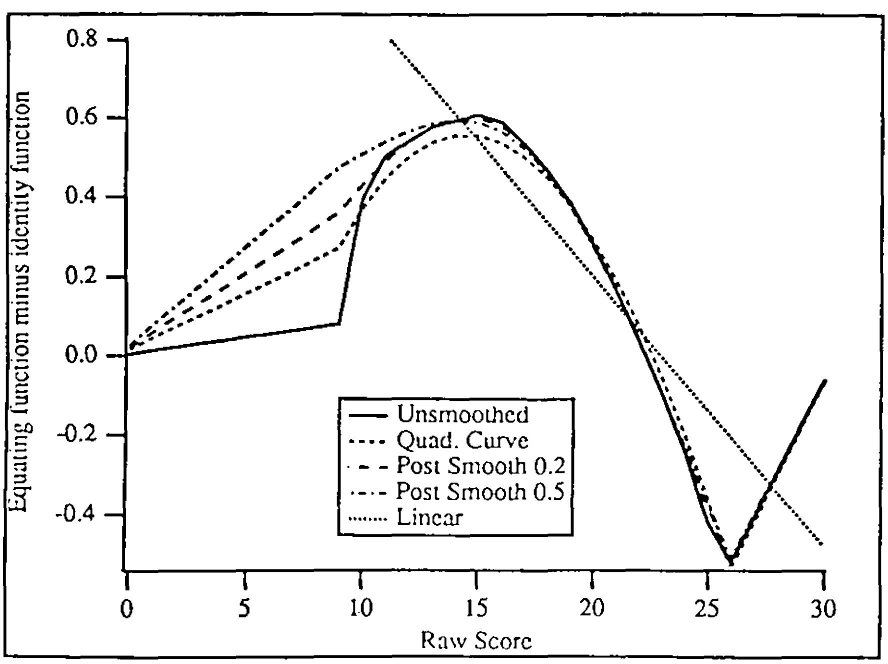

Figure 2. Equating functions for Licensure test using observed data.

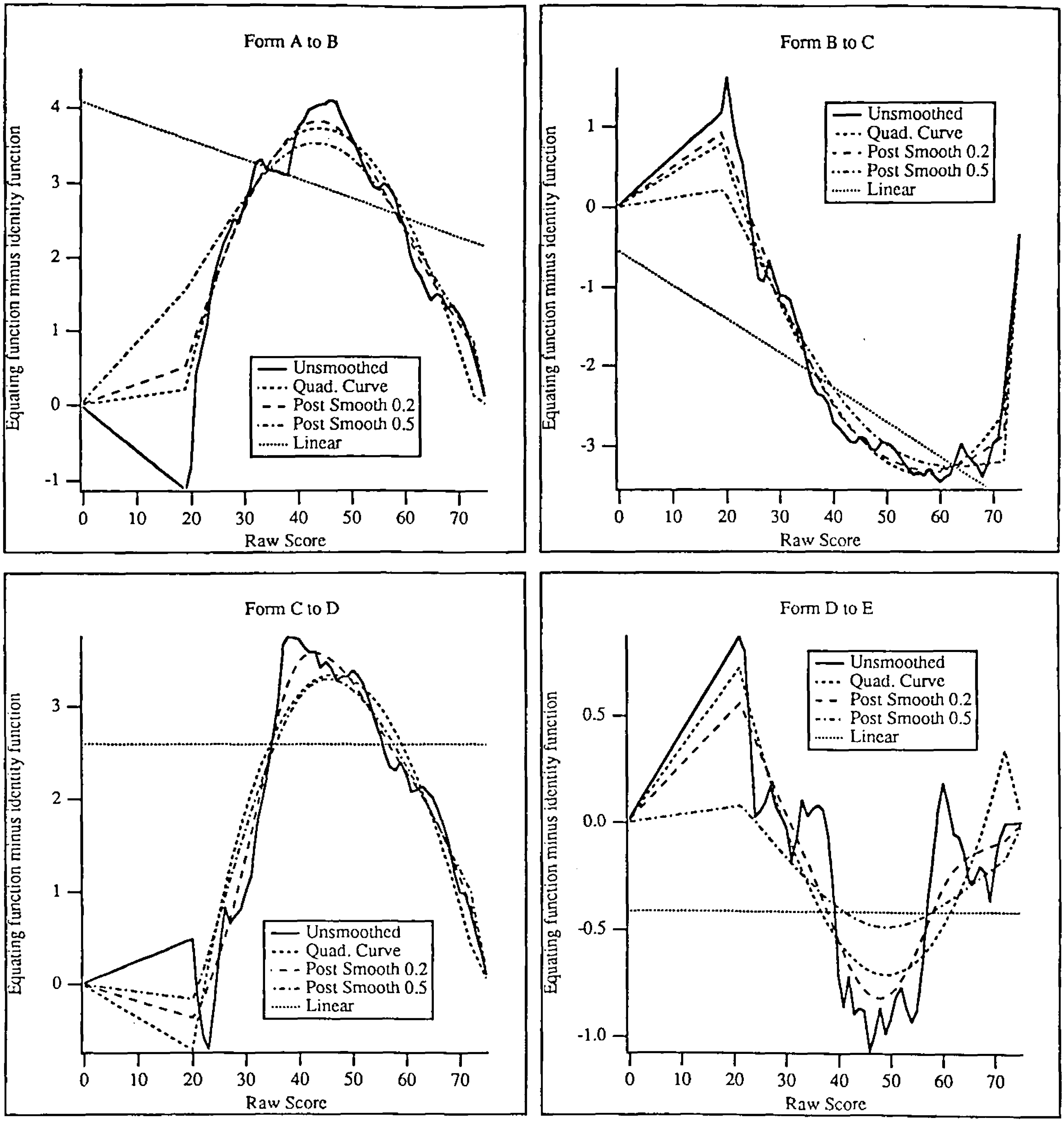

Figure 3. Equating functions for ACT English scores using observed data.

