Research Report

Notes on a General Framework for Observed Score Equating

Tim Moses

Paul Holland

November 2008

ETS RR-08-59

Notes on a General Framework for Observed Score Equating

Tim Moses and Paul Holland
ETS, Princeton, NJ

November 2008

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS's constituents and the field.

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

Abstract

The purpose of this paper is to extend von Davier, Holland, and Thayer's (2004b) framework of kernel equating so that it can incorporate raw data and traditional equipercentile equating methods. One result of this more general framework is that previous equating methodology research can be viewed more comprehensively. Another result is that the standard error of equated score difference (SEED) has a wider application than originally proposed. The methods described in this paper are empirically evaluated in an accompanying simulation study (Moses & Holland, 2007).

Key words: Equating, kernel, equipercentile, standard errors of equating

Acknowledgments

The authors thank Shelby Haberman and Alina von Davier for helpful suggestions and Kim Fryer for the editorial work.

Table of Contents

Introduction

Organizing frameworks for equating methodologies can be helpful for revealing relationships among studies of seemingly different scope as well as for generating new equating methodologies. For example, von Davier, Holland, and Thayer's (2004b) recent framework related chained and post-stratification/frequency estimation equating approaches in terms of a design function. This framework provided a useful way to reconsider the comparisons of chained and post-stratification methods that focused on equated score differences (Holland, von Davier, Sinharay, & Han, 2006; Livingston, 2004; Livingston, Dorans, & Wright, 1990; MacCann, 1990; von Davier, Holland, & Thayer, 2004a; Wang, Lee, Brennan, & Kolen, 2006) and on theoretical variability estimates (Liou & Cheng, 1995). In addition, a significance test for chained and post-stratification equated score differences was a straightforward development produced within the von Davier et al. framework.

The purpose of this paper is to add generality to von Davier et al.'s (2004b) framework of equating. The original framework focused on data that are pre-smoothed using loglinear models. This paper describes how raw data can be utilized throughout the steps of the original framework, promoting a broader consideration of the impact of smoothing on equating bias and empirical variability (Fairbank, 1987; Hanson, 1991; Hanson, Zeng, & Colton, 1994; Livingston, 1993b; Skaggs, 2004) and on theoretical standard error estimates (Liou & Cheng, 1995; Liou, Cheng, & Johnson, 1997). While the original framework focused on kernel equating, this paper shows how the traditional equipercentile methods can be incorporated, promoting a reconsideration of comparative evaluations of continuization methods (Grant, Zhang, Damiano, & Lonstein, 2006; Liou et al., 1997; Liu & Low, 2006; Livingston, 1993a). The result of adding generality to the original framework is that the tools developed in von Davier et al. for significance tests of equating function differences can be applied to raw data and to traditional equipercentile methodologies. This paper focuses on the technical details involved in raw and smoothed data and in traditional equipercentile and kernel equating for the nonequivalent groups with anchor test (NEAT) design. All of the procedures described in this paper are empirically evaluated in an accompanying report (Moses & Holland, 2007).

The Five-Step Equating Framework

The empirical decisions practitioners make to equate tests include a pre-smoothing step that estimates true distributions from a set of collected data, an estimation step that calculates the probabilities of the test scores used in the equating process, a continuization step that uses the discrete scores to determine "in-between" scores, a score equating step, and a standard error estimation step. This section discusses these five steps in detail, describing specific applications that are relevant to the different equating functions used for the NEAT design.

Step 1: Presmoothing

The equating process begins with a decision on how to utilize the raw data that are gathered from test administrations. For the NEAT design the data are collected as two samples from nonequivalent populations (*P* and *Q*) that take different total tests (*X* or *Y*) and one internal or external anchor (*A* in *P* and *A* in *Q*). The raw data take the form of two bivariate test score distributions (*X, A* in *P* and *Y*, *A* in *Q)*. The scores of the bivariate distributions are discrete and assumed to be integers. The sampling error of the raw data can reduce the precision of the equating function but is itself reducible through the use of smoothing methods. Because of the potential for smoothing methods to introduce bias while reducing sampling variability, discussions of whether to use smoothed or raw data are typically framed in terms of a bias-variability tradeoff (e.g., Kolen, 1991; Kolen & Brennan, 2004).

Loglinear models for presmoothing. One useful tool for addressing the bias-variability tradeoff in discrete test score distributions is loglinear modeling (Holland & Thayer, 1987, 2000). Loglinear modeling has been promoted in test equating as an extremely flexible and variability-reducing smoothing technique for discrete test score distributions (Kolen, 1991; Livingston, 1993b; Hanson, 1991; Rosenbaum & Thayer, 1987). The loglinear modeling framework for comparing the fits of several nested models makes it possible to select among strong and weak degrees of smoothing using structured significance tests (Agresti, 2002; Bishop, Feinberg & Holland, 1975; Haberman, 1974).

For the NEAT design, two loglinear models are used to fit the two bivariate distributions. The following model is for the (***X, A***) distribution in *P*. The model for the (***Y***, ***A***) distribution in *Q* is similar:

$$\log_e(p_{jl}) = \alpha + \sum_{i=1}^{I} \beta_{xi}(x_j)^i + \sum_{h=1}^{H} \beta_{ah}(a_l)^h + \sum_{g=1}^{G} \sum_{f=1}^{F} \beta_{gf}(x_j)^g (a_l)^f , \quad (1)$$

where p_{jl} is the joint score probability of the score (x_j, a_l) (score x_j on test X and score a_l on test A as taken by population P) and the α and β's are free parameters that are estimated in the model-fitting process. The fitting of this model produces a smoothed bivariate distribution that preserves I moments in the marginal distribution of X, H moments in the marginal distribution of A, and a number of cross-moments ($G \leq I,\ F \leq H$) in the bivariate XA distribution. Many values can be used for I, H, G, and F; larger values preserve more features of the raw data in the model than smaller values.

The estimated variance-covariance matrix of population P's probabilities ($\mathbf{\Sigma_P}$) modeled with (1) can be used for computing standard errors of kernel equating (von Davier et al., 2004b). An efficient root factorization of $\mathbf{\Sigma_P}$ is the so-called "C-matrix," defined as

$$\mathbf{\Sigma_p} = \mathbf{C_P C_P^t}, \quad (2)$$

where $\mathbf{C_P}$ is the JL (for an external anchor) by T (where $T = I + H + GF$ in (1)) matrix and $\mathbf{C_p^t}$ is its transpose. $\mathbf{C_p}$ can be efficiently computed as

$$\mathbf{C_P} = N_P^{-1/2} \mathbf{D_{\sqrt{p}} Q}. \quad (3)$$

The diagonal matrix $\mathbf{D_{\sqrt{p}}}$ has entries $\sqrt{p_{jl}}$ along its main diagonal. N_P is the total sample size in P, and $\mathbf{Q}$ is the JL by T orthogonal matrix that comes from the following ***QR***-factorization,

$$\left[\mathbf{D_{\sqrt{p}}} - \sqrt{\mathbf{p}}\mathbf{p}^{\mathbf{t}}\right]\mathbf{B} = \mathbf{QR}. \quad (4)$$

$\mathbf{Q}$ is a JL by T matrix with orthogonal columns, $\mathbf{R}$ is a T by T upper triangular matrix, $\sqrt{\mathbf{p}}$ and $\mathbf{p}$ are the column vectors of the $\sqrt{p_{jl}}$'s and p_{jl}'s respectively, and $\mathbf{B}$ is the matrix of the score functions of x_j and a_l from (1) (Holland & Thayer, 1987). A major advantage of the C-matrix is

storage efficiency for large bivariate problems. (Note that there are only as many columns in the C-matrix as there are parameters in the model.)

No presmoothing. The equating process can utilize data in their raw form without the use of loglinear modeling. The use of raw data and estimates of their sampling variability provides an important basis for evaluating the equating functions computed from some possibly under-parameterized and biased loglinear models and from other over-parameterized and highly variable loglinear models. When raw data are used, the original probabilities are used as estimates of the population distribution. The only issue is to estimate a raw data version of the C-matrix given in (3).

To do so, let $\mathbf{p} = (1/N_P)\mathbf{n_P}$ denote the vector of sample proportions based on the unsmoothed sample frequencies in vector $\mathbf{n_P}$, where $\mathbf{n_P}$ is the *JL* "column-vectorized" version of the original *J* by *L* matrix of bivariate frequencies. It is assumed that $\mathbf{n_P}$ has the multinomial distribution M(N_P, $\boldsymbol{\rho}$), where N_P denotes the total sample size and $\boldsymbol{\rho}$ is the column vector of population probabilities (Bishop et al., 1975, p. 469). The covariance of $\mathbf{p}$ can be computed as

$$\boldsymbol{\Sigma}_{\mathbf{P}} = (1/N_P)\,(\mathbf{D_p} - \mathbf{pp}^{\mathrm{t}}), \tag{5}$$

where $\mathbf{D_p}$ denotes the diagonal matrix with entries p_{jl} along its main diagonal, and $\mathbf{p}^{\mathrm{t}}$ denotes the transpose of $\mathbf{p}$. In comparison to (3), there is no loglinear presmoothing so that the dimensions of $\mathbf{C_p}$ will be *JL* by *JL* rather than *JL* by *T*. Note that there is no reduction in the range of applications for the raw C-matrix as compared to the smoothed C-matrix currently used in kernel equating.

$\mathbf{C_p}$ is found as follows: Write $\boldsymbol{\Sigma}_{\mathbf{P}}$ in (5) as

$$\boldsymbol{\Sigma}_{\mathbf{P}} = (1/N_P)\;\mathbf{D}_{\sqrt{\mathbf{p}}}(\mathbf{I}_{\mathrm{JL}} - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}^{\mathrm{t}}})\mathbf{D}_{\sqrt{\mathbf{p}}}, \tag{6}$$

where $\mathbf{D}_{\sqrt{\mathbf{p}}}$ is the diagonal matrix with entries $\sqrt{p_{jl}}$ along its main diagonal, $\sqrt{\mathbf{p}}$ is the (column) vector of the $\sqrt{p_{jl}}$'s, and $\mathbf{I}_{\mathrm{JL}}$ is a *JL* by *JL* identity matrix. Equation (6) is justified by factoring

out $\mathbf{D}_{\sqrt{\mathbf{p}}}$ from the $\mathbf{D}_{\mathbf{p}}$ and the $\mathbf{pp}^{\mathrm{t}}$ in (5). The "inner" matrix in (6), that is, $(\mathbf{I}_{\mathbf{JL}} - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}^{\mathrm{t}}})$, is *idempotent* so that it equals it own square. Thus, (6) can be expressed as

$$\mathbf{\Sigma}_{\mathbf{P}} = (1/N_P)\,\mathbf{D}_{\sqrt{\mathbf{p}}}(\mathbf{I}_{\mathrm{JL}} - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}^{\mathrm{t}}})(\mathbf{I}_{\mathrm{JL}} - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}^{\mathrm{t}}})\mathbf{D}_{\sqrt{\mathbf{p}}} = \mathbf{C}_{\mathbf{P}}\mathbf{C}_{\mathbf{P}}^{\mathrm{t}}, \tag{7}$$

$$\text{where } \mathbf{C}_{\mathbf{p}} = N_P^{-1/2}\,\mathbf{D}_{\sqrt{\mathbf{p}}}(\mathbf{I}_{\mathrm{JL}} - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}^{\mathrm{t}}}) = N_P^{-1/2}\,(\mathbf{D}_{\sqrt{\mathbf{p}}} - \mathbf{p}\sqrt{\mathbf{p}^{\mathrm{t}}}). \tag{8}$$

Similar developments of (5) in (7) and (8) have been given in Bishop et al. (1975, pp. 469–473).

Step 2: Score Distribution Estimation

The second step of equating requires an estimation of the distributions of the scores that are going to be used in the equating from the raw or smoothed data from Step 1. This estimation depends on a "design function" (von Davier et al., 2004b). For the NEAT design and observed score equating methods, there are two possibilities for utilizing anchor score information. One is an *X* to *Y* equating that is based on chaining the single-group links computed for *X* to *A* in *P* and for *A* to *Y* in *Q*. For chained equating, the needed score distributions are estimated by obtaining marginal score distributions for X_P, A_P , A_Q, and Y_Q, the probability vectors of which are denoted as $\mathbf{r}_P$, $\mathbf{t}_P$, $\mathbf{t}_Q$, and $\mathbf{s}_Q$.

