Research Report

Comparing Alternative Kernels for the Kernel Method of Test Equating: Gaussian, Logistic, and Uniform Kernels

Yi-Hsuan Lee

Alina A. von Davier

March 2008

ETS RR-08-12

Listening. Learning. Leading.®

Comparing Alternative Kernels for the Kernel Method of Test Equating: Gaussian, Logistic, and Uniform Kernels

Yi-Hsuan Lee and Alina A. von Davier
ETS, Princeton, NJ

March 2008

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Copyright © 2008 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, and LISTENING. LEARNING. LEADING. are
registered trademarks of Educational Testing Service (ETS).

Abstract

The kernel equating method (von Davier, Holland, & Thayer, 2004) is based on a flexible family of equipercentile-like equating functions that use a Gaussian kernel to continuize the discrete score distributions. While the classical equipercentile, or percentile-rank, equating method carries out the continuization step by linear interpolation, in principle the kernel equating methods could use various kernel smoothings to replace the discrete score distributions.

This paper expands the work of von Davier et al. (2004) in investigating alternative kernels for equating practice. To examine the influence of different kernel functions on the equating results, this paper focuses on two types of kernel functions: the logistic kernel and the continuous uniform distribution (known to be the same as the linear interpolation). The Gaussian kernel is used for reference. By employing an equivalent-groups design, the results of the study indicate that the tail properties of kernel functions have great impact on the continuized score distributions. However, the equated scores based on different kernel functions do not vary much, except for extreme scores.

The results presented in this paper not only support the previous findings on the efficiency and accuracy of the existing continuization methods, but also enrich the information on observed-score equating models.

Key words: Kernel equating, gaussian kernel, logistic kernel, uniform kernel, cumulants, equivalent-groups design

Acknowledgments

The authors would like to thank Dan Eignor, Shelby Haberman, Paul Holland, Frank Rijmen, and Sandip Sinharay for their helpful comments, and Kim Fryer for her editorial assistance.

Introduction

The need for test equating arises when there are two or more test forms that measure the same construct and that can yield different scores for the same examinee. The most common example involves multiple forms of a test within a testing program, as opposed to a single testing instrument. In a testing program, different test forms that are similar in content and format typically contain completely different test questions. Consequently, the tests can vary in difficulty depending on the degree of control available in the test development process. Examinees tested with the more difficult test form will receive lower scores than they would had they been tested with the easier form. Because testing programs often require comparability of the scores produced on these different forms, test-equating techniques were developed to adjust for these differences in test difficulty across test forms.

The goal of test equating is to allow the scores on different forms of the same test to be used and interpreted interchangeably. Test equating requires some type of control for differential examinee ability, or proficiency, in the assessment of, and adjustment for, differential test difficulty; the differences in abilities are controlled by employing an appropriate data collection design.

Many observed-score equating methods are based on the equipercentile equating function, which requires that the initial discrete score distribution functions have been continuized. Several important observed-score equating methods may be viewed as differing only in the way the continuization is achieved. The traditional equipercentile equating method (percentile-rank method) uses linear interpolation of the discrete distribution to make it piecewise linear and therefore continuous. The kernel equating (KE; von Davier, Holland, & Thayer, 2004) method uses Gaussian kernel smoothing to approximate the discrete histogram by a continuous density function.

Von Davier et al. (2004) introduced not only a continuization method for discrete score distributions but also a conceptual framework for the equating process. In this framework, five consecutive steps for manipulation of the raw data are developed in such a way that each step explicitly contributes to the equated scores and their accuracy. The five steps are (a) presmoothing of the discrete score distributions using loglinear models (Holland & Thayer, 2000); (b) estimating the marginal discrete score distributions by applying the design function, a mapping that reflects the data collection design; (c) continuization of the distributions; (d) computing the equating function and diagnosing it; and (e) computing several accuracy measures, such as the standard

error of equating (SEE) and the standard error of equating difference (SEED).

This paper expands the work of von Davier et al. (2004) by also looking at kernel functions other than the Gaussian kernel. Adopting the KE framework of von Davier et al. (2004), we will apply or adapt each of the five steps to incorporate the alternative kernels, the logistic kernel and the uniform kernel, along with the Gaussian kernel (GK).

The GK is the kernel function in common use, but it may not result in the best continuous approximation of the observed discrete score distribution in terms of cumulants due to the fact that the normal distribution has zero cumulants of orders higher than 2. The logistic kernel may work better in this regard since its cumulants are not all zero for orders higher than 2 (see Method section). The uniform kernel, on the other hand, is known to lead to the linear interpolation process adopted in the percentile-rank method. Thus, how it performs compared to the GK and the logistic kernel is of interest.

The rest of this section introduces basic notation. Two test forms are to be equated, X and Y, and a target population, T, on which this is to be done. The corresponding possible scores on T are X and Y, respectively. The data are collected in such a way that the differences in the difficulty of the test forms and the differences in the ability of the test-takers that take the two forms are not confounded. Two classes of data collection designs are used for equating: (a) designs that allow for *common people* (equivalent-groups, single-group, and counterbalanced designs) from a single target population of examinees T (see Livingston, 2004, for a slightly different view and definition of a target population); and (b) designs that allow for *common items* (the nonequivalent groups with an anchor test, or NEAT, design, also referred to as the common-item or anchor-test design) where the tests, X and Y, are given to two samples from two test populations (administrations), P and Q, respectively, and a set of common items (the *anchor test*) is given to samples from both these populations. As the name implies, in a NEAT design the samples from P and Q are not assumed to be of equivalent ability. The target population, T, for the NEAT design is assumed to be a weighted average of P and Q where P and Q are given weights that sum to 1. This is denoted by $T = wP + (1 - w)Q$.

The equipercentile equating function is defined on the target population, T, as $e_{Y;T}(x) = G_T^{-1}(F_T(x))$, where $F_T(x)$ and $G_T(y)$ are the cumulative distribution functions (CDFs), of X and Y, respectively, on T. In order for this definition to make sense and to insure that the inverse equating function exists, it is also assumed that $F_T(x)$ and $G_T(y)$ are

strictly increasing and have been made continuous (or continuized). Equipercentile equating leads to linear equating if one assumes that $F_T(x)$ and $G_T(y)$ are continuous and have the same shape while differing in mean and variance. The linear equating function is defined by $Lin_{Y;T}(x) = \mu_{YT} + \sigma_{YT}((x - \mu_{XT})/\sigma_{XT})$, where μ_{XT}, μ_{YT}, σ_{XT} and σ_{YT} are the means and standard deviations of X and Y on T, respectively.

This study will examine the effect of applying different kernel functions when using the equivalent-groups (EG) design. Equating that makes use of these kernel functions can be done with any of the other designs, such as the counterbalanced design and the NEAT design, with only slight modifications.

The next section describes the five-step process for kernel equating with a generic kernel function. Then, a detailed description is provided of the continuous distributions (including the cumulants) that are obtained through use of the newly investigated kernels (logistic and uniform). The subsequent section describes the results obtained by applying the kernel functions to the EG data given in Chapter 7 of von Davier et al. (2004). The last section of the paper discusses the results and draws conclusions.