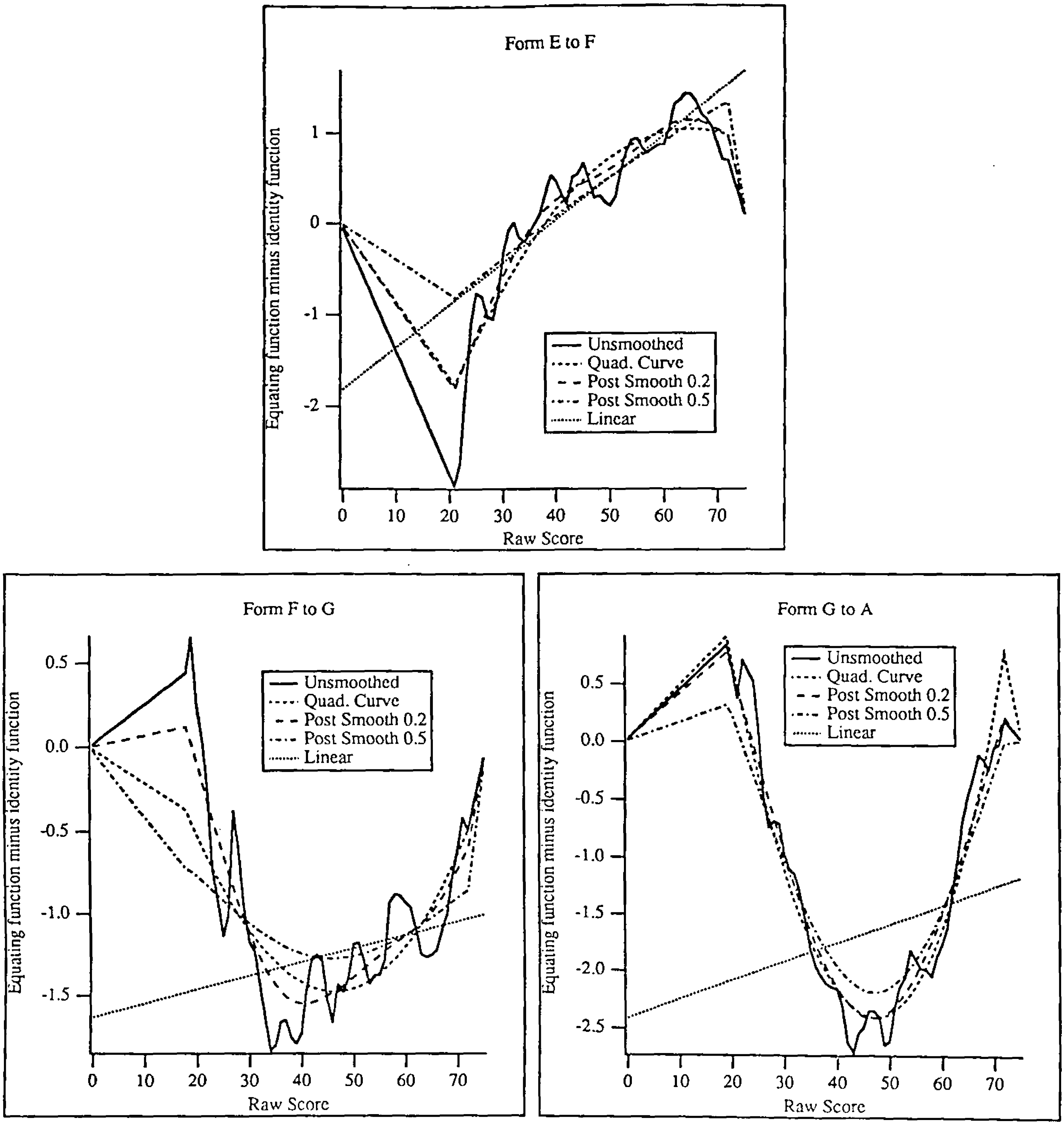

Figure 3 (continued). Equating functions for ACT English scores using observed data.

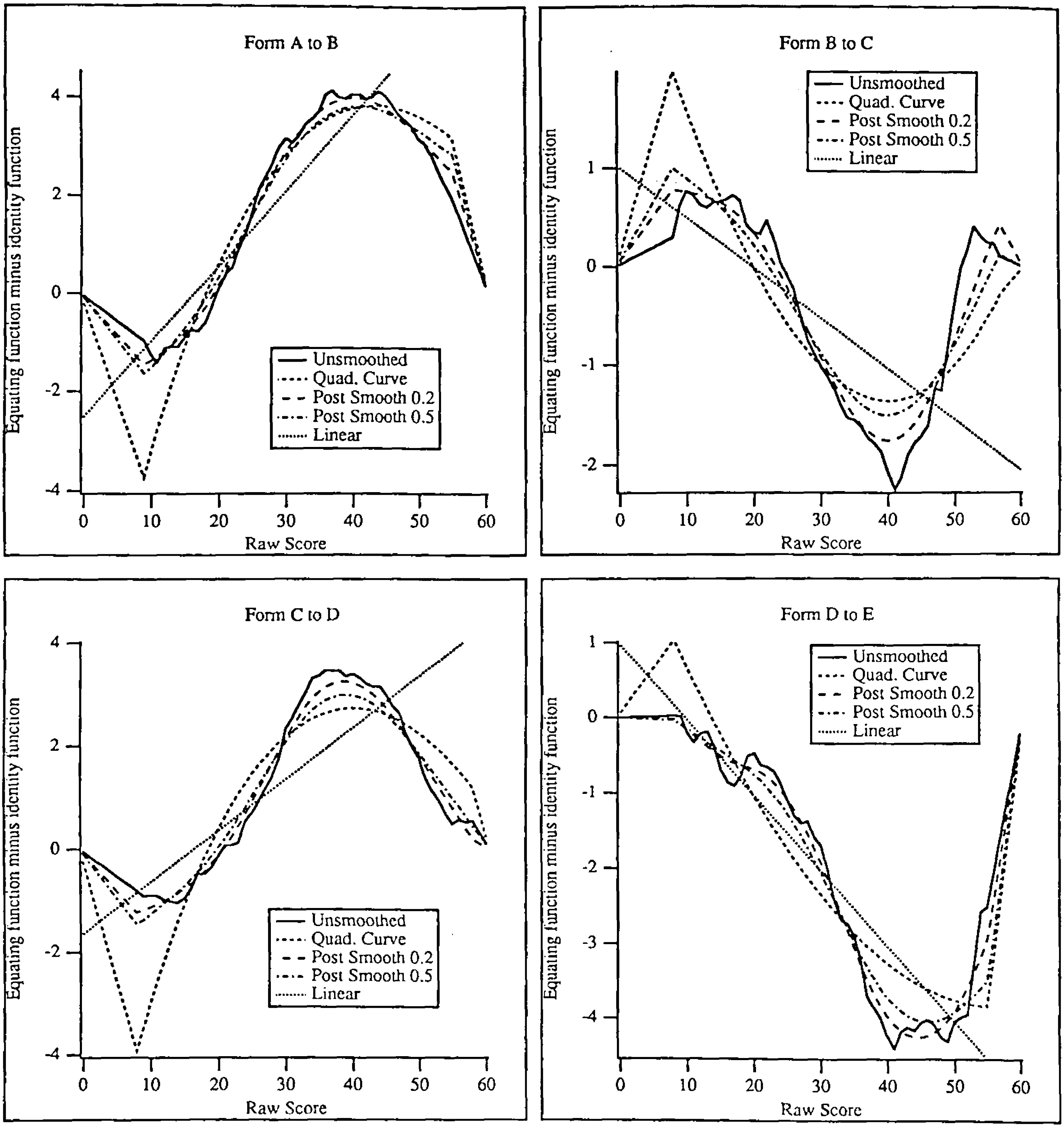

Figure 4. Equating functions for ACT Mathematics scores using observed data.