The other possibility is to use the anchor as a stratifying variable to estimate synthetic population score distributions on $\boldsymbol{X}$ and $\boldsymbol{Y}$ in *P* and *Q*. The synthetic score distributions are then used in post-stratification equating, also referred to as frequency estimation equipercentile equating (Jarjoura & Kolen, 1985; Kolen & Brennan, 2004). Score probabilities can be estimated in the synthetic populations for *X* and *Y* as

$$\mathbf{r}_{PSE} = \mathbf{r}_{wP+(1-w)Q},\ r_j = \sum_l \left[w + \frac{(1-w)t_{Ql}}{t_{Pl}} \right] p_{jl} \tag{9}$$

$$\mathbf{s}_{PSE} = \mathbf{s}_{wP+(1-w)Q},\ s_k = \sum_l \left[(1-w) + \frac{wt_{Pl}}{t_{Ql}} \right] q_{kl}, \tag{10}$$

where w is a value between 0 and 1 that specifies the proportion of *P* in the synthetic population, where the synthetic population is defined as $wP+(1-w)Q$, t_{Pl} and t_{Ql} are the marginal

probabilities at anchor score l in populations P and Q, and p_{jl} and q_{kl} are bivariate probabilities at scores X_j and A_l in population P and at scores Y_k and A_l in population Q.

Raw data complications. With raw data that have not been presmoothed, there is a potential for some marginal score probabilities to be zero. This creates problems for two aspects of equating. For the traditional equipercentile method, more than one score exists with the same cumulative probability, and there is no clear basis for assigning the cumulative probability to any one score. For the marginal anchor scores that are used in kernel and traditional equipercentile post-stratification equating, division-by-zero problems can occur [t_{Pl} and t_{Ql} in Equations (9) and (10)]. To avoid problems from scores with zero frequencies, many ad hoc rules have been proposed (Jarjoura & Kolen, 1986; Kolen & Brennan, 2004; Livingston, 2004).

The rule proposed in this paper is also ad hoc, but is proposed because it provides a direct basis for the estimation of equating errors in Step 5 while controlling for bias in Step 4, particularly for the post-stratification method. This rule is to average any marginal frequencies (or marginal synthetic probabilities multiplied by the sample size) less than one with the nearest lower and higher scores that had frequencies that were one or greater. These averaged frequencies and probabilities are then used in the rest of the equating processes.

For post-stratification, the averaging is done on the marginal anchor scores, and then, to preserve the bivariate distribution and make (9) and (10) correctly sum to one, on each of the total test score probabilities in the bivariate tables. For example, if $t_{P2} = 0$ so that the marginal anchor probabilities at scores l = 1, 2, and 3 are averaged, this means that $\frac{t_{P1} + t_{P2} + t_{P3}}{3}$ is used for t_{P1}, t_{P2} and t_{P3}, $\frac{p_{11} + p_{12} + p_{13}}{3}$ is used as the bivariate probability for p_{11}, p_{12} and p_{13}, $\frac{p_{21} + p_{22} + p_{23}}{3}$ is used as the bivariate probability for p_{21}, p_{22} and p_{23}, and so forth for all J. Because (9) and (10) rely on **p** and **q** vectors with averaged frequencies and the variance-covariances of **p** and **q** (5-8) are computed from the original frequencies, the discussions of (9) and (10) to follow in Step 5 and the appendix will notate the **p** and **q** that are produced from some frequency averaging as $\bar{\mathbf{p}}$ and $\bar{\mathbf{q}}$ to distinguish them from the original **p** and **q** vectors. Marginal probability averaging for the levels of j and k in **r** and **s** may also sometimes be necessary because

eliminating low frequencies in the column sums of **p** does not necessarily eliminate low frequencies in the row sums of $\overline{\mathbf{p}}$.

Step 3: Continuization

The discrete score distributions of Steps 1 and 2 have to be continuized into continuous cumulative density functions (cdfs) with defined inverses. The continuization process is based on an assertion of how "in-between" score values are distributed in the discrete distribution. Two continuization possibilities treated here are the traditional equipercentile method's percentile-rank method and the kernel method's kernel density estimation method. There are other continuization approaches that are not addressed here (e.g., Wang, 2004).

Traditional equipercentile continuization. The traditional equipercentile method of continuization is based on linearly interpolating between two scores and their cumulative percentages. The assumption is that scores and cumulative percentages are uniformly distributed between one-unit score intervals so that the cdf (i.e., "percentile rank"; Kolen & Brennan, 2004) for any x score, possible or impossible, can be computed by:

$$
\begin{aligned}
F(x; \mathbf{r}) &= 0 \ \text{ if } x \le x_1 - ½ \\
&= (x - (x_1 - ½))r_1 \ \text{ if } x_1 - ½ \le x \le x_1 + ½ , \\
&= \sum_{k=1}^{j-1} r_k + r_j (x - (x_j - ½)) \ \text{if } x_j - ½ \le x \le x_j + ½, \text{ for } j = 2, \ldots J, \\
&= 1 \ \text{if } x \ge x_J + ½. \qquad (11)
\end{aligned}
$$

The equipercentile continuization can be distinguished from the kernel continuization in two respects. First, the equipercentile's *degree* of continuization is independent of the sample data. Second, the equipercentile's continuization is based on the cumulative probabilities of pairs of consecutive scores in the score distribution.

Kernel continuization. The continuization method for kernel equating is kernel density estimation for univariate score distributions. In von Davier et al. (2004b, p. 56), a Gaussian kernel is used, with a parameter (h) that can be altered so that the resulting cdfs are more or less influenced by the Gaussian kernel:

$$F_{hX}(x;\mathbf{r}) = \sum_j r_j \Phi(R_{jX}(x)), \tag{12}$$

where $R_{jX}(x) = \dfrac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}$, $a_X = \sqrt{\dfrac{\sigma_X^2}{\sigma_X^2 + h_X^2}}$, and Φ denotes the standard Gaussian cdf.

The bandwidth parameter (h_x) can be set by the practitioner. Extremely large ($\geq 10\sigma_X$) bandwidths produce cdfs that retain only the mean and variance of the data, causing the equating function to resolve into a linear form. Smaller values of h_x produce cdfs that retain more of the data. von Davier et al. (2004b) recommend a rule for selecting h based on achieving a continuized cdf that closely approximates the discrete data (by minimizing the sum of the squared error) while having few modes (by avoiding rapid changes in the derivative of the cdf). One computational method that works well for meeting these two criteria in selecting h is known as Brent's method or parabolic interpolation (Press, Teukolsky, Vetterling, & Flannery, 1992).

Step 4: Equating

Traditional equipercentile and kernel equating functions are implemented by first computing cdf values for scores from one distribution and then finding scores on another distribution with cdfs that match those of the first (i.e., the inverse df). The equipercentile and kernel versions of the cdf computations were described in Step 3. The equipercentile and kernel versions of the inverse cdf computations are described here.

The traditional equipercentile inverse cdf. To compute an inverse cdf using the traditional equipercentile method, the percentile rank of one score on a score distribution, $F(x_j;\mathbf{r})$, is located on another score distribution such that $G(y_u + 1/2;\mathbf{s})$ is the smallest percentile rank at an interval boundary score (i.e., y +/- 1/2 for score intervals of 1) on $\mathbf{s}$ that is larger than $F(x_j;\mathbf{r})$. The inverse cdf is then computed as

$$e_y(x_j) = G^{-1}(F(x_j;\mathbf{r});\mathbf{s}) = y_u - 1/2 + \frac{F(x_j;\mathbf{r}) - G(y_{u-1} + 1/2;\mathbf{s})}{G(y_u + 1/2;\mathbf{s}) - G(y_{u-1} + 1/2;\mathbf{s})}. \tag{13}$$

The kernel inverse cdf. The inverse cdf for the kernel method can be computed using an iterative search process, such as Newton's method. Given a kernel continuized cdf, $F_{hX}(x_j;\mathbf{r})$, the

y score with a matching cdf value on $\mathbf{s}$ can be found by starting with a $G_{hY}(y^{*};\mathbf{s})$ that is based on an initial y^{*} value that is plausibly close to $e_{y}(x_{j})$. Then y^{*} is updated as $y^{**} = y^{*} + \frac{\left(F_{hX}(x_{j};\mathbf{r}) - G_{hY}(y^{*};\mathbf{s})\right)}{\left(\frac{\partial G_{hY}(y^{*};\mathbf{s})}{\partial y}\right)}$. When it is noted that both the derivative in the update's denominator and the difference in percentile ranks in (13)'s denominator are probabilities of y scores near $e_{y}(x_{j})$, the kernel method's use of Newton for the inverse cdf computation and the inverse cdf computation for the traditional equipercentile method can be understood as being almost identical. The Newton update is repeatedly calculated until a y^{**} is found such that $G_{hY}(y^{**};\mathbf{s})$ is acceptably close to $F_{hX}(x_{j};\mathbf{r})$. The result can be expressed as the kernel equating function

$$e_{y}(x_{j}) = G_{hY}^{-1}\left(F_{hX}(x_{j};\mathbf{r});\mathbf{s}\right). \tag{14}$$

Equating functions. The chained kernel and equipercentile equating functions can be expressed as two cdf matchings brought together based on the four marginal distributions:

$$e_{Y(CE)}(x) = G^{-1}(H(H^{-1}(F(x;\mathbf{r}_{\mathbf{P}});\mathbf{t}_{\mathbf{P}});\mathbf{t}_{\mathbf{Q}});\mathbf{s}_{\mathbf{Q}}). \tag{15}$$

The post-stratification kernel and equipercentile equating functions are based on matching the cdfs from the two synthetic populations:

$$e_{Y(PSE)}(x) = G^{-1}(F(x;\mathbf{r}_{\mathbf{(wP+(1-w)Q)}});\mathbf{s}_{\mathbf{(wP+(1-w)Q)}}). \tag{16}$$

Step 5: Standard Error Estimation

The delta method is frequently used to estimate the extent of sampling variation in the equating function (Jarjoura & Kolen, 1985; Liou & Cheng, 1995; Liou et al., 1997; Lord, 1982; von Davier et al., 2004b). With the delta method the assumption is made that the *X* to *Y* equating function is continuously differentiable across *X*, so that equating variability can be estimated using the equating function derivatives and the corresponding estimates of variability in the data.

A major difference between how the delta method is treated in this paper and how it was developed by Jarjoura and Kolen (1986), Liou and Cheng (1995), Liou et al. (1997), and Lord (1982), is that the differentiation and variance-covariance estimation used here are in terms of the probabilities *at every score level* rather than in terms of cumulative probabilities at a limited number of score levels. This orientation to score probabilities allows for a very flexible approach to equating function and standard error estimation that is straightforward to apply to all major equating designs (von Davier et al., 2004b), to smoothed and raw data, to traditional equipercentile and kernel equating functions, and to the estimated variability of equated score differences.

The pieces of the delta method standard errors are described below. They include the derivative of the equating function with respect to marginal score probabilities ($\mathbf{J_e}$, Steps 3 and 4), the derivative of the marginal score probabilities with respect to the original bivariate distribution ($\mathbf{J_{DF}}$, Steps 1 and 2), and the C-matrices of the two estimated distributions (Step 1). The general form of the standard error is given as the root sum of squares of the "SEE-vector," which is the product of the derivative vectors and the C-matrices:

$$\mathrm{SEE}_y(x) = \left\| \mathbf{J_e J_{DF} C} \right\| \tag{17}$$

In (17), $\|\mathbf{x}\| = \sqrt{\sum_j x_j^2}$ denotes the Euclidian length (norm) of vector $\mathbf{x}$.

Equating function derivatives. Kernel and equipercentile equating functions have the general form $e_Y(x;\mathbf{r},\mathbf{s}) = G^{-1}(F(x;\mathbf{r});\mathbf{s})$, so that they can be differentiated with respect to marginal score probabilities as

$$\frac{\partial e_Y}{\partial r_j} = \frac{1}{G'}\frac{\partial F(x;\mathbf{r})}{\partial r_j}, \tag{18}$$

$$\frac{\partial e_Y}{\partial s_k} = -\frac{1}{G'}\frac{\partial G(e_Y(x);\mathbf{s})}{\partial s_k}, \tag{19}$$

where $G' = \frac{\partial G(e_Y(x);\mathbf{s})}{\partial y}$ (von Davier et al., 2004b).

