Research Report

# Research on Standard Errors of Equating Differences

***Tim Moses***

***Wenmin Zhang***

*November 2010*

*ETS RR-10-25*

# Research on Standard Errors of Equating Differences

Tim Moses and Wenmin Zhang
ETS, Princeton, New Jersey

November 2010

**Technical Review Editor:** Daniel Eignor

**Technical Reviewers:** Frank Rijmen and Alina von Davier

## Abstract

In this paper, the *standard error of equating difference* (SEED) is described in terms of originally proposed kernel equating functions (von Davier, Holland, & Thayer, 2004) and extended to incorporate traditional linear and equipercentile functions. These derivations expand on prior developments of SEEDs and standard errors of equating and provide additional insight about the relationships of kernel and traditional equating functions. Simulations are used to evaluate the SEEDs' accuracies.

Key words: observed score equating, standard errors, SEEDs

**Acknowledgments**

The authors thank Alina von Davier for her initial ideas that inspired this project. The authors also thank Alina, Dan Eignor, and Frank Rijmen for helpful comments on an earlier version of this paper and Kim Fryer and Ruth Greenwood for their editorial work.

**Table of Contents**

Page

## List of Tables

## List of Figures

In equating research and practice, evaluations of equated score differences have informed the selection of linear and equipercentile equating functions, and nonequivalent groups with anchor test (NEAT) equating functions (Kolen & Brennan, 2004; Moses, Yang, & Wilson, 2007; von Davier, Holland, & Thayer, 2004; von Davier & Kong, 2005). Standard errors of equating functions (SEEs) are not immediately useful for evaluating equated score differences, but recent developments have applied SEE estimation approaches to the estimation of standard errors of equating differences (i.e., SEEDs; von Davier et al., 2004; von Davier & Kong, 2005). SEEDs have been mainly described for evaluating kernel equating functions based on whether their differences exceed sampling variability.

In this paper, the originally-proposed SEEDs for evaluating kernel equating functions (von Davier et al., 2004) are expanded so that the differences among kernel equating functions, traditional equipercentile equating functions, and traditional linear equating functions can be evaluated. The accuracies of the SEEDs are evaluated in simulations for six equated score difference situations. The final discussion addresses the use of the SEEDs in practice and the extensions that can be used to apply SEEDs to compare equating functions computed for the most frequently used equating designs.

**Standard Errors of Equating Functions and of Equating Differences**

The delta method (Kendall & Stuart, 1977) is often applied to estimate the variability of equating functions. Estimating the SEE of an *X*-to-*Y* equating function involves noting that this equating function is computed from a set of statistics, $\delta$, with associated covariance matrix $\Sigma_\delta$. Pre- and post-multiplying $\Sigma_\delta$ by the partial derivatives of the equating function at score $x_{j*}$ produces an estimated variance of the equating function. Taking the square root produces an estimated SEE,

$$SE(e_y(x_{j*})) = \sqrt{Var(e_y(x_{j*}))} = \sqrt{\left[\frac{\partial e_y(x_{j*})}{\partial \delta}\right] \Sigma_\delta \left[\frac{\partial e_y(x_{j*})}{\partial \delta}\right]^t}. \quad (1)$$

Traditional linear and equipercentile equating functions have been expressed and computed from different terms making up $\delta$. In particular, the traditional *X*-to-*Y* linear equating function is computed from *X* and *Y*’s means ( $\mu$ ) and standard deviations ( $\sigma$ ),

$$e_y(x_{j*}) = \mu_Y + \frac{\sigma_Y}{\sigma_X}\left(x_{j*} - \mu_X\right). \quad (2)$$

SEEs of (2) have been derived using (1) with tests' means and variances comprising $\delta$ (Braun & Holland, 1982; Hanson, Zeng, & Kolen, 1993; Kolen, 1985; Zeng, 1991; Zeng & Cope, 1995).

The basis of the traditional *X*-to-*Y* equipercentile equating function is tests' percentile ranks (Kolen & Brennan, 2004),

$$e_y(x_{j*}) = G^{-1}(F(x_{j*})) = y_u - 1/2 + \frac{F(x_{j*}) - G(y_{u-1} + 1/2)}{G(y_u + 1/2) - G(y_{u-1} + 1/2)}, \quad (3)$$

where $F(x_{j*})$ denotes $x_{j*}$'s percentile rank defined as

$$\begin{aligned} F(x_{j*}) &= 0 && \text{if } x_{j*} \le x_1 - 1/2 \\ &= \sum_{i=1}^{j-1} r_i + (x_{j*} - (x_j - 1/2))r_j && \text{if } x_j - 1/2 \le x_{j*} \le x_j + 1/2 \text{ for } j = 1, \ldots J \\ &= 1 && \text{if } x_{j*} \ge x_J + 1/2, \end{aligned} \quad (4)$$

$r_j$ is the probability at score $x_j$, and $y_{u-1}$ is the largest *Y* score where $y_{u-1} + 1/2$'s percentile rank, $G(y_{u-1} + 1/2)$, is smaller than $F(x_{j*})$. SEEs of (3) have been derived using (1) with the $F()$ and $G()$'s comprising $\delta$ (Jarjoura & Kolen, 1985; Liou & Cheng, 1995; Liou, Cheng, & Johnson, 1997; Lord, 1982). SEEDs for evaluating their equated score differences with respect to sampling variability have been unavailable due to the use of sets of statistics ($\delta$) that do not easily relate to each other.

Kernel equating (Holland, King, & Thayer, 1989; von Davier et al, 2004) was developed in ways that allow linear and equipercentile equating functions to be computed from the same set of statistics ($\delta$), the *J*- and *K*-column vectors of test *X* and *Y*'s score probabilities, $\mathbf{r}$ and $\mathbf{s}$,

$$e_y(x_{j*}) = G_{hY}^{-1}\left(F_{hX}(x_{j*};\mathbf{r});\mathbf{s}\right). \quad (5)$$