Method

Suppose the two tests, X and Y, have J and K possible raw-score values. Denote these possible scores of X and Y by $X = \{x_1, \ldots, x_J\}$ and $Y = \{y_1, \ldots, y_K\}$, respectively. In the case of concern, assume $x_1, \ldots, x_J$ to be consecutive integers; similarly for $y_1, \ldots, y_K$. As Braun and Holland (1982) emphasized, observed-score test equating always takes place on a specific population of examinees. We assume that this population is fixed and let $\mathbf{r} = \{r_j\}$ and $\mathbf{s} = \{s_k\}$ denote the score probabilities for this population, $r_j = P(X = x_j)$ and $s_k = P(Y = y_k)$. The CDFs of the score distributions for X and Y are

$$F(x) = \mathrm{P}(X \leq x) = \sum_{j:\, x_j \leq x} r_j \quad \text{and} \tag{1}$$

$$G(y) = \mathrm{P}(Y \leq y) = \sum_{k:\, y_k \leq y} s_k. \tag{2}$$

However, $\mathbf{r}$ and $\mathbf{s}$ are unobservable population parameters. In reality, the raw data obtained from an EG design are two sets of univariate frequencies $\{n_j\}$ and $\{m_k\}$, where n_j= number of examinees in sample one with $X = x_j$ and m_k= number of examinees in sample two with $Y = y_k$,

with sample sizes $N = \sum_j n_j$ and $M = \sum_k m_k$, respectively. As a result, the presmoothing step has to be carried out beforehand to estimate the population score probabilities $\mathbf{r}$ and $\mathbf{s}$.

Presmoothing

In the EG design, the score distributions X and Y are independent. As in Holland and Thayer (2000) and von Davier et al. (2004), a separate loglinear model is fitted to each univariate distribution using sample proportions as probabilities and design matrices (i.e., power moments of the scores) as covariates. The moments preserved in the loglinear models are $T_{\mathbf{r}}$ and $T_{\mathbf{s}}$ for $\mathbf{r}$ and $\mathbf{s}$, respectively. The score probabilities $\mathbf{r}$ and $\mathbf{s}$ are estimated by their maximum likelihood estimates (MLEs), $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$, respectively.

Let $\Sigma_{\hat{\mathbf{r}}}$ and $\Sigma_{\hat{\mathbf{s}}}$ denote the covariance matrices of $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$, respectively. Holland and Thayer (1987) proved that, for MLEs $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$, there exists a $J \times T_{\mathbf{r}}$ matrix $C_{\mathbf{r}}$ and a $K \times T_{\mathbf{s}}$ matrix $C_{\mathbf{s}}$ such that

$$\Sigma_{\hat{\mathbf{r}}} = C_{\mathbf{r}} C_{\mathbf{r}}^T \quad \text{and} \quad \Sigma_{\hat{\mathbf{s}}} = C_{\mathbf{s}} C_{\mathbf{s}}^T, \tag{3}$$

where T stands for the transpose of the matrix. The matrices $C_{\mathbf{r}}$ and $C_{\mathbf{s}}$ are called *C-matrices* (von Davier et al., 2004). The C-matrices are one of the key components in evaluating the standard error of equating (SEE) and the standard error of equating difference (SEED) of the equating functions, under the assumption that the loglinear models hold.

Continuization

The KE method is based on a flexible family of equipercentile-like equating functions. One score x on test X is said to be equivalent to one score y on test Y if x and y are at the same percentile in the population. If both score distributions X and Y were continuous, the equating function $e_Y(x)$ would have the form

$$e_Y(x) = G^{-1}(F(x)). \tag{4}$$

To apply Equation 4 when X and Y are discrete, continuous approximations of them can be found with means (and variances) remaining the same as their discrete alternatives. In the KE framework, this can be achieved by adding a continuous and independent random variable to both X and Y and by taking certain linear transformations afterwards. In the classical equipercentile method, continuization is achieved by linear interpolation, and only the means of the discrete distributions are preserved.

The kernel functions are the densities of the added continuous random variable. Consider $X(h_X)$ as a continuous transformation of X such that

$$X(h_X) = a_X(X + h_X V) + (1 - a_X)\mu_X, \tag{5}$$

where

$$a_X^2 = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_V^2 h_X^2} \tag{6}$$

and h_X is the bandwidth controlling the degree of smoothness. In this equation, V is a continuous (kernel) distribution. When h_X is *large*, the distribution of $X(h_X)$ approximates the distribution of V approximately; when h_X is small, $X(h_X)$ approaches X. In von Davier et al. (2004), V follows a standard normal distribution. In the exposition below, V is a generic continuous distribution.

It is easy to verify that $\mathrm{E}(X(h_X)) = \mu_X$ and $\mathrm{Var}(X(h_X)) = \sigma_X^2$. Similarly, the continuous approximation of Y is defined as

$$Y(h_Y) = a_Y(Y + h_Y V) + (1 - a_Y)\mu_Y, \tag{7}$$

where

$$a_Y^2 = \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_V^2 h_Y^2}, \tag{8}$$

and h_Y is the bandwidth.

In the next theorem, we illustrate a few limiting properties of $X(h_X)$ and a_X^2 that indicate their behavior as h_X takes on different values. This theorem represents a generalization of Theorem 4.1 in von Davier et al. (2004) to other kernel functions.

Theorem 1. The following statements hold:

(a) $\lim_{h_X \to 0} a_X = 1$;
(b) $\lim_{h_X \to \infty} a_X = 0$;
(c) $\lim_{h_X \to \infty} h_X a_X = \sigma_X / \sigma_V$;
(d) $\lim_{h_X \to 0} X(h_X) = X$; and
(e) $\lim_{h_X \to \infty} X(h_X) = (\sigma_X / \sigma_V)V + \mu_X$.

The next theorem shows the asymptotic form of the CDF when h varies. If h is small then the CDF of the continuized distribution will closely track the original discrete distribution, and

if h is large the CDF will approximate the distribution of the kernel, preserving the mean and variance of the original distribution.

Theorem 2. Let

$$R_{jX}(x) = \frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}. \tag{9}$$

It has the following approximate form when $h_X \to 0$ and when $h_X \to \infty$:

(a) $R_{jX}(x) = \dfrac{x - x_j}{h_x} + o(h_X)$ as $h_X \to 0$, and

(b) $R_{jX}(x) = \dfrac{x - \mu_X}{\sigma_X/\sigma_V} - \left(\dfrac{\sigma_X}{\sigma_V h_X}\right) \cdot \left(\dfrac{x - \mu_X}{\sigma_X/\sigma_V}\right) + o(\dfrac{\sigma_X}{\sigma_V h_X})$ as $h_X \to \infty$.

As mentioned before, it has been argued that the GK may not result in the best continuous approximation of the observed discrete score distribution in terms of cumulants due to the fact that the normal distribution has zero cumulants of orders higher than 2. This will be investigated in the rest of this section, where the use of the logistic kernel and of the uniform kernel are separated into two cases. Their properties of cumulants will be discussed, and their corresponding density functions of $X(h_X)$ and the penalty functions for the selection of bandwidths will be defined explicitly. At the end of this section, one way to examine the continuized score distributions will also be provided.

Case 1: Logistic Kernel Function

Suppose V is a logistic random variable and is independent of X and Y. Its probability density function (PDF) has the form:

$$h(v) = \frac{\exp\{-v/s\}}{s(1 + \exp\{-v/s\})^2}, \tag{10}$$

and its CDF is given by

$$H(v) = \frac{1}{1 + \exp\{-v/s\}}, \tag{11}$$

where s is the scale parameter. V has zero mean and variance $\sigma_V^2 = \pi^2 s^2/3$. Varying the scale parameter would expand or shrink the distribution. If $s = 1$, the distribution is called the *standard logistic* (SL), whose variance is $\pi^2/3$. We can rescale the distribution so that it has zero mean and identity variance, which can be accomplished by setting $s = \sqrt{3}/\pi$. It is called the *rescaled logistic* (RL) in this paper. (From now on, *SLK* stands for the cases where SL is used as the kernel

function, and *RLK* stands for those with RL kernel function. Without specification, *LK* will represent the logistic kernel in general.)

The next two theorems are generalizations of Theorems 4.2 and 4.3 of von Davier et al. (2004) to the logistic kernel. The CDF and PDF of $X(h_X)$ in Theorem 3 demonstrate that the LK function actually serves as a smoother on the discrete score distribution and that the degree of smoothness depends on the choice of bandwidth h_X.

