Research Report

Construction of Chained True Score Equipercentile Equatings Under the Kernel Equating (KE) Framework and Their Relationship to Levine True Score Equating

Haiwen (Henry) Chen

Paul Holland

June 2009

ETS RR-09-24

Construction of Chained True Score Equipercentile Equatings Under the Kernel Equating (KE) Framework and Their Relationship to Levine True Score Equating

Haiwen (Henry) Chen and Paul Holland[1]

ETS, Princeton, New Jersey

June 2009

Abstract

In this paper, we develop a new chained equipercentile equating procedure for the nonequivalent groups with anchor test (NEAT) design under the assumptions of the classical test theory model. This new equating is named chained true score equipercentile equating. We also apply the kernel equating framework to this equating design, resulting in a family of chained true score equipercentile equating functions, which include the Levine true score equating model as a special case.

Key words: NEAT design, chained true score equipercentile equating (CTSEE), Levine true score equating (LTSE), kernel equating (KE)

Acknowledgments

The authors are grateful to Neil Dorans, Tim Moses, and Dan Eignor for their in-depth reviews of the earlier drafts. Their many suggestions and comments were extremely helpful when we revised this report. The first author also thanks Alina von Davier for her encouragement and support in the writing of this paper. Special thanks go to Kimberlea Fryer for her editorial guidance and assistance.

Introduction

Non-IRT equating has been evolving for more than 80 years. At ETS, for a nonequivalent groups anchor test (NEAT) design, commonly used equating methods are linear and equipercentile chained equating (CE), linear and equipercentile post-stratification equating (PSE), the Tucker method, and the Levine true and Levine observed score methods (Angoff, 1982; Braun & Holland, 1982). Only the equipercentile CE and equipercentile PSE are nonlinear.

Kernel equating (KE) methodology (von Davier, Holland, & Thayer, 2004) gives a systematic treatment of several commonly used equating designs. One of its main properties is that it is able to connect a linear method with a nonlinear method by changing the values of the parameter known as the bandwidth.

Levine true score equating (LTSE) and Levine observed score equating (LOSE; see Kolen & Brennan, 2004), which have not been discussed to date in regard to the KE framework, have some favorable properties under certain circumstances. Recently, some work has been done to find nonlinear versions of LOSE, such as a hybrid Levine equipercentile equating (von Davier, Fournier-Zajac, & Holland, 2006), and a modified post-stratification equating (Wang & Brennan, 2007).

This paper constructs a new nonlinear equating method called chained true score equipercentile equating (CTSEE), based on the classical test theory model. Using the KE framework, we will construct a family of CTSEE functions and show that LTSE is a special member of this family.

There are six sections in this paper. After the introduction, the next section reviews the chained equating and Levine true score equating. The third section discusses KE process steps and their properties, and the fourth covers the construction of CTSEE. The fifth section has examples and discussion, and with the last section is the conclusion.

Review of Chained Equating and Levine True Score Equating

CE is a classical method that applies to a NEAT design (Angoff, 1971; Dorans, 1990; Livingston, Dorans, & Wright, 1990). For tests X and Y with anchor A, and populations P and Q taking X and Y, respectively, the chained equipercentile from X to Y is the composition of two equipercentile equatings from X to A with population P and from A to Y with population Q. If we assume that for each test the marginal distributions of all involved test scores are of similar shapes, then each equipercentile equating can be approximated by a suitable linear equating (von Davier et al., 2004, p. 12, Theorem 1.1.). Hence, the resulting chained equipercentile equating can be approximated by a suitable linear equating called a chained linear equating.