The $F_{hX}()$ in (5) denotes kernel smoothed cumulative densities,

$$F_{hX}(x_{j^*};\mathbf{r}) = \sum_j r_j \Phi\left(\frac{x_{j^*} - a_X x_j - (1-a_X)\mu_X}{a_X h_X}\right), \tag{6}$$

where $\Phi$ is the cumulative density of a standard normal distribution and $a_X = \sqrt{\frac{\sigma_X^2}{\sigma_X^2 + h_X^2}}$. $G_{hY}()$ is computed in a similar manner. Kernel linear and kernel equipercentile equating functions can be computed using (5) with small and large values for $h_X$ and $h_Y$. When kernel linear equating functions and kernel equipercentile equating functions are computed using (5), the equated score differences can be differentiated with respect to the same $\delta$, $\left[\frac{\partial e_{y1}(x_{j^*})}{\partial \delta} - \frac{\partial e_{y2}(x_{j^*})}{\partial \delta}\right]$ and SEEDs can be computed as a direct extension of (1),

$$SEED(e_{y1}(x_{j^*}) - e_{y2}(x_{j^*}))$$

$$= \sqrt{\left[\frac{\partial e_{y1}(x_{j^*})}{\partial \delta} - \frac{\partial e_{y2}(x_{j^*})}{\partial \delta}\right] \Sigma_\delta \left[\frac{\partial e_{y1}(x_{j^*})}{\partial \delta} - \frac{\partial e_{y2}(x_{j^*})}{\partial \delta}\right]^t}$$

$$= \sqrt{\left[\frac{\partial e_{y1}(x_{j^*})}{\partial \delta}\right] \Sigma_\delta \left[\frac{\partial e_{y1}(x_{j^*})}{\partial \delta}\right]^t + \left[\frac{\partial e_{y2}(x_{j^*})}{\partial \delta}\right] \Sigma_\delta \left[\frac{\partial e_{y2}(x_{j^*})}{\partial \delta}\right]^t - 2\left[\frac{\partial e_{y1}(x_{j^*})}{\partial \delta}\right] \Sigma_\delta \left[\frac{\partial e_{y2}(x_{j^*})}{\partial \delta}\right]^t}$$

$$= \sqrt{Var(e_{y1}(x_{j^*})) + Var(e_{y2}(x_{j^*})) - 2Covar(e_{y1}(x_{j^*}), e_{y2}(x_{j^*}))} \tag{7}$$

The results of (7) can be used to estimate SEEDs for the differences among traditional linear, traditional equipercentile, kernel linear, and kernel equipercentile functions, because all of these equating functions can be computed based on forming $\delta$ from *X* and *Y*'s score probabilities. Traditional equipercentile equating function can be computed directly from *X* and *Y*'s score probabilities (Moses & Holland, 2006; Wang, 2006), rather than the percentile ranks suggested in (3). Unlike prior works (Braun & Holland, 1982; Hanson et al., 1993; Kolen, 1985; Zeng, 1991; Zeng & Cope, 1995), traditional linear equating functions can also be computed from *X* and *Y* 's score probabilities. The derivatives of the kernel, traditional equipercentile, and traditional linear equating functions with respect to the *j*th and *k*th values of *X* and *Y*'s score probability vectors, $\mathbf{r}$ and $\mathbf{s}$ are given in Tables 1 and 2. Constructing (7)'s derivative vectors with respect to the *J* and *K* derivatives from Tables 1 and 2 results in SEEDs for the differences of kernel, traditional equipercentile, and traditional linear equating functions.

**Equating Functions, Derivatives, and Standard Errors of Equating Differences (SEEDs)**

The focus of this paper is six SEEDs that address comparisons among four equating functions (kernel linear, kernel equipercentile, traditional equipercentile, and traditional linear):

- kernel linear vs. kernel equipercentile
- kernel linear vs. traditional equipercentile
- kernel linear vs. traditional linear
- kernel equipercentile vs. traditional equipercentile
- kernel equipercentile vs. traditional linear
- traditional equipercentile vs. traditional linear

In this section, Tables 1 and 2's four equating functions' derivatives are used to provide new understanding of the equating functions' relationships and their SEEs.

The role of $\Sigma_\delta$ in equations (1) and (7) is important for understanding equating functions' SEEs and SEEDs. The $\delta$'s and $\Sigma_\delta$'s of interest are based on the probability vectors of the *X* and *Y* scores ($\mathbf{r}$ and $\mathbf{s}$). Well-known results involving computations with $\Sigma_\delta$ matrices based on probability vectors (Bishop, Fienberg, & Holland, 1975; Fienberg, 1979; Haberman, 1989), suggest that SEEs and SEEDs can be expressed in terms of the variances and covariances of the equating function derivative vectors, $\left[\frac{\partial e_y(x_{j*})}{\partial \delta}\right]$.

One SEED of interest is for evaluating differences between traditional linear and kernel linear equating functions. Because kernel linear equating functions are asymptotically equivalent to traditional linear equating functions (von Davier et al., 2004), the differences between kernel linear and traditional linear equating functions are expected to approach zero, and the variability of these differences (i.e., the SEED) is also expected to approach zero.

**Table 1**

***Derivatives of Kernel, Traditional Equipercentile, and Traditional Linear X-to-Y Equating Functions With Respect to X's jth Score Probability***

| *X*-to-*Y* equating function, $e_y(x_{j^*})$ | Derivatives with respect to *X's j*th score probability (*j* = 1 to *J*) |
|---|---|
| Kernel[a] | $$\left(\frac{1}{\frac{\partial G_{hY}(e_y(x_{j^*};\mathbf{r});\mathbf{s})}{\partial y}}\right)\frac{\partial F_{hX}(x_{j^*};\mathbf{r})}{\partial r_j}$$ |
| Traditional equipercentile | $\left(\frac{1}{s_u}\right)(1)$ when $j < j^*$ and $y_u - 1/2 < e_y(x_{j^*}) < y_u + 1/2$ |
| | $\left(\frac{1}{s_u}\right)(x_{j^*} - (x_j - 1/2))$ when $j = j^*$ and $y_u - 1/2 < e_y(x_{j^*}) < y_u + 1/2$ |
| | $\left(\frac{1}{s_u}\right)(0)$ when $j > j^*$ and $y_u - 1/2 < e_y(x_{j^*}) < y_u + 1/2$ |
| Traditional linear | $$\frac{\sigma_y}{\sigma_x}\left(\frac{(x_j - \mu_x)^2}{2\sigma_x^2}(\mu_x - x_{j^*}) - x_j\right)$$ |

[a]The kernel equating derivatives are described in more detail in von Davier et al. (2004).

