Center for Advanced Studies in Measurement and Assessment

CASMA Research Report

Number 11

An Alternative Continuization Method to the Kernel Method in von Davier, Holland and Thayer's (2004) Test Equating Framework

Tianyou Wang[†]

July 11, 2007

[†]Tianyou Wang is Research Scientist, Center for Advanced Studies in Measurement and Assessment (CASMA), 210 Lindquist Center, College of Education, University of Iowa, Iowa City, IA 52242 (email: tianyou-wang@uiowa.edu).

Center for Advanced Studies in
Measurement and Assessment (CASMA)
College of Education
University of Iowa
Iowa City, IA 52242
Tel: 319-335-5439
Web: www.education.uiowa.edu/casma

Contents

List of Tables

List of Figures

Abstract

von Davier, Holland and Thayer (2004) laid out a five-step framework of test equating which can be applied to various data collection designs and equating methods. In the continuization step, they present an adjusted Gaussian kernel method which preserves the first two moments. This paper proposes an alternative continuization method which directly uses the log-linear function from the smoothing step. This alternative method has the advantages of being simple and direct and producing smoother continuous distributions. When the number of score points is large, it can better preserve the moments. Two real test examples are provided to demonstrate these properties and to provide comparisons of equating functions based on different methods. The standard error of equating for this method is derived.

1 Introduction

von Davier, Holland and Thayer (2004) laid out a unified framework of test equating which can be applied to various data collection designs and equating methods. They presented this framework by decomposing the equating process into the following five steps (ch. 3. pp. 45-47):

> *Step 1: Pre-smoothing.* This step basically estimates score distributions for the populations from which test data were collected. Various statistical models can be used to fit the score distribution. Log-linear smoothing has proved to be a flexible method with desirable statistical properties such as preserving the moments.
>
> *step 2: Estimating score probabilities.* The difference between this step and the previous step is that in this step the estimation of score probabilities is for a common target population for both test forms. This step basically transforms the smoothed distribution from step 1 to obtain the marginal score probabilities for the target population. They used a term called Design Function to connotate this transformation process. For different data collection designs, different Design Functions will be used.
>
> *step 3: Continuization.* This step basically deals with the difficulty of performing equipercentile equating for discrete score distributions. Traditionally this problem is solved by using the percentile rank function and linear interpolation. von Davier et al. (2004) proposed using the Gaussian kernel method to fit a continuous distribution to the discrete score distribution.
>
> *step 4: Equating.* This step performs equipercentile equating using the continuous distributions from step 3. Denote the random variables for the test scores for Form X as X and for From Y as Y, and the target population cumulative distributions of X and Y as $F(X)$ and $G(Y)$, respectively. Then the equipercentile equating function $\hat{e}_Y(x)$ is
>
> $$\hat{e}_Y(X) = G^{-1}(F(X)), \tag{1}$$
>
> *step 5: Calculating the Standard Error of Equating.* With this general framework and the kernel continuization method, von Davier et al. (2004) derived an elegant general formula for estimating the standard error of equating (SEE) based on the $\delta-$method. This general formula can be applied to all equating designs and is composed of three components with each of them relating to a different step in this process.

The advantage of this framework is that it is applicable to most standard data collection designs and equating methods, except perhaps the IRT-based equating methods. Also, it modularizes the equating process so that different designs and methods only affect certain steps. For instance, different data collection designs will result in different Design Functions in step 2. For a random groups design, step 2 is usually omitted in the traditional description of the equating process, but in this framework, an identity Design Function is used. Likewise, different equating methods only affect step 4.

A minor problem in their presentation of this excellent framework is perhaps in step 3 (i.e. continuization). I suggest here that the authors put too much emphasis on the Guassian kernel method. It would be better to have a general continuization step and treat the kernel method as only one of a few possible continuization methods. The purpose of this short paper is to explore an alternative continuization method and make some comparisons of it with the Gaussian kernel method.

2 The Kernel Continuization Method

Traditionally the kernel method has been used as a smoothing method to fit score distributions. In von Davier et al. (2004), the Gaussian kernel method is used to fit a continuous distribution to an already smoothed discrete distribution. The kernel method is known to distort the moments higher than the first moment (i.e., the mean). In order to preserve the first two moments, von Davier et al. (2004) proposed an adjusted Guassian kernel method described below.

Let $x_j, j = 0, \ldots, J$ denote the score points for X, r_j be the discrete relative frequency at score x_j, Φ and ϕ denote the cdf and pdf, respectively, of the standard normal distribution, h_X be a bandwidth parameter, and μ_X and σ_X^2 denote the mean and variance of X over target population T. Their Gaussian kernel has a distribution with the cdf given by

$$F_{hX}(x) = \sum_j r_j \Phi(R_{jX}(x)), \tag{2}$$

where

$$R_{jX}(x) = \frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}, \tag{3}$$

and

$$a_X^2 = \frac{\sigma_X^2}{\sigma_X^2 + h_X^2}. \tag{4}$$

The pdf of this kernel distribution is

$$f_{hX}(x) = \sum_j r_j \phi(R_{jX}(x)) \frac{1}{a_X h_X}, \tag{5}$$

This kernel distribution has the same mean and standard deviation as the discrete distribution of X. The skewness and kurtosis, however, will be different by a factor of $(a_X)^j$, where $j = 3$ for skewness and $j = 4$ for kurtosis. Note that a_X is a number between 0 and 1. As h increases, a_X decreases. This means that a larger h will result in a more symmetric distribution with smaller kurtosis, which implies more distortion of shape in the kernel distribution compared to the discrete distribution. von Davier et al. (2004, pp. 62-63) introduced two penalty functions to find the optimal h parameter that mimimizes some weighted combination of these penalty functions.

Although the kernel continuization procedure has the advantage of being flexible, it has several drawbacks. First, it requires a rather arbitrary bandwidth parameter h_X. Although von Davier et al. (2004) introduced penalty functions to set the parameter, the procedure for determining the parameter can be complicated. Second, the kernel distribution does not preserve all the moments of the discrete distribution. Third, there is still some bumpiness if the parameter h_X is not set large enough, as illustrated with real data in a later part of this paper.