Derivatives for the equipercentile equating functions. The derivatives of the continuized percentile-rank cdf in (11) at score x with respect to each $x_{j'}$ marginal score probability requires that x_j be located in the range of $\boldsymbol{X}$ such that $x_j - ½ \leq x \leq x_j + ½$. Then

$$\frac{\partial F(x;\mathbf{r})}{\partial r_{j'}} = 1 \qquad j' < j,$$

$$= (x - (x_j - ½)) \qquad j' = j,$$

$$= 0 \qquad j' > j. \qquad (20)$$

The expression of $\frac{\partial G(e_y(x);\mathbf{s})}{\partial s_k}$ is similar. The next section of this paper gives a modification of (20) when raw and sparse data are used.

The derivatives of the equated scores with respect to the actual Y scores are:

$$\frac{\partial G(e_y(x);\mathbf{s})}{\partial y} = s_{yu} \text{ for } e_y(x) \text{ where } y_u - 1/2 < e_y(x) < y_u + 1/2$$

$$= \frac{s_{yu} + s_{yu-1}}{2} \text{ for } e_y(x) \text{ where } e_y(x) = y_u - 1/2 \qquad (21)$$

The averaging of two probabilities (i.e., derivatives) when the traditional equipercentile $e_y(x)$ is at a +/- 1/2 score boundary is a pragmatic proposal. It is not strictly correct because the delta method is intended for smooth and continuously differentiable functions. The percentile-rank derivative with respect to score probabilities is not continuous at +/- 1/2 score boundaries; it has two derivatives that depend on whether the rate of change in the percentile-rank function is approached from the upper or lower score around the score boundary. The precise way of dealing with equated scores that are exactly on score boundaries is to treat the derivative and resulting standard error estimate as undefined. The actual situations where equated scores are at score boundaries are rare except at the tails of score distributions, where standard error estimates are known to be poor for other reasons, namely sparse data (Jarjoura & Kolen, 1985; Liou & Cheng, 1995). Simulation results described in Moses and Holland (2007) suggest that the use of an average of the two possible derivatives for equated scores at +/- 1/2 score boundaries has minimally adverse effects on standard error estimates. The averaging of derivatives in (21) is

completely avoided in kernel equating functions because kernel cdfs are smooth and continuously differentiable across the entire score range.

Raw and sparse data. Equations (20) and (21) must be adjusted when frequency-averaging is done for sparse, marginal probabilities and traditional equipercentile equating. When frequency-averaging is used, the equating function is not based directly on the marginal probabilities. When the percentile-rank function, $F(x;\mathbf{r})$, is computed based on averaged marginal probabilities, it can be expressed as $F(x;\bar{\mathbf{r}})$, where $\bar{\mathbf{r}}$ is the vector of original marginal probabilities where some are averaged. For example, when the probabilities at scores x_2 and x_6 are zero, they are averaged with the probabilities at scores x_1 and x_3 and x_5 and x_7, respectively. Then the percentile rank at score 6 can be expressed as

$$
\begin{aligned}
F(x_6 = 5;\bar{\mathbf{r}}) &= \bar{r}_1 + \bar{r}_2 + \bar{r}_3 + r_4 + \bar{r}_5 + \bar{r}_6(5-(5-1/2)), \\
&= 3\left(\frac{r_1+r_2+r_3}{3}\right) + r_4 + \left(\frac{r_5+r_6+r_7}{3}\right) + \left(\frac{r_5+r_6+r_7}{3}\right)(5-(5-1/2)), \\
&= r_1 + r_2 + r_3 + r_4 + \left(\frac{r_5+r_6+r_7}{3}\right)\left(1+5-(5-1/2)\right) \\
&= r_1 + r_2 + r_3 + r_4 + \left(\frac{r_5+r_6+r_7}{3}\right)\left(6-(5-1/2)\right) \\
&= r_1 + r_2 + r_3 + r_4 + \left(\frac{r_5+r_6+r_7}{3}\right)\left(x_7-(x_6-1/2)\right) \\
&= r_1 + r_2 + r_3 + r_4 + \left(\frac{r_5+r_6+r_7}{3}\right)\left(x_6-(x_5-1/2)\right).
\end{aligned}
$$

Therefore,

$$
\frac{\partial F(x;\bar{\mathbf{r}})}{\partial r_{j'}} =
\begin{cases}
1 & j' < j, \\
(x-(x_j-1/2))(1/b) & \text{for any } j' \text{ for which } r_{j'} \text{ is averaged with } r_j, \\
0 & j' > j + b - 1.
\end{cases}
\tag{22}
$$

In (22), x_j is the lowest of the scores with averaged probabilities at x, and b is the number of score values with probabilities that are averaged together. Note that, when several scores have probabilities that are averaged together, b gets large, marginal probability distributions become uniform, equating functions approach the identity function, and the standard errors go to zero.

Derivatives for the kernel equating functions. With respect to marginal score probabilities and scores, the derivatives of the continuized kernel cdf in (12) are given in von Davier et al. (2004b) as

$$\frac{\partial F(x;\mathbf{r})}{\partial r_k} = \Phi(R_{jX}(x;\mathbf{r})) - M_{jX}(x;\mathbf{r})\frac{\partial F(x;\mathbf{r})}{\partial x} \tag{23}$$

and $$\frac{\partial F(x;\mathbf{r})}{\partial x} = \sum_j r_j \phi(R_{jX}(x;\mathbf{r}))\frac{1}{a_X h_X} \tag{24}$$

where $R_{jX}(x;\mathbf{r}) = \dfrac{x - a_X x_j - (1-a_X)\mu_X}{a_X h_X}$,

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

and ϕ denotes the standard Gaussian density function.

Equating function derivative vectors. When the vectors of score-level equating functions are computed, the equating function derivative vectors (i.e., Jacobians) can be merged into the $\mathbf{J}_\mathbf{e}$ that is used in (17). For post-stratification, the two Jacobians for each of the two synthetic score distributions are merged:

$$\mathbf{J}_{\mathbf{eY(PSE)}} = \left(\frac{\partial e_Y}{\partial r_{wP+(1-w)Q}}, \frac{\partial e_Y}{\partial s_{wP+(1-w)Q}}\right) = \left(\mathbf{J}_{\mathbf{rwP+(1-w)Q}}, \mathbf{J}_{\mathbf{swP+(1-w)Q}}\right). \tag{25}$$

For chained equating, the Jacobians based on the four marginal distributions are merged:

$$\mathbf{J}_{\mathbf{eY(CE)}}(x) = \left(\left(\frac{\partial e_Y}{\partial a}(e_a(x))\mathbf{J}_{\mathbf{ea}}(x)\right), \mathbf{J}_{\mathbf{eY}}(e_a(x))\right), \tag{26}$$

where $\mathbf{J}_{\mathbf{ea}}(x) = \left(\frac{\partial e_a(x)}{\partial \mathbf{r}_\mathbf{P}}, \frac{\partial e_a(x)}{\partial \mathbf{t}_\mathbf{P}} \right) = \left(\mathbf{J}_{\mathbf{rP}}, \mathbf{J}_{\mathbf{tP}} \right)$,

$$\mathbf{J}_{\mathbf{eY}}(e_a(x)) = \left(\frac{\partial e_Y(e_a(x))}{\partial \mathbf{t}_\mathbf{Q}}, \frac{\partial e_Y(e_a(x))}{\partial \mathbf{s}_\mathbf{Q}} \right) = \left(\mathbf{J}_{\mathbf{tQ}}, \mathbf{J}_{\mathbf{sQ}} \right),$$

and $\frac{\partial e_Y}{\partial a}(e_a(x)) = \frac{\partial H_Q(e_a(x); \mathbf{t}_\mathbf{Q})/\partial a}{\partial G_Q(e_Y(x); \mathbf{s}_\mathbf{Q})/\partial y}$.

Design function derivatives. The design function for chained equating involves summing up the bivariate probabilities to create two marginal probability distributions, one for the test and another for the anchor scores. For error estimation, the test and anchor entries of the C-matrix are summed in the same way as the bivariate probabilities. For a formal description of the design function derivatives for chained equating, refer to the discussion of the single-group design in the appendix.

The design function for post-stratification is nonlinear and can be written in terms of the two bivariate and four marginal probabilities used in (9) and (10):

$$\frac{\partial \boldsymbol{r}}{\partial \mathbf{p}_\mathbf{l}} = \frac{\partial \boldsymbol{r}}{\partial \overline{\mathbf{p}}_\mathbf{l}} \frac{\partial \overline{\mathbf{p}}_\mathbf{l}}{\partial \mathbf{p}_\mathbf{l}} = \left(w + (1-w)\left(\frac{t_{\bar{Q}l}}{t_{\bar{P}l}}\right)\mathbf{I}_\mathbf{J} - (1-w)\left(\frac{t_{\bar{Q}l}}{t_{\bar{P}l}}\right)\left[\left(t_{\bar{P}l}\right)^{-1}\overline{\mathbf{p}}_\mathbf{l}\right]\mathbf{1}_J^t \right)\frac{1}{\mathrm{b}_{\mathrm{Pl}}} \tag{27}$$

$$\frac{\partial \boldsymbol{s}}{\partial \mathbf{p}_\mathbf{l}} = \frac{\partial \boldsymbol{s}}{\partial \overline{\mathbf{p}}_\mathbf{l}} \frac{\partial \overline{\mathbf{p}}_\mathbf{l}}{\partial \mathbf{p}_\mathbf{l}} = \left(w\left[\left(t_{\bar{Q}l}\right)^{-1}\overline{\mathbf{q}}_\mathbf{l}\right]\mathbf{1}_J^t \right)\frac{1}{\mathrm{b}_{\mathrm{Pl}}} \tag{28}$$

$$\frac{\partial \boldsymbol{r}}{\partial \mathbf{q}_\mathbf{l}} = \frac{\partial \boldsymbol{r}}{\partial \overline{\mathbf{q}}_\mathbf{l}} \frac{\partial \overline{\mathbf{q}}_\mathbf{l}}{\partial \mathbf{q}_\mathbf{l}} = \left((1-w)\left[\left(t_{\bar{P}l}\right)^{-1}\overline{\mathbf{p}}_\mathbf{l}\right]\mathbf{1}_K^t \right)\frac{1}{\mathrm{b}_{\mathrm{Ql}}} \tag{29}$$

$$\frac{\partial \boldsymbol{s}}{\partial \mathbf{q}_\mathbf{l}} = \frac{\partial \boldsymbol{s}}{\partial \overline{\mathbf{q}}_\mathbf{l}} \frac{\partial \overline{\mathbf{q}}_\mathbf{l}}{\partial \mathbf{q}_\mathbf{l}} = \left(1 - w + w\left(\frac{t_{\bar{P}l}}{t_{\bar{Q}l}}\right)\mathbf{I}_\mathbf{K} - w\left(\frac{t_{\bar{P}l}}{t_{\bar{Q}l}}\right)\left[\left(t_{\bar{Q}l}\right)^{-1}\overline{\mathbf{q}}_\mathbf{l}\right]\mathbf{1}_K^t \right)\frac{1}{\mathrm{b}_{\mathrm{Ql}}}, \tag{30}$$

where the $\overline{\mathbf{p}}_\mathbf{l}$ and $\overline{\mathbf{q}}_\mathbf{l}$ terms denote computations involving the averaged form of the bivariate probabilities (Step 2), and the b's denote the number of score levels on the marginal anchor distributions in *P* or *Q* that were averaged to avoid divisions-by-zero to compute (9) and (10). For smoothed data, $\overline{\mathbf{p}}_\mathbf{l}$ and $\overline{\mathbf{q}}_\mathbf{l}$ equal $\mathbf{p}_\mathbf{l}$ and $\mathbf{q}_\mathbf{l}$ and all b's equal 1. For raw and sparse data, some

anchor scores' probabilities are averaged so that $\bar{\mathbf{p}}_{\mathbf{l}}$ and $\bar{\mathbf{q}}_{\mathbf{l}}$ are not always equal to $\mathbf{p}_{\mathbf{l}}$ and $\mathbf{q}_{\mathbf{l}}$ and some b's can be greater than 1.

C. The final **C** in (17) is given as a stacking of the *P* and *Q* C-matrices, and assumes that the two samples' probability distributions are estimated independently:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C_P} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_Q} \end{pmatrix}. \tag{31}$$

Computations for raw data. For raw data, the SEE-vector uses a square C-matrix and can become so large that it exceeds storage capacity. Additional multiplications that have a $\mathbf{(J_e J_{DF})C}$ orientation can result in SEE formulas that avoid the C-matrix in its large form. These are described in more detail in the appendix, with the main results presented here.