LTSE uses true scores to equate X to Y. By the classical test theory model (see Feldt & Brennan, 1989):

$$X = T_X + E_X,$$

where T_X is the true scores associated with X, and E_X is the errors with zero mean and zero correlation with the true scores (i.e., $\mu_X = \mu_{TX}$ and $\sigma^2_X = \sigma^2_{TX} + \sigma^2_{EX}$). The correlation of X with T_X,

$$\rho_{XT_X} = \frac{Cov(X,T_X)}{\sigma_X \sigma_{T_X}} = \frac{Cov(T_X,T_X) + Cov(E_X,T_X)}{\sigma_X \sigma_{T_X}} = \frac{\sigma^2_{T_X}}{\sigma_X \sigma_{T_X}} = \frac{\sigma_{T_X}}{\sigma_X},$$

is the square root of the reliability of test X (see Lord & Novick, 1968). Similar results also hold for A_P (A with population P), Y and A_Q. For convenience, for any univariate distribution X:

$$\rho_X = \frac{\sigma_{T_X}}{\sigma_X}. \tag{1}$$

One assumption of LTSE is that the true scores of the test and its anchor are perfectly correlated, which makes linking X to A_P in true score form a linear function that has this form:

$$a = \frac{\sigma_{T_{A_p}}}{\sigma_{T_X}}(x - \mu_{T_X}) + \mu_{T_{A_p}} = \frac{\sigma_{T_{A_p}}}{\sigma_{T_X}}(x - \mu_X) + \mu_{A_p}. \tag{2}$$

Similarly, there is a linear linking from A_Q to Y:

$$y = \frac{\sigma_{T_Y}}{\sigma_{T_{A_Q}}}(a - \mu_{T_{A_Q}}) + \mu_{T_Y} = \frac{\sigma_{T_Y}}{\sigma_{T_{A_Q}}}(a - \mu_{A_Q}) + \mu_Y. \tag{3}$$

Under the assumptions that (2) and (3) are population invariant, substitute (2) into (3) to get the formula for LTSE:

$$
\begin{aligned}
y \ = \ \mathrm{LTSE}(x) \ &= \ \frac{\sigma_{T_Y}}{\sigma_{T_{A_Q}}}(a-\mu_{A_Q})+\mu_Y \\
&= \ \frac{\sigma_{T_Y}}{\sigma_{T_{A_Q}}}\left(\frac{\sigma_{T_{A_P}}}{\sigma_{T_X}}(x-\mu_X)+\mu_{A_P}-\mu_{A_Q}\right)+\mu_Y \\
&= \ \frac{\sigma_{T_Y}\sigma_{T_{A_P}}}{\sigma_{T_X}\sigma_{T_{A_Q}}}(x-\mu_X)+\mu_Y+\frac{\sigma_{T_Y}}{\sigma_{T_{A_Q}}}(\mu_{A_P} - \mu_{A_Q}).
\end{aligned} \tag{4}
$$

Another assumption for Levine's methods is that error variances are invariant among all populations. This is used, along with the linearity between a test and its anchor on true scores, to calculate $\sigma_{T\bullet}$ (hence $\rho_{\bullet T_\bullet}$ and $\rho_\bullet$) for any distribution. See Brennan (1990) for detailed computations.

Review of Kernel Equating Procedures Applied to Chained Equating

KE, first proposed by Holland and Thayer (1989) and later fully developed by von Davier et al. (2004), gives a systematic treatment for many well-known equating designs to derive the equating functions. It consists of five basic steps:

1. Presmoothing the score probabilities by fitting a log-linear model. This step can be omitted or modified by using alternative models.
2. Estimating the score probabilities. This step estimates the score distributions for both test *X* and test *Y*, denoted as ***r*** and ***s***, respectively. The equating design will play a crucial role in this step.
3. Continuizing ***r*** and ***s***. Use kernel techniques to make continuized density functions and the related continuized distribution functions (CDFs) from the discrete density distributions created in Step 2. This is the unique step that defines KE. The kernel used here is the normal distribution function, also known as the Gaussian kernel.
4. Computing the equating. This step computes the equating function by composing two or more CDFs made in Step 3.
5. Calculating the standard error of equating. This step use C-matrices either generated in Step 1 or calculated within KE for nonpresmoothed data (Moses & Holland, 2006).

The book *The Kernel Method of Test Equating* (von Davier et al., 2004) explicitly details the five steps applied to a chained equating. This paper deals with only Steps 2–4 since the other steps are not relevant to the topic of this paper.