3 The Continuized Log-linear Method for the Equivalent Groups Design

For the equivalent groups (EG) design, the Design Function is the identity function. The distributions obtained from step 2 are the same as those from step 1. For this design, I propose an alternative continuization procedure which utilizes the polynomial log-linear function obtained in the log-linear smoothing step. I will call it the continuized log-linear (CLL) distribution. It's probability density function (pdf) is expressed as

$$f(x) = \frac{1}{D} exp(\boldsymbol{b^T \beta}), \tag{6}$$

where $\boldsymbol{b^T} = (1, x, x^2, ..., x^M)$ are a vector of polynomial terms of x, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, ..., \beta_M)^T$ is the vector of parameters, M is the order of polynomial, and D is a normalizing constant which ensures that $f(x)$ is a pdf and is expressed as

$$D = \int_l^u exp(\boldsymbol{b^T \beta}) dx,$$

where l and u are the lower and upper limit of integration. In this case, they are set to be -0.5 and $J + 0.5$, respectively, so that the probabilities of the end points of the discrete distribution are allowed to spread out in both directions. It is easy to show that all the moments of the CLL distribution are approximately equal to those of the smoothed discrete distribution by the following relationship between *i-th* non-central moments of the CLL distribution and the smoothed discrete distribution:

$$\frac{\int_l^u x^i exp(\boldsymbol{b^T \beta}) dx}{\int_l^u exp(\boldsymbol{b^T \beta}) dx} \approx \frac{1}{N} \sum_{x=0}^{J} x^i exp(\boldsymbol{b^T \beta}), \tag{7}$$

where N is the sample size. This approximation holds because the right side of the equation is actually an expression for numerical integration of the left side with equally spaced quadrature points. The numerator and denominator of the left side can be separately expressed as:

$$\int_l^u x^i exp(\boldsymbol{b^T\beta})dx \approx \sum_{x=0}^{J} x^i exp(\boldsymbol{b^T\beta}), \tag{8}$$

and

$$D = \int_l^u exp(\boldsymbol{b^T\beta})dx \approx \sum_{x=0}^{J} exp(\beta_0 + \beta_1 x + \ldots + \beta_M x^M) = N. \tag{9}$$

This means that the normalizing constant is approximately equal to the sample size which is known prior to equating. This result significantly simplifies the computation. The above expressions are very similar to the trapezoidal rule (see Thisted, 1988, p. 264. Note that the subinterval length equals 1). The range of the continuous distribution is set to be from -0.5 to $J+0.5$ so that in the quadrature the function is evaluated at the mid points of the subintervals rather than at the end points as in the regular trapezoidal rule. This range is consistent with the range of the percentile rank method in conventional equipercentile equating (Kolen & Brennan, 2004). Because of the smoothness of the log-linear function, the approximation can be quite close when the number of quadrature points (i.e., the score points J) gets large.

The proposed CLL continuization seems to have several advantages over kernel continuization. First, CLL continuization is simpler and more direct. Second, it is smoother and is guaranteed to be without the small bumpiness in the kernel continuization. Third, it preserves all the moments of the discrete distribution to the precision of equally spaced numerical integration with $J+1$ quadrature points. The next section illustrates these points with two data sets, one from von Davier et al. (2004), and the other from Kolen and Brennan (2004).

4 The Continuized Log-linear Method for Other Designs

The CLL approach for the EG design can be extended to other designs, such as the single group (SG) design, the single group counter balanced (CB) design, and the non-equivalent groups with anchor test (NEAT) design. Typically these designs require bivariate log-linear smoothing procedure in Step 1 of the test equating framework described earlier in this paper. With the Gaussian kernel continuization method, Step 2 is the step that applies the Design Functions, and Step 3 is the continuization step. With the CLL continuization method, because the continuization step must directly utilize the log-linear function from Step 1, it must carried out immediately after Step 1. So the Design Function must be applied after the continuization step and must be applied on continuous distribution functions rather than on discrete distributions as in the kernel method.

4.1 For the Single Group/Counter Balance Design

For the SG design, both Form X and Form Y are administered to the same group of examinees. For CB design, the whole group takes both Form X and From Y, however, approximately half of the group take Form X first and then Form Y, the other half takes Form Y first and then Form X. We will label the first half group as Group 1 and then second half as Group 2. The SG design can be viewed as a special case of CB design where there is only Group.

We directly take the the log-linear functions from Step 1 as if they are continuous functions and normalize them to be pdf's. For Group 1, the pdf can be expressed:

$$f_1(x,y) = \frac{1}{D_1} exp(\boldsymbol{b^T \beta}), \tag{10}$$

where $\boldsymbol{b^T} = (1, x, x^2, ..., x^{M_X}, y, y^2, ..., y^{M_Y}, xy, x^2y, xy^2, ..., x^{C_X}y^{C_Y})$ are a vector of polynomial terms of x and y,
$\boldsymbol{\beta} = (\beta_{00}, \beta_{01}, \beta_{02}, ..., \beta_{0M_X}, \beta_{10}, \beta_{20}, ..., \beta_{M_Y0}, \beta_{11}, \beta_{12}, \beta_{21}, ..., \beta_{C_XC_Y})^T$ is the vector of parameters, M_X, M_Y are the orders of marginal polynomial terms for X and Y, C_X, C_Y are the orders cross-product terms for X and Y, and D_1 is a normalizing constant which ensures that $f_1(x,y)$ is a pdf and is expressed as

$$D_1 = \int_{l_Y}^{u_Y} \int_{l_X}^{u_X} exp(\boldsymbol{b^T \beta}) dx dy,$$

again, it can be showed that the normalizing constant approximates the sample size.

The joint pdf of Group 2 $f_2(x,y)$ can be found in a similar fashion. Given the weights of X and Y for Group 1, w_X, and w_Y, the combined marginal distributions of X and Y can be expressed as:

$$f(x) = w_X \int_{l_Y}^{u_Y} f_1(x,y)dy + (1 - w_X) \int_{l_Y}^{u_Y} f_2(x,y)dy, \tag{11}$$

$$f(y) = w_y \int_{l_X}^{u_X} f_1(x,y)dx + (1 - w_y) \int_{l_X}^{u_X} f_2(x,y)dx. \tag{12}$$

Numerical integration method is used in carrying out the necessary integrations. The rest of the equating procedure is the same as in the EG design.