Theorem 3. If $X(h_X)$ is defined in Equation 5, its CDF is given by

$$F_{h_X}(x) = \sum_j r_j H(R_{jX}(x)) \tag{12}$$

with $R_{jX}(x)$ defined in Equation 9. The function $F_{h_X}(x)$ is the continuous approximation of $F(x)$, the CDF of X. In addition, the corresponding PDF is

$$f_{h_X}(x) = \frac{1}{a_X h_X} \sum_j r_j h(R_{jX}(x)). \tag{13}$$

One way to see how close the continuized CDFs are to the discrete CDFs is to compute the cumulants of $F_{h_X}(x)$ and $G_{h_Y}(y)$ and compare them with the cumulants of $F(x)$ and $G(y)$, respectively. The jth cumulant of $X(h_X)$, $\kappa_j(h_X)$, is defined by the coefficient of $(t)^j/j!$ in the cumulant-generating function $g(t)$,

$$g(t) = \log(E(\exp[tX(h_X)])) = \sum_{j=1}^{\infty} \frac{\kappa_j(h_X)t^j}{j!} \tag{14}$$

(Abramowitz & Stegun, 1972). Let $g^{(j)}(\cdot)$ denote the jth derivative of $g(\cdot)$. It is well known that $\kappa_1(h_X) = \mu_X = g^{(1)}(0)$ and $\kappa_2(h_X) = \sigma_X^2 = g^{(2)}(0)$, the mean and variance of $X(h_X)$, respectively. In general, $\kappa_j(h_X) = g^{(j)}(0)$. Two useful properties of cumulants are stated in the following.

Properties. Let $\kappa_{j,V}$ denote the jth cumulant of V. For any constant c,

1. $\kappa_{1,V+c} = c + \kappa_{1,V}$ but $\kappa_{j,V+c} = \kappa_{j,V}$ for $j \geq 2$; and
2. $\kappa_{j,cV} = c^j \cdot \kappa_{j,V}$ for $j \geq 1$.

Let $\kappa_{j,X}$ denote the jth cumulant of X. Holland and Thayer (1989) noted that, if GK is applied in the continuization step,

$$\kappa_j(h_X) = (a_X)^j \kappa_{j,X} \text{ for } j \geq 3. \tag{15}$$

The cumulants before and after continuization have a concise relationship due to the fact that a normal distribution has zero cumulants of orders higher than 2.

The heavier tails and sharper peak of a logistic distribution lead to larger cumulants of even orders than do those of a normal distribution. Apparently, the above relationship will no longer hold if LK is used. Suppose V is a logistic random variable with mean zero and variance $\pi^2/3$ (i.e., the case of SLK). For $|t| < 1$ the moment-generating function of V is given by

$$\begin{aligned} M_V(t) = \mathrm{E}(e^{tV}) &= \int_{-\infty}^{\infty} e^{tv} \cdot \frac{e^{-v}}{(1+e^{-v})^2}\, dv \\ &= \int_0^1 \xi^{-t}(1-\xi)^t\, d\xi \\ &= B(1-t, 1+t) \\ &= \Gamma(1-t)\cdot\Gamma(1+t), \end{aligned}$$

where $\xi = (1+e^v)^{-1}$, $B(\cdot,\cdot)$ is the beta function, and $\Gamma(\cdot)$ is the gamma function (Balakrishnan, 1992). The cumulant-generating function of V is

$$\log M_V(t) = \log\Gamma(1-t) + \log\Gamma(1+t). \tag{16}$$

Let $\Gamma^{(j)}(\cdot)$ be the jth derivative of $\Gamma(\cdot)$, for any positive integer j. The next theorem gives the mathematical expressions of the cumulants for the SLK. The results can be generalized to any logistically distributed random variable according to the properties of cumulants mentioned above.

Theorem 4. Define

$$\psi(u) = \frac{d\log\Gamma(u)}{du} = \frac{\Gamma^{(1)}(u)}{\Gamma(u)}, \tag{17}$$

and let $\psi^{(j)}(\cdot)$ be the jth derivative of $\psi(\cdot)$ for any positive integer j. Then the jth cumulant of a standard logistic random variable V is found to be

$$\kappa_{j,V} = \begin{cases} 0 & \text{if } j \text{ is odd} \\ 2\cdot\psi^{(j-1)}(1) & \text{if } j \text{ is even} \end{cases}. \tag{18}$$

For any $j \geq 1$ the value of $\psi^{(j-1)}(1)$ is given by

$$\begin{aligned} \psi^{(j-1)}(1) &= (-1)^j (j-1)!\, \zeta(j), \text{ and} \\ \psi(1) &= \Gamma^{(1)}(1) = -0.5772, \end{aligned}$$

where $\zeta(\cdot)$ is the Riemann zeta function. These numbers have been tabulated by Abramowitz and Stegun (1972), and the first six values of $\zeta(j)$ are $\zeta(1) = \infty$, $\zeta(2) = \pi^2/6$, $\zeta(3) \approx 1.2021$, $\zeta(4) = \pi^4/90$, $\zeta(5) \approx 1.0369$, and $\zeta(6) = \pi^6/945$. For example, we obtain

$$\mathrm{E}(V) = \Gamma^{(1)}(1) - \Gamma^{(1)}(1) = 0, \text{ and}$$

$$\mathrm{Var}(V) = 2 \cdot \psi^{(1)}(1) = \pi^2/3$$

since $\psi^{(1)}(1) = \pi^2/6$.

The continuization of the discrete $\mathbf{r}$ and $\mathbf{s}$ into continuous PDFs of $X(h_X)$ and $Y(h_Y)$ requires the selection of bandwidths. Take $X(h_X)$ for example. The optimal bandwidth is defined by von Davier et al. (2004) as the minimizer of the penalty function comprising two components. One is the least square term

$$\mathrm{PEN}_1(h_X) = \sum_j \left(\hat{r}_j - \hat{f}_{h_X}(x_j) \right)^2. \tag{19}$$

The other is the smoothness penalty term that avoids rapid fluctuations in the approximating density,

$$\mathrm{PEN}_2(h_X) = \sum_j A_j(1 - B_j), \tag{20}$$

where

$$A_j = \begin{cases} 1 & \text{if } f^{(1)}_{h_X}(x) < 0 \text{ at } x = x_j - 0.25 \\ 0 & \text{otherwise} \end{cases}, \tag{21}$$

$$B_j = \begin{cases} 0 & \text{if } f^{(1)}_{h_X}(x) > 0 \text{ at } x = x_j + 0.25 \\ 1 & \text{otherwise} \end{cases}, \tag{22}$$

and $f^{(1)}_{h_X}(x)$, the first derivative of $f_{h_X}(x)$, is defined as

$$f^{(1)}_{h_X}(x) = \frac{1}{s} \sum_j r_j \cdot h(R_{jX}) \cdot [1 - 2H(R_{jX})] \cdot \left(\frac{1}{a_X h_X} \right)^2. \tag{23}$$

Choices of h_X that allow a U-shaped $f_{h_X}(x)$ around the score value x_j would result in a penalty of 1. Combining PEN_1 and PEN_2 gives the complete penalty function

$$\mathrm{PEN} = \mathrm{PEN}_1 + \mathrm{PEN}_2, \tag{24}$$

which will keep the discrete distribution $\mathbf{r}$ and the continuized density $f_{h_X}(x)$ close to each other, while preventing $f_{h_X}(x)$ from having too many zero derivatives.

