Standard Errors of Levine Linear Equating

Bradley A. Hanson, Lingjia Zeng, and Michael J. Kolen

American College Testing

The delta method was used to derive standard errors (SEs) of the Levine observed score and Levine true score linear equating methods. SEs with a normality assumption as well as without a normality assumption were derived. Data from two forms of a test were used as an example to evaluate the derived SEs of equating. Bootstrap SEs also were computed for the purpose of comparison. The SEs derived without the normality assumption and the bootstrap SEs were very close. For the skewed score distributions, the SEs derived with the normality assumption differed from the SEs derived without the normality assumption and the bootstrap SEs. *Index terms: equating, delta method, linear equating, score equating, standard errors of equating.*

Test form equating requires that a design for data collection be specified and used. In the common-item-nonequivalent-groups (CING) design, one group of examinees is administered a new test form and another group of examinees is administered an old, previously equated test form. The two examinee groups are considered to be drawn from different populations. For example, one group might be administered the new form on one test date and the other group administered the old form on another test date. In this design, the alternate test forms have a subset of items in common. These common items, which are used in the equating process to identify differences between the two groups, are selected to be representative of the total test in content and statistical specifications. If examinee scores on the common items contribute to scores on the total test, then the common items are referred to as internal; otherwise, they are referred to as external.

Test form equating also requires that statistical methodology be specified and used. In linear equating, the relationship of the old form scores to the new form scores is described by a linear function. When using the CING design, it is also necessary to be able to estimate group differences on the items not common between the two forms. Strong statistical assumptions typically are required for this purpose (Petersen, Kolen, & Hoover, 1989). Levine (1955) developed two commonly used linear equating methods that are often referred to as the Levine equally reliable method and the Levine unequally reliable method (Petersen et al., 1989). The major distinction between these two methods is the different definitions of test score equivalence.

The Levine equally reliable equating method is a linear observed score equating method. In linear observed score equating, the scores on two forms are defined to be equivalent if their means and standard deviations (SDs) are the same in some specified population of examinees (Kolen & Brennan, 1987). The equating function is the linear function of the new form score that results in equal means and SDs of the converted new form score and the old form score in the population of interest.

The function used to relate the new form score to the old form score for the Levine unequally reliable method is the linear function that when applied to the new form *true* score makes the means and SDs of the converted new form true score and the old form true score equal. Although this function is a relationship between true scores, it is applied to observed scores. The equivalence of

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 17, No. 3, September 1993, pp. 225-237

0146-6216/93/030225-13$1.90

observed scores achieved by this method is denoted first-order equity (Hanson, 1991). The property of first-order equity holds for test scores on two forms if, after equating, the *expected* observed scores on the two forms are the same for all examinees in the population of interest.

Various authors (including Brennan, 1990; Kolen & Brennan, 1987; Woodruff, 1986, 1989) have shown that the Levine equally and unequally reliable equating methods can be derived without making reliability assumptions. For this reason, and because the major distinction between the two methods is the type of score equivalence achieved by each method, Levine's equally reliable equating method is referred to as Levine's observed score equating method, and Levine's unequally reliable equating method is referred to as Levine's true score equating method in the present paper, which is consistent with the terminology introduced by Brennan (1990).

The standard error (SE) of equating is useful for evaluating the precision of equating. A number of authors have studied SEs of linear equating methods. Lord (1950) presented SEs of linear equating under a variety of data collection designs, and many of these SEs were reported by Angoff (1971). All of these SEs were derived under the assumption that the score distributions were normal. Braun & Holland (1982) derived SEs of linear equating without making a restrictive normality assumption for the randomly equivalent populations equating design. Kolen (1985) derived SEs of Tucker equating using the delta method. Kolen also compared the SEs derived with a normality assumption to those derived without a normality assumption. He found that the SEs derived without the normality assumption were more accurate than those derived with a normality assumption in many situations, especially when the score distributions were skewed.

In this paper, equations for the large sample SEs of equating are derived for the Levine observed score (LOS) and Levine true score (LTS) equating methods for both internal and external common items, and with and without normality assumptions. The SEs derived here quantify the equating error that is due to examinee sampling. Systematic bias of an equating method in a particular setting (due, for example, to assumptions made by the method that do not hold) is not reflected in the SEs.

Levine Observed Score and True Score Equating Methods

In the CING design, a sample of examinees from one population (Population 1) is given the new form of the test and a sample of examinees from another population (Population 2) is given the old form of the test (the new form is to be equated to the old form). Let the test score of interest (here, the number of items answered correctly) on the items common to both test forms in Population i be denoted by the random variable V_i. Let random variables G_i and H_i represent the test scores on the items unique to the new form and old form of the test, respectively, for Population i. In the nonequivalent groups design, realizations of the random variables G_1, V_1, H_2, and V_2 are observed.

Let X_i and Y_i be the scores on the new and old form in Population i that are to be used for equating. If the common items are internal, then $X_i = G_i + V_i$ and $Y_i = H_i + V_i$. If the common items are external, then $X_i = G_i$ and $Y_i = H_i$.

To define a linear equating function in the CING design, in which the new and old test forms are administered to samples from different populations, the concept of a synthetic population is introduced. The synthetic population is conceived as containing two strata, Population 1 and Population 2, that are proportionally weighted by w_1 and w_2, where $w_1 + w_2 = 1$ and $w_1, w_2 \geq 0$. Detailed discussions of the synthetic population can be found in Braun & Holland (1982) and Kolen & Brennan (1987). Here, random variables in the synthetic population will be denoted with an s subscript. For example, if $P(X_1 = x)$ and $P(X_2 = x)$ are the probabilities that the score to be equated equals x (X_1 and X_2 are discrete random variables) in Populations 1 and 2, respectively, then the probability that the score to be equated equals x in the synthetic population is given by

$$P(X_s = x) = w_1 P(X_1 = x) + w_2 P(X_2 = x) \; . \tag{1}$$

In subsequent sections the mean and variance of a random variable (say Z) are denoted $\mu(Z)$ and $\sigma^2(Z)$, respectively, with $\sigma(Z) \equiv [\sigma^2(Z)]^{1/2}$. The covariance of two random variables (say Z and W) is denoted $\sigma(Z, W)$.