4.2 For the Non-Equivalent Groups with Anchor Test Design

For the NEAT design, Group 1 from Population 1 takes Form X plus the anchor set V, Group 2 from Population 2 takes Form Y plus the anchor set V. The continuous bivariate pdf's $f_1(x,v)$ for X and V, $f_2(y,v)$ for Y and V can be obtained in a similar fashion as described in the previous section for the CB design. There are essentially two equating methods under the NEAT design: the frequency estimation (also called post-stratification) and the chained equipercentile equating method. The frequency estimation makes the following assumption:

$$f_2(x|v) = f_1(x|v) = f_1(x, v)/f_1(v), \tag{13}$$

$$f_1(y|v) = f_2(y|v) = f_2(y, v)/f_2(v), \tag{14}$$

The marginal distributions can be found by the following expressions:

$$f_1(x) = \int_{l_V}^{u_V} f_1(x, v)dv, \tag{15}$$

$$f_2(y) = \int_{l_V}^{u_V} f_2(y, v)dv, \tag{16}$$

$$f_1(v) = \int_{l_X}^{u_X} f_1(x, v)dx, \tag{17}$$

$$f_2(v) = \int_{l_Y}^{u_Y} f_2(y, v)dy, \tag{18}$$

With this assumption and given the weight of Population 1 in the target population, w_1, we can have the marginal distributions of X and Y for the target population as:

$$f_T(x) = w_1 f_1(x) + (1 - w_1) \int_{l_V}^{u_V} f_1(x|v) f_2(v)dv, \tag{19}$$

$$f_T(y) = w_1 \int_{l_V}^{u_V} f_2(y|v) f_1(v)dv + (1 - w_1) f_2(y), \tag{20}$$

The rest of the equating procedure is the same as in the EG design.

The chained equipercentile equating method first equates X to V using $f_1(x)$ and $f_1(v)$, and then equates the V equivalent X scores to Y using $f_2(v)$ and $f_2(y)$. Given all the continuous marginal distributions in Equations 15 through 18, it takes twice applying Equation 1 to accomplish the chain equipercentile equating procedure.

5 Standard Error of Equating for CLL Continuization under the Equivalent Groups Design

von Davier et al. (2004) derived this general expression for the asymptotic standard error of equating (SEE):

$$SEE_Y(x) = \| \hat{J}_{eY} \hat{J}_{DF} C \| . \tag{21}$$

This expression is decomposed into three parts, each relating to a different stage of the equating process. $\hat{J}_{eY}$ is related to continuization (step 3) and equating (step 4). $\hat{J}_{DF}$ is related to estimation of score probabilities (step 2). And C

is related to pre-smoothing (step 1). Because the CLL method uses the log-linear function directly in the continuization step, the cumulative distribution functions (cdf) of Form X and Form Y depend on the the estimated parameter vectors $\hat{\boldsymbol{\beta}}_X$ and $\hat{\boldsymbol{\beta}}_Y$ of the log-linear models rather than on the estimated score probabilities $\hat{r}$ and $\hat{s}$ in von Davier et al (2004). Let F denote the cdf of Form X and G denote the cdf of Form Y. The equating function from X to Y can be expressed as

$$e_Y(x) = e_Y(x; \boldsymbol{\beta_X}, \boldsymbol{\beta_Y}) = G^{-1}(F(x; \boldsymbol{\beta_X}); \boldsymbol{\beta_Y}), \tag{22}$$

where

$$F(x; \beta_X) = \frac{\int_l^x exp(\boldsymbol{b_X^T \beta_X})dt}{\int_l^u exp(\boldsymbol{b_X^T \beta_X})dt}, \tag{23}$$

$$G(y; \beta_Y) = \frac{\int_l^y exp(\boldsymbol{b_Y^T \beta_Y})dt}{\int_l^u exp(\boldsymbol{b_Y^T \beta_Y})dt}, \tag{24}$$

Using the δ-method and following a similar approach as in Holland, King and Thayer (1989), the square of SEE can be expressed as

$$\sigma_Y^2(x) = (\partial e_Y)^t \Sigma (\partial e_Y) \tag{25}$$

where,

$$\Sigma = \begin{bmatrix} \Sigma_{\hat{\beta}_X} & \Sigma_{\hat{\beta}_X \hat{\beta}_Y} \\ \Sigma_{\hat{\beta}_X \hat{\beta}_Y} & \Sigma_{\hat{\beta}_Y} \end{bmatrix}, \tag{26}$$

and

$$(\partial e_Y) = \begin{bmatrix} \frac{\partial e_Y}{\partial \beta_X} \\ \frac{\partial e_Y}{\partial \beta_Y} \end{bmatrix}. \tag{27}$$

The elements of Σ are further obtained by the following equations:

$$\Sigma_{\hat{\beta}_X} = (B_X^t \Sigma_{mX} B_X)^{-1}, \tag{28}$$

where B_X is the design matrix for Form X in the log-linear model (see Holland and Thayer, 1987) and

$$\Sigma_{mX} = N(D_{p_X} - p_X p_X^t), \tag{29}$$

where p_X is the vector of probabilities in the multinomial categories for Form X and D_{p_X} is a diagonal matrix made from p_X. $\Sigma_{\hat{\beta}_Y}$ can be obtained in a similar fashion. We also have

$$\Sigma_{\hat{\beta}_X \hat{\beta}_Y} = \mathbf{0} \tag{30}$$

The elements of (∂e_Y) can be obtained from the following equations:

$$\frac{\partial e_Y}{\partial \beta_{Xi}} = \frac{1}{\frac{\partial G(y;\beta_Y)}{\partial y}}\Bigg|_{y=e_Y(x)} \frac{\partial F(x; \beta_X)}{\partial \beta_{Xi}} \tag{31}$$

$$\frac{\partial e_Y}{\partial \beta_{Yi}} = -\frac{1}{\frac{\partial G(y;\beta_Y)}{\partial y}}\Bigg|_{y=e_Y(x)} \frac{\partial G(y;\beta_Y)}{\partial \beta_{Yi}}\Bigg|_{y=e_Y(x)}. \quad (32)$$

Given Equations 23 and 24, the derivatives in the above equations can be derived straightforwardly. But their expressions can be quite messy and thus are omitted here.

The general expression of SEE in Equation 25 applies to all designs. However, for designs other than the EG design, it could be quite complicated in calculating expression in Equation 27, depending on the specific design and equating method, and is beyond the scope of this paper.

6 Illustration with Real Test Data

6.1 Comparison of the Continuization Procedures

Because the CLL method performs the continuization step before applying the Design Function, and the kernel method applies the Design Function before the continuization step, the two continuization procedures can only be compared under the EG design where the Design Function is the identity function and thus can be skipped. This section compares the CLL and kernel continuization methods using two real datasets in terms of the smoothness of the continuous distribution and preservation of moments.