Estimation of the Score Probabilities for Chained Equating

Given tests X and Y with common items known as anchor A, two bivariate score distributions are obtained. Then the marginal distributions are estimated for X and its anchor from the first data set and are denoted as X and A_P, respectively. Similarly, marginal distributions Y and A_Q are obtained from the second data set.

Continuization of the Marginal Distributions

Unlike other designs, CE needs four continuized score densities. The details on X are given below.

Let $\{(x_i, r_i)\}$ be the marginal distribution of X with probability r_i for each score x_i. Let μ_x and σ_x be the mean and standard deviation of X, respectively. For any positive number h_x, called the bandwidth, a_X is defined as:

$$a_X = \sqrt{\frac{\sigma_X^2}{\sigma_X^2 + h_X^2}}, \tag{5}$$

and $\eta = a_X h_x$.

Then for each x_i, $R_{iX}(x)$ is defined as

$$R_{iX}(x) = \frac{x - \mu_X - a_X(x_i - \mu_X)}{\eta}, \tag{6}$$

and the continuized distribution function (CDF) of X with bandwidth h_X is:

$$F_{hX}(x) = \sum_i r_i \Phi(R_{iX}(x)), \tag{7}$$

where Φ is the CDF of the standard normal function.

It is convenient to define a new random variable $X(h_X)$ to study the properties of $F_{hx}(x)$.

Let V be independent of X with the standard normal distribution, with given h_X:

$$X(h_X) \; = \; a_X (X + h_X V) + (1 - a_X) \mu_X . \tag{8}$$

It is shown in von Davier et al. (2004) that

$$F_{hX}(x) = \text{Prob}(X(h_X) \leq x). \tag{9}$$

Some properties of $X(h_X)$ are:

1. $\lim_{h_X \to 0} X(h_X) \; = \; X.$ (10)

2. $\lim_{h_X \to \infty} X(h_X) = \sigma_X V + \mu_X .$ (11)

3. $E(X(h_X)) \; = \; \mu_X$ and $E((X(h_X) - \mu_X)^2) \; = \; \sigma_X^2 ,$ for $\forall \; h_X .$ (12)

Property 1 shows that X can be approximated by $X(h_X)$ with small h_X (≤ 0.5). Property 2 indicates that $X(h_X)$, or any other distribution with a large bandwidth, is almost a normal distribution. In particular, all such modified distributions are (almost) of the same shape. Property 3 certifies that the mean and standard deviation of $X(h_X)$ will never change.

Similarly, $G_{hY}(y)$, the CDF of Y on population Q, $H_{P,hA_P}(a)$, the CDF of A on population P, and $H_{Q,hA_Q}(a)$, the CDF of A on population Q, can be defined with given bandwidths h_Y, h_{A_P}, and h_{A_Q}, respectively.

Computation of the Chained Equating Function $\hat{e}_{Y(CE)}(x)$

The computation of the chained equating function $\hat{e}_{Y(CE)}(x)$ is:

$$\hat{e}_{Y(CE)}(x) \; = \; G_{h_Y}^{-1}(H_{Q,h_{A_Q}}(H_{P,h_{A_P}}^{-1}(F_{h_X}(x)))). \tag{13}$$

When both h_X and h_{A_P} are small, by Property 1 of $X(h_X)$ (and $A(h_A)$), the composition function $H_{P,h_{A_P}}^{-1}(F_{h_X}(x))$ is an approximation of the equipercentile equating function from X to A. For very large h_X and h_{A_P}, since the density functions for both $H_{P,h_{A_P}}(a)$ and $F_{h_X}(x)$ are normal density functions by Property 2, $H_{P,h_{A_P}}^{-1}(F_{h_X}(x))$ becomes a linear equating (von Davier et al., 2004, p. 12, Theorem 1.1.), which has the form:

