

The Optimal Degree of Smoothing in Equipercentile Equating With Postsmoothing

Lingjia Zeng

American College Testing

The effects of different degrees of smoothing on the results of equipercentile equating in the random groups design using a postsmoothing method based on cubic splines were investigated. A computer-based procedure was introduced for selecting a desirable degree of smoothing. The procedure was based on two criteria: (1) that the equating function is reasonably smooth, as evaluated by the second derivatives of the cubic spline functions, and (2) that the equated score distributions are close to that of the old form. The equating functions obtained from smoothing the equipercentile equivalents by a fixed smoothing de-

gree and a degree selected by the computer-based procedure were evaluated in computer simulations for four tests. The results suggest that no particular fixed degree of smoothing always led to an optimal degree of smoothing. The degrees of smoothing selected by the computer-based procedure were better than the best fixed degrees of smoothing for two of the four tests studied; for one of the other two tests, the degrees selected by the computer procedure performed better or nearly as well as the best fixed degrees. *Index terms: computer simulation, cubic spline, equating, equipercentile equating, smoothing.*

Smoothing is an effective way to improve equipercentile equating results. Two types of analytic smoothing methods, presmoothing and postsmoothing, have been used in the random groups equating design. Presmoothing refers to methods in which the score distributions for the two test forms to be equated are smoothed before equipercentile equating is performed. Postsmoothing refers to methods in which the equipercentile equating functions obtained from the unsmoothed score distributions are smoothed. Hanson, Zeng, & Colton (1991) compared three analytic presmoothing methods (based on the beta binomial model, the four-parameter beta binomial model, and a loglinear model) to an analytic postsmoothing method based on cubic splines, linear equating, and unsmoothed equipercentile equating. They found that both analytic presmoothing and postsmoothing methods can improve estimation of the equipercentile equating function and that the most accurate presmoothing and postsmoothing methods resulted in comparable levels of equating error.

In using postsmoothing methods, a smoothing degree is selected by the investigator. The effectiveness of smoothing degrees depends on sampling error and the dissimilarities of the two test forms to be equated. Subjective judgment is required to select a smoothing degree in order to achieve the most desirable equating performance for a particular pair of test forms. The selection is usually based on a number of factors, such as the graphs of the smoothed equipercentile equating functions and comparisons of the moments of the score distributions. Because subjective judgment is involved in equipercentile equating with smoothing, the quality of the equating process is determined, to a certain extent, by the expertise and experience of the investigator who performs the equating.

This study was concerned with the effects of various different degrees of smoothing on the results of equipercentile equating in the random groups design using a postsmoothing method based on cubic smoothing splines. A programmable approach is introduced for selecting a desirable smoothing degree. The equating results obtained from a set of commonly used fixed degrees of smoothing and a computer-selected degree

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 19, No. 2, June 1995, pp. 177-190

© Copyright 1995 Applied Psychological Measurement Inc.

0146-6216/95/020177-14\$1.95

177

of smoothing were evaluated by comparing the equating errors estimated by a monte carlo simulation.

The Equipercntile Postsmoothing Method

A procedure for equipercntile equating with postsmoothing based on cubic splines described by Kolen (1984) was used here. The following description of the basic methodology is from Kolen. A more detailed description of the cubic spline smoothing technique can be found in de Boor (1978).

Given $k + 1$ ordered pairs of variables (x_i, y_i) , $i = 0, 1, \dots, k$, where $x_0 < x_1 < \dots < x_k$, and k is the maximum number of integer score points, the cubic spline function for the interval $x_i \leq x < x_{i+1}$ is

$$f_i(x) = a_{0i} + a_{1i}(x - x_i) + a_{2i}(x - x_i)^2 + a_{3i}(x - x_i)^3. \quad (1)$$

The coefficients of the cubic polynomial function a_{0i} , a_{1i} , a_{2i} , and a_{3i} are allowed to change from one interval to the next so that the overall cubic spline function has continuous second derivatives over the range x_0 to x_k . The function over the range x_0 to x_k is referred to as $f(x)$. The points at which the function is allowed to change (i.e., x_0, x_1, \dots, x_k) are referred to as "knots" or "ducks." A smoothing cubic spline function is defined as a function that minimizes

$$\int_{x_0}^{x_k} [f''(x)]^2 dx \quad (2)$$

among all the cubic spline functions that satisfy the condition of

$$\frac{1}{k} \sum_{i=0}^k \left[\frac{f(x_i) - y_i}{\delta(y_i)} \right]^2 \leq s, \quad (3)$$

where

f'' is the second derivative of $f(x)$,

$\delta(y_i)$ is a constant used to scale the deviations, and

s is a constant used to control the degree of smoothing.

Suppose two test forms, Form X and Form Y, are to be equated. Each form is administered to a separate randomly equivalent group of examinees. For each $k + 1$ possible integer score on Form X (x_i), a linearly interpolated equipercntile equivalent on Form Y (y_i) is calculated. A smoothing cubic spline function $f(x)$ is fit to y_i . Standard errors of y_i are used for the $\delta(y_i)$ values in Equation 3. A computer program written in the C language, based on the algorithm described by Reinsch (1967), was developed to solve $f(x)$. A second smoothing cubic spline function is fit to the equipercntile equating function relating scores on Form Y to Form X. The smoothed equipercntile equating function is taken as the average of the smoothing cubic spline relating scores on Form X to Form Y and the inverse of the smoothing cubic spline relating scores on Form Y to Form X. Thus, the smoothed equipercntile equating function is a symmetric function describing the equipercntile relationship between Form X and Form Y.

In smoothing equipercntile equivalents, the standard errors of the unsmoothed equivalents are used in Equation 3 to scale the deviation. Because the estimates of these standard errors are based on asymptotic results and few examinees typically obtain scores at the extreme ends of the score range, it is possible that these estimates are inaccurate. To avoid this problem, Kolen (1984) recommended smoothing those equivalents of integer scores with percentile ranks between .5 and 99.5. Equivalents of scores outside these limits are linearly interpolated.

Selecting the Optimal Degree of Smoothing

The goal of smoothing is to reduce the noise in data caused by random sampling errors. The degree of smoothing is determined by the value of s in Equation 3. Selecting $s = 0$ leaves the equipercntile function

computed from the observed score distributions unchanged. Selecting a large value for s has the potential of reducing the random errors but may introduce bias that may inflate the equating errors. For this reason, the selection of s is very important in implementing the postsmoothing method for equipercentile equating.