The first data set is taken from Table 7.1 of von Davier et al. (2004). It is a 20-item test data set. Only the Form X data are used here. First, the log-linear model is fitted with degree 2 to the raw frequency data. Then the kernel continuization is implemented with three different bandwidth parameter values: $h = 0.33, h = 0.622$, and $h = 1.0$. The h value of 0.622 represents the optimal h that minimizes the combined penalty function for this data set. The other two h values are sort of arbitrary, but were picked so that one is somewhat smaller than the optimal value and the other somewhat larger than the optimal value. Figure 1 contains plots of the raw frequency, the log-linear smoothed frequency, and the kernel continuized distributions with $h = .33$ and $h = .622$. The lower part of Figure 1 is a replication of Figure 4.2 in von Davier et al.

The continuized log-linear (CLL) distribution is plotted against the kernel distribution in Figure 2. This upper part shows that the kernel distributions are very close to the CLL distribution. In fact, the three lines almost coincide with each other, except with $h = 1$ making the kernel distribution depart slightly from the CLL distribution, especially at the ends of the score scale. As discussed previously, this departure reflects a distortion of the shape of the discrete distribution.

The lower part of Figure 2 plots the differences between the kernel distributions and the CLL distribution. It can be seen that with $h = .622$ the kernel distribution still has some bumps although they are too small to be seen in the upper part of the figure. (Note that the vertical scales for the upper and lower part of Figure 2 are very different.)

Table 1: The Moments and Differences in Moments for the 20-Item Data Set (with Kernel Moments Computed Based on Formula)

a. Moments

Form X	Raw Dist.	Log-linear	kernel(.33)	kernel(.622)	kernel(1.0)	CLL
mean	10.8183	10.8183	10.8183	10.8183	10.8183	10.8283
sd	3.8059	3.8059	3.8059	3.8059	3.8059	3.7909
skewness	0.0026	-0.0649	-0.0641	-0.0623	-0.0587	-0.0502
kurtosis	2.5322	2.6990	2.6588	2.5604	2.3616	2.6723

b. Difference in Moments with the Log-linear Discrete Distribution

Form X	kernel(.33)	kernel(.622)	kernel(1.0)	CLL
mean	0.0000	0.0000	0.0000	0.0100
sd	0.0000	0.0000	0.0000	-0.0150
skewness	0.0007	0.0025	0.0062	0.0147
kurtosis	-0.0401	-0.1386	-0.3373	-0.0267

The moments for different continuizations for this data set are in Table 1. Note that log-linear smoothing with degree 2 maintains the first two moments of the raw score distribution. The moments for the kernel distributions are computed based on the theoretical results in von Davier et al., namely, that the first two moments of kernel distribution are the same as the log-linear discrete distribution, but the skewness and kurtosis will differ by a factor of $(a_X)^3$ and $(a_X)^4$, respectively. The moments for CLL are empirically computed using numerical integration. For the kernel method, we can ignore the case of $h = .33$ since it produced unacceptably large bumps. It can be seen that all CLL moments approximate those of the log-linear distribution reasonably well, whereas the kernel methods have bigger differences in kurtosis. The kernel continuization did not distort the skewness of the distribution even when a large h was specified because the skewness of the discrete distribution is very small.

The same analyses were repeated for the 40-item ACT Mathematics data in Kolen and Brennan (2004). A log-linear model with a degree of 6 was fitted to the raw frequency. The same kernel and CLL continuization procedures were applied as for the first illustrative example. Three h parameter values were used for this data set: 0.33, 0.597, and 1.0. The value 0.597 represents the optimal h that minimizes the combined penalty function. (It turns out that in both data sets, the second penalty function PEN_2 does not have any effect on the combined penalty because there is no U-shaped distribution around any score point.) Results are plotted in Figure 3 and Figure 4. These plots show similar patterns of comparisons to those for the 20-item data set in the first example.

The moments of various distributions for this data set are in Table 2. With the kernel moments computed in either way, the CLL moments are slightly closer to the discrete distribution moments than the kernel moments although

Table 2: The Moments and Differences in Moments for the 40-Item Data Set (with Kernel Moments Computed Based on Formula)

a. Moments

Form X	Raw Dist	Log-linear	kernel(.33)	kernel(.597)	kernel(1.0)	CLL
mean	19.8524	19.8524	19.8524	19.8524	19.8524	19.8512
sd	8.2116	8.2116	8.2116	8.2116	8.2116	8.2105
skewness	0.3753	0.3753	0.3744	0.3722	0.3671	0.3751
kurtosis	2.3024	2.3024	2.2950	2.2783	2.2356	2.3023

b. Difference in Moments with the Log-linear Discrete Distribution

Form X	kernel(.33)	kernel(.597)	kernel(1.0)	CLL
mean	0.0000	0.0000	0.0000	-0.0012
sd	0.0000	0.0000	0.0000	-0.0012
skewness	-0.0009	-0.0031	-0.0082	-0.0001
kurtosis	-0.0074	-0.0241	-0.0668	-0.0001

both methods produce very close moments. The reason that the CLL method preserves moments better in this case is that the number of score points is larger and the approximation in Equation 7 is more accurate when the number of score points is larger.

Overall, these two illustrations confirm that the CLL continuization method has certain advantages over the kernel method with respect to simplicity, a smoother continuous distribution, and preserving moments better when the number of score points is relative large and the discrete distributions are highly skewed.

6.2 Comparisons of Equating Functions

The 40-item test data sets are also used to compare the equating functions under the EG based on three methods: (1) the traditional equipercentile equating method based on percentile ranks, (2) the kernel method, and (3) the CLL method. The optimal h parameters were used to compute the kernel continuous distributions. The traditional equipercentile method is also applied to the unsmoothed raw frequency data as a baseline. The results for the 40-item data set are in Table 3. The equating functions and their differences are plotted in Figure 5. The results in these tables and figures show that the equating functions based on these three methods are quite similar to each other. Except at the end points of the score scale, the differences are within 0.1.

Another real test data with a pairs of test forms are taken from Kolen and Brennan (2004, p. 147) to compare the CLL method with the kernel method under the NEAT design. The test has 36 items with a 12-item internal anchor test. The sample size is 1655 for the X group and 1638 for the Y group. A bivariate log-linear smoothing procedure is used for the smoothing step. The frequency estimation (FE) method is used for computing the equating function. The FE method under the NEAT design requires a rather complicated Design Function. Three continuization and equating methods are computed and compared: (1) the traditional equipercentile equating method based on percentile ranks, (2)the kernel method, and (3) the CLL method. The results are in Table 4. The equating functions and their differences are plotted in Figure 6. The results show that the CLL method produces very similar equating results as the kernel method and but slightly different than the traditional log-linear equipercentile method.