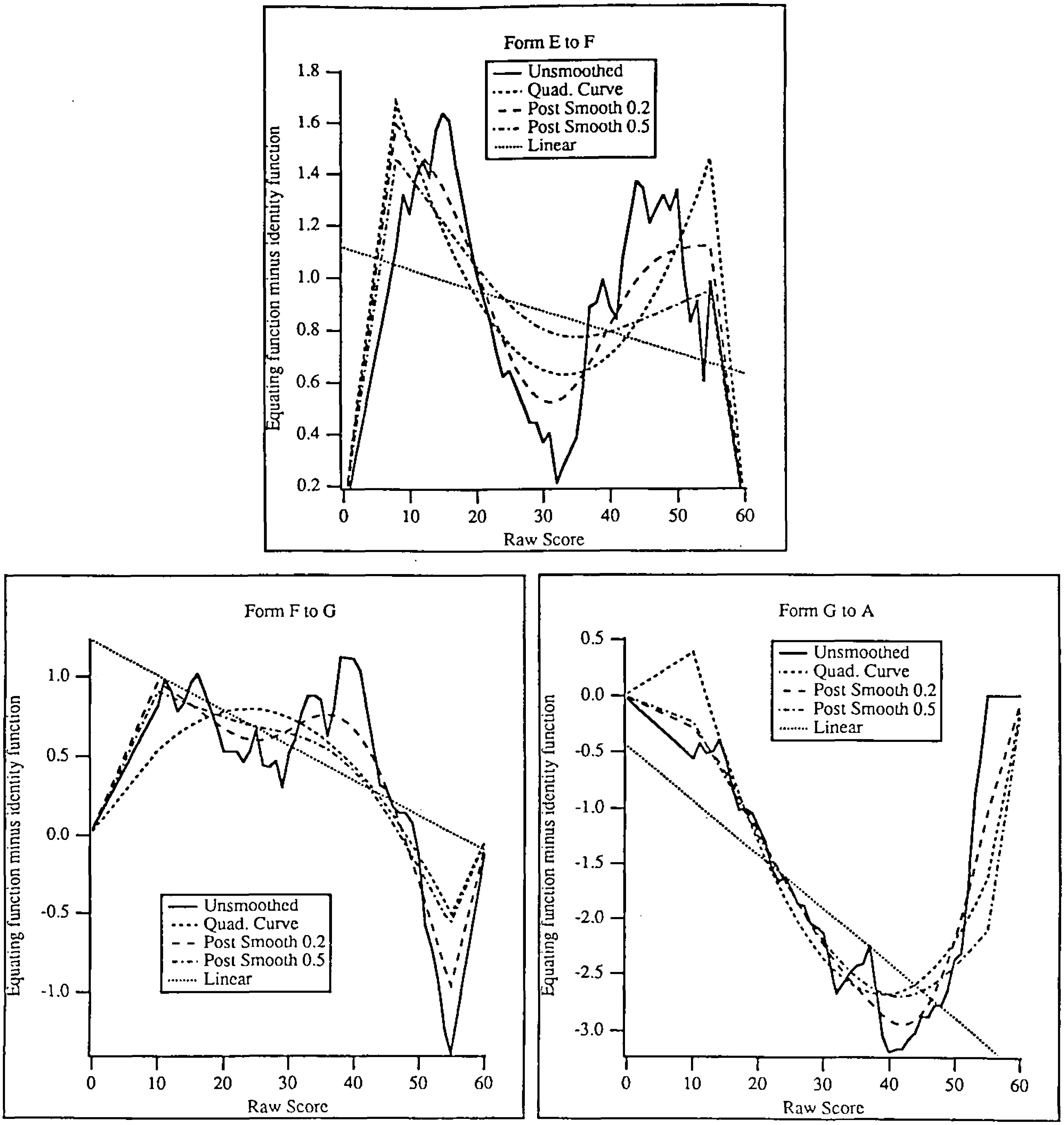

Figure 4 (continued). Equating functions for ACT Mathematics scores using observed data.

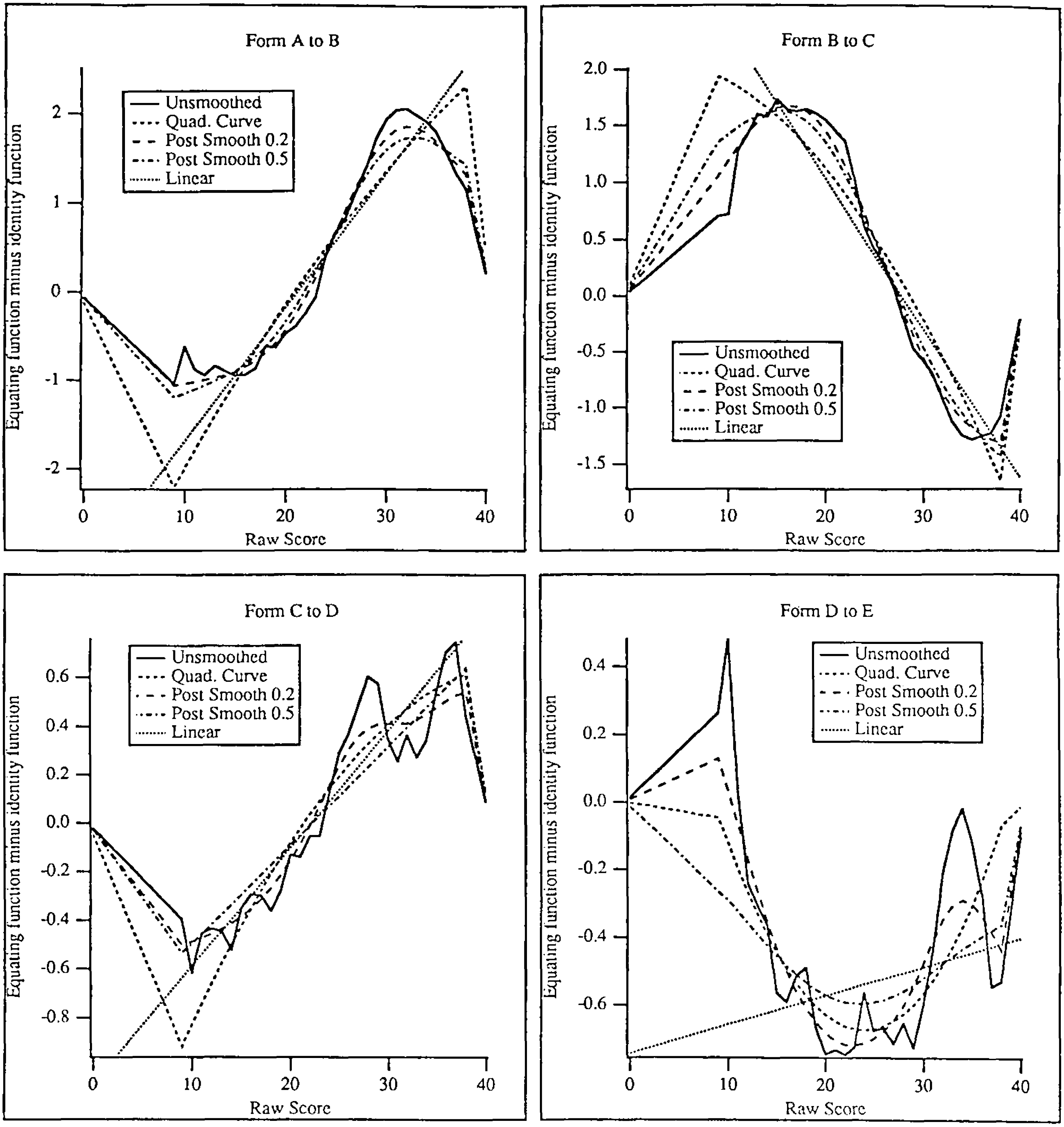

Figure 5. Equating functions for ACT Reading scores using observed data.