For chained equating, the variance of the raw equating function is given as the sum of the variances of the equating function derivatives in (25):

$$\mathrm{SEE}^2(x) = [\frac{\partial e_Y}{\partial a}(e_a(x))]^2 \mathrm{SEE}_{\mathrm{a}}^2(\mathrm{x}) + \mathrm{SEE}_{\mathrm{Y}}^2(\mathrm{e_a(x)}), \tag{32}$$

where $\mathrm{SEE}_{\mathrm{a}}^2(x) = (1/N_{\boldsymbol{P}}) \sum_{\mathrm{j,l}} (\mathrm{J_{r_j}} + \mathrm{J_{t_l}} - (\mu_{\mathbf{J_r P}} + \mu_{\mathbf{J_t P}}))^2 p_{jl}$,

$\mathrm{SEE}_{\mathrm{Y}}^2(\mathrm{e_a}(x)) = (1/N_{\boldsymbol{Q}}) \sum_{\mathrm{k,l}} (\mathrm{J_{s_k}} + \mathrm{J_{t_l}} - (\mu_{\mathbf{J_s Q}} + \mu_{\mathbf{J_t Q}}))^2 q_{kl}$, $\frac{\partial e_Y}{\partial a}(e_a(x))$ was defined in (26), and the μ's denote averages of particular derivative vectors.

For post-stratification equating, the variance of the raw equating function is given as the variances of the $\mathbf{JeJ_{DF}}$ products from (25) and (27-30):

$$\mathrm{SEE}^2(\mathrm{x}) = (1/N_{\boldsymbol{P}}) \sum_{\mathrm{jl}} (\mathrm{A_{jl}} - \mu_{\mathrm{AP}})^2 p_{jl} + (1/N_{\boldsymbol{Q}}) \sum_{\mathrm{kl}} (\mathrm{B_{kl}} - \mu_{\mathrm{BQ}})^2 q_{kl}, \tag{33}$$

where $\mathrm{A_{jl}} = \left(\mathrm{w_{l\bar{P}}}(\mathrm{J_{r_j}} - \mu_{\mathbf{J_r}\bar{\mathbf{P}}|\mathbf{l}}) + w(\mu_{\mathbf{J_r}\bar{\mathbf{P}}|\mathbf{l}} + \mu_{\mathbf{J_s}\bar{\mathbf{Q}}|\mathbf{l}}) \right) \frac{1}{\mathrm{b_{Pl}}}$,

$\mathrm{B_{kl}} = \left(\mathrm{w_{l\bar{Q}}}(\mathrm{J_{s_k}} - \mu_{\mathbf{J_s}\bar{\mathbf{Q}}|\mathbf{l}}) + (1-w)(\mu_{\mathbf{J_r}\bar{\mathbf{P}}|\mathbf{l}} + \mu_{\mathbf{J_s}\bar{\mathbf{Q}}|\mathbf{l}}) \right) \frac{1}{\mathrm{b_{Ql}}}$,

$w_{l\bar{P}} = w + (1-w)t_{\bar{Q}l}/t_{\bar{P}l} = (wt_{\bar{P}l} + (1-w)t_{\bar{Q}l})/t_{\bar{P}l} = t_{\bar{T}l}/t_{\bar{P}l}$,

and $w_{l\bar{Q}} = (1-w) + w(t_{\bar{P}l}/t_{\bar{Q}l}) = t_{\bar{T}l}/t_{\bar{Q}l}$.

SEEDs. Equating functions are often selected based on comparisons with other equating functions (e.g., $e_{1y}(x) - e_{2y}(x)$), such as curvilinear and linear differences, chained and post-stratification differences, post-stratification differences when *w* is varied, and differences between linear methods (von Davier & Kong, 2005). von Davier et al. (2004b) showed that through noting $\partial(e_1 - e_2) = \mathbf{J}_{e1} - \mathbf{J}_{e2}$, the SEE-vector formula in (17) can be generalized into a SEED-vector for the standard error of equating function differences:

$$\mathrm{SEED}_y(x) = \left\| \left(\mathbf{J}_{e1}\mathbf{J}_{DF1} - \mathbf{J}_{e2}\mathbf{J}_{DF2} \right) \mathbf{C} \right\|. \quad (34)$$

Two types of SEEDs are important in the NEAT design. One is for evaluating the significance of differences between curvilinear and linear kernel equating functions for a given design function (either chained or post-stratification). In the kernel framework, curvilinear and linear equating functions can be computed simply by selecting smaller and larger continuization bandwidths (*h's*). Because the design function for the curvilinear and linear comparison is the same, the SEED can be expressed as

$$\mathrm{SEED}_y(x) = \left\| \left(\mathbf{J}_{e1} - \mathbf{J}_{e2} \right) \mathbf{J}_{DF} \mathbf{C} \right\|. \quad (35)$$

Another type of SEED is useful for assessing the significance of differences between chained and post-stratification equated scores. For this type of comparison, there are separate design functions and derivative vectors, so (34) is used.

Raw SEEDs. The raw SEED analogues of (34) are discussed in detail in the appendix. The final results are as follows. For the curvilinear-linear comparison in chained kernel equating,

$$\mathrm{SEED}^2(x) = (1/N_P)\mathrm{SEED}_a^2(x) + (1/N_Q)\mathrm{SEED}_Y^2(e_a(x)), \quad (36)$$

where $\mathrm{SEED}_{\mathrm{a}}^{2}(x) =$

$$\sum_{\mathrm{j,l}}\left(\frac{\partial e1_{Y}(e1_{a}(x))}{\partial a}(\mathrm{J}_{\mathrm{r1}_{\mathrm{j}}}+\mathrm{J}_{\mathrm{t1}_{\mathrm{l}}})-\frac{\partial e2_{Y}(e2_{a}(x))}{\partial a}(\mathrm{J}_{\mathrm{r2}_{\mathrm{j}}}+\mathrm{J}_{\mathrm{t2}_{\mathrm{l}}})-\mu_{(\frac{\partial \mathbf{e1}}{\partial \mathbf{a}}\mathbf{J}_{\mathbf{r1}}-\frac{\partial \mathbf{e2}}{\partial \mathbf{a}}\mathbf{J}_{\mathbf{r2}})\mathbf{P}}-\mu_{(\frac{\partial \mathbf{e1}}{\partial \mathbf{a}}\mathbf{J}_{\mathbf{t1}}-\frac{\partial \mathbf{e2}}{\partial \mathbf{a}}\mathbf{J}_{\mathbf{t2}})\mathbf{P}}\right)^{2}p_{jl}$$ and

$$\mathrm{SEED}_{\mathrm{Y}}^{2}(\mathrm{e}_{\mathrm{a}}(x)) = \sum_{\mathrm{k,l}}(\mathrm{J}_{\mathrm{s1}_{\mathrm{k}}}-\mathrm{J}_{\mathrm{s2}_{\mathrm{k}}}+\mathrm{J}_{\mathrm{t1}_{\mathrm{l}}}-\mathrm{J}_{\mathrm{t2}_{\mathrm{l}}}-\mu_{(\mathbf{J_{s1}}-\mathbf{J_{s2}})\mathbf{Q}}-\mu_{(\mathbf{J_{t1}}-\mathbf{J_{t2}})\mathbf{Q}})^{2}q_{kl}.$$

For the curvilinear-linear comparison in post-stratification kernel equating,

$$\mathrm{SEED}^{2}(x)=(1/N_{\boldsymbol{P}})\sum_{\mathrm{jl}}(\mathrm{A}_{(\mathrm{e1\text{-}e2})\mathrm{jl}}-\mu_{\mathrm{A(e1\text{-}e2)}\mathbf{P}})^{2}p_{jl}+(1/N_{\boldsymbol{Q}})\sum_{\mathrm{kl}}(\mathrm{B}_{(\mathrm{e1\text{-}e2})\mathrm{kl}}-\mu_{\mathrm{B(e1\text{-}e2)}\mathbf{Q}})^{2}q_{kl}, \qquad (37)$$

where $\mathrm{A}_{(\mathrm{e1\text{-}e2})\mathrm{jl}}=\left(\mathrm{w}_{\mathrm{l}\bar{\mathbf{P}}}(\mathrm{J}_{\mathrm{r1}_{\mathrm{j}}}-\mathrm{J}_{\mathrm{r2}_{\mathrm{j}}}-\mu_{(\mathbf{J_{r1}}-\mathbf{J_{r2}})\bar{\mathbf{P}}|1})+\mathrm{w}(\mu_{(\mathbf{J_{r1}}-\mathbf{J_{r2}})\bar{\mathbf{P}}|1}+\mu_{(\mathbf{J_{s1}}-\mathbf{J_{s2}})\bar{\mathbf{Q}}|1})\right)\frac{1}{\mathrm{b}_{\mathrm{Pl}}}$,

and $\mathrm{B}_{(\mathrm{e1\text{-}e2})\mathrm{kl}}=\left(\mathrm{w}_{\mathrm{l}\bar{\mathbf{Q}}}(\mathrm{J}_{\mathrm{s1}_{\mathrm{k}}}-\mathrm{J}_{\mathrm{s2}_{\mathrm{k}}}-\mu_{(\mathbf{J_{s1}}-\mathbf{J_{s2}})\bar{\mathbf{Q}}|1})+(1-w)(\mu_{(\mathbf{J_{r1}}-\mathbf{J_{r2}})\bar{\mathbf{P}}|1}+\mu_{(\mathbf{J_{s1}}-\mathbf{J_{s2}})\bar{\mathbf{Q}}|1})\right)\frac{1}{\mathrm{b}_{\mathrm{Ql}}}$.

The raw SEED for assessing differences between chained and post-stratification equating functions is computed as

$$\mathrm{SEED}_{\mathrm{y}}(x)=\sqrt{\mathrm{SEED}_{\mathrm{y}\mathbf{P}}^{2}(x)+\mathrm{SEED}_{\mathrm{y}\mathbf{Q}}^{2}(x)}$$

$$\text{where}\ \mathrm{SEED}_{\mathrm{y}\mathbf{P}}^{2}(x)=\frac{1}{N_{P}}\sum_{\mathrm{jl}}\left(\frac{\partial e1_{Y}(e1_{a}(x))}{\partial a}(\mathrm{J}_{\mathrm{r_j}1,\mathbf{P}}+\mathrm{J}_{\mathrm{t_l}1,\mathbf{P}}-\mu_{\mathbf{J_{r1}P}}-\mu_{\mathbf{J_{t1}P}})-\mathrm{A}_{\mathrm{jl,2P}}+\boldsymbol{\mu}_{\mathrm{A_{2P}}}\right)^{2}p_{jl}, \qquad (38)$$

$\mathrm{SEED}_{\mathrm{y}\mathbf{Q}}^{2}(x)=\frac{1}{N_{Q}}\sum_{\mathrm{kl}}(\mathrm{J}_{\mathrm{s_k}1,\mathbf{Q}}+\mathrm{J}_{\mathrm{t_l}1,\mathbf{Q}}-\mu_{\mathbf{J_{s1}Q}}-\mu_{\mathbf{J_{t1}Q}}-\mathrm{B}_{\mathrm{kl,2P}}+\boldsymbol{\mu}_{\mathrm{A_{2Q}}})^{2}q_{kl}$, e_1 is the chained equating function, and e_2 is the post-stratification equating function.

Discussion

This paper describes how raw data and traditional equipercentile equating functions can be incorporated within von Davier et al.'s (2004b) kernel equating framework. Previous studies of theoretical standard errors of equating (Jarjoura & Kolen, 1985; Liou & Cheng, 1995; Liou et al., 1997; Lord, 1982) addressed many issues discussed in this paper, but in a somewhat disjointed manner. This paper's framework provides some additional structure to allow these previous issues

to be more easily related. It also provides a more general standard error of equating difference for post-stratification and chained equated scores.

This paper reveals many issues of equating methodology that deserve further study. More general forms of the kernel and linear interpolation continuization methods could be explored, such as the utilization of a uniform kernel that is not limited to discrete score intervals and the utilization of the suggested frequency averaging rule and linear interpolation over specifically defined score ranges. A revisiting of previously proposed rules for dealing with scores of zero frequencies in raw post-stratification equating (Jarjoura & Kolen, 1985; Hanson, 1991; Hanson et al., 1994; Kolen & Brennan, 2004; Livingston, 2004) and the implications of these rules on the accuracies of equating functions and variability is also warranted. The empirical evaluation of the described procedures and their relationships is another important issue that is addressed in an accompanying research report (Moses & Holland, 2007).