Levine Observed Score Method

The linear equating function estimated by the LOS equating method as a function of the new form score x is

$$l(x) = \frac{\sigma(Y_s)}{\sigma(X_s)}[x - \mu(X_s)] + \mu(Y_s) \; , \tag{2}$$

where by the definition of X_s and Y_s

$$\mu(X_s) = w_1\mu(X_1) + w_2\mu(X_2) \; , \tag{3}$$

$$\mu(Y_s) = w_1\mu(Y_1) + w_2\mu(Y_2) \; , \tag{4}$$

$$\sigma^2(X_s) = w_1\sigma^2(X_1) + w_2\sigma^2(X_2) + w_1 w_2[\mu(X_1) - \mu(X_2)]^2 \; , \tag{5}$$

and

$$\sigma^2(Y_s) = w_1\sigma^2(Y_1) + w_2\sigma^2(Y_2) + w_1 w_2[\mu(Y_1) - \mu(Y_2)]^2 \; . \tag{6}$$

Because G was not observed for any examinees in Population 2 and H was not observed for any examinees from Population 1, estimates of $\mu(X_2)$, $\mu(Y_1)$, $\sigma^2(X_2)$, and $\sigma^2(Y_1)$ are not directly available to use in Equations 3–6. Consequently, assumptions are needed to obtain estimates of the slope and intercept of the equating function in Equation 2. There are several forms of the LOS equating function that vary in regard to the assumptions used (Petersen et al., 1989). Here, the focus is on the most commonly used form of the LOS (and LTS) equating functions. This form of the LOS equating function can be derived by assuming that the test scores in the CING design follow a classical congeneric model (Angoff, 1953; Brennan, 1990; Feldt, 1975; Feldt & Brennan, 1989; Woodruff, 1986).

Under the classical congeneric model, the test scores—G, H, and V—are given as (Woodruff, 1986):

$$G_i = \lambda_G T_i + \delta_G + (\lambda_G)^{1/2} E_{G_i} \; , \tag{7}$$

$$H_i = \lambda_H T_i + \delta_H + (\lambda_H)^{1/2} E_{H_i} \; , \tag{8}$$

and

$$V_i = \lambda_V T_i + \delta_V + (\lambda_V)^{1/2} E_{V_i} \; , \tag{9}$$

for $i = 1, 2$, where it is assumed

$$\mu(E_{G_1}) = \mu(E_{H_1}) = \mu(E_{V_1}) = \mu(E_{G_2}) = \mu(E_{H_2}) = \mu(E_{V_2}) = 0 \; , \tag{10}$$

$$\sigma^2(E_{G_1}) = \sigma^2(E_{H_1}) = \sigma^2(E_{V_1}) = \sigma^2(E_{G_2}) = \sigma^2(E_{H_2}) = \sigma^2(E_{V_2}) \; , \tag{11}$$

$$\lambda_G + \lambda_H + \lambda_V = 1 \; , \tag{12}$$

and

$$\delta_G + \delta_H + \delta_V = 0 \; . \tag{13}$$

The variable T_i is the true score random variable in population i. The random variables E_{G_i}, E_{H_i}, and E_{V_i} are error random variables that are assumed to be uncorrelated with each other and uncorrelated with the true score random variables in all populations.

Woodruff (1986) showed that under the classical congeneric model $\mu(X_2)$, $\mu(Y_1)$, $\sigma^2(X_2)$, and $\sigma^2(Y_1)$ can be written as

$$\mu(X_2) = \mu(X_1) - \gamma_1[\mu(V_1) - \mu(V_2)] \ , \tag{14}$$

$$\mu(Y_1) = \mu(Y_2) + \gamma_2[\mu(V_1) - \mu(V_2)] \ , \tag{15}$$

$$\sigma^2(X_2) = \sigma^2(X_1) - \gamma_1^2[\sigma^2(V_1) - \sigma^2(V_2)] \ , \tag{16}$$

and

$$\sigma^2(Y_1) = \sigma^2(Y_2) + \gamma_2^2[\sigma^2(V_1) - \sigma^2(V_2)] \ , \tag{17}$$

where

$$\gamma_1 = \frac{\sigma^2(X_1)}{\sigma(X_1, V_1)} \tag{18}$$

and

$$\gamma_2 = \frac{\sigma^2(Y_2)}{\sigma(Y_2, V_2)} \tag{19}$$

if the common items are internal ($X_i = G_i + V_i$ and $Y_i = H_i + V_i$), and

$$\gamma_1 = \frac{\sigma^2(X_1) + \sigma(X_1, V_1)}{\sigma^2(V_1) + \sigma(X_1, V_1)} \tag{20}$$

and

$$\gamma_2 = \frac{\sigma^2(Y_2) + \sigma(Y_2, V_2)}{\sigma^2(V_2) + \sigma(Y_2, V_2)} \tag{21}$$

if the common items are external ($X_i = G_i$ and $Y_i = H_i$).