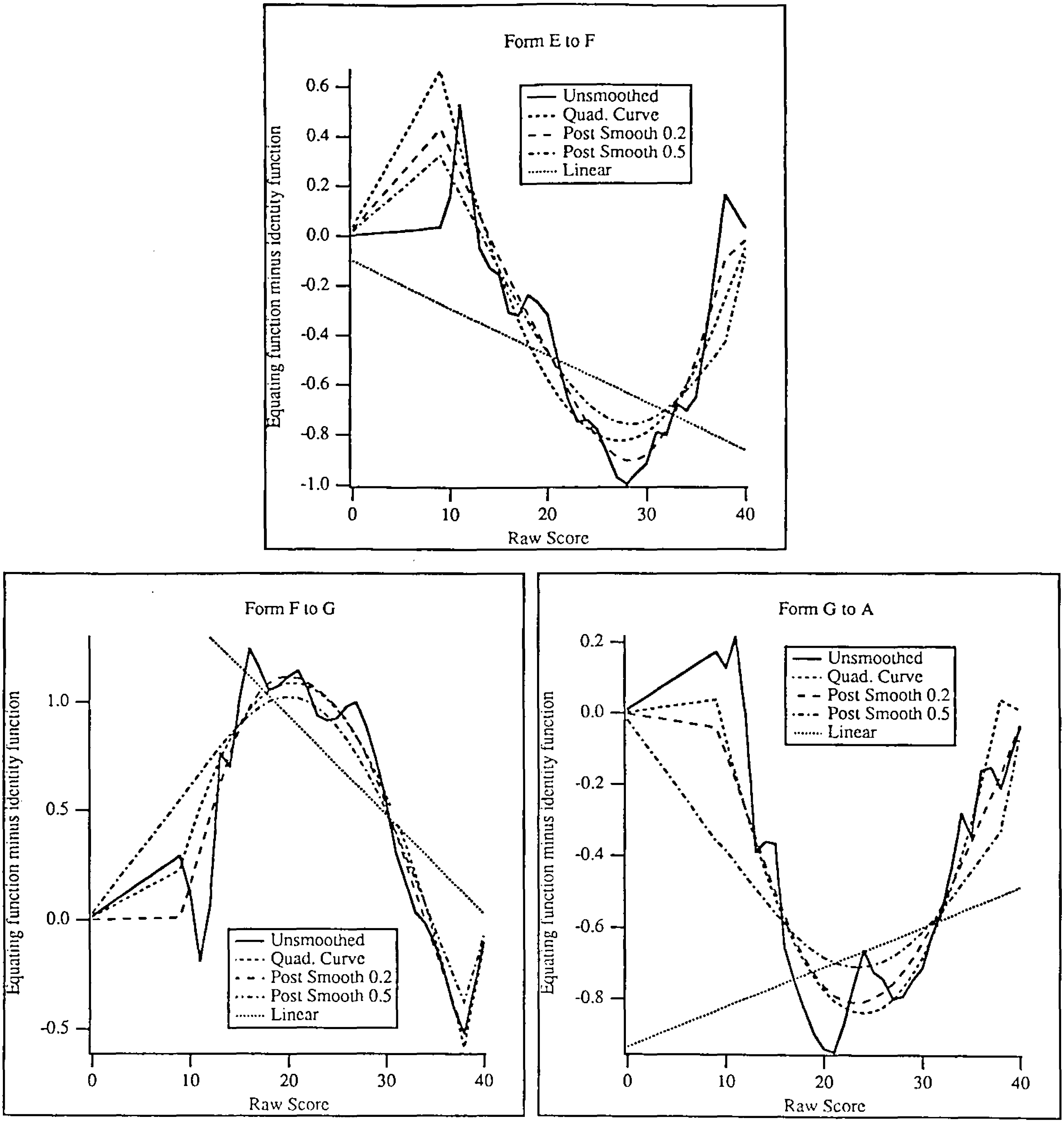

Figure 5 (continued). Equating functions for ACT Reading scores using observed data.

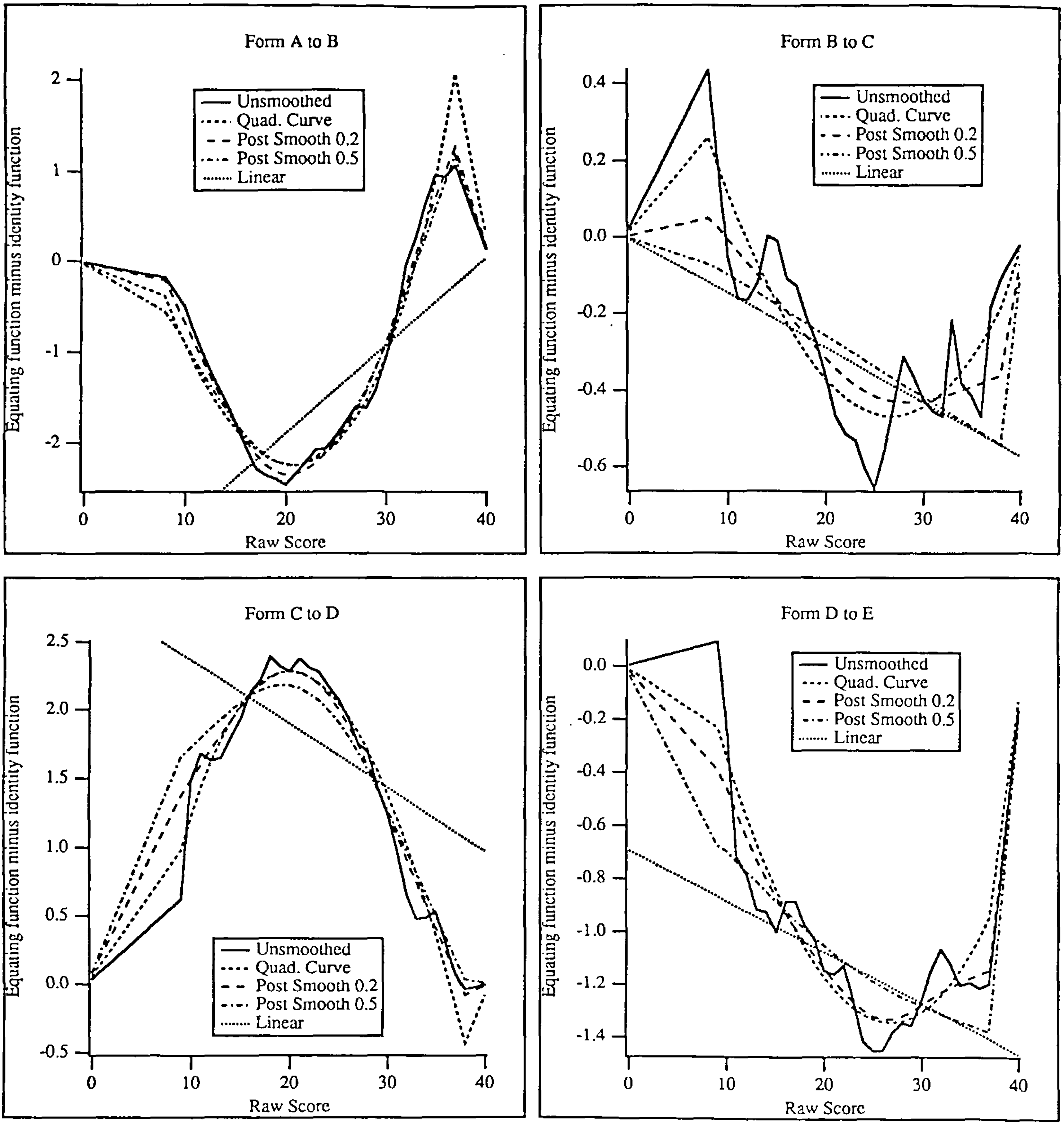

Figure 6. Equating functions for ACT Science scores using observed data.