Based on graphical inspection of American College Testing (ACT) Assessment Program (AAP) equatings, Kolen (1984) found that s values between .05 and 1.0 produced equatings that were reasonably smooth without deviating from the unsmoothed equipercentile relationship to a great extent. In a typical equating, a finite set of smoothings is performed on the equipercentile equivalents with a set of different s values. Then a decision is made to select a smoothed equating function with a particular s value. In observed score equating, the selection is based on two considerations: (1) the graphs of the smoothed equipercentile equating functions, and (2) comparisons of the score distributions of the old form and the equated form. An optimal degree of smoothing is expected to produce equipercentile equivalents that are reasonably smooth and that do not depart too much from the unsmoothed equivalents. Also, the moments of the equated score distribution of the new form are similar to the moments of the old form if the equating is satisfactory.

For the purpose of developing an algorithm for selecting the optimal smoothing degree from a set of smoothed equipercentile equating functions, an optimal smoothed equipercentile equating function is defined as one that satisfies two conditions: (1) that the equating function is reasonably smooth, and (2) that the distribution of the equated form is close to that of the old form. The methods described below were used to examine the smoothed equipercentile functions obtained with a set of s values and to identify the optimal smoothing degree of those investigated.

Evaluating the Smoothness of the Equating Function

A smoothed equipercentile equating function can be graphically represented by a curve. The smoothness of the equating function can be indexed by the changing rate of the curvature of the curve at various points. The curve of an inadequately smoothed equipercentile equating function is expected to have some points at which the curvature changes fast at two consecutive score points. As defined above, the postsmoothed equipercentile equating function of equating Form X to Form Y is the average of the smoothing cubic spline function relating scores on Form X to Form Y and the inverse of the smoothing cubic spline function relating scores on Form Y to Form X. The smoothness of an equipercentile equating function can be investigated by examining the curvature of the two cubic spline functions.

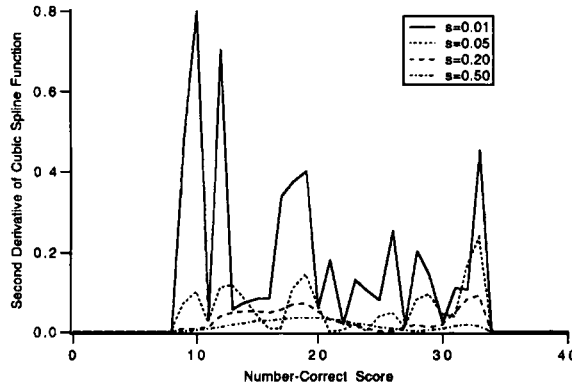
The second derivative of a twice differentiable function evaluated at a particular point reflects the curvature at that point. The second derivative of the cubic spline function for the interval $x_i \leq x < x_{i+1}$ (Equation 1) is

$$f_i''(x) = 2a_{2i} + 6a_{3i}(x - x_i). \quad (4)$$

Example. The following example illustrates the relationship between the second derivatives of a cubic spline function and its smoothness. A 40-item AAP science reasoning test was used. The second derivatives of four cubic spline functions (using $s = .01, .05, .20,$ and $.50$) that fit the equipercentile equivalents are shown in Figure 1. For $s = .01$ and $.05$, the second derivatives changed very fast at approximately score points 10, 16, 19, 27, and 32, which indicates that smoothing was inadequate. The second derivatives of the cubic spline function using $s = .20$ and $.50$ were quite stable, which indicates adequate smoothing.

The unsmoothed equipercentile equivalents and three cubic spline functions (using $s = .05, .20,$ and $.50$) are shown in Figure 2. In this example, smoothing was not performed for scores less than 9 and greater than 37 because of too few observations for these scores. The comparisons were focused on the score range between 9 and 37. Two obvious irregular sections (17–22 and 31–34) on the curve for $s = .05$ can be observed. These two sections corresponded to the large second derivatives for the cubic spline function with $s = .05$, as shown in Figure 1. The curves for $s = .20$ and $s = .50$ appear to be smoother than that for $s = .05$.

Figure 1
Second Derivatives of Four Cubic Spline Functions For an AAP Science Test



However, the curve for $s = .50$ obviously deviated from the unsmoothed plot around score points 10, 19, and 33, which indicates that the curve might have been oversmoothed.

A computing algorithm. The following algorithm can be easily programmed to allow a computer to evaluate the smoothness of a cubic spline function.

Step 1. Compute r_i , the ratio of the second derivatives at score points i and $i + 1$:

$$\text{if } f_i''(x) > f_{i+1}''(x) \text{ then } r_i = \frac{f_i''(x)}{f_{i+1}''(x)} \text{ else } r_i = \frac{f_{i+1}''(x)}{f_i''(x)}, \text{ for } i = n_1 \text{ to } n_2, \quad (5)$$

where n_1 and n_2 are the lowest and highest score points within the smoothing range, respectively.

Step 2. Compare r_i with a predetermined parameter τ (see below for an explanation of τ).

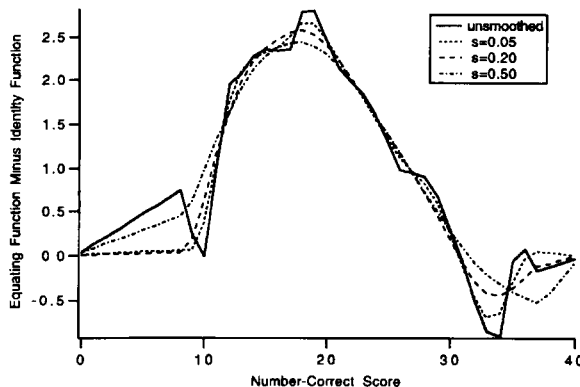
If r_i is greater than τ , then a bumping point (B) is detected at i :

Set $B = 0$.

if $r_i > \tau$ then increment B by 1, for $i = n_1$ to n_2 .

This procedure for evaluating smoothness is performed on the cubic spline function relating scores on Form X to Form Y and on the cubic spline function relating scores on Form Y to Form X. The total number of bumping points in the two spline functions is recorded by variable B for each smoothed equipercentile

Figure 2
Four Equipercentile Equating Functions For an AAP Science Test



equating function with a particular s . If $B = 0$, then the smoothed equipercentile equating function is considered adequately smoothed. The parameter τ is a criterion for determining a bumping point. A small value of τ tends to favor a large amount of smoothing and a large value of τ tends to favor a small amount of smoothing. A subjective judgment is required to select a τ value.