**Table 2**

***Derivatives of Kernel, Traditional Equipercentile, and Traditional Linear X-to-Y Equating Functions With Respect to Y's kth Score Probability***

| *X*-to-*Y* equating function, $e_y(x_{j*})$ | Derivatives with respect to *Y*'s *k*th score probability (*k* = 1 to *K*) |
|---|---|
| Kernel[a] | $\left( \dfrac{-1}{\dfrac{\partial G_{hY}(e_y(x_{j*};\mathbf{r});\mathbf{s})}{\partial y}} \right) \dfrac{\partial G_{hY}(e_y(x_{j*});\mathbf{s})}{\partial s_k}$ |
| Traditional equipercentile | $\left(\dfrac{-1}{s_u}\right)(1)$ when $k < u$ and $y_u - 1/2 < e_y(x_{j*}) < y_u + 1/2$ |
| | $\left(\dfrac{-1}{s_u}\right)(e_y(x_{j*}) - (y_k - 1/2))$ when $k = u$ and $y_u - 1/2 < e_y(x_{j*}) < y_u + 1/2$ |
| | $\left(\dfrac{-1}{s_u}\right)(0)$ when $k > u$ and $y_u - 1/2 < e_y(x_{j*}) < y_u + 1/2$ |
| Traditional linear | $y_k + \left(\dfrac{x_{j*} - \mu_x}{\sigma_x}\right) \dfrac{(y_k - \mu_y)^2}{2\sigma_y}$ |

[a]The kernel equating derivatives are described in more detail in von Davier et al. (2004).

These expectations can be verified by studying the variances of the kernel linear and traditional linear equating functions' derivatives (Tables 1 and 2).

Using von Davier et al.'s (2004) asymptotic results for kernel linear equating functions (p. 58), when $h_X$ and $h_Y$ approach positive infinity, the derivative vector of the kernel linear equating function with respect to *X*'s probability $r_j$ from Table 1 contains *J* elements, each of which approaches

$$\left( \frac{\sigma_Y \Phi(\frac{x_{j*} - \mu_X}{\sigma_X})}{\sum_k s_k \phi(\frac{e_y(x_{j*};\mathbf{r});\mathbf{s}) - \mu_Y}{\sigma_Y})} \right) + \left( \frac{\sigma_Y}{\sigma_X} \right) \left( \frac{(x_j - \mu_X)^2}{2\sigma_X^2} (\mu_X - x_{j*}) - x_j \right), \tag{8}$$

where the $\phi$ and $\Phi$ terms denote densities and cumulative densities from the standard normal distribution. Equation (8) is identical to Table 1's derivative of the traditional linear equating except for the first term that is constant for all *J* partial derivatives computed at $x_{j*}$. As $h_X$ and $h_Y$ approach positive infinity, the derivative of the kernel linear equating function with respect to *Y*'s probability $s_k$ from Table 2 contains *K* elements, each of which approaches

$$\left( \frac{-\sigma_Y \Phi(\frac{e_y(x_{j*};\mathbf{r});\mathbf{s}) - \mu_Y}{\sigma_Y})}{\sum_k s_k \phi(\frac{e_y(x_{j*};\mathbf{r});\mathbf{s}) - \mu_Y}{\sigma_Y})} \right) + y_k + \frac{(x_{j*} - \mu_X)}{\sigma_X} \frac{(y_k - \mu_Y)^2}{2\sigma_Y}. \tag{9}$$

Equation (9) is identical to Table 2's derivative of the traditional linear equating function except for the first term that is constant for all *K* partial derivatives at $e_y(x_{j*};\mathbf{r});\mathbf{s})$.

The derivatives of the traditional linear equating function (Tables 1 and 2) are equal to those of the kernel linear equating function (8 and 9) except for constants and their means. The kernel linear and traditional linear equating function derivatives' variances are equal, and the covariance of these equating function derivatives is equal to their variance(s). The SEE of a traditional linear equating function will be asymptotically equal to that of a kernel linear equating function because the SEE in (1) is a direct function of the variance or covariance of its derivatives (Bishop, et al., 1975; Fienberg, 1979; Haberman, 1989). When using the derivatives of the kernel linear and the traditional linear equating functions to calculate their SEED (7), the resulting SEED will approach zero because it is a function of the variances and covariance of the equating functions and equating functions' derivatives.

Kernel and traditional equipercentile equating functions are frequently noted to produce similar equating results and SEEs, with the traditional equipercentile function's SEE being slightly larger (e.g., Moses & Holland, 2006). The SEEs and SEEDs for kernel and traditional equipercentile equating functions can be understood in terms of the variances of these functions' derivative vectors. The traditional equipercentile function is computed from a limited number of the $\mathbf{r}$ and $\mathbf{s}$ entries, (4), and the kernel equipercentile equating function is computed from all of the $\mathbf{r}$ and $\mathbf{s}$ entries, (5) and (6). This results in derivative vectors for traditional equipercentile equating functions that often contain values as small as zero and as large as one (Tables 1 and 2) and derivative vectors for kernel equipercentile equating functions that often contain values greater than zero and less than one. The variance of the derivative vectors based on traditional equipercentile equating will usually be larger than the variance of the derivative vectors based on kernel equipercentile equating. Because the SEEs of the kernel and traditional equipercentile equating functions are direct functions of the variances and covariances of these equating functions' derivatives (Bishop, et al., 1975; Fienberg, 1979; Haberman, 1989), the SEE of the kernel equating function will be smaller than the SEE of the traditional equipercentile function. The SEED for kernel and traditional equipercentile equating functions is usually greater than zero because the kernel equating derivatives have variances that differ from those of the traditional equipercentile derivatives.

## Implications of Equating Functions and Their Derivatives for Standard Errors of Equating Differences (SEEDs)

The following implications for the six SEEDs of interest follow from the discussion of the previous section. Because the kernel linear and traditional linear equating functions and their SEEs are asymptotically equal, their SEEDs should approach zero. The SEED for evaluating kernel equipercentile versus kernel linear equating functions should be identical to the SEED for evaluating kernel equipercentile versus traditional linear equating functions. The SEED for evaluating kernel equipercentile and traditional equipercentile equating functions will be small, but usually not zero. The SEED for evaluating traditional equipercentile versus kernel linear equating functions should be identical to the SEED for evaluating traditional equipercentile versus traditional linear equating functions. The SEEDs involving kernel equipercentile equating functions should be smaller than the SEEDs involving traditional equipercentile equating functions. These implications can be studied using a real data example and simulations.

## SEEDs in a Real Data Example

The analyses of von Davier et al.'s (2004) are regenerated and expanded to assess the six SEEDs described in this paper. Two 20-item tests, *X* and *Y*, were administered to examinees from a common population. The characteristics of these data are summarized in Table 3, suggesting that test *X* is about 0.8 points harder than test *Y*.