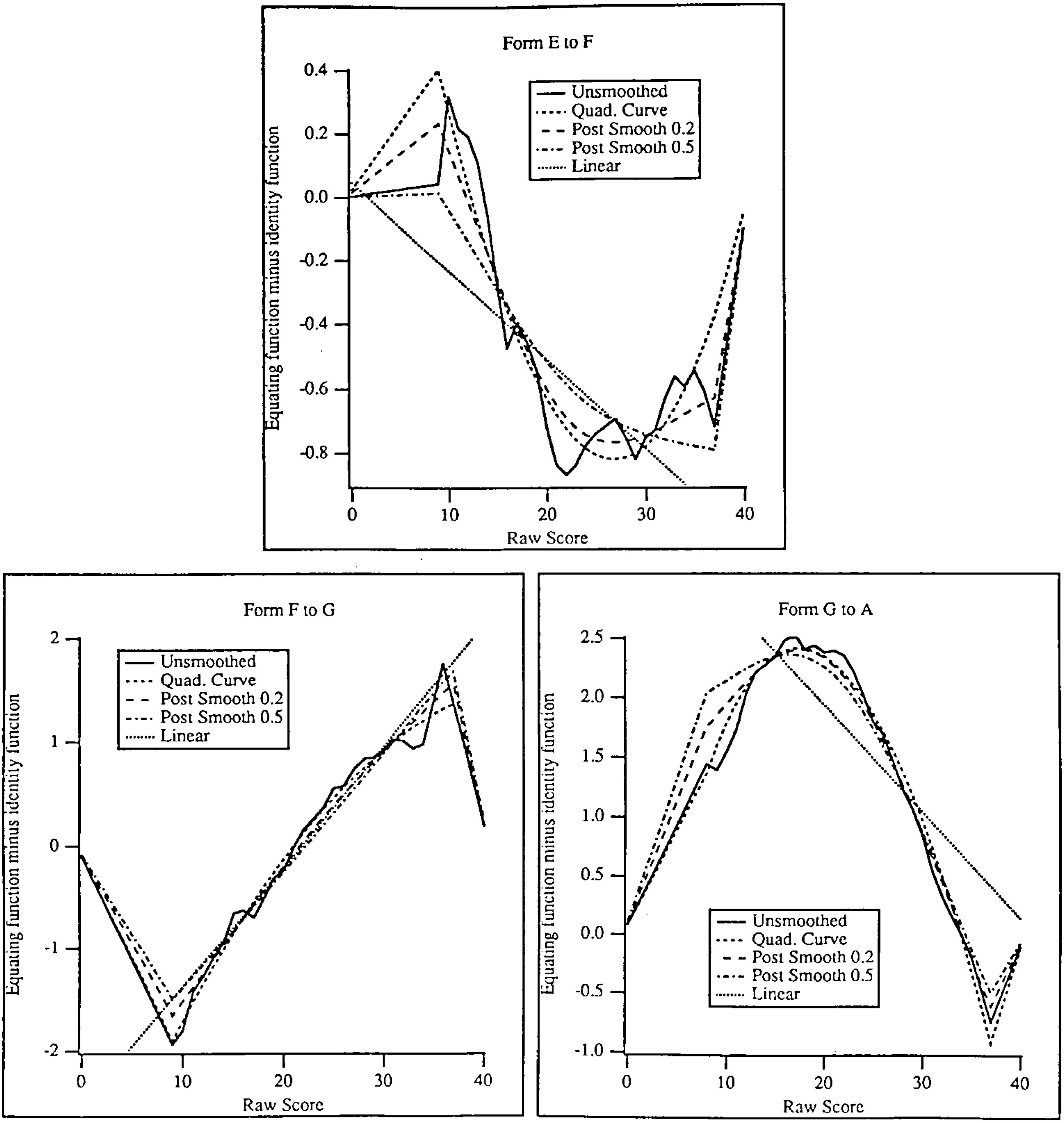

Figure 6 (continued). Equating functions for ACT Science scores using observed data.