Comparing the Score Distributions

A second criterion for selecting an optimal smoothed equating function is the closeness of the equated score distribution to that of the distribution of the old form. Let D_j be an index indicating the closeness of the moments of score distributions smoothed with s_j . Let $\theta_{ij}[\hat{e}_Y(X)]$, where X is the random variable for a score on Form X , be the estimate of a central moment of the equated score distribution smoothed with s_j , where $i = 1$ for mean, $i = 2$ for standard deviation, $i = 3$ for skewness, and $i = 4$ for kurtosis. Let $j = 1, \dots, J$ be the number of different smoothing degrees used. Let $\theta_i(Y)$ be the estimate of a central moment of the score distribution for the old form, where i is the same as defined for $\theta_{ij}[\hat{e}_Y(X)]$. The index D_j is computed using the following four-step algorithm for all combinations of $i = 1, \dots, 4$ and $j = 1, \dots, J$:

Step 1. Define

$$\delta_{ij} = |\theta_{ij}[\hat{e}_Y(X)] - \theta_i(Y)|, \quad \text{for } i = 1 \text{ to } 4, \quad j = 1 \text{ to } J. \quad (6)$$

Step 2. For an i , find \max_i among δ_{ij} .

Step 3. Define

$$\gamma_{ij} = 1 - \frac{\delta_{ij}}{\max_i}, \quad \text{for } i = 1 \text{ to } 4, \quad j = 1 \text{ to } J. \quad (7)$$

Step 4. Define

$$D_j = \sum_{i=1}^4 w_i \gamma_{ij}, \quad \text{for } i = 1 \text{ to } 4, \quad j = 1 \text{ to } J, \quad (8)$$

where w_i is a predetermined parameter.

The parameter w_i weights the importance of moment i in D_j . In the present study, $w_1 = .40$ (the weight used for the mean), $w_2 = .3$ for standard deviation, $w_3 = .15$ for skewness, and $w_4 = .15$ for kurtosis. The selection of these w_i s was intended to parameterize the nonanalytical rules used in AAP equatings for comparing equated score distributions with that of the old form.

Final Decision

Based on the indexes B_j and D_j , the optimal index O_j is computed by

$$O_j = D_j - .02B_j, \quad \text{for } j = 1 \text{ to } J, \quad (9)$$

where $.02$ was a constant used to make a selected degree of smoothing similar to what is typically used in AAP equatings under similar situations. The smoothed equating function with the largest O_j is selected. The following example illustrates how these indexes can be used to select a degree of smoothing.

For an AAP English test equating, Form X was equated to Form Y . The equipercentile equivalents were smoothed with eight different smoothing degrees ($s = .01, .05, .10, .15, .20, .30, .50, \text{ and } 2.00$). The first four central moments of the number-correct score and the equated score distributions are reported in Table 1. The indexes D , B , and O were computed and are also listed in Table 1.

According to index D , the equated score distribution smoothed with the smallest degree ($s = .01$) resembles the Form Y distribution the most. As the degree of smoothing increased, the equated score distribution deviated more from the Form Y distribution. According to index B , the cubic spline function had 13 bumping points for $s = .01$, 12 bumping points for $s = .05$, and so forth (see Table 1). The cubic spline

Table 1
 First Four Moments, *D*, *B*, and *O* From an AAP English Test Equating
 (*N* = 3,411 for Form Y and *N* = 3,129 for Form X)

| Distribution and Form/ Smoothing Degree | Mean | Standard Deviation | Skewness | Kurtosis | <i>D</i> | <i>B</i> | <i>O</i> |
|--|---------|-----------------------|----------|----------|----------|----------|----------|
| Number-Correct Score Distribution | | | | | | | |
| Form Y | 50.8877 | 11.1492 | -.2955 | 2.6248 | | | |
| Form X | 49.1934 | 12.5260 | -.2292 | 2.3732 | | | |
| Equated Number-Correct Score Distribution | | | | | | | |
| <i>s</i> = .01 | 50.8886 | 11.1486 | -.2950 | 2.6221 | 1.00 | 13 | .74 |
| <i>s</i> = .05 | 50.8894 | 11.1500 | -.2950 | 2.6217 | .99 | 12 | .75 |
| <i>s</i> = .10 | 50.8893 | 11.1522 | -.2951 | 2.6202 | .99 | 10 | .79 |
| <i>s</i> = .15 | 50.8894 | 11.1707 | -.2937 | 2.6057 | .92 | 7 | .78 |
| <i>s</i> = .20 | 50.8908 | 11.1916 | -.2924 | 2.5894 | .84 | 0 | .84 |
| <i>s</i> = .30 | 50.8895 | 11.2440 | -.2937 | 2.5589 | .79 | 0 | .79 |
| <i>s</i> = .50 | 50.8828 | 11.3047 | -.2941 | 2.5244 | .65 | 0 | .65 |
| <i>s</i> = 2.00 | 50.8534 | 11.3974 | -.2870 | 2.4702 | 0.00 | 0 | 0.00 |

functions were adequately smoothed for $s \geq .20$. The index *O* was the largest for $s = .20$. Therefore, .20 was the selected degree of smoothing for this equating process.

Computer Simulations

Method

Computer simulations were conducted to evaluate the performance of the proposed method for selecting a degree of smoothing. The simulations were conducted as follows:

1. A pair of population distributions was specified and an equipercentile equating was conducted with the population score distributions.
2. A pair of sample score distributions was randomly generated from the population distributions.
3. Equipercentile equating was performed with the score distributions of the generated random sample, the equated score distribution was smoothed with eight different smoothing degrees ($s = .01, .05, .10, .15, .20, .30, .50, \text{ and } 2.00$), and a smoothing degree was selected using the procedures described above.
4. Steps 2 and 3 were replicated 500 times.

Equating errors were identified by comparing the equating results obtained from the random samples with the equating results obtained from the population distributions. Three sample sizes (*N*) were used: $N = 500$, $N = 1,000$, and $N = 3,000$. Four AAP tests were used: English, mathematics, reading, and science reasoning.

τ was selected, based on inspection of previous AAP equatings, so that the degree of smoothing, s , selected by the proposed method was similar to that used in practice. Specifically, $\tau = 1.5$ was used for the 75-item AAP English test, $\tau = 1.7$ was used for the 60-item AAP mathematics test, and $\tau = 1.9$ was used for the 40-item reading and science reasoning tests.