Case 2: Uniform Kernel Function

Suppose V is a uniform random variable with PDF

$$h(v) = \begin{cases} \dfrac{1}{2b} & \text{for } -b < v < b \\ 0 & \text{otherwise} \end{cases}, \tag{25}$$

where b is a positive real number. The corresponding CDF is

$$H(v) = \begin{cases} 0 & \text{for } v < -b \\ \dfrac{v+b}{2b} & \text{for } -b \leq v < b \\ 1 & \text{for } v \geq b \end{cases}. \tag{26}$$

V has mean zero and variance $b^2/3$. Moreover, V is independent of X and Y. To see how V affects our kernel equating process, two cases are examined: V is said to follow the *standard uniform* (SU) distribution if $b = 1/2$. Its variance is $\sigma_V^2 = 1/12$. When V is rescaled to have identity variance (i.e., $b = \sqrt{3}$), the resulting distribution is called *rescaled uniform* (RU) here. SU and RU will be incorporated in the procedure of continuization and these methods will be denoted as *SUK* and *RUK*, respectively. Without specification, *UK* will stand for the uniform kernel. The following theorem gives the CDF and PDF of $X(h_X)$ when the uniform kernel is applied to Equations 5 and 7. Note that linear interpolation as it is achieved in existing equating practice does not involve rescaling, which leads to a continuous distribution that does not preserve the variance of the original discrete distribution.

Theorem 5. If $X(h_X)$ is defined as in Equation 5 with V following a uniform distribution, its CDF is given by

$$F_{h_X}(x) = \sum_{\substack{j\,:\\ R_{jX}(x) \geq b}} r_j + \sum_{\substack{j\,:\\ -b \leq R_{jX}(x) \leq b}} \left\{ r_j \cdot \frac{R_{jX}(x) + b}{2b} \right\}, \tag{27}$$

where $R_{jX}(x)$ is defined in Equation 9. In addition, the corresponding PDF is

$$f_{h_X}(x) = \frac{1}{a_X h_X} \sum_{\substack{j\,:\\ -b \leq R_{jX}(x) \leq b}} \left(\frac{r_j}{2b} \right). \tag{28}$$

Following the previous notation, $\kappa_{j,V}$ is the jth cumulant of V. Kupperman (1952) showed that all odd cumulants vanish and even cumulants are given by

$$\kappa_{j,V} = \frac{(2b)^j \cdot B_j}{j} \quad \text{for even number } j, \tag{29}$$

where $\{B_j\}$ are Bernoulli numbers. The first eleven Bernoulli numbers are $B_0 = 1$, $B_1 = -1/2$, $B_2 = 1/6$, $B_4 = -1/30$, $B_6 = 1/42$, $B_8 = -1/30$, $B_{10} = 5/66$, and $B_3 = B_5 = B_7 = B_9 = 0$.

As mentioned before, the degree of smoothness also relies on the choice of bandwidth. Similar to the case of LK, the optimal bandwidth minimizes the same penalty function given in Equation 24 with $f^{(1)}_{h_X}(x) = 0$ for all x satisfying $R_{jX}(x) \neq \pm b,\ j = 1, \ldots, J$ since $f_{h_X}(x)$ is piecewise constant.

Examination of the Continuized Score Distributions

The next theorem provides the relationship between the cumulants of the discrete score distributions and those of the corresponding continuized approximations.

Theorem 6. Let $\kappa_j(h_X)$ denote the jth cumulant of $X(h_X)$, $\kappa_{j,X}$ denote the jth cumulant of X, and $\kappa_{j,V}$ denote the jth cumulant of V. Then for $j \geq 3$,

$$\kappa_j(h_X) = (a_X)^j \cdot \Big(\kappa_{j,X} + (h_X)^j \cdot \kappa_{j,V}\Big). \tag{30}$$

As mentioned before, GK has $\kappa_{j,V} = 0$ for all $j \geq 3$, so $\kappa_j(h_X)$ must be smaller than $\kappa_{j,X}$ in magnitude for all j larger than 2. Meanwhile, LK has positive $\kappa_{j,V}$ for all even j greater than 3, which makes it possible that the latter could produce a better continuous approximation to X in terms of the cumulant when X also has positive even cumulants.

Equating

Once the KE continuized versions of $F(x)$ and $G(y)$, $F_{h_X}(x; \hat{\mathbf{r}})$ and $G_{h_Y}(y; \hat{\mathbf{s}})$, are in hand, the equating function defined in Equation 4 that transforms X to Y can be applied to $X(h_X)$ and $Y(h_Y)$,

$$\hat{e}_Y(x) = e_Y(x; \hat{\mathbf{r}}, \hat{\mathbf{s}}) = G^{-1}_{h_Y}(F_{h_X}(x; \hat{\mathbf{r}}); \hat{\mathbf{s}}). \tag{31}$$

Similarly, the equating function converting Y to X is given by

$$\hat{e}_X(y) = e_X(y; \hat{\mathbf{r}}, \hat{\mathbf{s}}) = F^{-1}_{h_X}(G_{h_Y}(y; \hat{\mathbf{s}}); \hat{\mathbf{r}}). \tag{32}$$

Evaluating $\hat{e}_Y(x)$ and $\hat{e}_X(y)$ at the possible raw-score values would give the equated scores from X to Y and those from Y to X, respectively. No matter how well the discrete CDFs are approximated by their continuized versions, the problem of concern is whether or not the distribution of the equated scores is similar to the target distribution (i.e., if X is transformed to Y, then Y is the target distribution). To diagnose the effectiveness of the transformation, the moments of Y and those of $e_Y(X)$ are compared. Denote the pth moment of Y and that of $e_Y(X)$ as $\mu_p(Y) = \sum_k (y_k)^p \cdot s_k$ and $\mu_p(e_Y(X)) = \sum_j (e_Y(x_j))^p \cdot r_j$. The percent relative error (PRE) in the pth moment from X to Y is defined as

$$\text{PRE}(p) = 100 \cdot \frac{\mu_p(e_Y(X)) - \mu_p(Y)}{\mu_p(Y)} \tag{33}$$

(von Davier et al., 2004). The PRE(p) from Y to X can be calculated in the same way.

Statistical Accuracy

Estimated equating functions are sample estimates of population quantities and therefore subject to sampling variability. The uncertainty can be measured by the SEE, the standard deviation of the asymptotic distribution of $\hat{e}_Y(x)$ if we are equating X to Y, or

$$\text{SEE}_Y(x) = \sqrt{\text{Var}(\hat{e}_Y(x))} \tag{34}$$

(von Davier et al., 2004). A standard way to evaluate this quantity is through use of the delta method. Because the score probabilities $\mathbf{r}$ and $\mathbf{s}$ are estimated independently by their MLEs, $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$, the asymptotic distribution of the estimates can be precisely determined, which facilitates the use of the delta method. Recall that the C-matrices are assumed to exist and satisfy Equation 3, so the variance-covariance matrix of $(\hat{\mathbf{r}}, \hat{\mathbf{s}})$ jointly for the EG design can be written as

$$\text{Cov}\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\mathbf{s}} \end{pmatrix} = \begin{pmatrix} C_{\mathbf{r}} C_{\mathbf{r}}^T & 0 \\ 0 & C_{\mathbf{s}} C_{\mathbf{s}}^T \end{pmatrix} = CC^T \tag{35}$$

with

$$C = \begin{pmatrix} C_{\mathbf{r}} & 0 \\ 0 & C_{\mathbf{s}} \end{pmatrix}. \tag{36}$$

Furthermore, the asymptotic distribution of the MLEs are known to be

$$\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\mathbf{s}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}, CC^T \right). \tag{37}$$

Then for each x, the estimated equating function given by Equation 31 is also approximately normally distributed,

$$e_Y(x;\hat{\mathbf{r}},\hat{\mathbf{s}}) \sim N(e_Y(x;\mathbf{r},\mathbf{s}), J_{e_Y} C C^T J_{e_Y}^T), \tag{38}$$

where J_{e_Y} is the $1 \times (J+K)$ Jacobian vector,

$$J_{e_Y} = \left(\frac{\partial e_Y}{\partial \mathbf{r}} \ , \ \frac{\partial e_Y}{\partial \mathbf{s}} \right) = \left(\frac{\partial e_Y}{\partial r_1}, \cdots, \frac{\partial e_Y}{\partial r_J}, \ \frac{\partial e_Y}{\partial s_1}, \cdots, \frac{\partial e_Y}{\partial s_K} \right). \tag{39}$$

As a result,

$$\mathrm{SEE}_Y(x) = \|\hat{J}_{e_Y} C\|, \tag{40}$$

where $\|\mathbf{v}\| = \sqrt{\sum_j v_j^2}$ denotes the Euclidian norm of the vector $\mathbf{v}$.