**Table 3**

***Von Davier et al.'s (2004) Equivalent Groups Equating Data: Descriptive Statistics***

| | *X* | *Y* |
|---|---|---|
| *N* | 1,453 | 1,455 |
| Mean | 10.82 | 11.59 |
| Standard deviation | 3.81 | 3.94 |
| Skew | 0.0026 | -0.0626 |
| Kurtosis | 2.53 | 2.50 |

Kernel equating was used to determine the most appropriate equating function for equating the *X* scores to the *Y* scale (von Davier et al., 2004). First the *X* and *Y* score distributions were presmoothed using loglinear models. Then linear and equipercentile kernel equating functions were estimated and the differences in the equated scores were computed. Finally, the SEED for the equating function differences was estimated as in (7) with a $\Sigma_\delta$ matrix based on the loglinear presmoothed $\mathbf{r}$ and $\mathbf{s}$ vectors and the derivatives of the kernel equating functions (Tables 1 and 2). Figure 1 plots this evaluation of the kernel functions' equated score differences, showing the differences between the kernel equipercentile and kernel linear functions across all *X* scores. When these differences exceed ± 2 SEEDs, they are considered statistically significant.

Broadening von Davier et al.'s (2004) analyses, (7) was used with Tables 1 and 2's derivatives and von Davier et al.'s loglinear presmoothed $\Sigma_{\delta}$ matrix to obtain equated score differences and SEEDs for traditional linear and traditional equipercentile equating functions. Figure 2 plots the equated score differences between the kernel equipercentile and traditional equipercentile functions and the ± 2 SEEDs. Figure 3 plots the equated score differences between traditional equipercentile and traditional linear functions and the ± 2 SEEDs. The other three equated score differences reflect the implications listed in the previous section, meaning that the differences and SEEDs between kernel linear and traditional linear equating functions differ from zero only beyond the ninth decimal place, the differences and SEEDs between kernel equipercentile and traditional linear equating functions are identical to Figure 2's differences and SEEDs, and the differences and SEEDs between traditional equipercentile and kernel linear equating functions are identical to Figure 3's differences and SEEDs. The results of Figures 1 through 3 show that the differences between the kernel equipercentile and traditional equipercentile equating functions are within ± 2 SEEDs across all *X* scores (Figure 2) and that the differences between the linear, traditional equipercentile, and kernel equipercentile equating functions are within ± 2 SEEDs for most *X* scores (Figures 1 and 3).

## Simulations for Evaluating Standard Errors of Equating Differences (SEEDs)

To evaluate the accuracy of Figures 1 through 3's SEEDs, simulations were conducted. Population distributions for X and Y were defined as von Davier et al.'s (2004) presmoothing models. Random samples of particular sizes were obtained from the population distributions: 1,000 samples at the original sample sizes ($N_X$ = 1,453 and $N_Y$ = 1,455), 1,000 obtained at smaller sample sizes ($N_X = N_Y$ = 200), and 1,000 obtained at much larger sizes ($N_X = N_Y$ = 10,000). The four equating functions and six SEEDs were estimated in each sample. In order to evaluate the SEEDs of the kernel equipercentile and kernel linear functions (Figure 1), the means of 1,000 SEEDs (the solid lines) can be compared to the ± 2 standard deviations of the 1,000 equated score differences (the dashed lines).