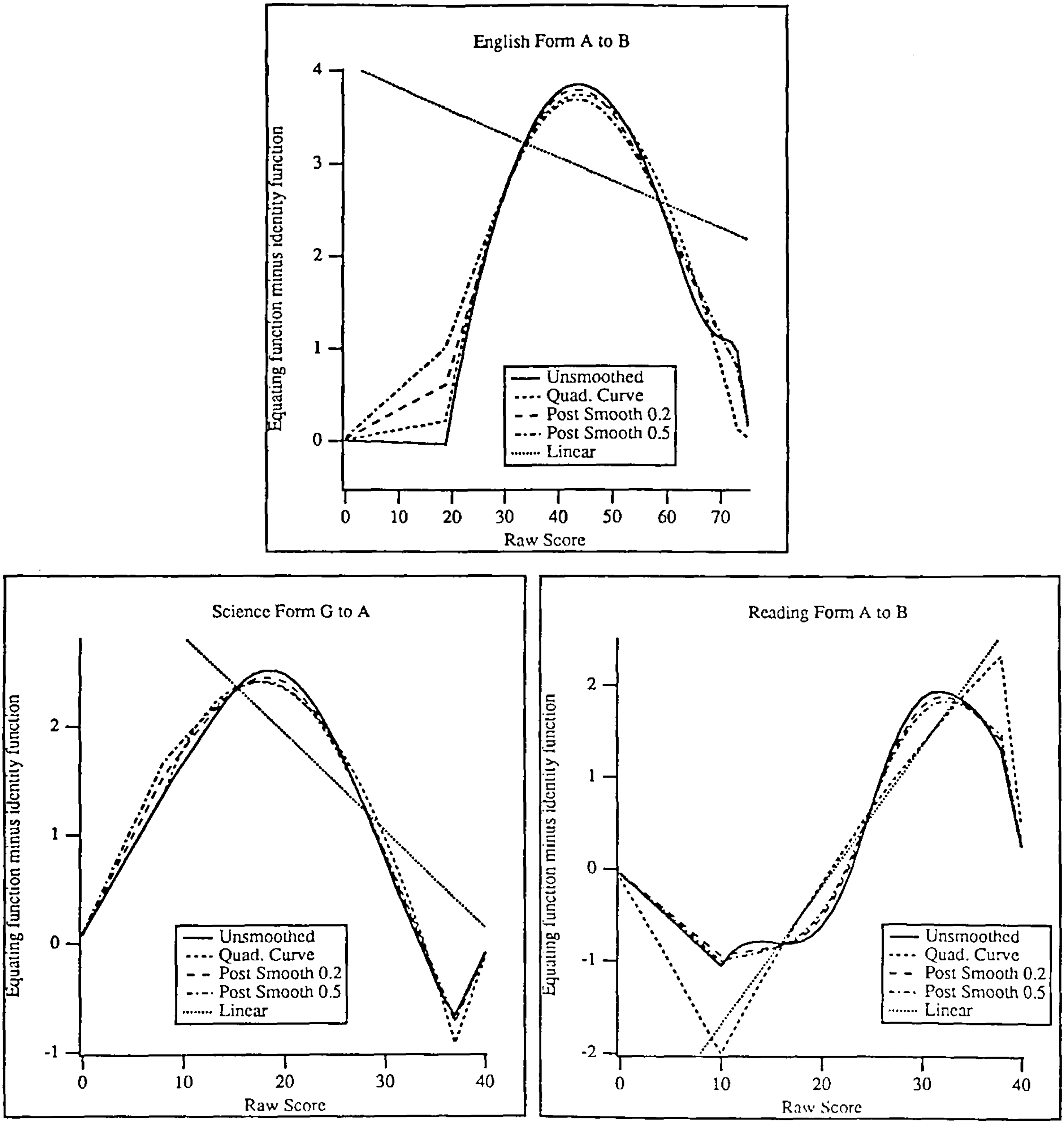

Figure 7. Population equatings for three pairs of distributions used in simulation.

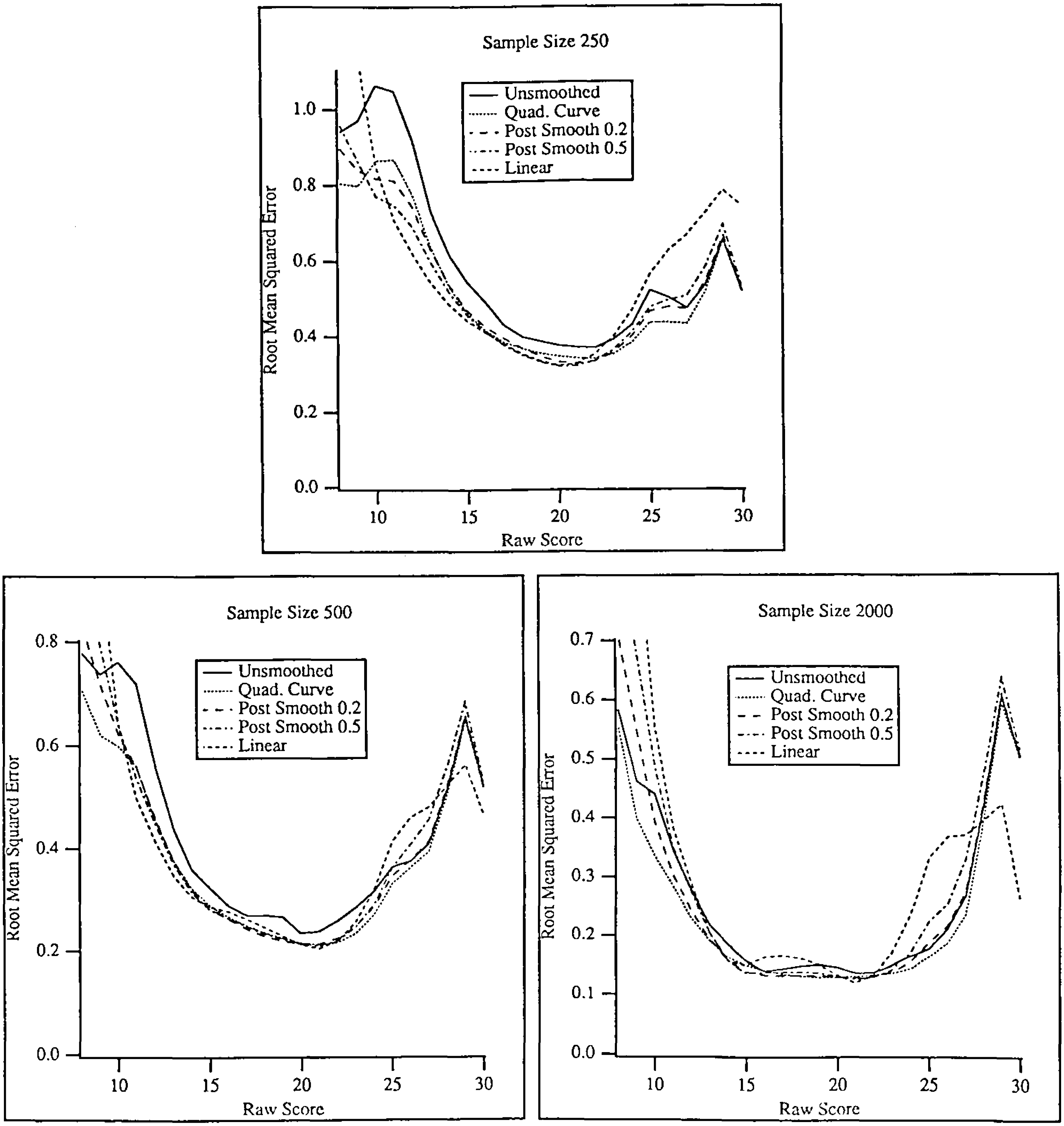

Figure 8. Root mean squared error of equating methods for Licensure test.