Substituting the expressions for $\mu(X_2)$, $\mu(Y_1)$, $\sigma^2(X_2)$, and $\sigma^2(Y_1)$ from Equations 14–17 into Equations 3–6 gives

$$\mu(X_s) = \mu(X_1) - w_2\gamma_1[\mu(V_1) - \mu(V_2)] \ , \tag{22}$$

$$\mu(Y_s) = \mu(Y_2) + w_1\gamma_2[\mu(V_1) - \mu(V_2)] \ , \tag{23}$$

$$\sigma^2(X_s) = \sigma^2(X_1) - w_2\gamma_1^2[\sigma^2(V_1) - \sigma^2(V_2)] + w_1w_2\gamma_1^2[\mu(V_1) - \mu(V_2)]^2 \ , \tag{24}$$

and

$$\sigma^2(Y_s) = \sigma^2(Y_2) + w_1\gamma_2^2[\sigma^2(V_1) - \sigma^2(V_2)] + w_1w_2\gamma_2^2[\mu(V_1) - \mu(V_2)]^2 \ . \tag{25}$$

Substituting the expressions for the synthetic population moments in Equations 22–25 into Equation 2 gives the equating function for the LOS method in terms of moments of observed random variables. Estimates of these observed score moments are used in Equation 2 to produce an estimated equating function.

Levine True Score Method

The linear equating function estimated by the LTS equating method as a function of the new form score x is

$$l(x) = \frac{\sigma(\tilde{Y}_s)}{\sigma(\tilde{X}_s)}[x - \mu(\tilde{X}_s)] + \mu(\tilde{Y}_s) \ , \tag{26}$$

where $\tilde{X}_s$ and $\tilde{Y}_s$ are the true score random variables corresponding to the observed score random variables X_s and Y_s, respectively. The LTS equating function can, like the LOS equating function, be derived by assuming that the test scores G, H, and V follow a classical congeneric model.

Because $\mu(\tilde{X}_s) = \mu(X_s)$ and $\mu(\tilde{Y}_s) = \mu(Y_s)$, the estimates for $\mu(X_s)$ and $\mu(Y_s)$ obtained from Equations 22 and 23 can be used for estimates of $\mu(\tilde{X}_s)$ and $\mu(\tilde{Y}_s)$, respectively. Brennan (1990) showed that under the classical congeneric model $\sigma^2(\tilde{X}_s)$ and $\sigma^2(\tilde{Y}_s)$ can be written as

$$\sigma^2(\tilde{X}_s) = \gamma_1^2\{w_1\sigma^2(\tilde{V}_1) + w_2\sigma^2(\tilde{V}_2) + w_1w_2[\mu(V_1) - \mu(V_2)]^2\} \tag{27}$$

and

$$\sigma^2(\tilde{Y}_s) = \gamma_2^2\{w_1\sigma^2(\tilde{V}_1) + w_2\sigma^2(\tilde{V}_2) + w_1w_2[\mu(V_1) - \mu(V_2)]^2\} \ , \tag{28}$$

where $\tilde{V}_t$ is the true score random variable corresponding to the observed random variable V_t, and γ_1 and γ_2 are given by Equations 18 and 19 when the common items are internal or by Equations 20 and 21 when the common items are external. Substituting expressions for $\sigma^2(\tilde{X}_s)$ and $\sigma^2(\tilde{Y}_s)$ from Equations 27 and 28 and expressions for $\mu(X_s)$ and $\mu(Y_s)$ from Equations 22 and 23 into Equation 26 gives the LTS equating function as

$$l(x) = \frac{\gamma_2}{\gamma_1}x + \mu(Y_2) - \frac{\gamma_2}{\gamma_1}\mu(X_1) + \gamma_2[\mu(V_1) - \mu(V_2)] \ . \tag{29}$$

For the LTS equating method, the equating function does not involve w_1 and w_2. The equating function for the LOS method (substituting Equations 22–25 into Equation 2) is a more complicated expression that involves w_1 and w_2.

Large Sample Standard Errors

From Equations 2 through 29 it can be seen that for both the LOS and LTS methods, the linear function of the new form score x involves 10 moments—$\mu(X_1)$, $\mu(V_1)$, $\sigma^2(X_1)$, $\sigma^2(V_1)$, $\sigma(X_1, V_1)$, $\mu(Y_2)$, $\mu(V_2)$, $\sigma^2(Y_2)$, $\sigma^2(V_2)$, and $\sigma(Y_2, V_2)$. All 10 moments have directly estimable sampling variances and covariances. Kendall & Stuart (1977) described a general method, often called the delta method, for approximating the sampling variance of a function of random variables. The delta method is derived from the first-order Taylor expansion. A number of authors have used the delta method to derive SEs of equating under a variety of different data collection designs (e.g., Braun & Holland, 1982; Kolen, 1985; Lord, 1950).

For the delta method, let $\theta_1, \theta_2, \ldots, \theta_{10}$ represent the 10 moments, $\mu(X_1)$, $\mu(V_1)$, $\sigma^2(X_1)$, $\sigma^2(V_1)$, $\sigma(X_1, V_1)$, $\mu(Y_2)$, $\mu(V_2)$, $\sigma^2(Y_2)$, $\sigma^2(V_2)$, and $\sigma(Y_2, V_2)$, respectively, and let their estimates be $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{10}$. Define $\hat{l}(x)$ as the estimate of the linear equating function $l(x)$ for either the LOS or LTS methods using $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{10}$ in place of $\theta_1, \theta_2, \ldots, \theta_{10}$. Then, according to the delta method described by Kendall & Stuart (1977), the sampling variance of $\hat{l}(x)$ can be expressed approximately as

$$\sigma^2[\hat{l}(x)] \approx \sum_{i=1}^{10}\left(\frac{\partial l}{\partial\theta_i}\right)^2\sigma^2(\hat{\theta}_i) + \sum_{i=1}^{10}\sum_{j\neq i=1}^{10}\frac{\partial l}{\partial\theta_i}\frac{\partial l}{\partial\theta_j}\sigma(\hat{\theta}_i, \hat{\theta}_j) \ . \tag{30}$$