6.3 Comparison of SEE Estimates

The SEE's for the CLL method were computed for the 20-item data set using Equation 25 and are contained in Table 5. The SEE's for the kernel method were also computed and are presented in the same table. It can be seen that the SEE's for the two methods are very similar.

7 Discussion

This paper proposes an alternative continuization method for the test equating framework constructed by von Davier et al. (2004). With this new continuization method, there are two major differences between the proposed CLL method and the kernel method: First, the proposed CLL method directly uses the function form from the log-linear smoothing step and makes it into a pdf. Second, the continuization step occurs before the Design Function is applied. The illustration with real test data shows that with a relatively long test length, the CLL method produces smoother continuous score distributions and preserve the moments better than the kernel method. The equating results from the CLL method are quite similar to the kernel methods under both the EG design and the NEAT design. The similarity of the equating results make it difficult to make any recommendation about which method is the best choice under real testing situations. The comparisons are not comprehensive and lack objective criteria to evaluate the equating errors. A more thorough simulation study is needed to compare the kernel and CLL methods in order to make some practical recommendations.

A few differences between the CLL method and the kernel methods merit discussion. First, because the CLL method requires that the continuization step occurs before the Design Function is applied, the Design Function is applied on continuous distribution. This makes the expression of the Design Function easier to describe and program than the kernel method. For example, for the FE method under the NEAT design, the kerel methods applies a complicated set of matrix/vector operations in order to estimate the marginal distributions for the target population. For the CLL method, the Design Function is expressed nicely in Equations 13 through 20 .

Seond, the kernel method appear to have close mathematical form at the continuization and equating steps, whereas the CLL method requires numerical integration. A closer look finds that computing normal cdf in the kernel method also requires numerical integration or some approximation algorithm. Therefore, computationally speaking, both methods requires some kind of numerical method in the computation although the CLL method does require more frequent use of the numerical integration.

Finally, the kernel method requires a band-width parameter h for the Gaussian kernel. There are some advantage and disadvantages of having this parameter. The advantage is that it allow user to manipulate this parameter to achieve some goals. For example, when h is set very large, then the kernel method becomes a linear equating method. The disadvantage is that it is kind of arbitrary. Although von Davier et. al (2004) offered a penalty function to compute the optimal band-width, this penalty function is also arbitrary in some sense. The CLL method, on the other hand, does not have such parameter, thus saves a step in the computation.

References

Holland, P. W., & Thayer, D. T. (1987). Notes on the use of the log-linear models for fitting discrete probability distributions. *Technical Report, No. 87-79.* Princeton, NJ: Educational Testing Service.

Holland, P. W., King, B. F., & Thayer, D. T. (1989). The standard error of equating for the kernel method of equating score distributions. *Technical Report, No. 89-83.* Princeton, NJ: Educational Testing Service.

Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices (2nd Ed.)* New York: Springer-Verlag.

Thisted, R. A. (1988). *Elements of statistical computing: numerical computation.* New York: Champman & Hall/CRC.

von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating.* New York: Springer-Verlag.

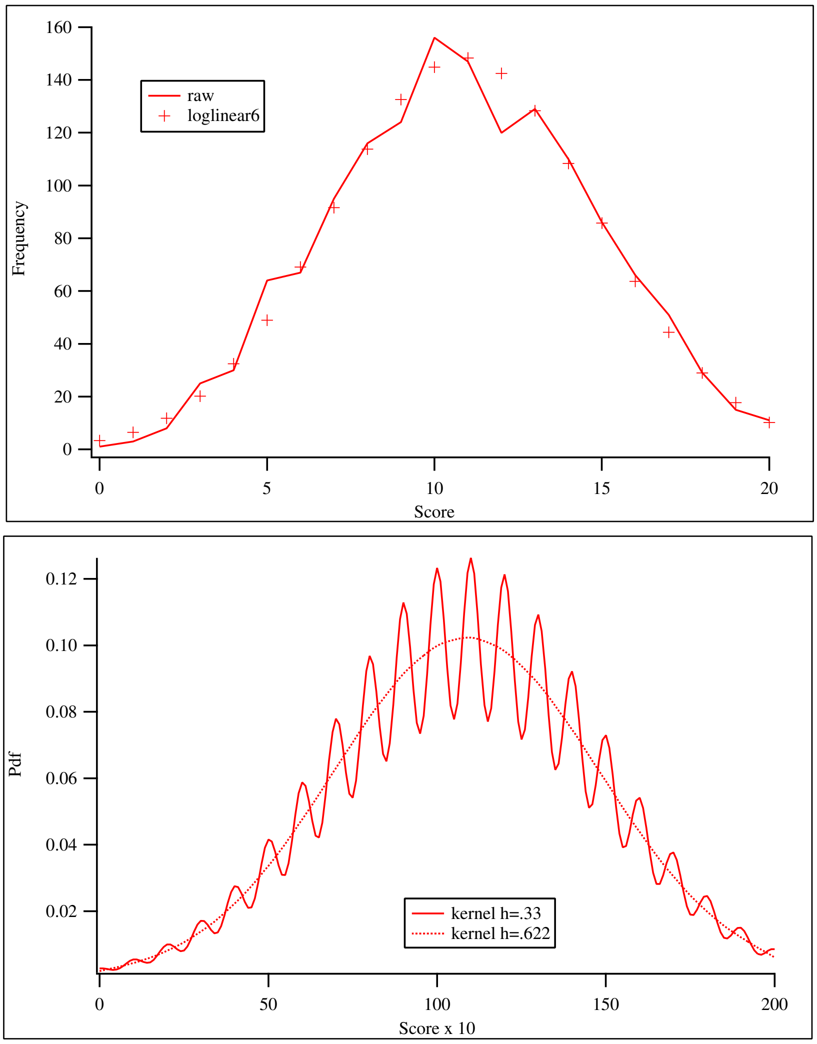


Figure 1: Plots of Raw Frequency, Log-linear Smoothing, and Kernel Continuization for the 20-Item Data Set