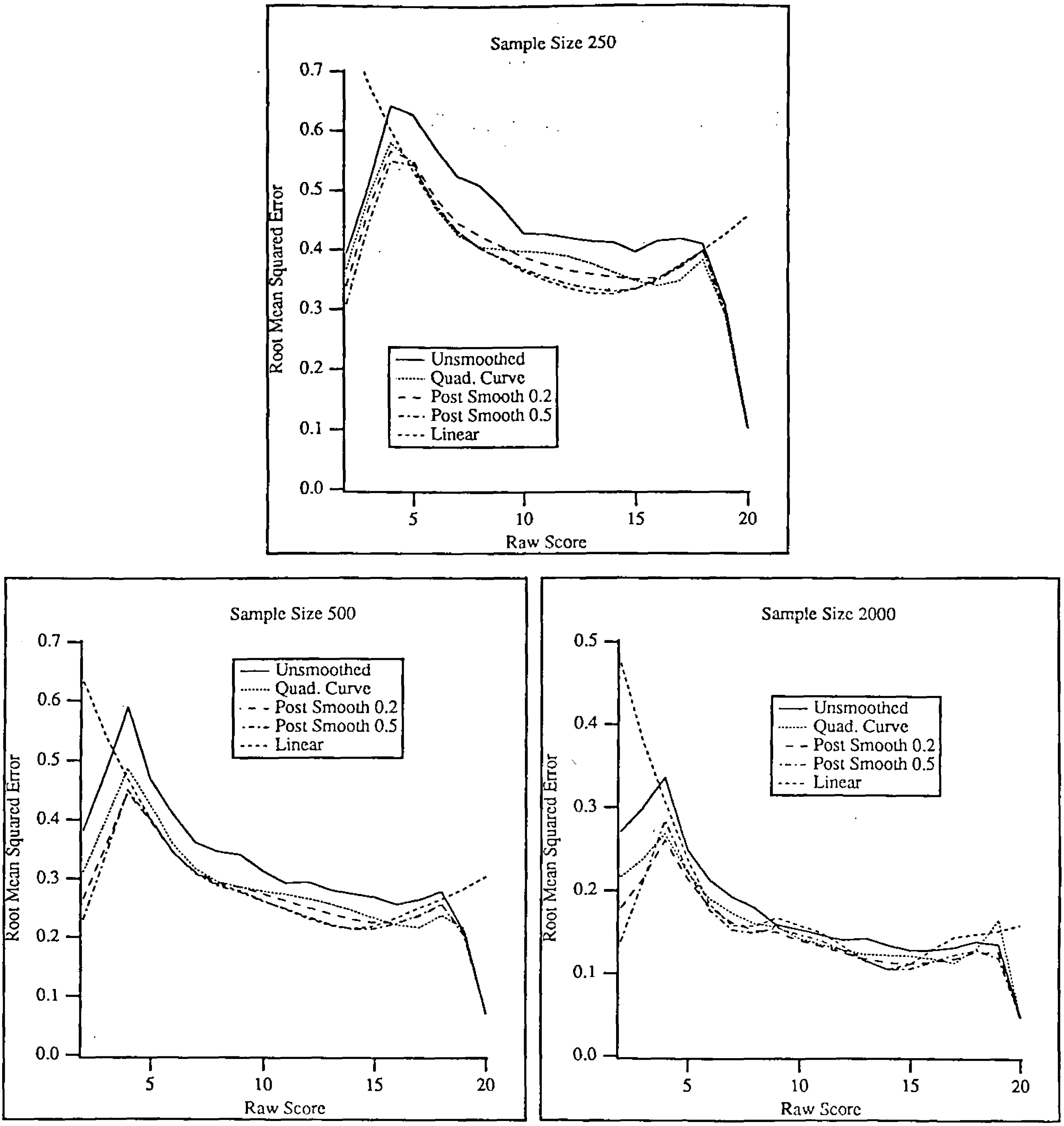

Figure 9. Root mean squared error of equating methods for ACT Reading subtest.

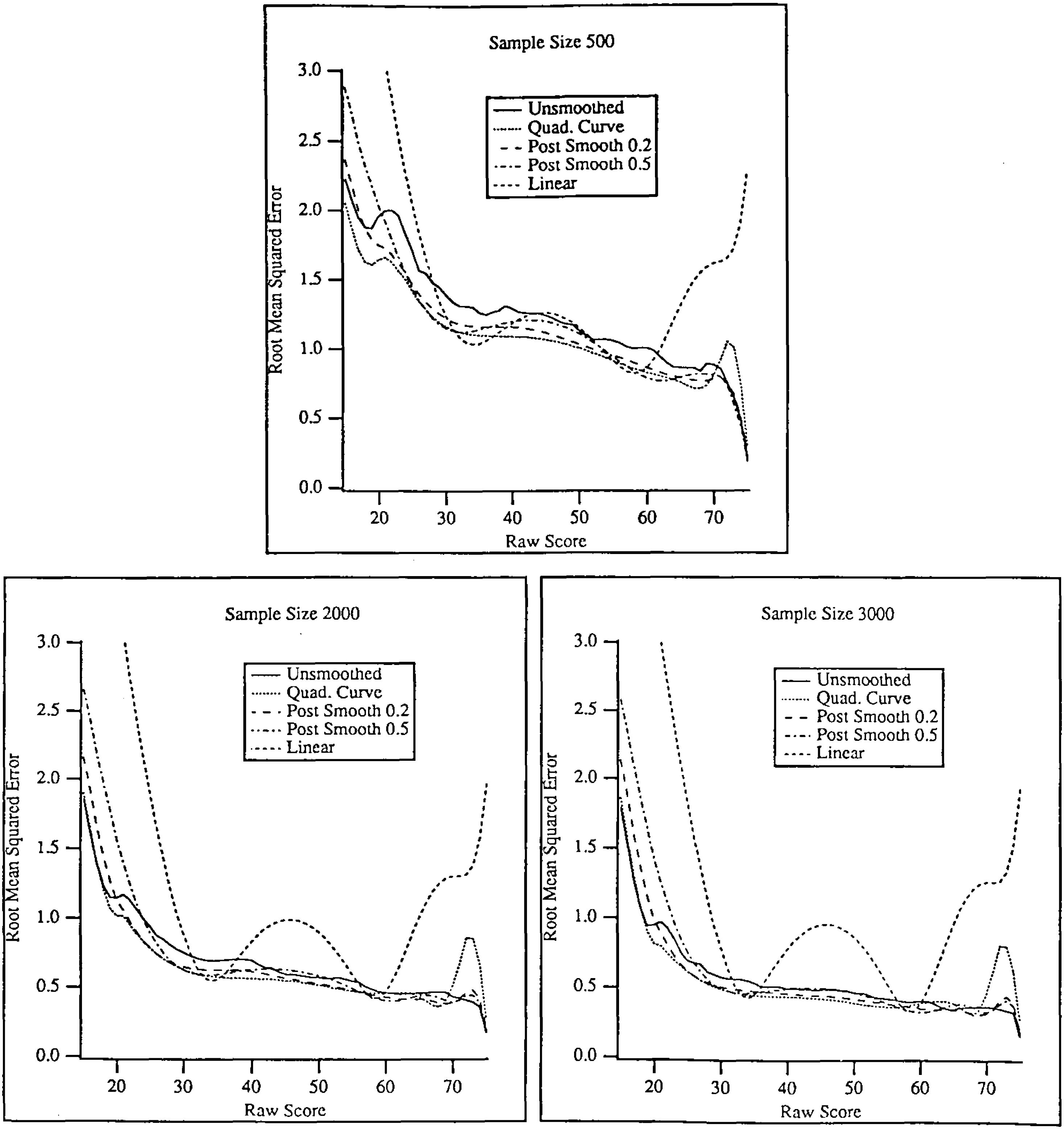

Figure 10. Root mean squared error of equating methods for ACT English test (A to B).