Note that the main difference between the formulas described for the GK in von Davier et al. (2004) and the formulas below is in the expression of the Jacobians, which reflects the difference in the type of kernel function that was used in the continuization. When the score distributions have been approximated by sufficiently smoothed continuized CDFs, the derivatives of the equating functions can be computed,

$$\frac{\partial e_Y}{\partial r_j} = \frac{1}{G^{(1)}} \cdot \frac{\partial F_{h_X}(x;\mathbf{r})}{\partial r_j}, \tag{41}$$

$$\frac{\partial e_Y}{\partial s_k} = -\frac{1}{G^{(1)}} \cdot \frac{\partial G_{h_Y}(e_Y(x);\mathbf{s})}{\partial s_k}, \text{ with} \tag{42}$$

$$G^{(1)} = \frac{\partial G_{h_Y}(y;\mathbf{s})}{\partial y} \text{ evaluated at } y = e_Y(x). \tag{43}$$

Because the partial derivatives of $F_{h_X}(x;\mathbf{r})$ with respect to components of $\mathbf{r} = \{r_j, 1 \le j \le J\}$ are needed in the calculation of SEE, some calculus will lead to the result

$$\frac{\partial F_{h_X}(x;\mathbf{r})}{\partial r_j} = H(R_{jX}) - M_{jX}(x;\mathbf{r}) \cdot f_{h_X}(x), \tag{44}$$

where

$$M_{jX}(x;\mathbf{r}) = \frac{1}{2}(x-\mu_X)(1-a_X^2)\left(\frac{x_j-\mu_X}{\sigma_X}\right)^2 + (1-a_X)x_j, \tag{45}$$

$H(R_{jX})$ is the CDF of LK or UK evaluated at R_{jX}, and $f_{h_X}(x)$ is the continuized PDF.

To compare two equating functions that depend on the same parameters, their difference, $R(x)$, can be evaluated, along with the SEED that provides guidelines for statistical significance

(as shown in von Davier et al., 2004). Suppose $\hat{e}_1(x)$ and $\hat{e}_2(x)$ are the two equating functions of interest and both convert X to Y, resulting from the use of two different kernel functions. Then

$$R(x) = \hat{e}_1(x) - \hat{e}_2(x) \tag{46}$$

and

$$\text{SEED}_Y(x) = \sqrt{\text{Var}(\hat{e}_1(x) - \hat{e}_2(x))} = \|(\hat{J}_{e_1} - \hat{J}_{e_2})C\|. \tag{47}$$

Equating Results

The data we will be using are results from two 20-item mathematics tests given in von Davier et al. (2004). The tests, both number-right scored tests, were administered independently to two samples from a national population of examinees. The two sample sizes are $N = 1,453$ and $M = 1,455$. The observed sample proportions are shown in Figure 1.

Presmoothing

As mentioned before, the score probabilities for the population are estimated by fitting loglinear models that have power moments of the sample proportions for their sufficient statistics. The moments preserved in the final models are the first two and three for X and Y, respectively. That is, the mean and variance of the X distribution and the mean, variance, and skewness of the Y distribution are preserved. The fit of the models is examined by the likelihood ratio tests and the Freeman-Tukey residuals, and the results show no evidence of lack of fit. See von Davier et al. (2004) for more details about the score probability estimation. The fitted score probabilities, $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$, are shown in Figure 1 as well.

Continuization

The optimal bandwidths using SLK, RLK, SUK, and RUK are listed in Table 1. Results for GK are shown as reference. It is clear that the ratio of the optimal bandwidths for the same distribution with different scale parameters (i.e., s in LK and b in UK) reflects exactly their scale difference. For example, the optimal h_X for SLK and that for RLK are 0.5117 and 0.9280, respectively; their ratio is $0.5117/0.9280 \approx \sqrt{3}/\pi$, which is equal to the inverse of the ratio of the corresponding scale parameters. In general, if two kernel functions are from the same family of distributions, both have zero mean, and their standard deviations are σ_1 and σ_2, respectively.

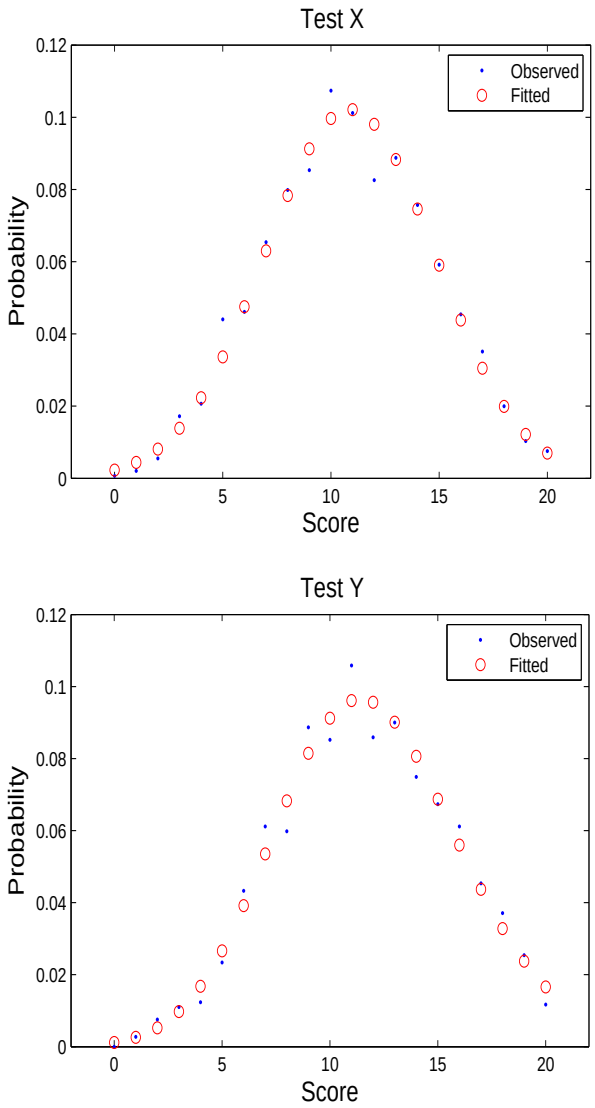

Figure 1. **Observed sample proportions and fitted score probabilities of X and Y.**

Table 1
Optimal Bandwidths for Standard Logistic Kernel (SLK), Rescaled Logistic Kernel (RLK), Standard Uniform Kernel (SUK), Rescaled Uniform Kernel (RUK), and Gaussian Kernel (GK)

	SLK	RLK	SUK	RUK	GK
h_X	0.5117	0.9280	1.0029	0.2895	0.6223
h_Y	0.4462	0.8094	1.0027	0.2895	0.5706
a_X	0.9715	0.9715	0.9971	0.9971	0.9869
a_Y	0.9795	0.9795	0.9973	0.9973	0.9896

Then their corresponding optimal bandwidths, h_1 and h_2, for test X or test Y, satisfy the following equality:

$$\sigma_1 h_1 = \sigma_2 h_2. \tag{48}$$

In addition, they have identical a_X and a_Y values, and, therefore, their resulting continuized score distributions are identical.

When RLK, RUK, and GK are applied, all kernel functions have zero mean and unit variance, so the difference in their optimal bandwidths is purely due to the distribution characteristics of the kernel functions. The kurtosis of a distribution says how heavy its tails are; the larger the kurtosis the heavier the tails. It is known that the kurtoses of the logistic distribution, uniform distribution, and normal distribution are 1.2, -1.2, and 0, respectively. Table 1 indicates that the heavier the tails of the kernel function, the smaller the resulting a_X and a_Y.