Because the two samples used in the equating function are independently drawn from Populations 1 and 2, the sampling covariances involving moments from different samples are 0. Thus, Equation 30 can be rewritten as

$$\sigma^2[\hat{l}(x)] \approx \sum_{i=1}^{10} \left(\frac{\partial l}{\partial \theta_i}\right)^2 \sigma^2(\hat{\theta}_i) + \sum_{i=1}^{5} \sum_{j \neq i=1}^{5} \frac{\partial l}{\partial \theta_i} \frac{\partial l}{\partial \theta_j} \sigma(\hat{\theta}_i, \hat{\theta}_j) + \sum_{i=6}^{10} \sum_{j \neq i=6}^{10} \frac{\partial l}{\partial \theta_i} \frac{\partial l}{\partial \theta_j} \sigma(\hat{\theta}_i, \hat{\theta}_j) \ . \tag{31}$$

The SE of equating is the square root of $\sigma^2[\hat{l}(x)]$. To compute $\sigma^2[\hat{l}(x)]$ using Equation 31 requires obtaining the 10 first partial derivatives with respect to each of the 10 moments, the sampling variances of the 10 moments, and the 20 unique nonzero covariances of the 10 moments. The first partial derivatives of $l(x)$ with respect to each of the 10 moments used in $l(x)$ are listed in Tables 1 through 4 for the LOS method [internal (Table 1) and external (Table 2) common items] and for the LTS method [internal (Table 3) and external (Table 4) common items]. The notation used in Tables 1 through 4 is the same as that previously defined. The variances and covariances used in Equation 31 are given in Kolen (1985, Table 2, p. 215) and Zeng (1993, Table 1).

A formula for computing $\sigma^2[\hat{l}(x)]$ for the Levine methods can be obtained by substituting the partial derivatives from Tables 1 through 4 and the variances and covariances from the "general" column in Kolen's Table 2 into Equation 31. {The "normal" column of Kolen's Table 2 is used to obtain the formula for computing $\sigma^2[\hat{l}(x)]$ under the normality assumption.} Note that the expressions are presented in terms of population parameters. In actual computations, the sample estimates of the parameters are substituted in the formulas. The method of computing the SE of equating for the Levine methods described here is basically the same as the method described in Kolen (1985) for the Tucker equating method, but the first partial derivatives differ.

Example

The SEs of LOS and LTS equating for internal common items were computed for data from two forms of a 125-item multiple-choice test. The new form, designated as Form X, was administered to 773 examinees sampled from Population 1; the old form, designated as Form Y, was administered to 795 examinees sampled from Population 2. The two forms contained a common set of 30 items, referred to as V. The scores on V contributed to the total scores of both forms. The two forms were administered one year apart. The data used in this example also were used in Kolen (1985) for computing the SEs of Tucker equating.

Univariate and bivariate moments for these data are listed in Tables 5 and 6, respectively. For Score X, the population skewness was defined as

$$S = \mathrm{E}\{[X - \mu(x)]^3\}/[\sigma(X)]^3 \tag{32}$$

and the population kurtosis as

$$K = \mathrm{E}\{[X - \mu(x)]^4\}/[\sigma(X)]^4 \tag{33}$$

(S and K for Scores Y and V were defined analogously). For Scores X and V,

$$\sigma_{ij} = E\{[X - \mu(X)]^i[V - \mu(V)]^j\} \tag{34}$$

(σ_{ij} for Scores Y and V was defined analogously).

The means of the number-correct scores indicated that the average examinee correctly answered approximately 75% of the total items. The score distributions for both forms were considerably skewed, and the distributions were more peaked than a normal distribution.

Tables 5 and 6 give the statistics needed for computing the SEs of equating using Equation 31