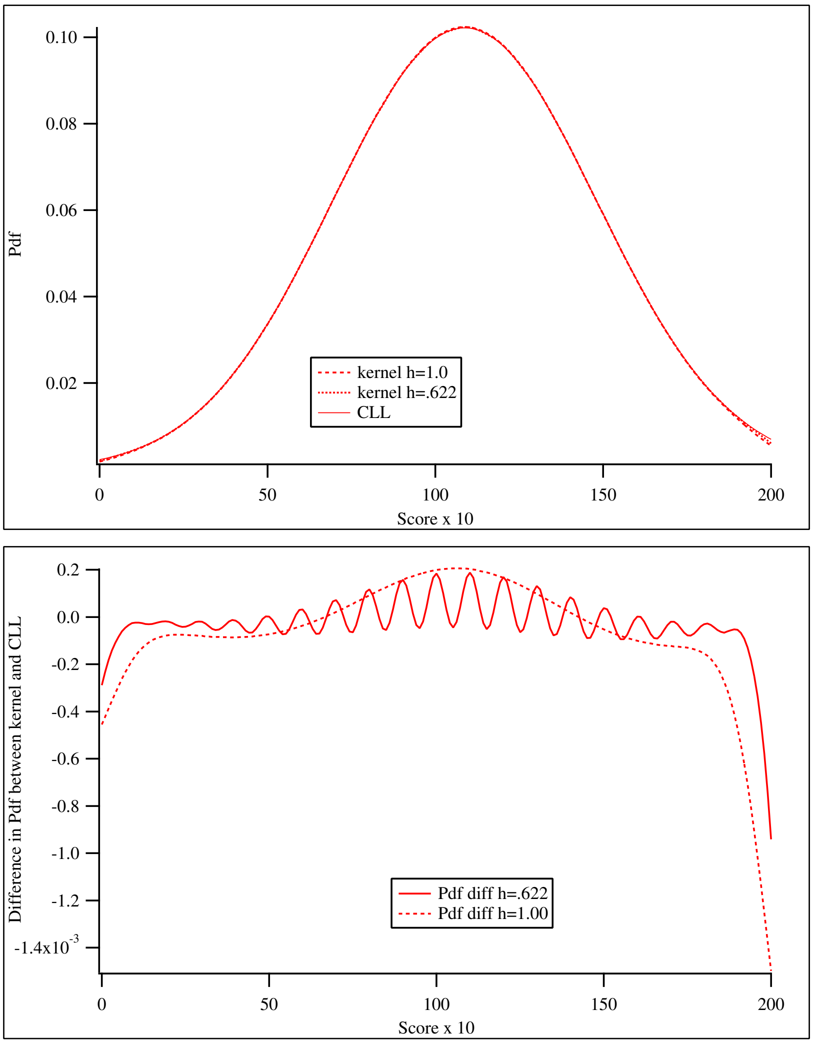


Figure 2: Comparisons of Kernel Continuization and CLL for the 20-Item Data Set

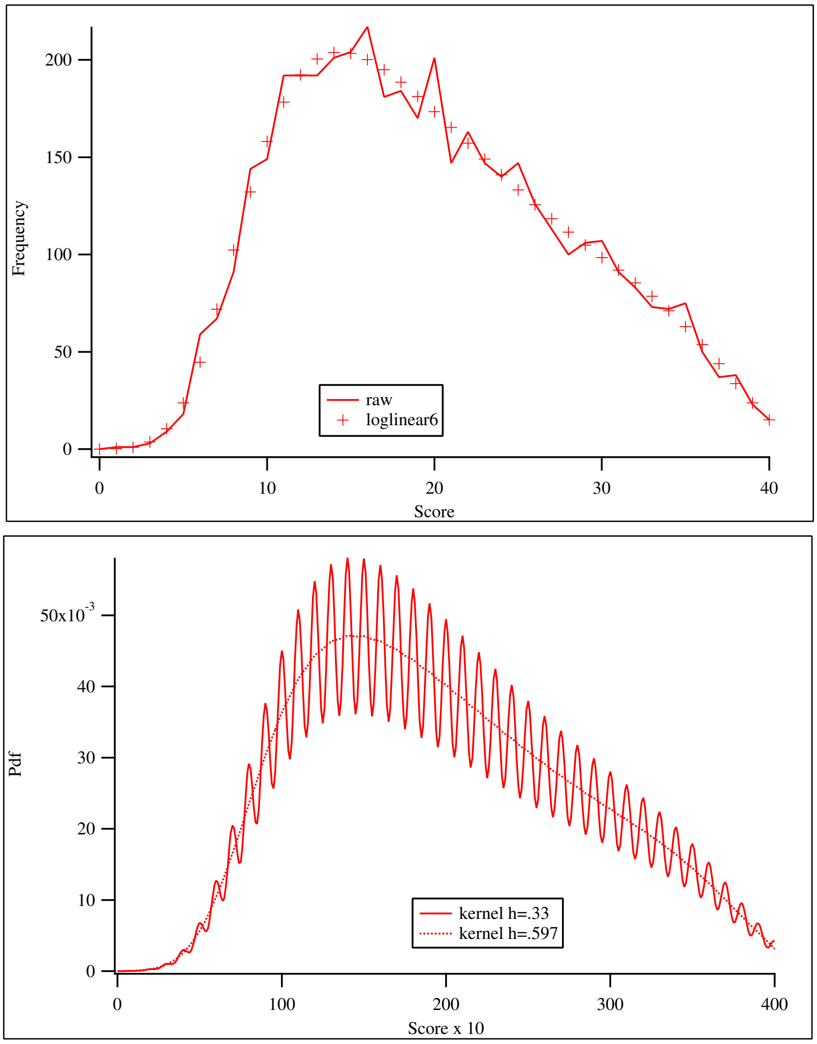


Figure 3: Plots of Raw Frequency, Log-linear Smoothing, and Kernel Continuization for the 40-Item Data Set