Figure 2 shows the continuized PDFs and CDFs for LK, UK, and GK. The graph in the left panel indicates that the continuized PDFs for LK and GK are smooth functions, and it is hard to distinguish between these two curves. The continuized PDF for UK is now piecewise constant. The right panel only presents part of the continuized CDFs within the range of -1 to 1.5 because the difference between curves may not easily be seen when graphed against the whole score range. Apparently, the tail of LK is heavier than that of GK, which corresponds to the fact that the logistic distribution has heavier tails than the normal distribution. The use of UK results in a piecewise linear CDF, which is how linear interpolation functions in the percentile-rank method. It is clear that the distribution characteristics of kernel functions are inherited by the corresponding continuous approximations.

Numerically, it is easier to calculate the moments of a distribution than its cumulants. Because of the close relationship between the moment-generating function and the cumulant-generating

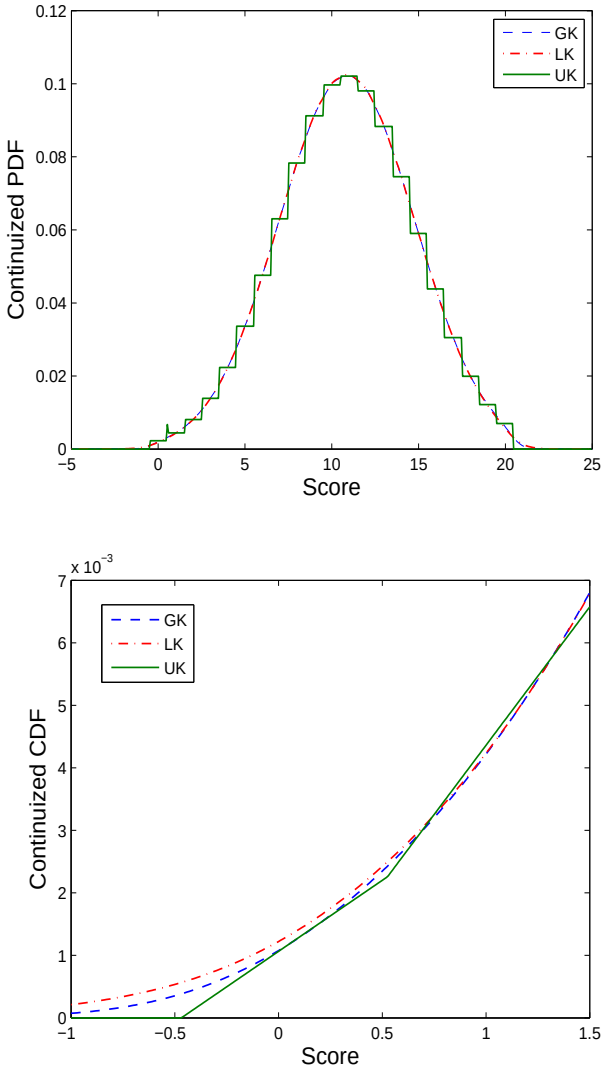

Figure 2. **Continuized probability density function (PDF) and cumulative distribution function (CDF) of X.**

Note. GK = Gaussian kernel, LK = logistic kernel, UK = uniform kernel.

function, certain connections can be built between moments and cumulants.

Suppose ν_j is the jth moment of $X(h_X)$. The cumulants are related to the moments by the following recursion formula (Smith, 1995),

$$\kappa_j(h_X) = \nu_j - \sum_{n=1}^{j-1} \begin{pmatrix} j-1 \\ n-1 \end{pmatrix} \kappa_n(h_X) \cdot \nu_{j-n} \tag{49}$$

Table 2
Cumulants for Logistic Kernel (LK), Uniform Kernel (UK), and Gaussian Kernel (GK), With Optimal Bandwidths

	Order	Discrete	LK	UK	GK
$F(X)$	1	10.82	10.82	10.82	10.82
	2	14.48	14.48	14.48	14.48
	3	-3.57	-3.28	-3.56	-3.44
	4	-63.16	-55.49	-63.10	-59.91
	5	23.17	20.06	22.81	21.71
	6	510.69	432.12	501.86	471.77
$G(Y)$	1	11.59	11.59	11.59	11.59
	2	15.48	15.48	15.48	15.48
	3	-3.82	-3.59	-3.79	-3.70
	4	-102.49	-93.87	-101.36	-98.31
	5	-102.55	-92.47	-101.94	-97.17
	6	3,539.4	3,127.0	3,493.8	3,325.0

with

$$\begin{pmatrix} j \\ n \end{pmatrix} = \frac{j!}{n!\,(j-n)!}. \tag{50}$$

In this paper, the first six cumulants of $X(h_X)$ were computed, using SLK, RLK, SUK, RUK, and GK. We first estimated $\{\nu_j\}$'s by definition,

$$\hat{\nu}_j = \int_{-\infty}^{\infty} (x - \mu_X)^j \, d\hat{F}_{h_X}(x), \tag{51}$$

and then converted them to $\kappa_j(h_X)$ via the recursion formula given in Equation 49. The results summarized in Table 2 agree perfectly with the mathematical findings in Equation 30 for LK and UK and in Equation 15 for GK. Moreover, the scale difference in the same family of kernel function does not influence the cumulants. The value of $(h_X)^j \cdot \kappa_{j,V}$ in Equation 30 is so small for $j > 2$ for both LK and UK that $\kappa_j(h_X) \approx (a_X)^j \cdot \kappa_{jX}$. As a result, when the same order of cumulants are compared in absolute value, UK has the largest one (for orders higher than 2) while LK has the smallest one, depending on the corresponding a_X or a_Y value.

The cumulants for the three kernel functions with fixed bandwidths were also compared, and the results are summarized in Table 3. For each kernel function, cumulants were computed for small h_X ($h_X = 0.2895$), median h_X ($h_X = 0.6223$), and large h_X ($h_X = 0.9280$). Each h_X is optimal for a certain kernel function; a_X is fixed once h_X is fixed at certain value. From Equation 30, the difference in $\kappa_j(h_X)$'s is due to different $\kappa_{j,V}$'s for fixed h_X. The odd cumulants for all

Table 3
Cumulants for Logistic Kernel (LK), Uniform Kernel (UK), and Gaussian Kernel (GK), X to Y, With Fixed Bandwidth

Order	Discrete	LK	UK	GK
		$h_X = 0.2895$		
3	-3.57	-3.54	**-3.56**	-3.54
4	-63.16	-62.42	**-63.10**	-62.43
5	23.17	22.83	**22.81**	22.83
6	510.69	501.88	**501.86**	501.85
		$h_X = 0.6223$		
3	-3.57	-3.44	-3.44	**-3.44**
4	-63.16	-59.74	-60.09	**-59.91**
5	23.17	21.70	21.75	**21.71**
6	510.69	472.19	471.72	**471.77**
		$h_X = 0.9280$		
3	-3.57	**-3.28**	-3.27	-3.28
4	-63.16	**-55.49**	-57.13	-56.27
5	23.17	**20.06**	20.06	20.05
6	510.69	**432.12**	431.89	429.45

Note. Boldface indicates cumulants of a certain kernel function with its optimal bandwidth.

three kernel functions are 0, so numerically $\kappa_j(h_X)$ for odd integer j should be almost identical for a fixed h_X. It is clear that, the larger the h_X, the more the cumulants of the continuized distributions deviate from the fitted discrete score distributions. However, the cumulants do not vary much for different kernel functions with a fixed bandwidth. LK does not outperform GK in terms of cumulants under respective optimal bandwidths. UK performs best here since its optimal bandwidth is the smallest.