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

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

Fairbank, B. A. (1987). The use of presmoothing and postsmoothing to increase the precision of equipercentile equating. *Applied Psychological Measurement, 11(3),* 245–262.

Grant, M. C., Zhang, Y., Damiano, M., & Lonstein, L. (2006, April). *An evaluation of the kernel equating method: Small sample equating in non-equivalent groups design.* Paper presented at the annual meeting of the American Educational Research Association Annual Meeting, San Francisco, CA.

Haberman, S. J. (1974). *The analysis of frequency data.* Chicago: University of Chicago Press.

Hanson, B. A. (1991). A comparison of bivariate smoothing methods in common-item equipercentile equating. *Applied Psychological Measurement, 15*(4), 391–408.

Hanson, B. A., Zeng, L., & Colton, D. (1994). *A comparison of presmoothing and postsmoothing methods in equipercentile equating.* (ACT Technical Rep. No. 94-4). Iowa City, IA: ACT.

Holland, P. W., von Davier, A. A., Sinharay, S., & Han, N. (2006). *Testing the untestable assumptions of the chain and poststratification equating methods for the NEAT design* (ETS Research Rep. No. RR-06-17). Princeton, NJ: ETS.

Holland, P. W., & Thayer, D. T. (1987). *Notes on the use of log-linear models for fitting discrete probability distributions* (ETS Research Rep. No. TR-87-31). 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.

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.

Kolen, M. J. (1991). Smoothing methods for estimating test score distributions. *Journal of Educational Measurement, 28*(3)*,* 257–282.

Kolen, M. J., & Brennan, R. J. (2004). *Test equating: Methods and practices.* (2nd ed.)*.* New York: 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*(4)*,* 349–369.

Liu, J., & Low, A. (2006, April). *An exploration of kernel equating using SAT data.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Livingston, S. A. (1993a). *An empirical tryout of kernel equating* (ETS Research Rep. No. RR-93-33). Princeton, NJ: ETS.

Livingston, S. (1993b). Small-sample equatings with log-linear smoothing. *Journal of Educational Measurement, 30,* 23–39.

Livingston, S. A. (2004). *Equating test scores (without IRT).* Princeton, NJ: ETS.

Livingston, S. A., Dorans, N. J., & Wright, N. K. (1990). What combination of sampling and equating methods works best? *Applied Measurement in Education, 3*(1), 73–95.

Lord, F. (1982). The standard error of equipercentile equating. *Journal of Educational Statistics, 7*(3), 165–174.

MacCann, R. G. (1990). Derivations of observed score equating methods that cater to populations differing in difficulty. *Journal of Educational Statistics, 15,* 146–170.

Moses, T., & Holland, P. (2007). *Kernel and traditional equipercentile equating with degrees of presmoothing* (ETS Research Rep. No. RR-07-15). Princeton, NJ: ETS.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: The art of scientific computing* (2nd ed.). New York: Cambridge University Press.

Rosenbaum, P. R., & Thayer, D. (1987). Smoothing the joint and marginal distributions of scored two-way contingency tables in test equating. *British Journal of Mathematical and Statistical Psychology, 40,* 43–49.

Skaggs, G. (2004, April). *Passing score stability when equating with very small samples.* Paper presented at the meeting of the American Educational Research Association, San Diego, CA.

von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004a). The chain and post-stratification methods for observed-score equating: Their relationship to population invariance. *Journal of Educational Measurement, 41,* 15–32.

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

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

Wang, T. (2004). *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, Center for Advanced Studies in Measurement and Assessment.

Wang, T., Lee, W., Brennan, R. L., & Kolen, M. J. (2006, April). *A comparison of the frequency estimation and chained equipercentile methods under the common-item non-equivalent groups design.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Appendix

Deriving the Standard Errors of Equating for Raw Data

This appendix gives the derivations for efficient raw SEE and SEED computations that avoid directly forming *JL* by *JL* and *KL* by *KL* C-matrices. First a basic result is given for the raw C-matrix and the product of anything multiplied by the raw C-matrix. Then the derivations of $(\mathbf{J}_e\mathbf{J}_{DF})\mathbf{C}$ are given for the single-group design (*X*-to-*A* in *P*). The single-group results are then easily applied to chained equating. Then the $(\mathbf{J}_e\mathbf{J}_{DF})\mathbf{C}$ is described for post-stratification. Finally, the SEED equations are given.

The Basic Result

In getting to the SEE-vector for raw data, we first get the row vector $\mathbf{J}_e\mathbf{J}_{DF}$ and then premultiply it into **C** of the form in (8). Let **v** denote a row vector of the appropriate dimension for the multiplication to make sense, then

$$\mathbf{v}\,\mathbf{C} = \mathbf{v}\;(1/\sqrt{N})(\mathbf{D}_{\sqrt{\mathbf{p}}} - \mathbf{p}\sqrt{\mathbf{p}^{\mathbf{t}}}) \;=\; (1/\sqrt{N})(\mathbf{v}\mathbf{D}_{\sqrt{\mathbf{p}}} - \mathbf{v}\mathbf{p}\sqrt{\mathbf{p}^{\mathbf{t}}})\,. \tag{A1}$$

Since **p** is a column vector, **vp** is the mean of the elements of **v** using the probabilities in **p**, so denote this by $\mu_{\mathbf{vP}}$. Furthermore, $\mathbf{vD}_{\sqrt{\mathbf{p}}}$ is a row vector with coordinates $v_j\sqrt{p_j}$. Thus, **vC** is the row vector with elements

$$(1/\sqrt{N})(v_j - \mu_{\mathbf{vP}})\sqrt{p_j}\,. \tag{A2}$$

Finally, the squared length of **vC** will be the sum of squares of the elements in (A2) or 1/*N* times the variance of the elements of the vector **v**, that is,

$$\|\mathbf{v}\,\mathbf{C}\|^2 = (1/N)\sigma^2_{\mathbf{vP}}\,. \tag{A3}$$

As this paper will show, the SEEs for the various designs when there is no presmoothing will always involve one or more quantities of the form (A3).

The Single-Group (SG) Design (*X*-to-*A* in *P*)

Following von Davier et al. (2004b), the array of interest is $\mathbf{P} = (1/N)\mathbf{n}$, the two-way table of bivariate cell proportions. Furthermore, **P** is regarded as vectorized into a column vector denoted as **p**. If **P** is *J* by *L*, then **p** is *JL* by 1. It is now **p** that is assumed to have the multinomial distribution.

Continuing to follow von Davier et al., the vector of raw score proportions for the two tests, **X** and **A**, that is, $(\mathbf{r}^t, \mathbf{t}^t)^t$, is computed as,

$$(\mathbf{r}^t, \mathbf{t}^t)^t = \begin{pmatrix} \mathbf{r} \\ \mathbf{t} \end{pmatrix} = \begin{pmatrix} \mathbf{M} \\ \mathbf{O} \end{pmatrix} \mathbf{p}, \tag{A4}$$

where **M** and **N** are defined by,

$$\mathbf{M} = (\mathbf{I}_J, \ldots \mathbf{I}_J), \tag{A5}$$

$$\mathbf{O} = \begin{pmatrix} \mathbf{1}_J^t & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_J^t & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1}_J^t \end{pmatrix}, \tag{A6}$$

and where $\mathbf{I}_J$ denotes a *J* by *J* identity matrix and $\mathbf{1}_J^t$ denotes a row J-vector of all 1's. **M** is *J* by *JL* and **O** has *L* rows and is *L* by *JL*. Hence, by definition,

$$\mathbf{J}_{DF} = \begin{pmatrix} \mathbf{M} \\ \mathbf{O} \end{pmatrix}, \tag{A7}$$

a $J + L$ by *JL* matrix.

The C-matrix for this case is exactly as before,

$$\mathbf{C} = (1/\sqrt{N})(\mathbf{D}_{\sqrt{\mathbf{p}}} - \mathbf{p}\sqrt{\mathbf{p}^t}).$$

But **p** is now a *JL*-vector, partitioned as,

$$\mathbf{p} = \begin{pmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_L \end{pmatrix},$$

where, $\mathbf{p}_l$, a column J-vector, is the l^{th} column of $\mathbf{P}$.

To compute $\mathbf{J}_{e(x)}\,\mathbf{J}_{DF} = (\mathbf{J_r}, \mathbf{J_t})\,\mathbf{J}_{DF}$, observe that

$$(\mathbf{J_r}, \mathbf{J_t})\,\mathbf{J}_{DF} = (\mathbf{J_r}, \mathbf{J_t})\begin{pmatrix} \mathbf{M} \\ \mathbf{O} \end{pmatrix} = \mathbf{J_r}\,\mathbf{M} + \mathbf{J_t}\,\mathbf{O}.$$

Then,

$$\mathbf{J_r}\,\mathbf{M} = \mathbf{J_r}\,(\mathbf{I}_J, \ldots, \mathbf{I}_J) = (\mathbf{J_r}, \ldots, \mathbf{J_r}), \tag{A8}$$

and

$$\mathbf{J_t}\,\mathbf{O} = \mathbf{J_t}\begin{pmatrix} \mathbf{1}_J^t & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_J^t & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1}_J^t \end{pmatrix} = (\mathrm{J}_{t_1}\,\mathbf{1}_J^t,\ \mathrm{J}_{t_2}\,\mathbf{1}_J^t, \ldots,\ \mathrm{J}_{t_L}\,\mathbf{1}_J^t) \tag{A9}$$

Hence, the $(j, l)^{th}$ element of $\mathbf{J_r}\,\mathbf{M} + \mathbf{J_t}\,\mathbf{O}$ is

$$\mathrm{J}_{r_j} + \mathrm{J}_{t_l}. \tag{A10}$$

Thus, applying (A2), the $(j, l)^{th}$ element of the SEE-vector for the SG design is

$$(1/\sqrt{N})(\mathrm{J}_{r_j} + \mathrm{J}_{t_l} - \mu_{(\mathbf{J_r}+\mathbf{J_t})\mathbf{P}})\sqrt{p_{jl}}\,, \tag{A11}$$

where,

$$\mu_{(\mathbf{J_r}+\mathbf{J_t})\mathbf{P}} = \sum_{j,l} (\mathrm{J}_{r_j} + \mathrm{J}_{t_l})\,p_{jl}\,,$$

so that the SEE for the SG design is given by the square root of

$$\mathrm{SEE}^2(x) = (1/N)\ \sigma^2_{(\mathbf{J_r}+\mathbf{J_t})\mathbf{P}}. \tag{A12}$$

The NEAT Design—Chain Equating

Chain equating (CE) uses the results of two embedded single-group designs. In the NEAT design there are two bivariate arrays of probabilities, **P** and **Q**, where **P** is *J* by *L* and **Q** is *K* by *L*. Again, it is assumed that they have been vectorized into **p** and **q**, respectively, and again the sample and population proportions are mixed. In CE, the quantities that are used are the four probability vectors $\mathbf{r}_P$, $\mathbf{t}_P$, $\mathbf{t}_Q$, $\mathbf{s}_Q$, defined in von Davier et al (2004b). In both this and the next section, j is used for the coordinate subscript for **r**, k for the subscript for **s** and l for the subscript for **t**, that is, r_j, s_k and t_l.