Table 1
Partial Derivatives of LOS Equating With Internal Common Items

Variable (θ_i)	Partial Derivative ($\partial l/\partial\theta_i$)
$\mu(X_1)$	$-\dfrac{\sigma(Y_s)}{\sigma(X_s)}$
$\mu(V_1)$	$\dfrac{w_1 w_2 Z_x[\sigma^2(Y_2)]^2[\mu(V_1)-\mu(V_2)]}{[\sigma(Y_2,V_2)]^2\sigma(Y_s)} - \dfrac{w_1 w_2 Z_x \sigma(Y_s)[\sigma^2(X_1)]^2[\mu(V_1)-\mu(V_2)]}{[\sigma(X_1,V_1)]^2\sigma^2(X_s)} + \dfrac{w_2\sigma(Y_s)\sigma^2(X_1)}{\sigma(X_1,V_1)\sigma(X_s)} + \dfrac{w_1\sigma^2(Y_2)}{\sigma(Y_2,V_2)}$
$\sigma^2(X_1)$	$-\dfrac{\sigma(Y_s)Z_x}{2\sigma^2(X_s)}\left\{1+\dfrac{2\sigma^2(X_1)[\sigma^2(V_s)-\sigma^2(V_1)]}{[\sigma(X_1,V_1)]^2}\right\} + \dfrac{w_2\sigma(Y_s)[\mu(V_1)-\mu(V_2)]}{\sigma(X_1,V_1)\sigma(X_s)}$
$\sigma^2(V_1)$	$\dfrac{w_1 Z_x[\sigma^2(Y_2)]^2}{2[\sigma(Y_2,V_2)]^2\sigma(Y_s)} + \dfrac{w_2 Z_x\sigma(Y_s)[\sigma^2(X_1)]^2}{2[\sigma(X_1,V_1)]^2\sigma^2(X_s)}$
$\sigma(X_1,V_1)$	$\dfrac{Z_x\sigma(Y_s)[\sigma^2(X_1)]^2[\sigma^2(V_s)-\sigma^2(V_1)]}{[\sigma(X_1,V_1)]^3\sigma^2(X_s)} - \dfrac{w_2\sigma(Y_s)\sigma^2(X_1)[\mu(V_1)-\mu(V_2)]}{[\sigma(X_1,V_1)]^2\sigma(X_s)}$
$\mu(Y_2)$	1
$\mu(V_2)$	$-\dfrac{w_1 w_2 Z_x[\sigma^2(Y_2)]^2[\mu(V_1)-\mu(V_2)]}{[\sigma(Y_2,V_2)]^2\sigma(Y_s)} + \dfrac{w_1 w_2 Z_x\sigma(Y_s)[\sigma^2(X_1)]^2[\mu(V_1)-\mu(V_2)]}{[\sigma(X_1,V_1)]^2\sigma^2(X_s)} - \dfrac{w_2\sigma(Y_s)\sigma^2(X_1)}{\sigma(X_1,V_1)\sigma(X_s)} - \dfrac{w_1\sigma^2(Y_2)}{\sigma(Y_2,V_2)}$
$\sigma^2(Y_2)$	$\dfrac{Z_x}{2\sigma(Y_s)}\left\{1+\dfrac{2\sigma^2(Y_2)[\sigma^2(V_s)-\sigma^2(V_2)]}{[\sigma(Y_2,V_2)]^2}\right\} + \dfrac{w_1[\mu(V_1)-\mu(V_2)]}{\sigma(Y_2,V_2)}$
$\sigma^2(V_2)$	$-\dfrac{w_1 Z_x[\sigma^2(Y_2)]^2}{2[\sigma(Y_2,V_2)]^2\sigma(Y_s)} - \dfrac{w_2 Z_x\sigma(Y_s)[\sigma^2(X_1)]^2}{2[\sigma(X_1,V_1)]^2\sigma^2(X_s)}$
$\sigma(Y_2,V_2)$	$-\dfrac{Z_x[\sigma^2(Y_2)]^2[\sigma^2(V_s)-\sigma^2(V_2)]}{[\sigma(Y_2,V_2)]^3\sigma(Y_s)} - \dfrac{w_1\sigma^2(Y_2)[\mu(V_1)-\mu(V_2)]}{[\sigma(Y_2,V_2)]^2}$

where $Z_x = \left[x-\mu(X_s)\right]/\sigma(X_s)$, and $\sigma^2(V_s) = w_1\sigma^2(V_1) + w_2\sigma^2(V_2) + w_1 w_2[\mu(V_1)-\mu(V_2)]^2$.

for LOS equating and LTS equating. The computed SEs, with $w_1 = 1$ and $w_2 = 0$, are given in Table 7. For comparison, bootstrap SEs also were computed and are presented in Table 7. The bootstrap (Efron, 1982) is a nonparametric method of computing SEs. To compute the bootstrap estimates of SEs, data were simulated from the observed distributions of equating item and nonequating item scores. For the results reported in Table 7, 20,000 bootstrap replications were used. The procedure described by Kolen (1985) was followed to compute the bootstrap SEs.

Table 2
Partial Derivatives of LOS Equating With External Common Items

Variable (θ_i)	Partial Derivative ($\partial l/\partial\theta_i$)
$\mu(X_1)$	$-\frac{\sigma(Y_s)}{\sigma(X_s)}$
$\mu(V_1)$	$\frac{w_2\sigma(Y_s)[\sigma^2(X_1)+\sigma(X_1,V_1)]}{\sigma^2(X_s)[\sigma^2(V_1)+\sigma(X_1,V_1)]^2}\Big\{\sigma(X_s)[\sigma^2(V_1)+\sigma(X_1,V_1)] - w_1 Z_x[\mu(V_1)-\mu(V_2)][\sigma^2(X_1)+\sigma(X_1,V_1)]\Big\} + \frac{w_1[\sigma^2(Y_2)+\sigma(Y_2,V_2)]}{\sigma(Y_s)[\sigma^2(V_2)+\sigma(Y_2,V_2)]^2}\Big\{\sigma(Y_s)[\sigma^2(V_2)+\sigma(Y_2,V_2)] + w_2 Z_x[\mu(V_1)-\mu(V_2)][\sigma^2(Y_2)+\sigma(Y_2,V_2)]\Big\}$
$\sigma^2(X_1)$	$\frac{Z_x\sigma(Y_s)}{\sigma^2(X_s)}\left\{\frac{1}{2} + \frac{w_2[\sigma^2(X_1)+\sigma(X_1,V_1)]\{\sigma^2(V_2)-\sigma^2(V_1)+w_1[\mu(V_2)-\mu(V_1)]^2\}}{[\sigma^2(V_1)+\sigma(X_1,V_1)]^2}\right\} + \frac{w_2\sigma(Y_s)[\mu(V_1)-\mu(V_2)]}{\sigma(X_s)[\sigma^2(V_1)+\sigma(X_1,V_1)]}$
$\sigma^2(V_1)$	$\frac{w_2\sigma(Y_s)[\sigma^2(X_1)+\sigma(X_1,V_1)]}{2\sigma^2(X_s)[\sigma^2(V_1)+\sigma(X_1,V_1)]^3}\Big\{-2\sigma(X_s)[\mu(V_1)-\mu(V_2)][\sigma^2(V_1)+\sigma(X_1,V_1)] - Z_x[\sigma^2(X_1)+\sigma(X_1,V_1)]\big\{\sigma^2(V_1)-2\sigma^2(V_2)-\sigma(X_1,V_1) - 2w_1[\mu(V_2)-\mu(V_1)]^2\big\}\Big\} + \frac{w_1 Z_x[\sigma^2(Y_2)+\sigma(Y_2,V_2)]^2}{2\sigma(Y_s)[\sigma^2(V_2)+\sigma(Y_2,V_2)]^2}$
$\sigma(X_1, V_1)$	$\frac{w_2\sigma(Y_s)[-\sigma^2(X_1)+\sigma^2(V_1)]}{\sigma^2(X_s)[\sigma^2(V_1)+\sigma(X_1,V_1)]^3}\Big\{\sigma(X_s)[\mu(V_1)-\mu(V_2)][\sigma^2(V_1)+\sigma(X_1,V_1)] - Z_x[\sigma^2(X_1)+\sigma(X_1,V_1)]\big\{\sigma^2(V_2)-\sigma^2(V_1)+w_1[\mu(V_2)-\mu(V_1)]^2\big\}\Big\}$
$\mu(Y_2)$	1
$\mu(V_2)$	$-\frac{w_2\sigma(Y_s)[\sigma^2(X_1)+\sigma(X_1,V_1)]}{\sigma^2(X_s)[\sigma^2(V_1)+\sigma(X_1,V_1)]^2}\Big\{\sigma(X_s)[\sigma^2(V_1)+\sigma(X_1,V_1)] - w_1 Z_x[\mu(V_1)-\mu(V_2)][\sigma^2(X_1)+\sigma(X_1,V_1)]\Big\} - \frac{w_1[\sigma^2(Y_2)+\sigma(Y_2,V_2)]}{\sigma(Y_s)[\sigma^2(V_2)+\sigma(Y_2,V_2)]^2}\Big\{\sigma(Y_s)[\sigma^2(V_2)+\sigma(Y_2,V_2)] + w_2 Z_x[\mu(V_1)-\mu(V_2)][\sigma^2(Y_2)+\sigma(Y_2,V_2)]\Big\}$