The population score distributions used in the simulations were estimated by fitting the observed score distributions obtained from the AAP test administration with a loglinear model (Agestri, 1990; Hanson, 1991). The reason for using the smoothed distributions instead of the observed number-correct score distributions is that the smoothed distributions usually represent the population distributions better than the observed number-correct score distributions (Kolen, 1991). χ^2 tests were performed to determine if the model fit the data. The χ^2 values of the tests of fit, their degrees of freedom, and *p* values are reported in Table 2. These data suggest that the model fit the data very well. The fitted distributions were used as population distributions for the simulation.

The average squared bias (ASB) and mean square error (MSE) were computed. Let $e_{\gamma}(i)$ be the Form Y observed score equivalent to the Form X observed score *i* given by the equating function obtained from the

Table 2
Goodness-of-Fit Test Statistics From Equatings of
English, Mathematics, Reading, and Science Tests

| Test and Form | Number of Moments Fit | χ^2 | Degrees of Freedom | p |
|--------------------|--------------------------|----------|-----------------------|-----|
| English | | | | |
| Form Y | 5 | 58.43 | 70 | .84 |
| Form X | 5 | 54.81 | 70 | .91 |
| Mathematics | | | | |
| Form Y | 5 | 54.29 | 55 | .50 |
| Form X | 5 | 54.53 | 55 | .82 |
| Reading | | | | |
| Form Y | 4 | 24.86 | 36 | .92 |
| Form X | 5 | 42.88 | 35 | .17 |
| Science | | | | |
| Form Y | 4 | 34.64 | 36 | .53 |
| Form X | 4 | 39.97 | 36 | .30 |

population score distributions, and $\hat{e}_Y(i)$ be the estimated observed score equivalent given by a particular equating function obtained from a pair of random samples. The MSE for that particular equating function at observed score i is

$$MSE(i) = \frac{1}{500} \sum_{v=1}^{500} [\hat{e}_{Y_v}(i) - e_Y(i)]^2, \tag{10}$$

where v indicates a replication. The MSE for a particular equating function is

$$MSE = \sum_{i=0}^k MSE(i) P(X = i), \tag{11}$$

where $P(X = i)$ is the proportion of examinees in the population that had an observed score i on Form X, and k is the maximum integer score. The squared bias (SB) at observed score i is

$$SB(i) = \left\{ \left[\frac{1}{500} \sum_{v=1}^{500} \hat{e}_{Y_v}(i) \right] - e_Y(i) \right\}^2. \tag{12}$$

The ASB for a particular equating function is

$$ASB = \sum_{i=0}^k SB(i) P(X = i). \tag{13}$$

For comparison, the minimum MSE degree of smoothing was computed. The minimum MSE degree of smoothing is defined as the smoothing degree with the lowest MSE. The minimum MSE degree of smoothing was selected at each replication by comparing the equating functions obtained from the various smoothing degrees with the equating function obtained from the population score distributions. Note that when equating is conducted in practice, the minimum MSE degree of smoothing cannot be used because the true equating relationship is unknown.

Because the population score distributions used in the simulation were obtained by fitting the score distributions of a pair of real samples with a loglinear model, $e_Y(i)$ is similar to a loglinear presmoothing equivalent that would be obtained in practice. However, there is a distinction between $e_Y(i)$ and a loglinear presmoothing equivalent. In practice, a loglinear model is used to smooth the score distributions of a pair of samples. In the simulation, the smoothed score distributions were treated arbitrarily as population distributions for generating artificial data; however, they may differ from the real populations from which the samples were drawn.

Because $e_v(i)$ was directly computed from the population distributions, no smoothing was involved in this process; therefore, $e_v(i)$ should not be considered to be a loglinear presmoothing equivalent.

Results

English Test Equating

The first four central moments of the population score distributions for the old form (Form Y) and the new form (Form X) and the equated score distributions for the 75-item AAP English test are listed in Table 3. Also shown are equating errors (ASB and MSE) for the different fixed degrees of smoothing (including unsmoothed), the linear equating, the degree selected by the computer-based procedure (CBP), and the minimum MSE degree of smoothing.

Table 3 shows that the linear equating function was associated with the largest ASBs and MSEs, which suggests that the equipercentile equating is preferable. For example, for $N = 500$, the ASB was .8586 for the linear equating, which was larger than the ASBs for all the smoothed and unsmoothed equipercentile equivalents; the MSE was 1.7909 for the linear equating, which was larger than the MSEs for all the smoothed and unsmoothed equipercentile equivalents.

Among the fixed degrees of smoothing, the smaller s values were associated with smaller ASBs, and larger s values were associated with larger ASBs. For example, in Table 3 for $N = 500$, the ASBs were .0045, .0044, and .0049 for s values of .01, .05, and .10, respectively, and .1102 for $s = .50$ and .7791 for $s = 2.00$. This observation suggests that smoothing introduces bias.

The unsmoothed equivalents had very small ASBs, but the means were less similar to those of the old form and the MSEs were larger than those with moderate smoothing. For example, in Table 3 for $N = 500$, the mean was 47.794 for the unsmoothed equivalents, which differed more from the mean (48.482) of the old form (Form Y) than the means for the smoothed equivalents with $s = .20$ and .30. The MSE was 1.4425 for the unsmoothed equivalents, which was larger than the MSE for the smoothed equivalents with $s = .50$ or less. The MSE became smaller as the degree of smoothing increased until it reached a lowest point; then it became larger as the degree of smoothing increased (see the MSEs in Table 3 for all three sample sizes).

This result suggests that moderate smoothing can reduce equating error that is due to sampling error without introducing too much bias. An optimal degree of smoothing was expected to result in an equated score distribution similar to the population distribution of the old form with reasonably small equating errors. Among the fixed degrees of smoothing, $s = .30$ worked best for $N = 500$ and $N = 1,000$ (as indicated by the smallest MSEs). For $N = 3,000$, $s = .20$ seemed to work best. Smaller degrees of smoothing tended to result in smaller bias but not enough smoothing. Larger degrees of smoothing tended to oversmooth the equipercentile equivalents, which resulted in larger bias and MSE.

For example, in Table 3 for $N = 500$ the smoothed equivalents with $s = .30$ had ASB = .0286 and the smallest MSE (1.1111). Although the equivalents with smaller s values had smaller ASBs, the random error components of the MSEs were not reduced enough by smoothing, as indicated by the larger MSEs. The equivalents with larger s values (.30, .50, and 2.00) had much larger ASBs, which could also increase the MSEs.