For CE, the equating function, $e_Y(x; \mathbf{r}_P, \mathbf{t}_P, \mathbf{t}_Q, \mathbf{s}_Q)$ is the composition of two simpler functions, i.e.,

$$e_Y(x; \mathbf{r}_P, \mathbf{t}_P, \mathbf{t}_Q, \mathbf{s}_Q) = e_Y(e_a(x; \mathbf{r}_P, \mathbf{t}_P); \mathbf{t}_Q, \mathbf{s}_Q)$$

where $e_A(x; \mathbf{r}_P, \mathbf{t}_P)$ is the single-group link from **X** to **A** and $e_Y(a; \mathbf{t}_Q, \mathbf{s}_Q)$ is the single-group link from **A** to **Y**. Then, as shown in von Davier et al. (p. 84), the $SEE^2(x)$ for CE is the sum of two parts corresponding to the two SG links, that is,

$$\mathrm{SEE}^2(x) = [\frac{\partial e_Y}{\partial a}(e_a(x))]^2 \mathrm{SEE}_{\mathrm{a}}^2(x) + \mathrm{SEE}_{\mathrm{Y}}^2(e_a(x)). \tag{A13}$$

Now apply (A12):

$$\mathrm{SEE}_{\mathrm{a}}^2(x) = (1/N_{\mathbf{P}}) \sum_{\mathrm{j,l}} (\mathrm{J}_{\mathrm{r_j}} + \mathrm{J}_{\mathrm{t_l}} - (\mu_{\mathbf{J_r P}} + \mu_{\mathbf{J_t P}}))^2 p_{jl} = (1/N_{\mathbf{P}}) \sum_{\mathrm{j,l}} (\mathrm{J}_{\mathrm{r_j}} + \mathrm{J}_{\mathrm{t_l}} - \mu_{\mathbf{(J_r+J_t)P}})^2 p_{jl} \tag{A14}$$

and

$$\mathrm{SEE}_{\mathrm{Y}}^2(e_a(x)) = (1/N_{\mathbf{Q}}) \sum_{\mathrm{k,l}} (\mathrm{J}_{\mathrm{s_k}} + \mathrm{J}_{\mathrm{t_l}} - (\mu_{\mathbf{J_s Q}} + \mu_{\mathbf{J_t Q}}))^2 q_{kl}. =$$

$$(1/N_{\mathbf{Q}}) \sum_{\mathrm{k,l}} (\mathrm{J}_{\mathrm{s_k}} + \mathrm{J}_{\mathrm{t_l}} - \mu_{\mathbf{(J_s+J_t)Q}})^2 q_{kl}. \tag{A15}$$

In (A15), the Jacobians and their mean values need to be evaluated at the converted x-scores via $e_a(x; \mathbf{r}_P, \mathbf{t}_P)$. The SEE for CE in the NEAT design is the square root of the quantity in (A13).

The NEAT Design—Post-Stratification Equating

The use of the data from the NEAT design for post-stratification equating assumes that the arrays, **P** and **Q**, have been vectorized into **p** and **q**. The C-matrix for the vector of unsmoothed and unaveraged proportions $(\mathbf{p}^t, \mathbf{q}^t)^t$ is given by

$$\mathbf{C} = \begin{pmatrix} \mathbf{C_P} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_Q} \end{pmatrix}, \tag{A16}$$

where

$$\mathbf{C_P} = (1/\sqrt{N_P})(\mathbf{D}_{\sqrt{\mathbf{p}}} - \mathbf{p}\sqrt{\mathbf{p}^t})$$

and

$$\mathbf{C_Q} = (1/\sqrt{N_Q})(\mathbf{D}_{\sqrt{\mathbf{q}}} - \mathbf{q}\sqrt{\mathbf{q}^t}).$$

Furthermore, the Jacobian of the design function for post-stratification equating in the NEAT design is given by

$$\mathbf{J}_{\mathrm{DF}} = \begin{pmatrix} \dfrac{\partial \mathbf{r}}{\partial \mathbf{p}} & \dfrac{\partial \mathbf{r}}{\partial \mathbf{q}} \\ \dfrac{\partial \mathbf{s}}{\partial \mathbf{p}} & \dfrac{\partial \mathbf{s}}{\partial \mathbf{q}} \end{pmatrix}, \tag{A17}$$

and, as before,

$$\mathbf{J}_e = (\mathbf{J}_r, \mathbf{J}_s).$$

Hence,

$$\mathbf{J}_e \mathbf{J}_{\mathrm{DF}} = (\mathbf{A}, \mathbf{B}), \tag{A18}$$

where

$$\mathbf{A} = \mathbf{J}_r \frac{\partial \mathbf{r}}{\partial \mathbf{p}} + \mathbf{J}_s \frac{\partial \mathbf{s}}{\partial \mathbf{p}}, \tag{A19}$$

$$\mathbf{B} = \mathbf{J}_{\mathbf{r}} \frac{\partial \mathbf{r}}{\partial \mathbf{q}} + \mathbf{J}_{\mathbf{s}} \frac{\partial \mathbf{s}}{\partial \mathbf{q}}. \tag{A20}$$

Using (A16) and (A18),

$$\text{SEE-vector} = (\mathbf{A}, \mathbf{B}) \begin{pmatrix} \mathbf{C_P} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_Q} \end{pmatrix} = (\mathbf{A}\,\mathbf{C_P}, \mathbf{B}\,\mathbf{C_Q}). \tag{A21}$$

From (A21), the squared length of the SEE-vector in this case will be $||\mathbf{AC_P}||^2 + ||\mathbf{BC_Q}||^2$, and from (A3) each of these components will be $1/N$ times the variance of the components of **A** or **B**. A notation can be used that is like the notation used earlier:

$$\mathrm{EE}^2(x) = (1/N_P)\,\sigma^2_{\mathbf{A}\,\mathbf{P}} + (1/N_Q)\,\sigma^2_{\mathbf{B}\,\mathbf{Q}}, \tag{A22}$$

where $\sigma^2_{\mathbf{A}\,\mathbf{P}}$ denotes the variance of the elements of **A** with respect to the bivariate distribution **P**, and $\sigma^2_{\mathbf{B}\,\mathbf{Q}}$ denotes the variance of the elements of **B** with respect to the bivariate distribution **Q**.

Hence, all that remains is to work out what the elements of **A** and **B** are. This is more complicated than the previous cases because of the form of the Jacobian of the design function for post-stratification equating in the NEAT design.

To proceed further it may be useful to get the dimensions of the arrays in (A18) to (A20) clear. **A** is 1 by *JL*, and **B** is 1 by *KL*, so they are the same shapes as $\mathbf{p}^t$ and $\mathbf{q}^t$, respectively. More importantly, **A** and **B** may be partitioned *L* blocks in the same was that $\mathbf{p}^t$ and $\mathbf{q}^t$ are, that is

$$\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_l, \ldots, \mathbf{A}_L),$$

$$\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_l, \ldots, \mathbf{B}_L).$$

Next, the expressions for the elements of the partitions for **A** and **B** are obtained. These are:

$$\mathbf{A}_l = \mathbf{J}_{\mathbf{r}} \frac{\partial \mathbf{r}}{\partial \mathbf{p}_l} + \mathbf{J}_{\mathbf{s}} \frac{\partial \mathbf{s}}{\partial \mathbf{p}_l} = \mathbf{J}_{\mathbf{r}} \frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_l} \frac{\partial \overline{\mathbf{p}}_l}{\partial \mathbf{p}_l} + \mathbf{J}_{\mathbf{s}} \frac{\partial \mathbf{s}}{\partial \overline{\mathbf{p}}_l} \frac{\partial \overline{\mathbf{p}}_l}{\partial \mathbf{p}_l}$$

$$= \left(\mathbf{J}_{\mathbf{r}} \frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} + \mathbf{J}_{\mathbf{s}} \frac{\partial \mathbf{s}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} \right) \frac{\partial \overline{\mathbf{p}}_{\mathrm{l}}}{\partial \mathbf{p}_{\mathrm{l}}}, \tag{A23}$$

$$\mathbf{B}_{\mathrm{l}} = \mathbf{J}_{\mathbf{r}} \frac{\partial \mathbf{r}}{\partial \mathbf{q}_{\mathrm{l}}} + \mathbf{J}_{\mathbf{s}} \frac{\partial \mathbf{s}}{\partial \mathbf{q}_{\mathrm{l}}} = \mathbf{J}_{\mathbf{r}} \frac{\partial \mathbf{r}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} \frac{\partial \overline{\mathbf{q}}_{\mathrm{l}}}{\partial \mathbf{q}_{\mathrm{l}}} + \mathbf{J}_{\mathbf{s}} \frac{\partial \mathbf{s}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} \frac{\partial \overline{\mathbf{q}}_{\mathrm{l}}}{\partial \mathbf{q}_{\mathrm{l}}}$$

$$= \left(\mathbf{J}_{\mathbf{r}} \frac{\partial \mathbf{r}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} + \mathbf{J}_{\mathbf{s}} \frac{\partial \mathbf{s}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} \right) \frac{\partial \overline{\mathbf{q}}_{\mathrm{l}}}{\partial \mathbf{q}_{\mathrm{l}}}. \tag{A24}$$

In (A23) and (A24), the bivariate probability vectors with averaged probabilities ($\overline{\mathbf{p}}$ and $\overline{\mathbf{q}}$) are distinguished from the original bivariate probability vectors (**p** and **q**). This distinction is important, because the raw post-stratification equating function, synthetic marginal probabilities (**r** and **s**), and algebraic development of $\mathbf{J}_{\mathrm{e}}\,\mathbf{J}_{\mathrm{DF}}$ that follows are based on $\overline{\mathbf{p}}$ and $\overline{\mathbf{q}}$, while the final computation of $(\mathbf{J}_{\mathbf{e}}\mathbf{J}_{\mathbf{DF}})\mathbf{C}$ utilizes a **C** that is based on **p** and **q**. Additional multiplications of $\mathbf{J}_{\mathrm{e}}\mathbf{J}_{\mathrm{DF}}$ are needed for derivatives that are in terms of the original bivariate probabilities. They are of the form $\frac{\partial \overline{\mathbf{p}}_{\mathrm{l}}}{\partial \mathbf{p}_{\mathrm{l}}} = \partial \left(\left(\mathbf{p}_{\mathrm{l}} + \mathbf{p}_{\mathrm{l+1}} \ldots . + \right) \frac{1}{\mathrm{b}_{\mathrm{Pl}}} \right) = \frac{1}{\mathrm{b}_{\mathrm{Pl}}}$, where b_{Pl} is the number of anchor score levels with marginal probabilities that are averaged together.

von Davier et al. (2004b, p. 76) provide formulas for the matrices, $\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}}$, and so on, that can be exploited for the $\mathbf{J}_{\mathrm{e}}\,\mathbf{J}_{\mathrm{DF}}$ of post-stratification. These are:

$$\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} = \mathrm{w}_{\mathrm{l}\overline{\mathbf{P}}} \mathbf{I}_{\mathrm{J}} - (1-w)(t_{\bar{Q}l}/t_{\bar{P}l})[(t_{\bar{P}l})^{-1} \overline{\mathbf{p}}_{\mathrm{l}}] \mathbf{1}_{\mathrm{J}}^{t}, \tag{A25}$$

$$\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} = w[(t_{\bar{Q}l})^{-1} \overline{\mathbf{q}}_{\mathrm{l}}] \mathbf{1}_{\mathrm{J}}^{t}, \tag{A26}$$

$$\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} = (1-w)[(t_{\bar{P}l})^{-1} \overline{\mathbf{p}}_{\mathrm{l}}] \mathbf{1}_{\mathrm{K}}^{t}, \tag{A27}$$

$$\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} = \mathrm{w}_{\mathrm{l}\overline{\mathbf{Q}}}\mathbf{I}_{\mathrm{K}} - w(t_{\overline{P}l}/t_{\overline{Q}l})[(t_{\overline{Q}l})^{-1}\overline{\mathbf{q}}_{\mathrm{l}}]\mathbf{1}_{\mathrm{K}}^{t}. \tag{A28}$$

In (A25) to (A28), the quantities $\mathrm{w}_{\mathrm{l}\overline{\mathbf{P}}}$ and $\mathrm{w}_{\mathrm{l}\overline{\mathbf{Q}}}$ are defined as

$$\mathrm{w}_{\mathrm{l}\overline{\mathbf{P}}} = w + (1-w)t_{\overline{Q}l}/t_{\overline{P}l} = (wt_{\overline{P}l} + (1-w)t_{\overline{Q}l})/t_{\overline{P}l} = t_{\overline{T}l}/t_{\overline{P}l}, \tag{A29}$$

and $$\mathrm{w}_{\mathrm{l}\overline{\mathbf{Q}}} = (1-w) + w(t_{\overline{P}l}/t_{\overline{Q}l}) = t_{\overline{T}l}/t_{\overline{Q}l}. \tag{A30}$$