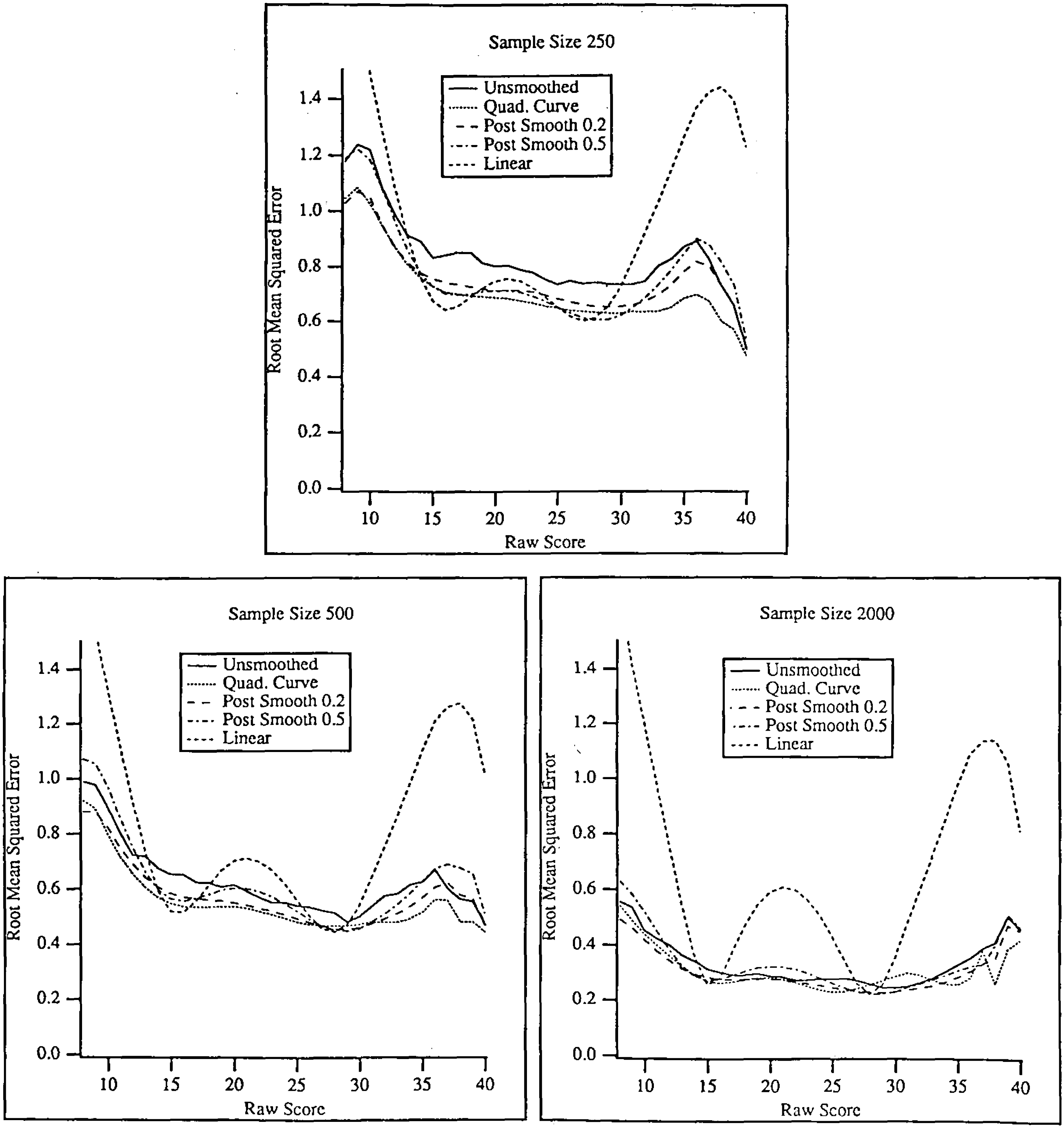

Figure 11. Root mean squared error of equating methods for ACT Science test (G to A).

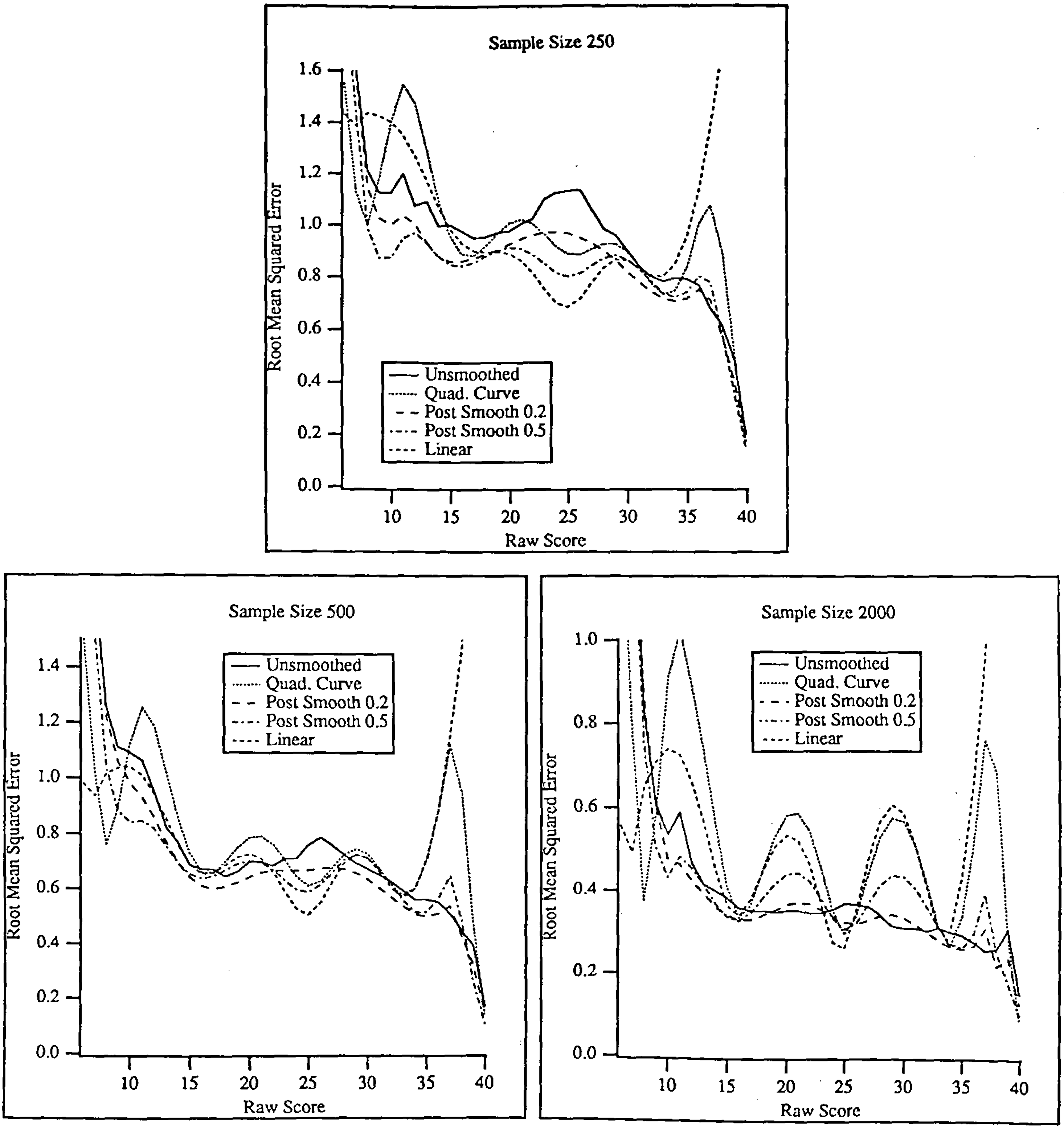

Figure 12. Root mean squared error of equating methods for ACT Reading test (A to B).