The degree of smoothing selected by the CBP varied from one sample to another based on the second derivatives of the cubic spline functions and the comparisons of the equated score distributions. The MSEs and ASBs for the computer-selected degree of smoothing were smaller than those of the corresponding best fixed degrees of smoothing for all three sample sizes. For example, the MSE was .1824 and the ASB was .0040 for the computer-selected degree of smoothing for $N = 3,000$. For the corresponding best fixed degree of smoothing ($s = .20$), the MSE was .1844 and the ASB was .0060. The moments of the equated score distribution for the computer-selected degree of smoothing were reasonably close to the old form population score distribution.

The equated score distributions obtained from larger sample sizes were closer to the old form population distribution than those obtained from smaller sample sizes. For example, the means of the equated

score distributions varied from 47.793 to 48.302 for $N = 500$ and from 47.943 to 48.320 for $N = 1,000$; the latter mean was closest to 48.482.

The MSEs for the CBP were larger than those for the minimum MSE degree of smoothing in the simulations for all three sample sizes (see Table 3). This suggests that the CBP failed to select the best smoothing degree in terms of MSE at every replication of the simulations. However, the ASBs for the CBP were much

Table 3
First Four Moments, ASB, and MSE From Simulations Based on an AAP
English Test for Three Sample Sizes and s From .01 to 2.00

| Distribution, N , and Form/Smoothing Degree | Mean | Standard Deviation | Skewness | Kurtosis | ASB | MSE |
|--|--------|-----------------------|----------|----------|-------|--------|
| Population Distribution | | | | | | |
| Form Y | 48.482 | 13.088 | -.089 | 2.187 | | |
| Form X | 51.325 | 12.755 | -.322 | 2.423 | | |
| Equated Score Distribution | | | | | | |
| $N = 500$ | | | | | | |
| Unsmoothed | 47.794 | 12.190 | -.109 | 2.205 | .0041 | 1.4425 |
| $s = .01$ | 47.793 | 12.194 | -.110 | 2.209 | .0045 | 1.4140 |
| $s = .05$ | 47.794 | 12.193 | -.110 | 2.209 | .0044 | 1.3519 |
| $s = .10$ | 47.796 | 12.192 | -.112 | 2.211 | .0049 | 1.2831 |
| $s = .15$ | 47.803 | 12.188 | -.115 | 2.216 | .0065 | 1.2202 |
| $s = .20$ | 47.817 | 12.182 | -.122 | 2.223 | .0104 | 1.1694 |
| $s = .30$ | 47.858 | 12.170 | -.139 | 2.240 | .0286 | 1.1111 |
| $s = .50$ | 47.956 | 12.153 | -.176 | 2.274 | .1102 | 1.1199 |
| $s = 2.00$ | 48.302 | 12.145 | -.294 | 2.378 | .7791 | 1.6029 |
| Linear | 47.878 | 12.191 | -.331 | 2.437 | .8586 | 1.7909 |
| CBP | 47.837 | 12.171 | -.132 | 2.234 | .0201 | 1.0961 |
| Minimum MSE | 47.873 | 12.168 | -.145 | 2.245 | .0383 | .9578 |
| $N = 1,000$ | | | | | | |
| Unsmoothed | 47.944 | 12.858 | -.077 | 2.198 | .0011 | .6887 |
| $s = .01$ | 47.943 | 12.862 | -.077 | 2.201 | .0012 | .6723 |
| $s = .05$ | 47.943 | 12.861 | -.077 | 2.201 | .0012 | .6363 |
| $s = .10$ | 47.946 | 12.858 | -.079 | 2.205 | .0017 | .5968 |
| $s = .15$ | 47.955 | 12.852 | -.084 | 2.212 | .0041 | .5669 |
| $s = .20$ | 47.968 | 12.844 | -.091 | 2.221 | .0096 | .5481 |
| $s = .30$ | 48.002 | 12.831 | -.107 | 2.239 | .0317 | .5419 |
| $s = .50$ | 48.071 | 12.817 | -.138 | 2.269 | .1061 | .5963 |
| $s = 2.00$ | 48.320 | 12.820 | -.248 | 2.355 | .6920 | 1.1186 |
| Linear | 47.966 | 12.821 | -.297 | 2.442 | .8571 | 1.3073 |
| CBP | 47.971 | 12.841 | -.092 | 2.223 | .0122 | .5328 |
| Minimum MSE | 47.996 | 12.834 | -.104 | 2.236 | .0270 | .4714 |
| $N = 3,000$ | | | | | | |
| Unsmoothed | 48.032 | 12.986 | -.129 | 2.227 | .0012 | .2286 |
| $s = .01$ | 48.032 | 12.990 | -.130 | 2.229 | .0013 | .2230 |
| $s = .05$ | 48.032 | 12.989 | -.130 | 2.229 | .0011 | .2093 |
| $s = .10$ | 48.035 | 12.986 | -.132 | 2.234 | .0011 | .1951 |
| $s = .15$ | 48.041 | 12.982 | -.137 | 2.242 | .0024 | .1868 |
| $s = .20$ | 48.049 | 12.977 | -.143 | 2.251 | .0060 | .1844 |
| $s = .30$ | 48.067 | 12.971 | -.155 | 2.267 | .0180 | .1889 |
| $s = .50$ | 48.100 | 12.968 | -.176 | 2.291 | .0511 | .2147 |
| $s = 2.00$ | 48.243 | 13.003 | -.262 | 2.364 | .3386 | .4966 |
| Linear | 48.030 | 13.007 | -.365 | 2.485 | .8570 | 1.0039 |
| CBP | 48.045 | 12.979 | -.140 | 2.246 | .0040 | .1824 |
| Minimum MSE | 48.058 | 12.976 | -.149 | 2.257 | .0102 | .1598 |

smaller than those for the minimum MSE degree of smoothing. For example, the ASB was .0040 for the CBP and .0102 for the minimum MSE degree of smoothing for $N = 3,000$. The smaller ASBs suggest that the CBP was less biased than the minimum MSE degree of smoothing.

Mathematics Test Equating

The results for the 60-item AAP mathematics test (Table 4) were similar to those observed for the English test. In terms of MSE, the best fixed degrees of smoothing were $s = .30$ for $N = 500$, $s = .20$ for $N = 1,000$, and $s = .15$ for $N = 3,000$. The MSEs for the CBP were smaller than those for the corresponding best fixed degrees of smoothing. As for the English test, the MSEs were larger for the CBP than those for the minimum MSE degree of smoothing, and the ASBs were smaller for the CBP.