Next,

$$\mathbf{J}_{\mathbf{r}}\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} = \mathrm{w}_{\mathrm{l}\overline{\mathbf{P}}}\mathbf{J}_{\mathbf{r}} - (1-w)(t_{\overline{Q}l}/t_{\overline{P}l})[(t_{\overline{P}l})^{-1}\mathbf{J}_{\mathbf{r}}\overline{\mathbf{p}}_{\mathrm{l}}]\mathbf{1}_{\mathrm{J}}^{t},$$

$$\mathbf{J}_{\mathbf{s}}\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} = w[(t_{\overline{Q}l})^{-1}\mathbf{J}_{\mathbf{s}}\overline{\mathbf{q}}_{\mathrm{l}}]\mathbf{1}_{\mathrm{J}}^{t},$$

and

$$\mathbf{J}_{\mathbf{r}}\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} = (1-w)[(t_{\overline{P}l})^{-1}\mathbf{J}_{\mathbf{r}}\overline{\mathbf{p}}_{\mathrm{l}}]\mathbf{1}_{\mathrm{K}}^{t},$$

$$\mathbf{J}_{\mathbf{s}}\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} = \mathrm{w}_{\mathrm{l}\overline{\mathbf{Q}}}\mathbf{J}_{\mathbf{s}} - w(t_{\overline{P}l}/t_{\overline{Q}l})[(t_{\overline{Q}l})^{-1}\mathbf{J}_{\mathbf{s}}\overline{\mathbf{q}}_{\mathrm{l}}]\mathbf{1}_{\mathrm{K}}^{t}.$$

Quantities of the form $(t_{\overline{P}l})^{-1}\mathbf{J}_{\mathbf{r}}\overline{\mathbf{p}}_{\mathrm{l}}$ occur in each of the above expressions so it is useful to have a separate notation for them and in fact they are all conditional means. They are defined as

$$\mu_{\mathbf{J}_{\mathbf{r}}\overline{\mathbf{P}}|\mathbf{l}} = \sum_{\mathrm{j}} \mathrm{J}_{\mathrm{r_j}}(\overline{p}_{jl}/t_{\overline{P}l}) = (t_{\overline{P}l})^{-1}\mathbf{J}_{\mathbf{r}}\overline{\mathbf{p}}_{\mathrm{l}}, \tag{A31}$$

$$\mu_{\mathbf{J}_{\mathbf{s}}\overline{\mathbf{Q}}|\mathbf{l}} = \sum_{\mathrm{k}} \mathrm{J}_{\mathrm{s_k}}(\overline{q}_{kl}/t_{\overline{Q}l}) = (t_{\overline{Q}l})^{-1}\mathbf{J}_{\mathbf{s}}\overline{\mathbf{q}}_{\mathrm{l}}. \tag{A32}$$

Because $\{\overline{p}_{jl}/t_{\overline{P}l}\}$ is the conditional probability distribution of **X** given **A** in $\overline{\mathbf{P}}$, it is clear that $\mu_{\mathbf{J}_{\mathbf{r}}\overline{\mathbf{P}}|\mathbf{l}}$ is the conditional mean of $\mathbf{J}_{\mathbf{r}}$ given **A** over $\overline{\mathbf{P}}$. A similar interpretation holds for $\mu_{\mathbf{J}_{\mathbf{s}}\overline{\mathbf{Q}}|\mathbf{l}}$, thus:

$$\mathbf{J}_{\mathbf{r}}\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} = \mathrm{w}_{\mathrm{l}\bar{\mathbf{P}}}\mathbf{J}_{\mathbf{r}} - (1-w)(t_{\bar{Q}l}/t_{\bar{P}l})\mu_{\mathbf{J}_{\mathbf{r}}\bar{\mathbf{P}}|\mathrm{l}}\mathbf{1}_{\mathrm{J}}^{t}, \tag{A33}$$

$$\mathbf{J}_{\mathbf{s}}\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} = w\mu_{\mathbf{J}_{\mathbf{s}}\bar{\mathbf{Q}}|\mathrm{l}}\mathbf{1}_{\mathrm{J}}^{t}, \tag{A34}$$

$$\mathbf{J}_{\mathbf{r}}\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} = (1-w)\mu_{\mathbf{J}_{\mathbf{r}}\bar{\mathbf{P}}|\mathrm{l}}\mathbf{1}_{\mathrm{K}}^{t}, \tag{A35}$$

$$\mathbf{J}_{\mathbf{s}}\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} = \mathrm{w}_{\mathrm{l}\bar{\mathbf{Q}}}\mathbf{J}_{\mathbf{s}} - w(t_{\bar{P}l}/t_{\bar{Q}l})\mu_{\mathbf{J}_{\mathbf{s}}\bar{\mathbf{Q}}|\mathrm{l}}\mathbf{1}_{\mathrm{K}}^{t}. \tag{A36}$$

Hence, $\mathbf{A}_{\mathrm{l}}$ and $\mathbf{B}_{\mathrm{l}}$ in (A23) and (A24) can be examined in more detail:

$$\mathbf{A}_{\mathrm{l}} = \left(\mathbf{J}_{\mathbf{r}}\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} + \mathbf{J}_{\mathbf{s}}\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}}\right)\frac{\partial \overline{\mathbf{p}}_{\mathrm{l}}}{\partial \mathbf{p}_{\mathrm{l}}} = \left(\mathrm{w}_{\mathrm{l}\bar{\mathbf{P}}}\mathbf{J}_{\mathbf{r}} - (1-w)(t_{\bar{Q}l}/t_{\bar{P}l})\mu_{\mathbf{J}_{\mathbf{r}}\bar{\mathbf{P}}|\mathrm{l}}\mathbf{1}_{\mathrm{J}}^{t} + w\mu_{\mathbf{J}_{\mathbf{s}}\bar{\mathbf{Q}}|\mathrm{l}}\mathbf{1}_{\mathrm{J}}^{t}\right)\frac{1}{\mathrm{b}_{\mathrm{Pl}}}, \tag{A37}$$

and

$$\mathbf{B}_{\mathrm{l}} = \left(\mathbf{J}_{\mathbf{r}}\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}} + \mathbf{J}_{\mathbf{s}}\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{q}}_{\mathrm{l}}}\right)\frac{\partial \overline{\mathbf{q}}_{\mathrm{l}}}{\partial \mathbf{q}_{\mathrm{l}}} = \left((1-w)\mu_{\mathbf{J}_{\mathbf{r}}\bar{\mathbf{P}}|\mathrm{l}}\mathbf{1}_{\mathrm{K}}^{t} + \mathrm{w}_{\mathrm{l}\bar{\mathbf{Q}}}\mathbf{J}_{\mathbf{s}} - w(t_{\bar{P}l}/t_{\bar{Q}l})\mu_{\mathbf{J}_{\mathbf{s}}\bar{\mathbf{Q}}|\mathrm{l}}\mathbf{1}_{\mathrm{K}}^{t}\right)\frac{1}{\mathrm{b}_{\mathrm{Ql}}}. \tag{A38}$$

Equations (A29) and (A30) give the following relationships,

$$(1-w)t_{\bar{Q}l}/t_{\bar{P}l} = \mathrm{w}_{\mathrm{l}\bar{\mathbf{P}}} - w, \text{ and } w(\mathrm{t}_{\bar{\mathbf{P}}\mathrm{l}}/\mathrm{t}_{\bar{\mathbf{Q}}\mathrm{l}}) = \mathrm{w}_{\mathrm{l}\bar{\mathbf{Q}}} - (1-w). \tag{A39}$$

Equations (A37) and (A38) are simplified using (A39) as follows:

$$\mathbf{A}_{\mathrm{l}} = \left(\mathbf{J}_{\mathbf{r}}\frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}} + \mathbf{J}_{\mathbf{s}}\frac{\partial \mathbf{s}}{\partial \overline{\mathbf{p}}_{\mathrm{l}}}\right)\frac{\partial \overline{\mathbf{p}}_{\mathrm{l}}}{\partial \mathbf{p}_{\mathrm{l}}} = \left(\mathrm{w}_{\mathrm{l}\bar{\mathbf{P}}}\mathbf{J}_{\mathbf{r}} - (\mathrm{w}_{\mathrm{l}\bar{\mathbf{P}}} - w)\mu_{\mathbf{J}_{\mathbf{r}}\bar{\mathbf{P}}|\mathrm{l}}\mathbf{1}_{\mathrm{J}}^{t} + w\mu_{\mathbf{J}_{\mathbf{s}}\bar{\mathbf{Q}}|\mathrm{l}}\mathbf{1}_{\mathrm{J}}^{t}\right)\frac{1}{\mathrm{b}_{\mathrm{Pl}}} \tag{A40}$$

and

$$\mathbf{B}_l = \left(\mathbf{J}_\mathbf{r} \frac{\partial \mathbf{r}}{\partial \overline{\mathbf{q}}_l} + \mathbf{J}_\mathbf{s} \frac{\partial \mathbf{s}}{\partial \overline{\mathbf{q}}_l} \right) \frac{\partial \overline{\mathbf{q}}_l}{\partial \mathbf{q}_l} = \left((1-w)\mu_{\mathbf{J}_\mathbf{r}\bar{\mathbf{P}}|l} \mathbf{1}_K^t + \mathrm{w}_{l\bar{\mathbf{Q}}} \mathbf{J}_\mathbf{s} - (\mathrm{w}_{l\bar{\mathbf{Q}}} - (1-w))\mu_{\mathbf{J}_\mathbf{s}\bar{\mathbf{Q}}|l} \mathbf{1}_K^t \right) \left(\frac{1}{\mathrm{b}_{Ql}} \right). \quad \text{(A41)}$$

Hence,

$$\mathbf{A}_l = \left(\mathbf{J}_\mathbf{r} \frac{\partial \mathbf{r}}{\partial \overline{\mathbf{p}}_l} + \mathbf{J}_\mathbf{s} \frac{\partial \mathbf{s}}{\partial \overline{\mathbf{p}}_l} \right) \frac{\partial \overline{\mathbf{p}}_l}{\partial \mathbf{p}_l} = \left(\mathrm{w}_{l\bar{\mathbf{P}}} (\mathbf{J}_\mathbf{r} - \mu_{\mathbf{J}_\mathbf{r}\bar{\mathbf{P}}|l} \mathbf{1}_J^t) + w(\mu_{\mathbf{J}_\mathbf{r}\bar{\mathbf{P}}|l} + \mu_{\mathbf{J}_\mathbf{s}\bar{\mathbf{Q}}|l}) \mathbf{1}_J^t \right) \frac{1}{\mathrm{b}_{Pl}}, \quad \text{(A42)}$$

and

$$\mathbf{B}_l = \left(\mathbf{J}_\mathbf{r} \frac{\partial \mathbf{r}}{\partial \overline{\mathbf{q}}_l} + \mathbf{J}_\mathbf{s} \frac{\partial \mathbf{s}}{\partial \overline{\mathbf{q}}_l} \right) \frac{\partial \overline{\mathbf{q}}_l}{\partial \mathbf{q}_l} = \left(\mathrm{w}_{l\bar{\mathbf{Q}}} (\mathbf{J}_\mathbf{s} - \mu_{\mathbf{J}_\mathbf{s}\bar{\mathbf{Q}}|l} \mathbf{1}_K^t) + (1-w)(\mu_{\mathbf{J}_\mathbf{r}\bar{\mathbf{P}}|l} + \mu_{\mathbf{J}_\mathbf{s}\bar{\mathbf{Q}}|l}) \mathbf{1}_K^t \right) \frac{1}{\mathrm{b}_{Ql}}. \quad \text{(A43)}$$

All that remains is to give the individual coordinates of the vector partitions, $\mathbf{A}_l$ and $\mathbf{B}_l$. These are:

$$\mathrm{A}_{jl} = \left(\mathrm{w}_{l\bar{\mathbf{P}}} (\mathrm{J}_{r_j} - \mu_{\mathbf{J}_\mathbf{r}\bar{\mathbf{P}}|l}) + w(\mu_{\mathbf{J}_\mathbf{r}\bar{\mathbf{P}}|l} + \mu_{\mathbf{J}_\mathbf{s}\bar{\mathbf{Q}}|l}) \right) \frac{1}{\mathrm{b}_{Pl}}, \quad \text{(A44)}$$

$$\text{and } \mathrm{B}_{kl} = \left(\mathrm{w}_{l\bar{\mathbf{Q}}} (\mathrm{J}_{s_k} - \mu_{\mathbf{J}_\mathbf{s}\bar{\mathbf{Q}}|l}) + (1-w)(\mu_{\mathbf{J}_\mathbf{r}\bar{\mathbf{P}}|l} + \mu_{\mathbf{J}_\mathbf{s}\bar{\mathbf{Q}}|l}) \right) \frac{1}{\mathrm{b}_{Ql}}. \quad \text{(A45)}$$

Finally, we use (A44) and (A45) to obtain the two variances in (A22) to get the SEE for the post-stratification method of equating in the NEAT design.