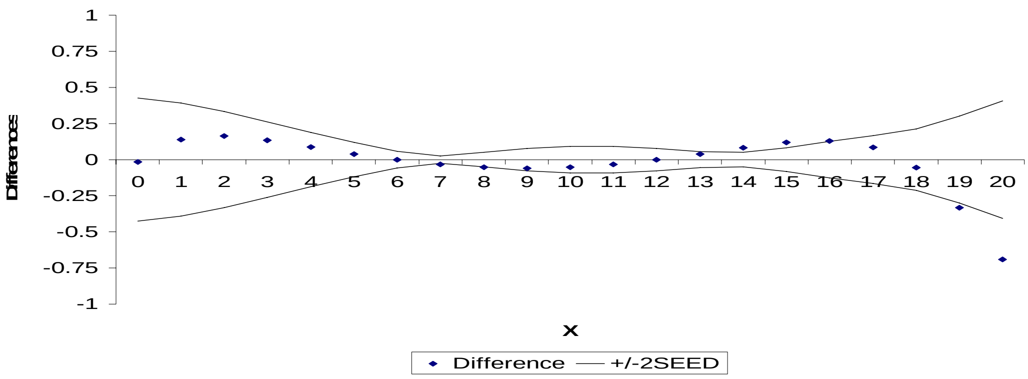


***Figure 1*. Kernel equipercentile—kernel linear equated score differences and their ± 2 SEEDs.**

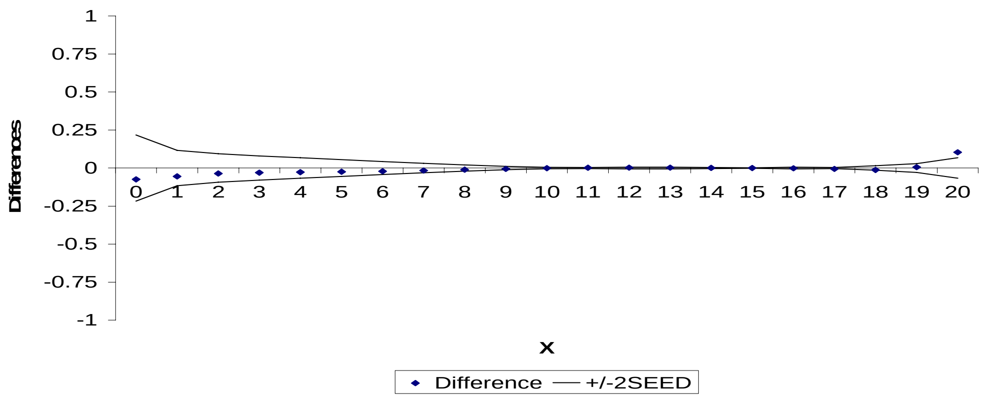


***Figure 2*. Kernel equipercentile—traditional equipercentile equated score differences and their ± 2 SEEDs.**

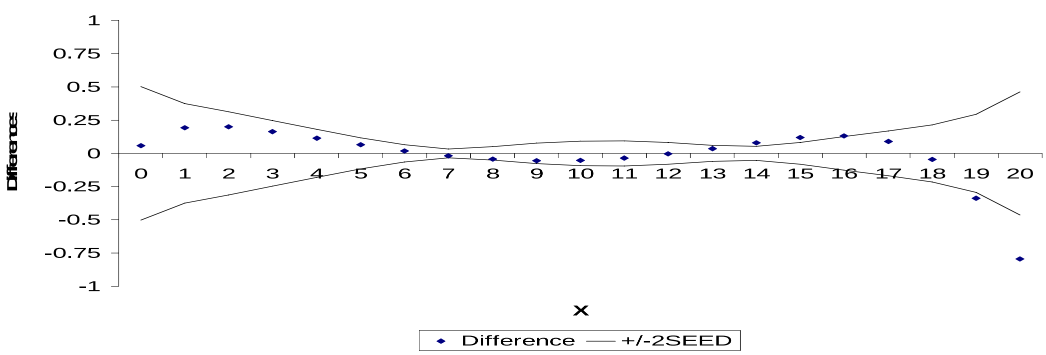


***Figure 3*. Traditional equipercentile—traditional linear equated score differences and their ± 2 SEEDs.**

The SEEDs' accuracies are summarized for sample sizes of $N_X = N_Y = 200$ (Figure 4), of $N_X$ = 1,453 and $N_Y$ = 1,455 (Figure 5), and of $N_X = N_Y$ = 10,000 (Figure 6). The means of the equated score differences are also plotted. Figures 4 through 6 show that the ± 2 SEED lines are narrower and more accurate (i.e., closer to the ± 2 standard deviation lines) for large sample sizes than for smaller sample sizes. Several of the equated score differences would be considered statistically significant for sample sizes of 10,000 (Figure 6), and none would be significant for sample sizes of 200 (Figure 4).

In order to evaluate the accuracy of the SEED of the kernel equipercentile and traditional equipercentile equating functions (Figure 2), the means of 1,000 ± 2 SEEDs (the solid lines) can be compared to the ± 2 standard deviations of the 1,000 equated score differences (the dashed lines). The SEEDs' accuracies are shown for sample sizes of $N_X = N_Y = 200$ (Figure 7), of $N_X$ = 1,453 and $N_Y$ = 1,455 (Figure 8), and of $N_X = N_Y$ = 10,000 (Figure 9). The means of the equated score differences are also displayed. The SEEDs for the differences between the kernel equipercentile and traditional equipercentile functions have quite small variability (i.e., narrow ± 2 lines, Figures 7 through 9). The SEEDs are also fairly inaccurate estimates of the actual standard deviations, where at the lowest $X$ scores, the SEEDs overestimate the actual standard deviations by a factor of about 3 for $N_X = N_Y = 200$ sample sizes (Figure 7), and overestimate the actual standard deviations by a factor of about 2 for $N_X$ = 1,453 and $N_Y$ = 1,455 sample sizes (Figure 8).

In order to evaluate the accuracy of the SEED for estimating the variability of differences between the traditional equipercentile and traditional linear equating functions (Figure 3), the means of 1,000 ± 2 SEEDs (the solid lines) can be compared to the ± 2 standard deviations of the 1,000 equated score differences (the dashed lines). The SEEDs' accuracies are shown for sample sizes of $N_X = N_Y = 200$ (Figure 10), of $N_X$ = 1,453 and $N_Y$ = 1,455 (Figure 11), and of $N_X = N_Y$ = 10,000 (Figure 12). The means of the equated score differences are also displayed. The results shown in Figures 10 through 12 look like those based on the kernel equated score differences (Figures 4 through 6), in that the ± 2 SEED lines have larger ranges and greater inaccuracy when based on sample sizes of 200 (Figures 4 and 10) than when based on sample sizes of 10,000 (Figures 6 and 12). The differences between

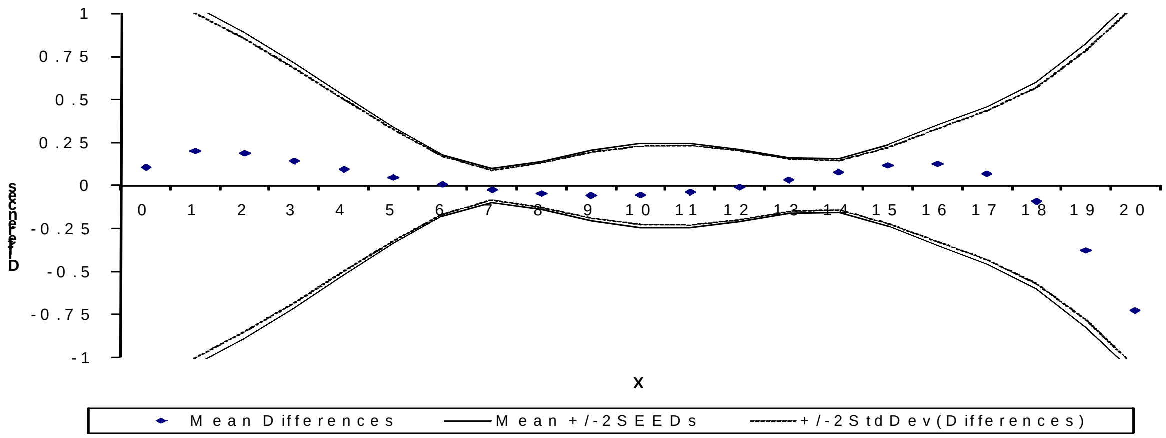


***Figure 4*. Means and standard deviations of kernel equipercentile—kernel linear equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = *Ny* = 200).**

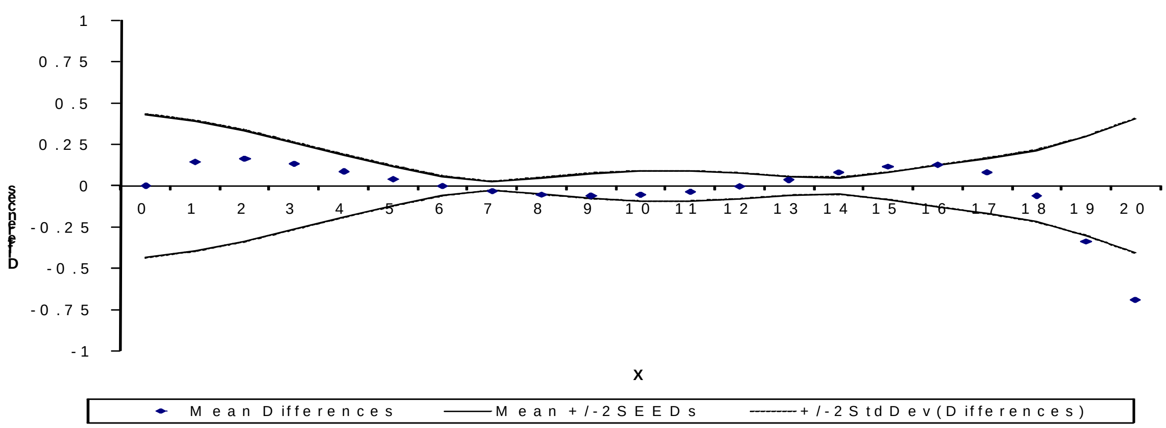


***Figure 5*. Means and standard deviations of kernel equipercentile—kernel linear equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = 1,453, *Ny* = 1,455).**