continued on the next page

Table 2, continued
Partial Derivatives of LOS Equating With External Common Items

Variable (θ_i)	Partial Derivative ($\partial l/\partial\theta_i$)
$\sigma^2(Y_2)$	$\frac{Z_x}{\sigma(Y_s)}\left\{\frac{1}{2}+\frac{w_1[\sigma^2(Y_2)+\sigma(Y_2,V_2)]\{\sigma^2(V_1)-\sigma^2(V_2)+w_2[\mu(V_2)-\mu(V_1)]^2\}}{[\sigma^2(V_2)+\sigma(Y_2,V_2)]^2}\right\} + \frac{w_1[\mu(V_1)-\mu(V_2)]}{\sigma^2(V_2)+\sigma(Y_2,V_2)}$
$\sigma^2(V_2)$	$\frac{w_1[\sigma^2(Y_2)+\sigma(Y_2,V_2)]}{2\sigma(Y_s)[\sigma^2(V_2)+\sigma(Y_2,V_2)]^3}\Big\{-2\sigma(Y_s)[\mu(V_1)-\mu(V_2)][\sigma^2(V_2)+\sigma(Y_2,V_2)] - Z_x[\sigma^2(Y_2)+\sigma(Y_2,V_2)]\{-\sigma^2(V_2)+2\sigma^2(V_1)+\sigma(Y_2,V_2) + 2w_2[\mu(V_2)-\mu(V_1)]^2\}\Big\} - \frac{w_2 Z_x\sigma(Y_s)[\sigma^2(X_1)+\sigma(X_1,V_1)]^2}{2\sigma^2(X_s)[\sigma^2(V_1)+\sigma(X_1,V_1)]^2}$
$\sigma(Y_2, V_2)$	$\frac{w_1[-\sigma^2(Y_2)+\sigma^2(V_2)]}{\sigma(Y_s)[\sigma^2(V_2)+\sigma(Y_2,V_2)]^3}\Big\{\sigma(Y_s)[\mu(V_1)-\mu(V_2)][\sigma^2(V_2)+\sigma(Y_2,V_2)] + Z_x[\sigma^2(Y_2)+\sigma(Y_2,V_2)]\{\sigma^2(V_1)-\sigma^2(V_2)+w_2[\mu(V_2)-\mu(V_1)]^2\}\Big\}$

where $Z_x = [x - \mu(X_s)]/\sigma(X_s)$.

Consider a Form X score of 80 (see the first column of Table 7). Based on LOS equating, this score has a Form Y LOS equivalent of 84.6, and the SE of this equivalent is .526 under the normality assumption, .582 without this assumption, and .582 using the bootstrap method. A ± 1 SE band for the Form Y equivalent of a Form X score of 80 is 84.6 $\pm$.582 for the SEs derived without the normality assumption.

Based on LTS equating, the Form Y equivalent for this score is 84.3, and the SE of this equivalent is .501 under the normality assumption, .566 without the normality assumption, and .564 using the bootstrap. A ± 1 SE band for the LTS Form Y equivalent is 84.3 $\pm$.566 for the SEs derived without the normality assumption.

To compute the SE of LTS equating at $X = 80$ without the normality assumption, as reported in Table 7, each of the 30 terms in the summations in Equation 31 must be computed and all the terms summed. For example, using results from Table 3 and Kolen's (1985) Table 2 the first term of the second summation of Equation 31 can be written as

$$\frac{\partial l}{\partial\theta_1}\frac{\partial l}{\partial\theta_2}\sigma(\hat\theta_1,\hat\theta_2) = \frac{\partial l}{\partial\mu(X_1)}\frac{\partial l}{\partial\mu(V_1)}\sigma[\hat\mu(X_1),\hat\mu(V_1)] = \frac{-\sigma^2(Y_2)\ \sigma(X_1,V_1)}{\sigma^2(X_1)\ \sigma(Y_2,V_2)}\frac{\sigma^2(Y_2)}{\sigma(Y_2,V_2)}\frac{\sigma(X_1,V_1)}{n_1}, \quad (35)$$

where n_1 is the sample size for the group taking the new form. Substituting estimates of the moments in Equation 35 from Tables 5 and 6 gives

$$\frac{-178.823\times 47.944}{179.039\times 51.105}\ \frac{178.823}{51.105}\ \frac{47.944}{773} = -.203\ . \quad (36)$$