Reading Test Equating

Table 5 shows the results for the 40-item AAP reading test. In terms of MSE, the best fixed degrees of smoothing were $s = .50$ for $N = 500$, $s = .30$ for $N = 1,000$, and $s = .20$ for $N = 3,000$. The MSEs for the CBP were larger than those for the best fixed and minimum MSE degrees of smoothing for all three sample sizes. For $N = 3,000$, the MSE for the CBP was only .0002 larger than that for the best fixed degree of smoothing. The CBP was less biased than the best fixed and minimum MSE degrees of smoothing for all three sample sizes as indicated by the smaller ASBs.

Science Reasoning Test Equating

The results for the 40-item AAP science reasoning test are presented in Table 6. The MSEs indicate that the best fixed degrees of smoothing were $s = .20$ for $N = 500$ and for $N = 1,000$, and $s = .15$ for $N = 3,000$. The MSEs and ASBs for the CBP were smaller than those for the corresponding best fixed degrees of smoothing for $N = 500$. For $N = 1,000$, the MSE for the CBP was slightly larger than that for the best fixed degree of smoothing (.1466 versus .1464). However, the ASB for the CBP was smaller than that for the best fixed degree of smoothing (.0057 versus .0064). For $N = 3,000$, the MSE for the CBP was slightly smaller than that for the best fixed degree of smoothing (.0517 versus .0518). The ASB for the CBP was slightly larger than that for the best fixed degree of smoothing ($N = 3,000$; .0039 versus .0036). From a practical point of view, the CBP performed at least as well as the best fixed degrees. As for the simulations of the English and mathematics tests, the MSEs were smaller for the minimum MSE degrees of smoothing than those for the CBP and the best fixed degrees of smoothing, and the ASBs were generally larger for the minimum MSE degrees of smoothing.

Discussion and Conclusions

The results suggest that there is no single fixed degree of smoothing that always leads to the optimal degree of smoothing, based on the observations that (1) the equating errors for the minimum MSE degrees of smoothing were smaller than those of the best fixed degrees, and (2) the best fixed degrees were different for different sample sizes or tests. For example, in the simulation based on the English test, the fixed degree of $s = .15$ resulted in the smallest MSE (.1868) among all the fixed degrees for $N = 3,000$. This does not mean that for each individual equating in the simulation the equating errors for the smoothed equating function with $s = .15$ were the smallest among all the smoothed functions. For some equatings in the simulation, the smoothed equating functions with $s = .10$, .20, or other values had smaller equating errors. In most cases, the CBP selected an s value that led to smaller equating errors than the fixed degree of $s = .15$.

When computing MSEs and ASBs for the minimum MSE degrees of smoothing, the smoothed equating function with an s value that had the least difference from the population equating function was used during each replication of the simulation. This minimum MSE degree of smoothing can only be defined in

Table 4
 First Four Moments, ASB, and MSE From Simulations Based on an AAP
 Mathematics Test for Three Sample Sizes and s From .01 to 2.00

| Distribution, N , and Form/Smoothing Degree | Mean | Standard Deviation | Skewness | Kurtosis | ASB | MSE |
|--|--------|-----------------------|----------|----------|-------|--------|
| Population Distribution | | | | | | |
| Form Y | 28.463 | 10.569 | .481 | 2.535 | | |
| Form X | 30.301 | 12.187 | .276 | 2.148 | | |
| Equated Score Distribution | | | | | | |
| $N = 500$ | | | | | | |
| Unsmoothed | 27.715 | 9.514 | .395 | 2.423 | .0070 | .9258 |
| $s = .01$ | 27.714 | 9.519 | .395 | 2.426 | .0076 | .9044 |
| $s = .05$ | 27.715 | 9.520 | .394 | 2.424 | .0073 | .8599 |
| $s = .10$ | 27.718 | 9.528 | .391 | 2.414 | .0064 | .8130 |
| $s = .15$ | 27.724 | 9.545 | .386 | 2.395 | .0070 | .7774 |
| $s = .20$ | 27.734 | 9.566 | .379 | 2.371 | .0115 | .7556 |
| $s = .30$ | 27.760 | 9.609 | .364 | 2.324 | .0330 | .7427 |
| $s = .50$ | 27.819 | 9.678 | .336 | 2.253 | .1042 | .7740 |
| $s = 2.00$ | 28.032 | 9.830 | .247 | 2.112 | .5122 | 1.0154 |
| Linear | 27.765 | 9.634 | .232 | 2.080 | .6471 | 1.2221 |
| CBP | 27.733 | 9.565 | .378 | 2.371 | .0129 | .7363 |
| Minimum MSE | 27.757 | 9.607 | .366 | 2.328 | .0302 | .6508 |
| $N = 1,000$ | | | | | | |
| Unsmoothed | 27.962 | 10.209 | .440 | 2.385 | .0060 | .4482 |
| $s = .01$ | 27.962 | 10.214 | .440 | 2.388 | .0063 | .4371 |
| $s = .05$ | 27.962 | 10.214 | .440 | 2.386 | .0062 | .4121 |
| $s = .10$ | 27.964 | 10.220 | .437 | 2.377 | .0067 | .3876 |
| $s = .15$ | 27.968 | 10.231 | .432 | 2.360 | .0097 | .3737 |
| $s = .20$ | 27.974 | 10.245 | .426 | 2.340 | .0160 | .3693 |
| $s = .30$ | 27.989 | 10.271 | .414 | 2.302 | .0367 | .3792 |
| $s = .50$ | 28.021 | 10.309 | .392 | 2.244 | .0927 | .4246 |
| $s = 2.00$ | 28.183 | 10.395 | .300 | 2.101 | .4652 | .7281 |
| Linear | 27.943 | 10.220 | .281 | 2.060 | .6457 | .9211 |
| CBP | 27.970 | 10.238 | .429 | 2.350 | .0136 | .3625 |
| Minimum MSE | 27.980 | 10.259 | .421 | 2.320 | .0237 | .3274 |
| $N = 3,000$ | | | | | | |
| Unsmoothed | 28.052 | 10.290 | .410 | 2.401 | .0058 | .1545 |
| $s = .01$ | 28.052 | 10.294 | .409 | 2.401 | .0059 | .1502 |
| $s = .05$ | 28.052 | 10.294 | .408 | 2.399 | .0057 | .1409 |
| $s = .10$ | 28.055 | 10.297 | .406 | 2.391 | .0063 | .1336 |
| $s = .15$ | 28.060 | 10.303 | .402 | 2.379 | .0086 | .1315 |
| $s = .20$ | 28.064 | 10.309 | .399 | 2.367 | .0125 | .1329 |
| $s = .30$ | 28.073 | 10.320 | .392 | 2.344 | .0228 | .1407 |
| $s = .50$ | 28.090 | 10.338 | .381 | 2.309 | .0471 | .1631 |
| $s = 2.00$ | 28.194 | 10.396 | .310 | 2.181 | .2510 | .3628 |
| Linear | 28.048 | 10.281 | .239 | 2.095 | .6437 | .7344 |
| CBP | 28.060 | 10.302 | .402 | 2.379 | .0089 | .1310 |
| Minimum MSE | 28.064 | 10.308 | .399 | 2.368 | .0116 | .1187 |