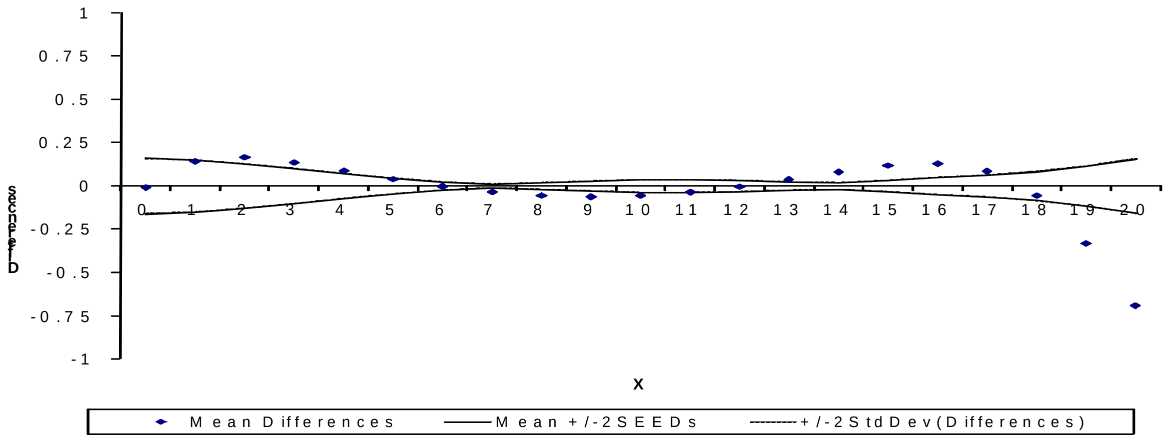


*Figure 6*. **Means and standard deviations of kernel equipercentile—kernel linear equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = *Ny* = 10,000).**

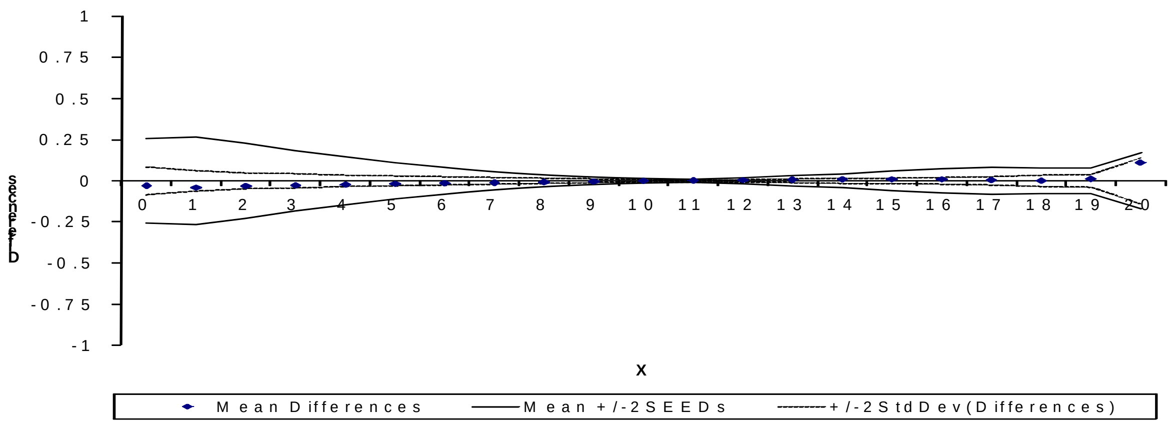


*Figure 7*. **Means and standard deviations of kernel equipercentile—traditional equipercentile equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = *Ny* = 200).**

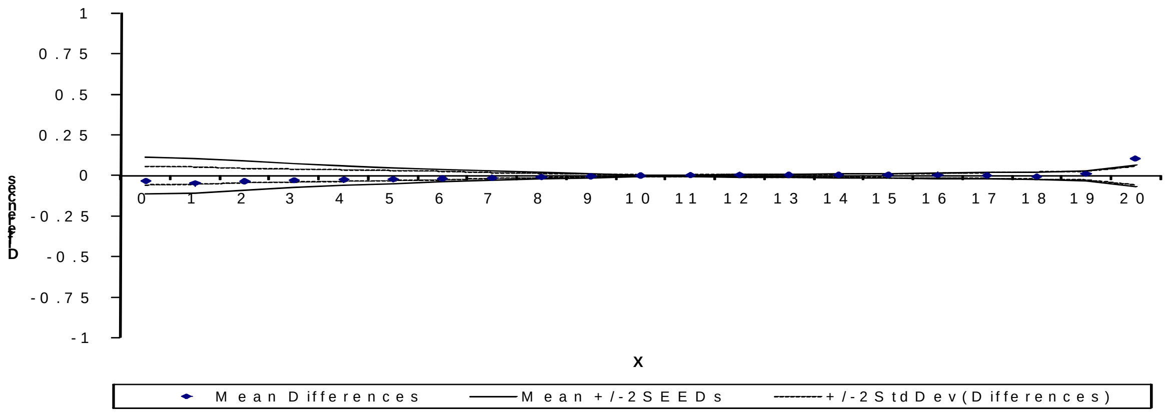


***Figure 8*. Means and standard deviations of kernel equipercentile—traditional equipercentile equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = 1,453, *Ny* = 1,455).**