Other terms in the sums given in Equation 31 are computed in an analogous manner. These terms

Table 3
Partial Derivatives of LTS Equating With Internal Common Items

Variable (θ_i)	Partial Derivative ($\partial l/\partial\theta_i$)
$\mu(X_1)$	$-\dfrac{\sigma^2(Y_2)\,\sigma(X_1, V_1)}{\sigma^2(X_1)\,\sigma(Y_2, V_2)}$
$\mu(V_1)$	$\dfrac{\sigma^2(Y_2)}{\sigma(Y_2, V_2)}$
$\sigma^2(X_1)$	$-\dfrac{[x-\mu(X_1)]\,\sigma^2(Y_2)\,\sigma(X_1, V_1)}{[\sigma^2(X_1)]^2\,\sigma(Y_2, V_2)}$
$\sigma^2(V_1)$	0
$\sigma(X_1, V_1)$	$\dfrac{[x-\mu(X_1)]\,\sigma^2(Y_2)}{\sigma^2(X_1)\,\sigma(Y_2, V_2)}$
$\mu(Y_2)$	1
$\mu(V_2)$	$-\dfrac{\sigma^2(Y_2)}{\sigma(Y_2, V_2)}$
$\sigma^2(Y_2)$	$\dfrac{\sigma(X_1, V_1)[x-\mu(X_1)] + \sigma^2(X_1)[\mu(V_1)-\mu(V_2)]}{\sigma^2(X_1)\,\sigma(Y_2, V_2)}$
$\sigma^2(V_2)$	0
$\sigma(Y_2, V_2)$	$\dfrac{\sigma^2(Y_2)\{-\sigma(X_1, V_1)[x-\mu(X_1)] + \sigma^2(X_1)[\mu(V_2)-\mu(V_1)]\}}{\sigma^2(X_1)\,[\sigma(Y_2, V_2)]^2}$

are summed to produce the result reported in Table 7 of .566 (the SE of the Form Y equivalent based on LTS equating without the normality assumption).

Table 7 shows that the SEs are smallest near the mean and become larger farther away from the mean. The SEs derived without the normality assumption were very close to the bootstrap SEs for both equating methods. The SEs derived with the normality assumption tended to overestimate the SEs of equating for scores above the mean and underestimate them at scores below the mean, relative to the bootstrap SEs and the SEs derived without the normality assumption. This bias in the SEs derived with the normality assumption relative to the other two methods may be the result of the nonnormality (especially negative skewness) of the score distributions of Forms X and Y in the example.

Discussion

For the skewed score distributions, the SEs derived with the normality assumption differed from those derived without the normality assumption and the bootstrap SEs. This suggests that the

Table 4
Partial Derivatives of LTS Equating With External Common Items

Variable (θ_i)	Partial Derivative ($\partial l/\partial\theta_i$)
$\mu(X_1)$	$-\dfrac{[\sigma^2(Y_2)+\sigma(Y_2,V_2)][\sigma^2(V_1)+\sigma(X_1,V_1)]}{[\sigma^2(X_1)+\sigma(X_1,V_1)][\sigma^2(V_2)+\sigma(Y_2,V_2)]}$
$\mu(V_1)$	$\dfrac{\sigma^2(Y_2)+\sigma(Y_2,V_2)}{\sigma^2(V_2)+\sigma(Y_2,V_2)}$
$\sigma^2(X_1)$	$-\dfrac{[x-\mu(X_1)][\sigma^2(Y_2)+\sigma(Y_2,V_2)][\sigma^2(V_1)+\sigma(X_1,V_1)]}{[\sigma^2(X_1)+\sigma(X_1,V_1)]^2[\sigma^2(V_2)+\sigma(Y_2,V_2)]}$
$\sigma^2(V_1)$	$\dfrac{[x-\mu(X_1)][\sigma^2(Y_2)+\sigma(Y_2,V_2)]}{[\sigma^2(X_1)+\sigma(X_1,V_1)][\sigma^2(V_2)+\sigma(Y_2,V_2)]}$
$\sigma(X_1,V_1)$	$\dfrac{[x-\mu(X_1)][\sigma^2(X_1)-\sigma^2(V_1)][\sigma^2(Y_2)+\sigma(Y_2,V_2)]}{[\sigma^2(X_1)+\sigma(X_1,V_1)]^2[\sigma^2(V_2)+\sigma(Y_2,V_2)]}$
$\mu(Y_2)$	1
$\mu(V_2)$	$-\dfrac{\sigma^2(Y_2)+\sigma(Y_2,V_2)}{\sigma^2(V_2)+\sigma(Y_2,V_2)}$
$\sigma^2(Y_2)$	$\dfrac{[x-\mu(X_1)][\sigma^2(V_1)+\sigma(X_1,V_1)]+[\mu(V_1)-\mu(V_2)][\sigma^2(X_1)+\sigma(X_1,V_1)]}{[\sigma^2(X_1)+\sigma(X_1,V_1)][\sigma^2(V_2)+\sigma(Y_2,V_2)]}$
$\sigma^2(V_2)$	$\dfrac{[\sigma^2(Y_2)+\sigma(Y_2,V_2)]}{[\sigma^2(X_1)+\sigma(X_1,V_1)][\sigma^2(V_2)+\sigma(Y_2,V_2)]^2}\Big\{[\mu(X_1)-x][\sigma^2(V_1)+\sigma(X_1,V_1)] -[\mu(V_1)-\mu(V_2)][\sigma^2(X_1)+\sigma(X_1,V_1)]\Big\}$
$\sigma(Y_2,V_2)$	$\dfrac{[\sigma^2(V_2)-\sigma^2(Y_2)]}{[\sigma^2(X_1)+\sigma(X_1,V_1)][\sigma^2(V_2)+\sigma(Y_2,V_2)]^2}\Big\{[x-\mu(X_1)][\sigma^2(V_1)+\sigma(X_1,V_1)] +[\mu(V_1)-\mu(V_2)][\sigma^2(X_1)+\sigma(X_1,V_1)]\Big\}$