Equating

Test X and test Y are equated as in Equation 31 using LK, UK, and GK under corresponding optimal bandwidths. The equated scores are shown in Table 4. They are very close for all three kernel functions, except for the following two situations: the equated y score when $x = 20$ in test X, and the equated x score when $y = 0$ in test Y. The explanation is that the UK has finite range while the LK and GK do not, so the equated scores could be quite different when they are close to the boundaries of the continuized distributions of UK. The PREs defined in Equation 33 are computed for the first 10 moments, and the results are summarized in Table 5.

Table 4
Equated Scores for Logistic Kernel (LK), Uniform Kernel (UK), and Gaussian Kernel (GK)

	X to Y			Y to X		
Score	LK	UK	GK	LK	UK	GK
0	0.4474	0.4392	0.3937	-0.4125	-0.2268	-0.3216
1	1.5732	1.6387	1.5813	0.4857	0.5565	0.4965
2	2.6285	2.6783	2.6404	1.3958	1.4289	1.3862
3	3.6353	3.6762	3.6443	2.3653	2.3888	2.3558
4	4.6253	4.6604	4.6316	3.3673	3.3917	3.3604
5	5.6137	5.6434	5.6177	4.3793	4.4052	4.3749
6	6.6079	6.6313	6.6100	5.3894	5.4151	5.3870
7	7.6116	7.6280	7.6120	6.3918	6.4151	6.3912
8	8.6269	8.6361	8.6260	7.3840	7.4030	7.3847
9	9.6549	9.6576	9.6530	8.3646	8.3785	8.3662
10	10.6960	10.6937	10.6935	9.3332	9.3423	9.3354
11	11.7497	11.7448	11.7471	10.2902	10.2954	10.2925
12	12.8147	12.8097	12.8126	11.2364	11.2390	11.2386
13	13.8879	13.8849	13.8869	12.1736	12.1748	12.1752
14	14.9633	14.9638	14.9641	13.1046	13.1051	13.1052
15	16.0306	16.0349	16.0339	14.0343	14.0334	14.0334
16	17.0717	17.0805	17.0781	14.9710	14.9669	14.9680
17	18.0578	18.0729	18.0677	15.9297	15.9197	15.9236
18	18.9520	18.9702	18.9607	16.9389	16.9216	16.9288
19	19.7339	19.7072	19.7183	18.0578	18.0356	18.0477
20	20.4613	20.2781	20.3930	19.3692	19.3993	19.4153

Statistical Accuracy

The C-matrices are obtained in the presmoothing step so that they are not affected by the choice of kernel. Let $J_{e_{LK}}$ denote the Jacobian vector of the equating function that equates X to Y using LK, $J_{e_{UK}}$ denote the Jacobian vector of UK, and $J_{e_{GK}}$ denote the Jacobian vector of GK. Their SEEs are given by Equation 40 with $\hat{J}_{e_Y} = J_{e_{LK}}$, $\hat{J}_{e_Y} = J_{e_{UK}}$, and $\hat{J}_{e_Y} = J_{e_{GK}}$, respectively; the results are shown in the left panel of Figure 3. The SEEs for equating functions that transform Y to X can be computed analogously, as illustrated in the right panel of Figure 3. These graphs reveal that the SEEs for LK and GK differ only a bit at extreme scores and that their *curves* have similar shape. However, the SEEs for UK do not exhibit the same pattern. In addition, they have greater variations from test to test.

In Figure 4 the difference $R(x)$ between two estimated equating functions is plotted, converting X to Y, $R(x) = \hat{e}_{LK}(x) - \hat{e}_{GK}(x)$ for the left panel and $R(x) = \hat{e}_{UK}(x) - \hat{e}_{GK}(x)$ for

Table 5
Percent Relative Errors (PREs) for Logistic Kernel (LK), Uniform Kernel (UK), and Gaussian Kernel (GK)

	X to Y			Y to X		
Moments	LK	UK	GK	LK	UK	GK
1	0.0073	0.0467	0.0059	-0.0094	0.0553	-0.0063
2	0.0186	0.0353	0.0122	-0.0314	0.0212	-0.0228
3	0.0344	0.0099	0.0220	-0.0718	-0.0528	-0.0550
4	0.0611	-0.0175	0.0398	-0.1446	-0.1661	-0.1139
5	0.1100	-0.0402	0.0729	-0.2691	-0.3277	-0.2135
6	0.1911	-0.0550	0.1276	-0.4628	-0.5487	-0.3665
7	0.3130	-0.0600	0.2091	-0.7398	-0.8391	-0.5838
8	0.4818	-0.0539	0.3213	-1.1104	-1.2065	-0.8739
9	0.7022	-0.0358	0.4672	-1.5811	-1.6561	-1.2426
10	0.9773	-0.0052	0.6489	-2.1548	-2.1902	-1.6934

the right panel. Two curves representing ± 2 times of the $\text{SEED}_Y(x)$ are also provided as the upper and lower bounds of the 95% confidence interval. For the comparison between LK and GK, $R(0)$ and $R(20)$ are significantly different at level 0.05 since they are out of bounds, but the scale of SEED is so small (less than 0.1 raw-score point) that the difference may still be negligible in practice. The absolute values of $R(x)$ and $\text{SEED}_Y(x)$ increase as x approaches its boundaries (0 and 20) because the continuized CDFs for LK differ the most from the continuized CDFs for GK at both tails. The right panel in Figure 4 shows that the difference between $\hat{e}_{UK}(x)$ and $\hat{e}_{GK}(x)$ is much larger than that of $\hat{e}_{LK}(x)$ and $\hat{e}_{GK}(x)$ for all score values except for median score values. The difference is nonsignificant at level 0.05, however, since the corresponding SEEDs are larger in scale.

The KE function closely approximates the standard linear equating function when the bandwidths h_X and h_Y are both large. Here we computed the $e_{Llin}(x)$ and $e_{Ulin}(x)$ for LK and UK by choosing $h_X = h_Y = 20$ in Equation 31. The difference between the LK (UK) estimated equating function and the LK (UK) estimated linear equating function is defined as $R(x) = e_{LK}(x) - e_{Llin}(x)$ ($R(x) = e_{UK}(x) - e_{Ulin}(x)$), and the SEED is given by Equation 47. Figure 5 exhibits the values of $R(x)$ for LK and UK at each possible score along with their 95% confidence intervals, indicating the two equating functions are useful alternatives for all scores except for the highest two values, $x = 19,\ 20$.

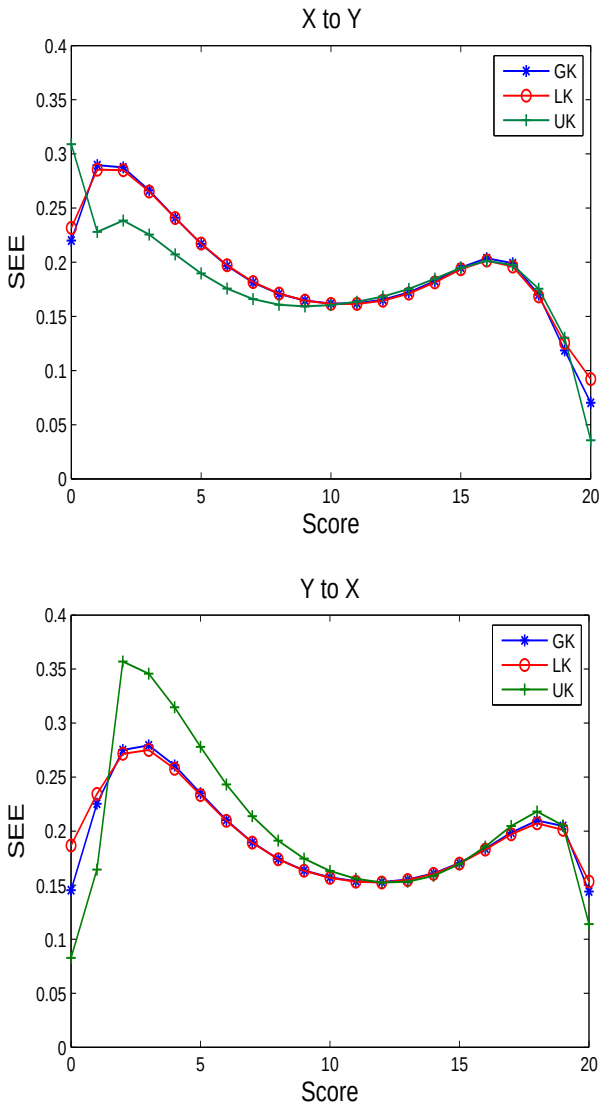

Figure 3. **Standard errors of equating (SEEs) for logistic kernel (LK), uniform kernel (UK), and Gaussian kernel (GK).**