a simulation in which the population score distributions are available. The optimal degree of smoothing varied from one simulation to another. The s value, a parameter to control the degree of smoothing, seems to be related to sample size. In the simulations based on a given test, a larger value of s was needed for a small sample size than for a larger sample size in order to minimize the MSE. For example, for the 60-item mathematics test (see Table 4), for $N = 500$ the best fixed s value was .30 (MSE = .7740), for $N = 1,000$ the

Table 5
First Four Moments, ASB, and MSE From Simulations Based on an
AAP Reading Test for Three Sample Sizes and s From .01 to 2.00

| Distribution, N , and Form/Smoothing Degree | Mean | Standard Deviation | Skewness | Kurtosis | ASB | MSE |
|---|--------|--------------------|----------|----------|-------|-------|
| Population Distribution | | | | | | |
| Form Y | 24.804 | 6.584 | -.024 | 2.375 | | |
| Form X | 25.669 | 6.578 | -.150 | 2.466 | | |
| Equated Score Distribution | | | | | | |
| $N = 500$ | | | | | | |
| Unsmoothed | 24.507 | 6.089 | -.052 | 2.361 | .0015 | .3680 |
| $s = .01$ | 24.506 | 6.096 | -.053 | 2.366 | .0019 | .3571 |
| $s = .05$ | 24.506 | 6.095 | -.054 | 2.367 | .0019 | .3371 |
| $s = .10$ | 24.508 | 6.095 | -.056 | 2.369 | .0020 | .3155 |
| $s = .15$ | 24.514 | 6.094 | -.061 | 2.374 | .0023 | .2972 |
| $s = .20$ | 24.522 | 6.094 | -.068 | 2.379 | .0030 | .2831 |
| $s = .30$ | 24.540 | 6.093 | -.085 | 2.390 | .0067 | .2665 |
| $s = .50$ | 24.571 | 6.092 | -.114 | 2.407 | .0189 | .2571 |
| $s = 2.00$ | 24.627 | 6.094 | -.161 | 2.433 | .0556 | .2635 |
| Linear | 24.532 | 6.096 | -.168 | 2.446 | .0549 | .2949 |
| CBP | 24.516 | 6.092 | -.063 | 2.378 | .0025 | .2862 |
| Minimum MSE | 24.541 | 6.094 | -.087 | 2.390 | .0072 | .2349 |
| $N = 1,000$ | | | | | | |
| Unsmoothed | 24.584 | 6.478 | -.023 | 2.352 | .0008 | .1765 |
| $s = .01$ | 24.583 | 6.484 | -.024 | 2.355 | .0009 | .1708 |
| $s = .05$ | 24.584 | 6.483 | -.024 | 2.356 | .0009 | .1598 |
| $s = .10$ | 24.586 | 6.481 | -.026 | 2.360 | .0010 | .1490 |
| $s = .15$ | 24.590 | 6.480 | -.031 | 2.366 | .0014 | .1407 |
| $s = .20$ | 4.597 | 6.478 | -.038 | 2.372 | .0024 | .1356 |
| $s = .30$ | 24.610 | 6.476 | -.055 | 2.384 | .0065 | .1319 |
| $s = .50$ | 24.634 | 6.474 | -.082 | 2.401 | .0190 | .1366 |
| $s = 2.00$ | 24.676 | 6.474 | -.130 | 2.426 | .0564 | .1593 |
| Linear | 24.591 | 6.474 | -.139 | 2.441 | .0547 | .1704 |
| CBP | 24.591 | 6.479 | -.032 | 2.368 | .0015 | .1372 |
| Minimum MSE | 24.607 | 6.477 | -.051 | 2.382 | .0053 | .1179 |
| $N = 3,000$ | | | | | | |
| Unsmoothed | 24.629 | 6.561 | -.069 | 2.384 | .0008 | .0588 |
| $s = .01$ | 24.629 | 6.567 | -.070 | 2.386 | .0011 | .0573 |
| $s = .05$ | 24.630 | 6.567 | -.070 | 2.388 | .0010 | .0534 |
| $s = .10$ | 24.631 | 6.566 | -.073 | 2.393 | .0010 | .0497 |
| $s = .15$ | 24.634 | 6.564 | -.078 | 2.401 | .0012 | .0475 |
| $s = .20$ | 24.638 | 6.564 | -.085 | 2.408 | .0019 | .0466 |
| $s = .30$ | 24.646 | 6.563 | -.098 | 2.421 | .0042 | .0471 |
| $s = .50$ | 24.660 | 6.564 | -.121 | 2.438 | .0113 | .0525 |
| $s = 2.00$ | 24.700 | 6.574 | -.183 | 2.472 | .0503 | .0848 |
| Linear | 24.629 | 6.571 | -.198 | 2.486 | .0548 | .0926 |
| CBP | 24.634 | 6.564 | -.077 | 2.401 | .0012 | .0468 |
| Minimum MSE | 24.641 | 6.563 | -.090 | 2.414 | .0026 | .0407 |

best fixed s value was .20 (MSE = .3693), and for $N = 3,000$ the best fixed s value was .15 (MSE = .1315). This seems to be reasonable, because the smaller samples, with potentially greater sampling error, required larger amounts of smoothing to overcome the sample-dependent fluctuation.