Table 5
Univariate Moments for Number-Correct Scores

Form	Group	Mean	SD	Skewness	Kurtosis
X	1	95.74515	13.38053	-1.02526	3.91213
Y	2	96.84025	13.37249	-.99786	3.88848
V	1	23.18499	4.05214	-.83668	3.47899
V	2	22.54214	4.31428	-.79337	3.46673

Table 6
Bivariate Moments for Number-Correct Scores

Forms	$\hat{\sigma}_{11}$	$\hat{\sigma}_{21}$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$	$\hat{\sigma}_{31}$	$\hat{\sigma}_{13}$
X, V	47.94365	-666.08021	-189.10793	9,660.23934	33,745.31466	2,900.56414
Y, V	51.10547	-648.08195	-193.88898	10,470.94569	34,509.99565	3,402.31716

SEs derived with the normality assumption were biased. Because negatively skewed score distributions are typical in many testing programs, the SEs derived without the normality assumption are recommended.

The SEs derived without the normality assumption were very close to the SEs computed using the bootstrap method. An advantage of the SEs based on Equation 31 is that they can be calculated directly. Bootstrap SEs require simulation that is more expensive computationally than using Equation 31.

An alternative to using Equation 31 with the derivatives given in Tables 1 through 4 is to use numerical derivatives (Zeng, 1993). A potential advantage of using numerical derivatives instead of the analytical derivatives given in Tables 1 through 4 is that calculations using numerical derivatives may be easier to program.

Table 7
Y Equivalent Scores From LOS and LTS Equating for *X* Scores of 50 to 125, and Their Standard Errors With the Normality Assumption (Norm), Without the Normality Assumption (Nonorm), and From the Bootstrap (Boot)

			Standard Errors					
	Y Equivalent		LOS			LTS		
X	LOS	LTS	Norm	Nonorm	Boot	Norm	Nonorm	Boot
125	126.0	126.5	.884	.804	.796	.833	.780	.762
120	121.4	121.8	.762	.684	.675	.721	.665	.649
115	116.8	117.1	.645	.570	.561	.615	.557	.542
110	112.2	112.4	.536	.468	.458	.517	.460	.446
105	107.6	107.8	.442	.385	.375	.432	.383	.370
100	103.0	103.1	.373	.338	.328	.371	.338	.328
95	98.4	98.4	.346	.341	.334	.346	.340	.333
90	93.8	93.7	.369	.393	.389	.364	.388	.383
85	89.2	89.0	.434	.478	.476	.420	.468	.465
80	84.6	84.3	.526	.582	.582	.501	.566	.564
75	80.0	79.7	.634	.696	.698	.597	.675	.672
70	75.4	75.0	.750	.817	.820	.703	.790	.787
65	70.8	70.3	.871	.941	.945	.813	.909	.905
60	66.2	65.6	.996	1.067	1.072	.927	1.030	1.026
55	61.6	60.9	1.123	1.196	1.201	1.044	1.153	1.148
50	56.9	56.2	1.252	1.325	1.331	1.162	1.277	1.271

References

Angoff, W. H. (1953). Test reliability and effective test length. *Psychometrika, 18,* 1–14.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 508–600). Washington DC: American Council on Education.

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

Brennan, R. L. (1990). *Congeneric models and Levine's* *linear equating procedures* (American College Testing Research Rep. No. 90-12). Iowa City IA: American College Testing.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans.* Philadelphia PA: Society for Industrial and Applied Mathematics.

Feldt, L. S. (1975). Estimation of the reliability of a test divided into two parts of unequal length. *Psychometrika, 40,* 557-561.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 105-146). New York: Macmillan.

Hanson, B. A. (1991). A note on Levine's formula for equating unequally reliable tests using data from the common item nonequivalent groups design. *Journal of Educational Statistics, 16,* 93-100.

Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics* (4th ed., Vol. 1). New York: Macmillan.

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

Kolen, M. J., & Brennan, R. L. (1987). Linear equating models for the common-item nonequivalent-populations design. *Applied Psychological Measurement, 11,* 263-277.

Levine, R. (1955). *Equating the score scales of alternative forms administered to samples of different ability* (RB-55-23). Princeton NJ: Educational Testing Service.

Lord, F. M. (1950). *Notes on comparable scales for test scores* (RB-50-48). Princeton NJ: Educational Testing Service.

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 221-262). New York: Macmillan.

Woodruff, D. J. (1986). Derivations of observed score linear equating methods based on test score models for the common item nonequivalent population design. *Journal of Educational Statistics, 11,* 245-257.

Woodruff, D. J. (1989). A comparison of three linear equating methods for the common-item nonequivalent population design. *Applied Psychological Measurement, 13,* 353-369.

Zeng, L. (1993). A numerical approach for computing standard errors of linear equating. *Applied Psychological Measurement, 17,* 177-186.

Acknowledgments

The authors thank Robert L. Brennan and David Woodruff for helpful comments on an earlier version of this paper.

Author's Address

Send requests for reprints or further information to Brad Hanson, ACT, P.O. Box 168, Iowa City IA 52243, U.S.A. Internet: bhanson@umaxa.weeg.uiowa.edu.