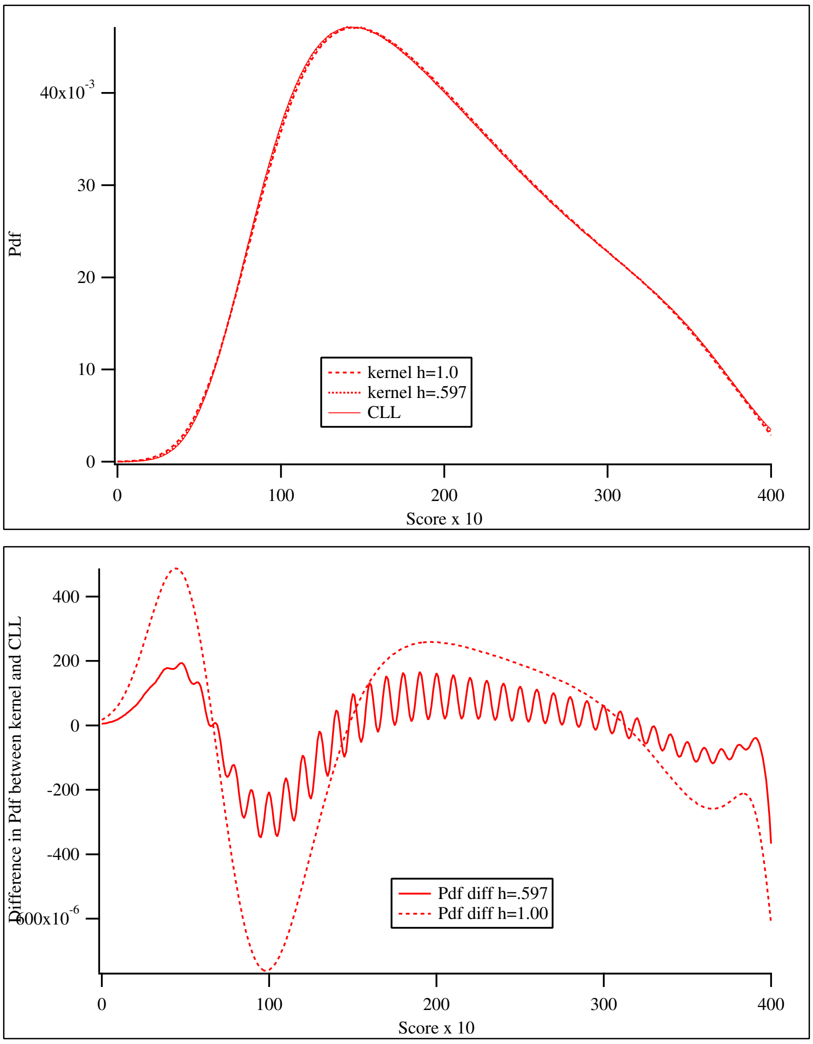
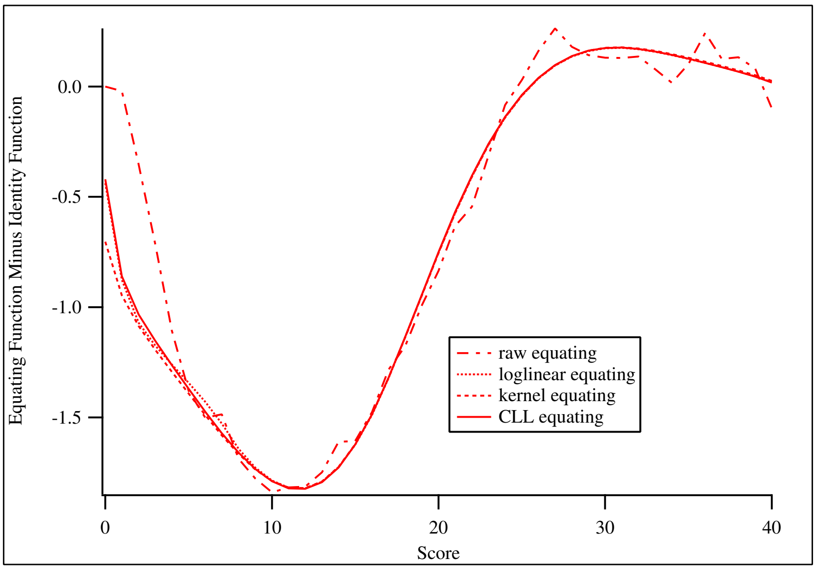


Figure 4: Comparisons of Kernel Continuization and CLL for the 40-Item Data Set

Table 3: The Equating Functions for the 40-Item Data Set under a EG Design

Score	Raw Equating	Log-linear	kernel(.597)	CLL Equating
0	0.0000	-0.4384	-0.7031	-0.4199
1	0.9796	0.1239	0.0537	0.1406
2	1.6462	0.9293	0.9143	0.9664
3	2.2856	1.8264	1.8069	1.8473
4	2.8932	2.7410	2.7072	2.7369
5	3.6205	3.6573	3.6082	3.6300
6	4.4997	4.5710	4.5112	4.5266
7	5.5148	5.4725	5.4191	5.4291
8	6.3124	6.3577	6.3355	6.3411
9	7.2242	7.2731	7.2648	7.2668
10	8.1607	8.2143	8.2119	8.2111
11	9.1827	9.1819	9.1819	9.1792
12	10.1859	10.1790	10.1798	10.1762
13	11.2513	11.2092	11.2101	11.2067
14	12.3896	12.2750	12.2761	12.2734
15	13.3929	13.3764	13.3784	13.3770
16	14.5240	14.5111	14.5147	14.5146
17	15.7169	15.6784	15.6790	15.6801
18	16.8234	16.8638	16.8623	16.8647
19	18.0092	18.0566	18.0541	18.0580
20	19.1647	19.2469	19.2449	19.2497
21	20.3676	20.4262	20.4263	20.4312
22	21.4556	21.5911	21.5916	21.5961
23	22.6871	22.7368	22.7365	22.7404
24	23.9157	23.8595	23.8588	23.8623
25	25.0292	24.9594	24.9586	24.9616
26	26.1612	26.0374	26.0369	26.0394
27	27.2633	27.0954	27.0955	27.0973
28	28.1801	28.1357	28.1364	28.1375
29	29.1424	29.1606	29.1621	29.1625
30	30.1305	30.1729	30.1750	30.1746
31	31.1297	31.1749	31.1777	31.1765
32	32.1357	32.1691	32.1726	32.1705
33	33.0781	33.1576	33.1618	33.1588
34	34.0172	34.1424	34.1470	34.1434
35	35.1016	35.1250	35.1300	35.1257
36	36.2426	36.1064	36.1118	36.1068
37	37.1248	37.0873	37.0929	37.0873
38	38.1321	38.0676	38.0729	38.0670
39	39.0807	39.0462	39.0514	39.0448
40	39.9006	40.0202	40.0256	40.0177