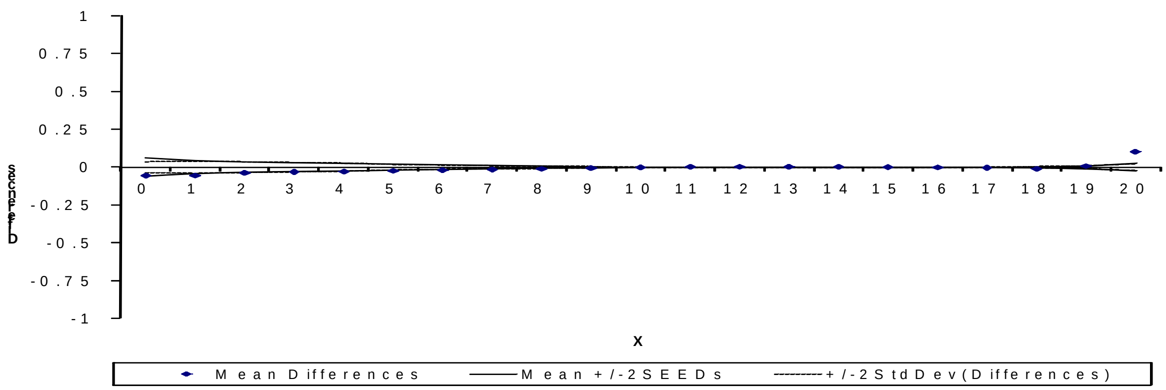


***Figure 9*. Means and standard deviations of kernel equipercentile—traditional equipercentile equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = *Ny* = 10,000).**

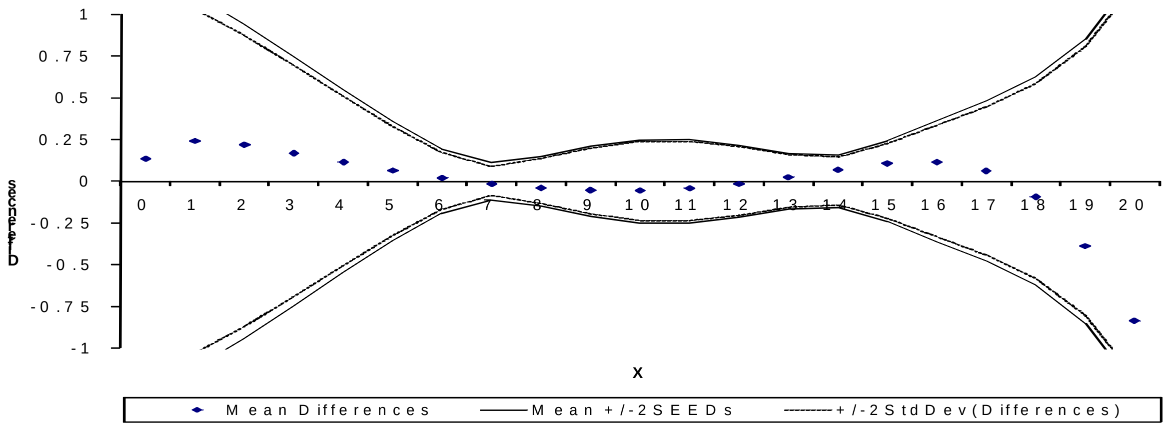


***Figure 10*. Means and standard deviations of traditional equipercentile—traditional linear equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = *Ny* = 200).**

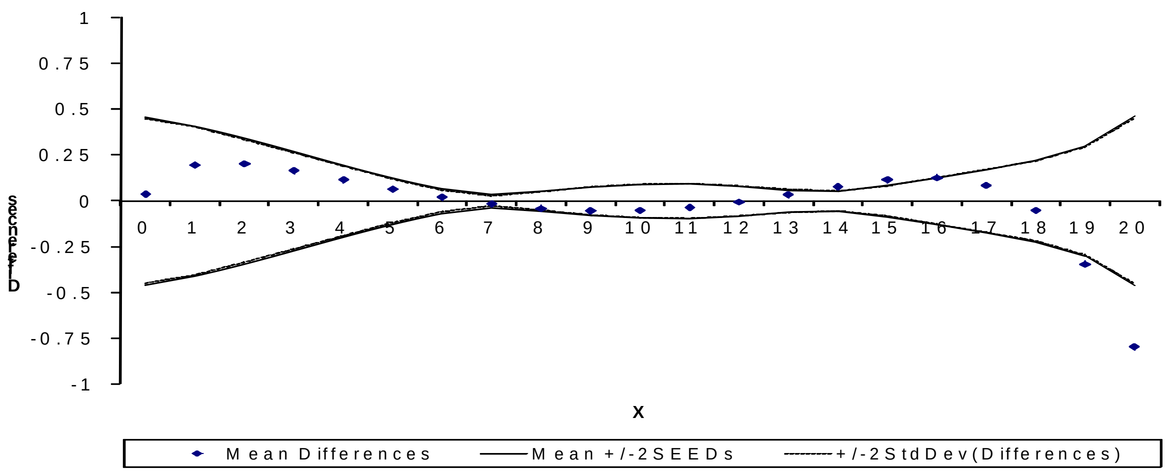


***Figure 11*. Means and standard deviations of traditional equipercentile—traditional linear equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = 1,453, *Ny* = 1,455).**

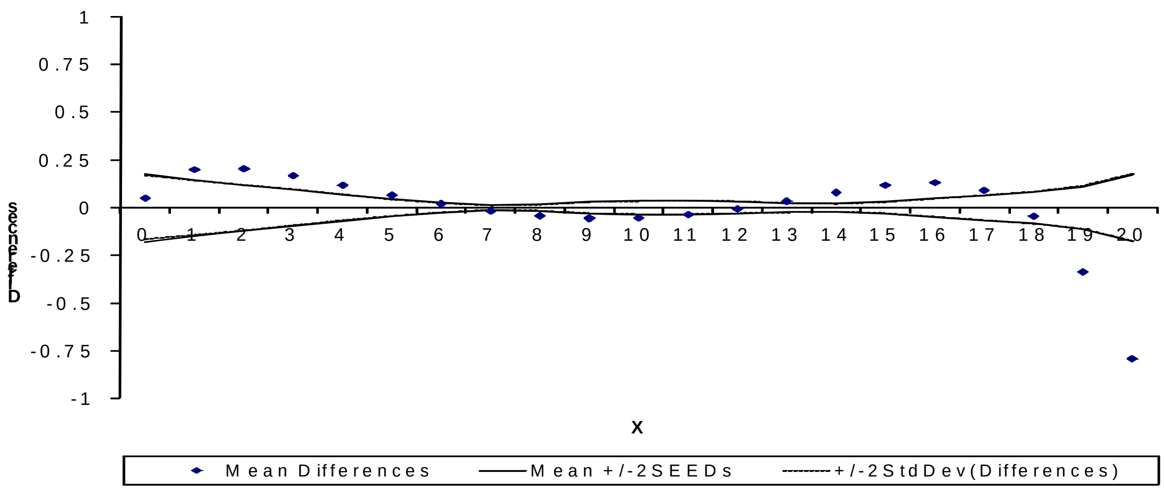


*Figure 12.* **Means and standard deviations of traditional equipercentile—traditional linear equated score differences and the mean ± 2 SEEDs for 1,000 samples of *X* and *Y* data (*Nx* = *Ny* = 10,000).**

the traditional equipercentile and traditional linear equating functions are statistically insignificant when based on sample sizes of 200, and several are significant for sample sizes of 10,000.