Conclusions

In this paper we introduce two new equating models within the KE framework, namely, LK and UK. The choice of LK was motivated by the criticism of GK that its use might lead to a continuous distribution that does not preserve the higher moments of the original discrete distribution. The choice of UK was motivated by its similarity to the linear interpolation that is widely used in practice. It is worth noting that in this paper we improved upon the linear interpolation by rescaling it such that the continuous distribution preserves the mean and the

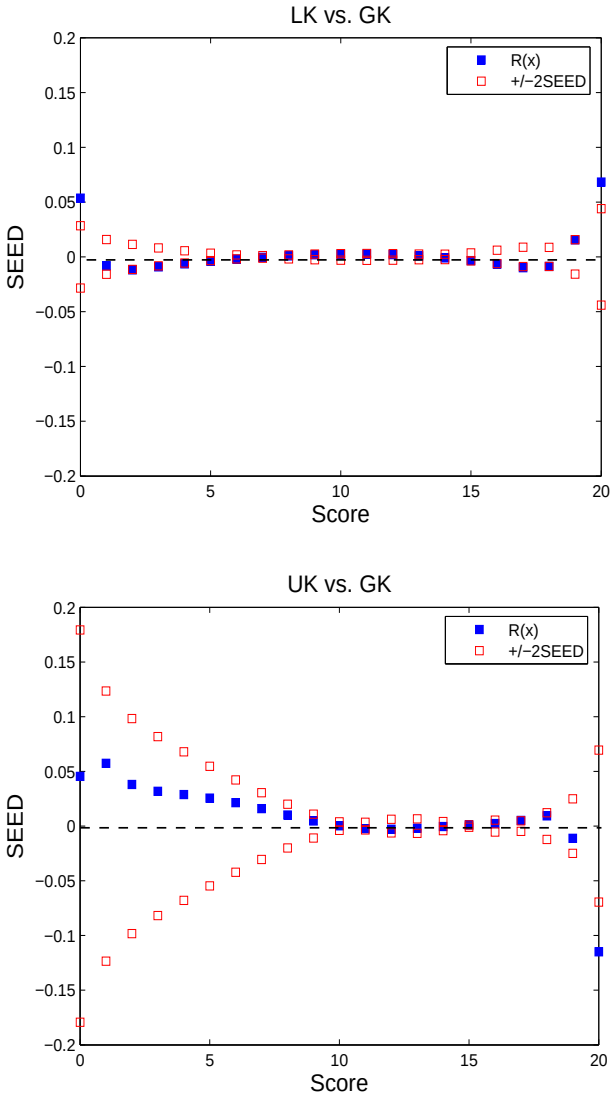

Figure 4. **Standard errors of equating differences (SEEDs) between logistic kernel (LK) and Gaussian kernel (GK), uniform kernel (UK) and Gaussian kernel (GK), from X to Y.**

variance of the discrete distribution.

This study suggests that the three kernels (with the various versions due to rescaling) provide very similar equating results and that despite the criticism, GK does well in preserving the higher order cumulants, the PRE, and the level of accuracy. The main differences between the three kernels seem to be their out-of-range characteristics (i.e., GK and LK have strictly positive density on the whole real line, but UK does not).

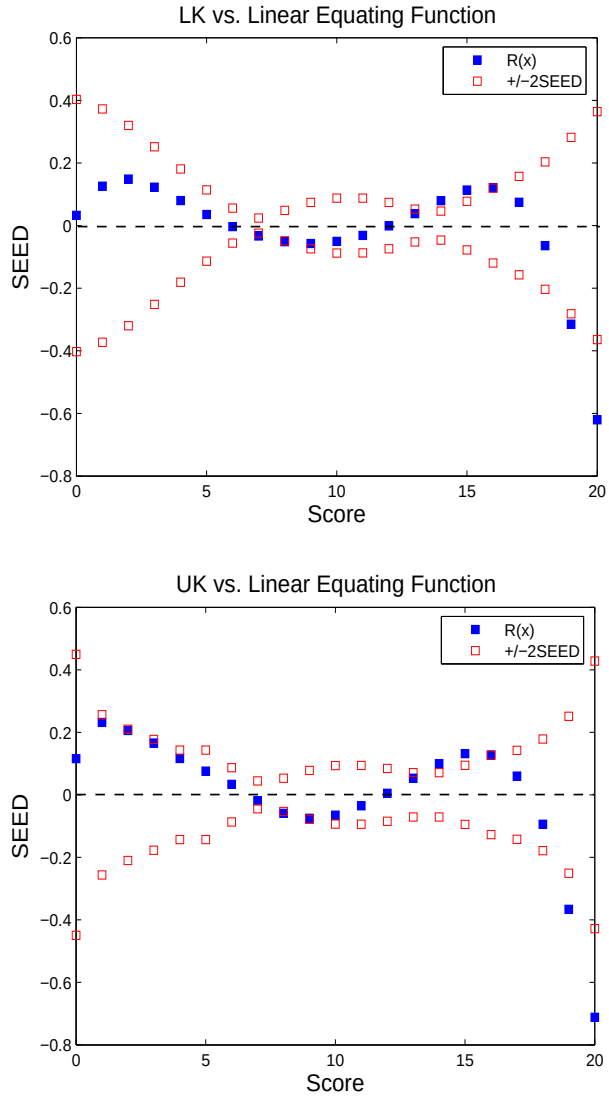

Figure 5. **Standard errors of equating differences (SEEDs) between logistic kernel (LK) and uniform kernel (UK), compared with linear equating functions, from X to Y.**

The main limitation of this study lies in its data example. In some future research, to stay in the KE framework these methods should be applied to distributions that have shapes that depart significantly from the normal distributions (more pronounced skewness and kurtosis, as well as other characteristics). Other smoothing techniques such as spline smoothing could be adopted in the continuization step. The presmoothing step and the continuization step could also be combined before two tests with discrete score distributions are equated (e.g., Wang, 2007).

In addition, note that UK (and in consequence, linear interpolation) is not differentiable at the score points, which leads to some problems in estimating the SEE. Technically, the continuized functions are still differentiable on the score space except on some finite points with total probability of zero. This means that this requirement of the delta method is actually met in this situation. Future simulation studies could shed more light on this statement.

References

Abramowitz, M., & Stegun, I. A. (Eds.). (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* New York: Dover.

Balakrishnan, N. (1992). *Handbook of the logistic distribution.* New York: Marcel Dekker.

Braun, H. I., & Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 9-49). New York: Academic.

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

Holland, P. W., & Thayer, D. T. (1989). *The kernel method of equating score distributions* (Technical Rep. No. 89-84). Princeton, NJ: ETS.

Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics, 25*, 133-183.

Kupperman, M. (1952). On exact grouping correlations to moments and cumulants. *Biometrika, 39*, 429-434.

Livingston, S. A. (2004). *Introduction to test equating (without IRT).* Princeton, NJ: ETS.

Smith, P. (1995). A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. *The American Statistician, 49*, 217-218.

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

Wang, T. (2007). *An alternative continuization method to the kernel method in von Davier, Holland and Thayer's (2004) test equating framework* (CASMA Research Rep. No. 11). Iowa City, IA: University of Iowa.