The NEAT Design—Standard Errors of Equating Differences (SEEDs)

Curvilinear-Linear SEEDs

For kernel equating functions, curvilinear and linear equating functions for a given method (either post-stratification or chained) can be computed by varying the continuization bandwidths (*h's*). The derivative of the (e_1 - e_2) difference is simply the difference between the equating functions' derivative vectors, $\partial(e_1 - e_2) = \mathbf{J}_{\mathbf{e1}} - \mathbf{J}_{\mathbf{e2}}$, so that,

$$\mathrm{SEED}_y(x) = \left\| (\mathbf{J}_{\mathbf{e1}} - \mathbf{J}_{\mathbf{e2}}) \mathbf{J}_{\mathbf{DF}} \mathbf{C} \right\| = \left\| \mathbf{J}_{\mathbf{e1-e2}} \mathbf{J}_{\mathbf{DF}} \mathbf{C} \right\| \quad \text{(A46)}$$

(von Davier et al., 2004b, p. 80).

The implication of (A46) is that raw curvilinear-linear SEEDs can be computed by utilizing all of the raw SEE results with the substitution of $\mathbf{J_{e1} - J_{e2}}$ for $\mathbf{J_e}$.

The SEED analogues of (A1) and (A2) are:

$$\text{SEED - Vector}_y(x) = \frac{1}{\sqrt{N}}\left(\mathbf{(J_{e1} - J_{e2})J_{DF}D_{\sqrt{p}}} - \mathbf{(J_{e1} - J_{e2})J_{DF}p\sqrt{p^t}}\right), \text{ and} \tag{A47}$$

$$\text{SEED}_y(x) = \sqrt{\frac{1}{N}\sum_j \left(\mathbf{(J_{e1j} - J_{e2j})J_{DF}} - \mu_{(J_{e1}-J_{e2})J_{DF}}\right)^2 p_j}\,. \tag{A48}$$

Equation (A48) makes it clear that the raw SEED is the standard deviation of the difference between equating functions' derivatives for a given design and data set.

To compute the raw SEEDs for kernel post-stratification, simply use $\mathbf{J_{r1} - J_{r2}}$ and $\mathbf{J_{s1} - J_{s2}}$ in place of $\mathbf{J_r}$ and $\mathbf{J_s}$ in (A22), (A31-A45). The curvilinear-linear post-stratification SEED elements are:

$$\text{A}_{(e1\text{-}e2)jl} = \left(\text{w}_{l|\bar{\mathbf{P}}}(\text{J}_{r1_j} - \text{J}_{r2_j} - \mu_{(\mathbf{J_{r1}-J_{r2}})\bar{\mathbf{P}}|l}) + w(\mu_{(\mathbf{J_{r1}-J_{r2}})\bar{\mathbf{P}}|l} + \mu_{(\mathbf{J_{s1}-J_{s2}})\bar{\mathbf{Q}}|l})\right)\frac{1}{\text{b}_{Pl}}, \tag{A49}$$

$$\text{and } \text{B}_{(e1\text{-}e2)kl} = \left(\text{w}_{l|\bar{\mathbf{Q}}}(\text{J}_{s1_k} - \text{J}_{s2_k} - \mu_{(\mathbf{J_{s1}-J_{s2}})\bar{\mathbf{Q}}|l}) + (1-w)(\mu_{(\mathbf{J_{r1}-J_{r2}})\bar{\mathbf{P}}|l} + \mu_{(\mathbf{J_{s1}-J_{s2}})\bar{\mathbf{Q}}|l})\right)\frac{1}{\text{b}_{Ql}}. \tag{A50}$$

The raw SEED for chained kernel requires a little more than the simple substitution of $J_{e1} - J_{e2}$ for J_e. There are two $\frac{\partial e_Y}{\partial a}(e_a(x))$'s that are needed, and these must be incorporated into specific parts of the $J_{e1} - J_{e2}$ computations:

$$\text{SEED}^2(x) = (1/N_P)\text{SEED}_a^2(x) + (1/N_Q)\text{SEED}_Y^2(e_a(x)), \tag{A51}$$

where $\text{SEED}_a^2(x) =$

$$\sum_{j,l}\left(\frac{\partial e1_y(e1_a(x))}{\partial a}(\text{J}_{r1_j} + \text{J}_{t1_l}) - \frac{\partial e2_y(e2_a(x))}{\partial a}(\text{J}_{r2_j} + \text{J}_{t2_l}) - \mu_{(\frac{\partial \mathbf{e1}}{\partial \mathbf{a}}\mathbf{J_{r1}} - \frac{\partial \mathbf{e2}}{\partial \mathbf{a}}\mathbf{J_{r2}})\mathbf{P}} - \mu_{(\frac{\partial \mathbf{e1}}{\partial \mathbf{a}}\mathbf{J_{t1}} - \frac{\partial \mathbf{e2}}{\partial \mathbf{a}}\mathbf{J_{t2}})\mathbf{P}}\right)^2 p_{jl} \tag{A52}$$

$$\text{SEED}_Y^2(e_a(x)) = \sum_{k,l}\left(\text{J}_{s1_k} - \text{J}_{s2_k} + \text{J}_{t1_l} - \text{J}_{t2_l} - \mu_{(\mathbf{J_{s1}-J_{s2}})\mathbf{Q}} - \mu_{(\mathbf{J_{t1}-J_{t2}})\mathbf{Q}}\right)^2 q_{kl}. \tag{A53}$$

Chained-Post-Stratification SEEDs

The SEED vector can also be expressed more generally as

$$\mathrm{SEED}_y(x) = \left\| (\mathbf{J}_{e1}\mathbf{J}_{DF1} - \mathbf{J}_{e2}\mathbf{J}_{DF2})\mathbf{C} \right\|, \tag{A54}$$

Equation (A54) can be used to estimate the variability of the difference between equating functions based on different design functions, such as two post-stratification equating functions where *w* is varied, or between chained (*e1*) and post-stratification (*e2*) equating functions (von Davier et al., 2004b p. 81).

Based on the definition of C in (3), the raw SEED in (A54) can be computed as follows:

$$\mathrm{SEED}_y(x) = \left\| \left\| (\mathbf{J}_{e1P}\mathbf{J}_{DF1P} - \mathbf{J}_{e2P}\mathbf{J}_{DF2P})\mathbf{C}_P \right\|^2 + \left\| (\mathbf{J}_{e1Q}\mathbf{J}_{DF1Q} - \mathbf{J}_{e2Q}\mathbf{J}_{DF2Q})\mathbf{C}_Q \right\|^2 \right\|. \tag{A55}$$

From (A55), there are two major parts to the SEED:

$$\mathrm{SEED}_y(x) = \sqrt{\mathrm{SEED}^2_{yP}(x) + \mathrm{SEED}^2_{yQ}(x)}, \tag{A56}$$

which are now separately simplified.

From (A1),

$$\mathrm{SEED}^2_{yP}(x) = \frac{1}{N_P} \left\| (\mathbf{J}_{e1P}\mathbf{J}_{DF1P} - \mathbf{J}_{e2P}\mathbf{J}_{DF2P})\mathbf{D}_{\sqrt{\mathbf{p}}} - (\mathbf{J}_{e1P}\mathbf{J}_{DF1P} - \mathbf{J}_{e2P}\mathbf{J}_{DF2P})\mathbf{p}\sqrt{\mathbf{p}^t} \right\|^2. \tag{A57}$$

$$\mathrm{SEED}^2_{yQ}(x) = \frac{1}{N_Q} \left\| (\mathbf{J}_{e1Q}\mathbf{J}_{DF1Q} - \mathbf{J}_{e2Q}\mathbf{J}_{DF2Q})\mathbf{D}_{\sqrt{\mathbf{q}}} - (\mathbf{J}_{e1Q}\mathbf{J}_{DF1Q} - \mathbf{J}_{e2Q}\mathbf{J}_{DF2Q})\mathbf{q}\sqrt{\mathbf{q}^t} \right\|^2. \tag{A58}$$

Factoring out all $\sqrt{\mathbf{p}^t}$'s and $\sqrt{\mathbf{q}^t}$'s and noting that products of row and column vectors are sums,

$$\mathrm{SEED}^2_{yP}(x) = \frac{1}{N_P} \left(\mathbf{J}_{e1P}\mathbf{J}_{DF1P} - \mathbf{J}_{e2P}\mathbf{J}_{DF2P} - (\mathbf{J}_{e1P}\mathbf{J}_{DF1P} - \mathbf{J}_{e2P}\mathbf{J}_{DF2P})\mathbf{p} \right)^2 \mathbf{p}. \tag{A59}$$

$$\mathrm{SEED}^2_{yQ}(x) = \frac{1}{N_Q} \left(\mathbf{J}_{e1Q}\mathbf{J}_{DF1Q} - \mathbf{J}_{e2Q}\mathbf{J}_{DF2Q} - (\mathbf{J}_{e1Q}\mathbf{J}_{DF1Q} - \mathbf{J}_{e2Q}\mathbf{J}_{DF2Q})\mathbf{q} \right)^2 \mathbf{q}. \tag{A60}$$

From (A2), the elements multiplied by $\mathbf{p}$ and $\mathbf{q}$ vectors are means:

$$\mathrm{SEED}^2_{\mathbf{yP}}(x) = \frac{1}{N_P}\left(\mathbf{J_{e1P}J_{DF1P}} - \mathbf{J_{e2P}J_{DF2P}} - (\boldsymbol{\mu}_{J_{e1\mathbf{P}}J_{DF1\mathbf{P}}} - \boldsymbol{\mu}_{J_{e2\mathbf{P}}J_{DF2\mathbf{P}}})\right)^2 \mathbf{p} \quad .(A61)$$

$$\mathrm{SEED}^2_{\mathbf{yQ}}(x) = \frac{1}{N_Q}\left(\mathbf{J_{e1Q}J_{DF1Q}} - \mathbf{J_{e2Q}J_{DF2Q}} - (\boldsymbol{\mu}_{J_{e1\mathbf{Q}}J_{DF1\mathbf{Q}}} - \boldsymbol{\mu}_{J_{e2\mathbf{Q}}J_{DF2\mathbf{Q}}})\right)^2 \mathbf{q} \,. \quad (A62)$$

From (25), (26), (A13), (A14), (A15), (A44), and (A45),

$$\mathrm{SEED}^2_{\mathbf{yP}}(x) = \frac{1}{N_P}\left(\frac{\partial el_Y(el_a(x))}{\partial a}(\mathrm{J}_{\mathrm{r1,P}} + \mathrm{J}_{\mathrm{t1,P}} - \mu_{\mathbf{J_{r1}P}} - \mu_{\mathbf{J_{t1}P}}) - \mathrm{A}_{\mathbf{2P}} + \boldsymbol{\mu}_{\mathrm{A}_{2\mathbf{P}}}\right)^2 \mathbf{p} \,. \quad (A63)$$

$$\mathrm{SEED}^2_{\mathbf{yQ}}(x) = \frac{1}{N_Q}\left(\mathrm{J}_{\mathrm{s1,Q}} + \mathrm{J}_{\mathrm{t1,Q}} - \mu_{\mathbf{J_{s1}Q}} - \mu_{\mathbf{J_{t1}Q}} - \mathrm{B}_{\mathbf{2P}} + \boldsymbol{\mu}_{\mathrm{A}_{2\mathbf{Q}}}\right)^2 \mathbf{q} \,. \quad (A64)$$

Then the elements of (A56) can be computed as

$$\mathrm{SEED}^2_{\mathbf{yP}}(x) = \frac{1}{N_P}\sum_{\mathrm{jl}}\left(\frac{\partial el_Y(el_a(x))}{\partial a}(\mathrm{J}_{\mathrm{r_j1,P}} + \mathrm{J}_{\mathrm{t_l1,P}} - \mu_{\mathbf{J_{r1}P}} - \mu_{\mathbf{J_{t1}P}}) - \mathrm{A}_{\mathrm{jl,2P}} + \boldsymbol{\mu}_{\mathrm{A}_{2\mathbf{P}}}\right)^2 p_{jl} \,. \quad (A65)$$

$$\mathrm{SEED}^2_{\mathbf{yQ}}(x) = \frac{1}{N_Q}\sum_{\mathrm{kl}}\left(\mathrm{J}_{\mathrm{s_k1,Q}} + \mathrm{J}_{\mathrm{t_l1,Q}} - \mu_{\mathbf{J_{s1}Q}} - \mu_{\mathbf{J_{t1}Q}} - \mathrm{B}_{\mathrm{kl,2P}} + \boldsymbol{\mu}_{\mathrm{A}_{2\mathbf{Q}}}\right)^2 q_{kl} \,. \quad (A66)$$