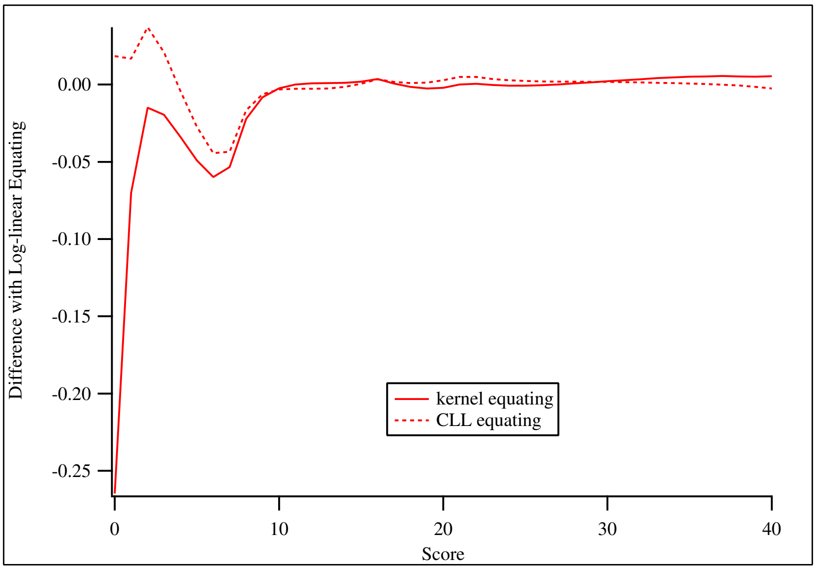


Figure 5: Comparisons of Equating Functions for the 40-Item Data Set under a EG Design

Table 4: The Equating Functions for the 36-Item Data Set under a NEAT Design

Score	Log-linear FE	kernel FE	CLL FE
0	-0.0129	0.0313	0.0059
1	1.0242	1.0827	1.0723
2	2.0988	2.1552	2.1357
3	3.1986	3.2375	3.2256
4	4.3091	4.3239	4.3132
5	5.4200	5.4122	5.4031
6	6.5194	6.5007	6.4929
7	7.5971	7.5880	7.5828
8	8.6759	8.6729	8.6688
9	9.7542	9.7546	9.7515
10	10.8305	10.8322	10.8303
11	11.9035	11.9053	11.9048
12	12.9721	12.9733	12.9738
13	14.0353	14.0357	14.0370
14	15.0924	15.0923	15.0941
15	16.1426	16.1423	16.1447
16	17.1854	17.1854	17.1881
17	18.2200	18.2209	18.2238
18	19.2460	19.2483	19.2513
19	20.2627	20.2668	20.2698
20	21.2694	21.2756	21.2783
21	22.2653	22.2740	22.2770
22	23.2495	23.2612	23.2636
23	24.2209	24.2362	24.2386
24	25.1784	25.1982	25.2004
25	26.1207	26.1464	26.1480
26	27.0466	27.0804	27.0813
27	27.9550	27.9997	27.9998
28	28.8454	28.9046	28.9050
29	29.7179	29.7964	29.7949
30	30.5739	30.6769	30.6760
31	31.4295	31.5496	31.5461
32	32.2939	32.4192	32.4163
33	33.1700	33.2920	33.2908
34	34.0732	34.1764	34.1807
35	35.0130	35.0836	35.1035
36	35.9983	36.0326	36.0800

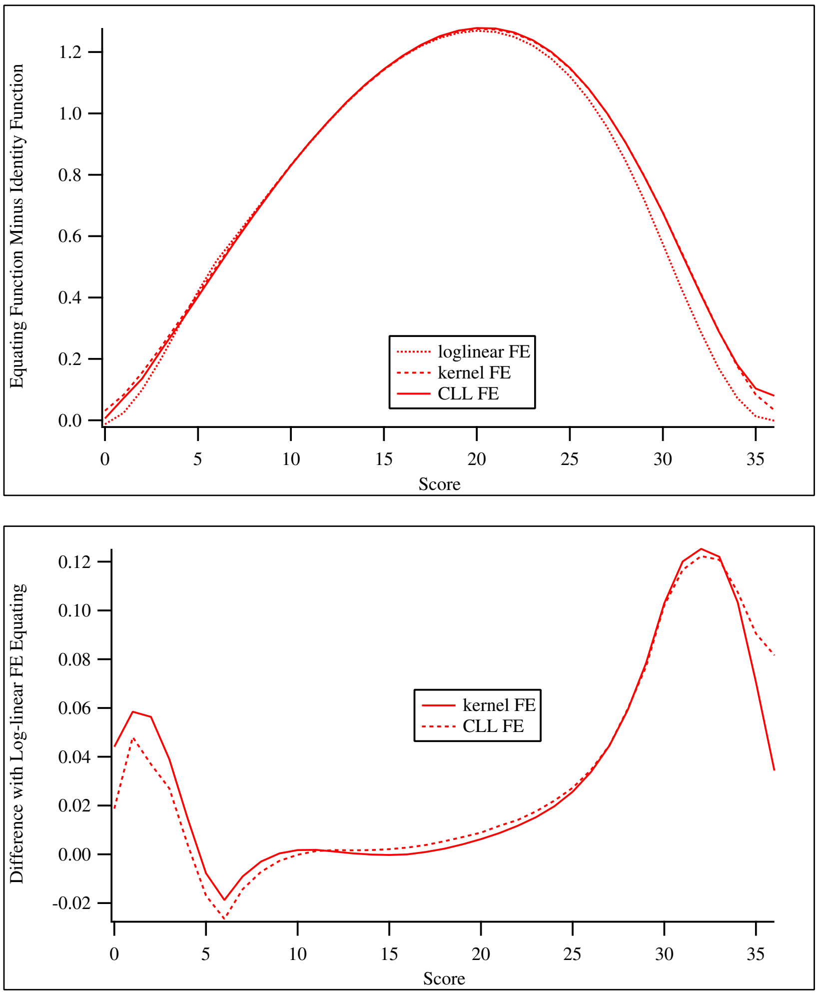


Figure 6: Comparisons of Equating Functions for the 36-Item Data Set under a NEAT Design

Table 5: The Standard Errors of Equating for the 20-Item Data Set

Score	kernel SEE	CLL SEE
0	0.2200	0.2100
1	0.2895	0.2933
2	0.2875	0.2904
3	0.2664	0.2682
4	0.2410	0.2418
5	0.2170	0.2169
6	0.1967	0.1963
7	0.1812	0.1808
8	0.1708	0.1705
9	0.1646	0.1646
10	0.1619	0.1622
11	0.1621	0.1627
12	0.1653	0.1661
13	0.1721	0.1731
14	0.1827	0.1839
15	0.1951	0.1969
16	0.2038	0.2064
17	0.1990	0.2028
18	0.1700	0.1747
19	0.1186	0.1170
20	0.0703	0.0396