Five subjectively selected parameters, τ and w_1 to w_4 were used in the simulations. The values of w_1 to w_4 were the same for all simulations. Different values of τ were used for simulations based on different

Table 6
 First Four Moments, ASB, and MSE From Simulations Based on an AAP
 Science Reasoning Test for Three Sample Sizes and s From .01 to 2.00

| Distribution, N , and Form/Smoothing Degree | Mean | Standard Deviation | Skewness | Kurtosis | ASB | MSE |
|--|--------|-----------------------|----------|----------|-------|-------|
| Population Distribution | | | | | | |
| Form Y | 24.153 | 6.439 | -.192 | 2.553 | | |
| Form X | 22.659 | 7.064 | .170 | 2.373 | | |
| Equated Score Distribution | | | | | | |
| $N = 500$ | | | | | | |
| Unsmoothed | 23.881 | 5.948 | -.226 | 2.552 | .0031 | .3662 |
| $s = .01$ | 23.880 | 5.954 | -.227 | 2.559 | .0037 | .3553 |
| $s = .05$ | 23.879 | 5.955 | -.226 | 2.557 | .0035 | .3348 |
| $s = .10$ | 23.875 | 5.957 | -.218 | 2.547 | .0035 | .3158 |
| $s = .15$ | 23.866 | 5.961 | -.205 | 2.531 | .0051 | .3036 |
| $s = .20$ | 23.856 | 5.965 | -.190 | 2.514 | .0088 | .2979 |
| $s = .30$ | 23.832 | 5.969 | -.160 | 2.483 | .0221 | .3002 |
| $s = .50$ | 23.788 | 5.973 | -.110 | 2.441 | .0622 | .3316 |
| $s = 2.00$ | 23.618 | 5.965 | .061 | 2.375 | .3614 | .6031 |
| Linear | 23.816 | 5.947 | .126 | 2.353 | .4651 | .6961 |
| CBP | 23.859 | 5.962 | -.194 | 2.521 | .0077 | .2973 |
| Minimum MSE | 23.843 | 5.968 | -.172 | 2.494 | .0159 | .2628 |
| $N = 1,000$ | | | | | | |
| Unsmoothed | 23.942 | 6.322 | -.177 | 2.510 | .0023 | .1782 |
| $s = .01$ | 23.942 | 6.328 | -.178 | 2.513 | .0027 | .1726 |
| $s = .05$ | 23.941 | 6.328 | -.176 | 2.511 | .0025 | .1615 |
| $s = .10$ | 23.938 | 6.329 | -.170 | 2.503 | .0025 | .1521 |
| $s = .15$ | 23.933 | 6.332 | -.161 | 2.491 | .0037 | .1475 |
| $s = .20$ | 23.927 | 6.334 | -.150 | 2.478 | .0064 | .1464 |
| $s = .30$ | 23.914 | 6.339 | -.128 | 2.455 | .0153 | .1506 |
| $s = .50$ | 23.889 | 6.343 | -.091 | 2.424 | .0401 | .1711 |
| $s = 2.00$ | 23.783 | 6.341 | .052 | 2.368 | .2465 | .3718 |
| Linear | 23.933 | 6.341 | .168 | 2.314 | .4644 | .5761 |
| CBP | 23.929 | 6.333 | -.152 | 2.484 | .0057 | .1466 |
| Minimum MSE | 23.923 | 6.336 | -.143 | 2.471 | .0085 | .1307 |
| $N = 3,000$ | | | | | | |
| Unsmoothed | 23.984 | 6.407 | -.232 | 2.549 | .0023 | .0606 |
| $s = .01$ | 23.983 | 6.413 | -.232 | 2.550 | .0026 | .0587 |
| $s = .05$ | 23.982 | 6.413 | -.231 | 2.548 | .0024 | .0550 |
| $s = .10$ | 23.980 | 6.413 | -.226 | 2.542 | .0027 | .0525 |
| $s = .15$ | 23.977 | 6.414 | -.219 | 2.532 | .0036 | .0518 |
| $s = .20$ | 23.974 | 6.415 | -.212 | 2.523 | .0050 | .0523 |
| $s = .30$ | 23.968 | 6.416 | -.198 | 2.506 | .0092 | .0550 |
| $s = .50$ | 23.956 | 6.415 | -.174 | 2.481 | .0196 | .0640 |
| $s = 2.00$ | 23.903 | 6.402 | -.077 | 2.413 | .1032 | .1453 |
| Linear | 23.984 | 6.398 | .130 | 2.351 | .4648 | .5013 |
| CBP | 23.977 | 6.414 | -.217 | 2.531 | .0039 | .0517 |
| Minimum MSE | 23.975 | 6.414 | -.212 | 2.525 | .0049 | .0471 |

tests. The same set of parameters was used in the simulations based on the same test using the three different sample sizes. These parameters were selected to ensure that the CBP would select similar s values to those that had typically been selected in AAP equatings. For equatings with different characteristics, a different set of parameter values may be needed. Based on the inspection of AAP equatings, a smaller τ is needed for longer tests and a larger τ is needed for shorter tests. A general rule for selecting

these parameters is to have the procedure select degrees of smoothing that are typically selected in a similar situation in practice.

Conclusions

Based on the simulation results presented here, the following conclusions seem justified. Equipercentile equating is preferable to linear equating for the data studied. A moderate amount of smoothing can improve equating results by reducing the random errors due to sampling errors, without introducing too much bias. In terms of equating errors, there is no single fixed s value that always leads to an optimal degree of smoothing. Equating conducted with smaller sample sizes requires more smoothing to reach an optimal degree of smoothing. The equated score distributions obtained from larger sample sizes were generally closer to the old form population score distributions than those obtained from smaller sample sizes. The degrees of smoothing selected by the proposed CBP were found to be better in minimizing MSE than the best fixed degrees of smoothing for two of the four tests (English and mathematics) studied. For the science reasoning test, the CBP performed better at minimizing MSE or nearly as well as the best fixed degrees. The CBP was found to be a promising procedure for helping to select a smoothing degree. However, this procedure should be used in conjunction with informed psychometric judgment when selecting a degree of smoothing in practice.

References

- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
- de Boor, C. (1978). *A practical guide to splines*. New York: Springer-Verlag.
- Hanson, B. A. (1991). *Using the MLogLin function*. Unpublished manuscript, American College Testing, Iowa City IA.
- Hanson, B. A., Zeng, L., & Colton, D. (1991, April). *A comparison of presmoothing and postsmoothing methods in equipercentile equating*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.
- Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics*, 9, 25-44.
- Kolen, M. J. (1991). Smoothing methods for estimating test score distributions. *Journal of Educational Measurement*, 28, 277-282.
- Reinsch, C. H. (1967). Smoothing by spline functions. *Numerische Mathematik*, 10, 177-183.

Acknowledgments

Computer programs written by Bradley Hanson were used to estimate the population distributions and to generate the data used here. Michael Kolen, Tim Davey, Mark Reckase, and two anonymous reviewers provided comments and suggestions for improving the manuscript.

Author's Address

Send requests for reprints or further information to Lingjia Zeng, ACT, P.O. Box 168, Iowa City IA 52243, U.S.A.