Simulations were also used to assess the other SEEDs of interest, the results of which reflect previous discussions and the results shown in Figures 4 through 12. To be specific, the SEED evaluation results for the kernel equipercentile and traditional linear functions agreed with the SEED evaluation results for the kernel equipercentile and kernel linear functions (Figures 4 through 6). Similarly, the SEED evaluation results for the traditional equipercentile and kernel linear functions agreed with the SEED evaluation results for the traditional equipercentile and traditional linear equating functions (Figures 10 through 12). Lastly, the SEED evaluation results for the kernel linear and traditional linear functions differed from zero, but only beyond the ninth decimal place.

## Discussion

In this paper, the SEEDs originally proposed for kernel equating functions (von Davier et al., 2004) were expanded into applications suitable for traditional linear and traditional equipercentile equating functions. The derivations provided in this paper for the traditional linear equating functions differ from those given in prior works on SEEs (Braun & Holland, 1982; Hanson, Zeng, & Kolen, 1993; Kolen, 1985; Zeng, 1991; Zeng & Cope, 1995). In addition to supporting SEEDs, these derivations provided new insights into the relationships between kernel and traditional equating functions. Specifically, the derivatives of kernel and traditional equating functions with respect to the same probability vectors can be directly related to the equating functions' SEEs and SEEDs, through their variances and covariances. The relationships between variances of equating functions and variances and covariances of their derivatives provide explanations for the kernel equipercentile equating function's smaller SEE relative to the traditional equipercentile equating function and for the asymptotically equivalent SEEs of the kernel linear and traditional linear equating functions.

Real data and simulation results were used to assess the accuracies of the SEEDs. The simulation results confirmed this paper's analytical suggestions of equating functions' SEEDs and also supported previous research of SEEs by showing that the SEEDs based on traditional equipercentile equating functions are similar to, but slightly larger than, those based on kernel equipercentile equating functions (Liou, Cheng, & Johnson, 1997; Moses & Holland, 2006). Prior studies' results about the effects of small sample sizes on SEEDs' accuracy and size were also found in this study, as well as the tendency for SEEDs to be less accurate at the lowest and highest scores of *X*. In conclusion, the major results of this study show that SEEDs agree with prior SEE studies.

One unanticipated result from the simulations was that the SEED for evaluating differences between kernel equipercentile and traditional equipercentile equating functions appeared to have accuracy problems. In sample sizes less than 10,000, the SEEDs overestimated the actual variability of the equated score differences. Follow-up analyses and evaluations of the results shown in Figures 4 through 6 and Figures 10 through 12 suggest that the source of the inaccuracy may be that the traditional equipercentile method is slightly less accurate than the kernel equipercentile method (Figures 10–12 vs. Figures 4–6). These slight accuracy differences are contributing to the inaccuracies of the SEED for evaluating differences between kernel equipercentile and traditional equipercentile equating functions.

### Extensions

By using $\Sigma_{\delta}$ matrices described in von Davier et al. (2004) with the derivatives described in Tables 1 and 2 of this paper, SEEDs can be computed to evaluate differences between equating functions for the single group, counterbalanced, and nonequivalent groups with anchor test data collection designs. Extensions to other designs are expected to produce results that are conceptually similar to those reported in this study, particularly the result that SEEs of non–presmoothed equating functions can be computed from the variances of the equating function's derivatives (e.g., Moses & Holland, 2006). Simulations that evaluate the accuracies of SEEDs for different equating designs across wide ranges of sample sizes and equating conditions would be useful.

## References


Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis.* Cambridge, MA: MIT Press.

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, NY: Academic.

Fienberg, S. E. (1979). The use of chi-squared statistics for categorical data problems. *Journal of the Royal Statistical Society, B, 41,* 54–64.

Haberman, S. J. (1989). Concavity and estimation. *The Annals of Statistics, 17,* 1631–1661.

Hanson, B. A., Zeng, L., & Kolen, M. J. (1993). Standard errors of Levine linear equating. *Applied Psychological Measurement, 17,* 225–237.

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.

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

Jarjoura, D., & Kolen, M. J. (1985). Standard errors of equipercentile equating for the common item nonequivalent populations design. *Journal of Educational Statistics, 10,* 143–160.

Kendall, M. G., & Stuart, A. (1977). *The advanced theory of statistics* (Vol. I; 4th ed.). New York, NY: Macmillan.

Kolen, M. J. (1985). Standard errors of Tucker equating. *Applied Psychological Measurement, 9,* 209–223.

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

Liou, M., & Cheng, P. E. (1995). Asymptotic standard error of equipercentile equating. *Journal of Educational and Behavioral Statistics, 20,* 259–286.

Liou, M., Cheng, P. E., & Johnson, E. G. (1997). Standard errors of the kernel equating methods under the common-item design. *Applied Psychological Measurement, 21,* 349–369.

Lord, F. M. (1982). The standard error of equipercentile equating. *Journal of Educational Statistics, 7,* 165–174.

Moses, T., & Holland, P. (2006). *Kernel equating without presmoothing*. Paper presented at the American Educational Research Association, San Francisco, CA.

Moses, T., Yang, W., & Wilson, C. (2007). Using kernel equating to check the statistical equivalence of nearly identical test editions*. Journal of Educational Measurement*, *44*, 157–178.

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

von Davier, A. A., & Kong, N. (2005). A unified approach to linear equating for the nonequivalent groups design. *Journal of Educational and Behavioral Statistics, 30,* 313–342.

Wang, T. (2006). *Standard errors of equating for equipercentile equating with log-linear presmoothing using the delta method* (CASMA Research Report, No. 14). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment.

Zeng, L. (1991). *Standard errors of linear equating for the single-group design* (Research Rep. No. 91-4). Iowa City, IA: American College Testing.

Zeng, L., & Cope, R. T. (1995). Standard error of linear equating for the counterbalanced design*. Journal of Educational and Behavioral Statistics, 20,* 337–348.