$$a = \frac{\sigma_{A_p}}{\sigma_X}(x - \mu_X) + \mu_{A_p}. \tag{14}$$

The same arguments apply to test Y with anchor A_Q. Hence, for small h_X , $h_{Y,}$ h_{A_P}, and h_{A_Q}, $\hat{e}_{Y(CE)}(x)$ is the chained equipercentile equating, but for large bandwidths, it is the chained linear equating, whose function is:

$$y = \text{Lin}_{\text{CE}}(x) = \frac{\sigma_Y \sigma_{A_P}}{\sigma_X \sigma_{A_Q}}(x - \mu_X) + \mu_Y + \frac{\sigma_Y}{\sigma_{A_Q}}(\mu_{A_P} - \mu_{A_Q}). \tag{15}$$

Constructing a Chained True Score Equipercentile Equating via the Kernel Equating (KE) Framework

From the previous discussion of chained equipercentile equating, it is easy to see how to construct a CTSEE; that is, to make the linkings on the true scores under the assumption that the linking from T_X to T_A and from T_A to T_Y are both population invariant. Now the problem is how to get the true scores when an observed score distribution is given. Theoretically, the problem is unsolvable. In practice, however, there are many ways to approximate the true scores.

To find a true score X_T for X when σ_{TX} is given, the basic criteria are:

1. $E(X_T) = \mu_X.$ (16)
2. $E((X_T - \mu_X)^2) = \sigma_{T_X}^2.$ (17)

By the discussion in the previous section on KE, the goal is to construct a distribution defined in (18):

$$X_T(h_X) = a_{T_X}(X + h_X V) + b_{T_X}, \tag{18}$$

for a given number $h_{X,}$ and solve (18) through (20) for both a_{T_X} and b_{T_X}:

$$E(X_T(h_X)) = \mu_X \tag{19}$$

$$E((X_T(h_X) - \mu_X)^2) = \sigma_{T_X}^2. \tag{20}$$

Substitute (18) into (19) to get

$$E(X_T(h_X)) \ = \ E(a_{T_X}X) + E(a_{T_X}h_X V) + E(b_{T_X}) = a_{T_X}\mu_X + 0 + b_{T_X} \xrightarrow{set\,=} \mu_X. \quad (21)$$

Hence,

$$b_{T_X} \ = \ (1 - a_{T_X})\ \mu_X. \quad (22)$$

Substitute (18) and (22) into (20) to get

$$\begin{aligned} E((X_T(h_X)\ -\ \mu_X)^2) \ &= \ E((a_{T_X}(X + h_X V) + (1 - a_{T_X})\mu_X\ -\ \mu_X)^2) \\ &= \ E((a_{T_X}(X - \mu_X + h_X V))^2) \\ &= \ a_{T_X}^2\{E((X - \mu_X)^2) + E((h_X V)^2)\} \\ &= \ a_{T_X}^2(\sigma_X^2 + h_X{}^2) \xrightarrow{Set\,=} \sigma_{T_X}^2. \end{aligned} \quad (23)$$

The solution for a_{T_X} is

$$a_{T_X} \ = \ \sqrt{\frac{\sigma_{T_X}^2}{\sigma_X^2 + h_X^2}}. \quad (24)$$

Since h_X is an arbitrary number, a family of true scores $\{X_T(h_X)\}$ related to X is created. Substitute (22) and (24) into (18), with the properties that $\lim_{h_X \to \infty} a_{T_X} = 0$ and $\lim_{h_X \to \infty} a_{T_X} h_X = \sigma_{T_X}$, to get

$$\lim_{h_X \to \infty} X_T(h_X) \ = \ \sigma_{T_X} V + \mu_X. \quad (25)$$

Families of true scores can also be made for A_P, Y, and A_Q, with similar properties in (19), (20), and (25), respectively. Then, for very large bandwidths, the CTSEE is virtually the LTSE. The result can be stated as a theorem:

Theorem. For a NEAT design having test X with anchor A_P, test Y with anchor A_Q, let $T_{()}$ be the true score of, and $h_{()}$ be the bandwidth associated with the specified test, $\mu_{()}$, and $\sigma_{()}$ be the mean and standard deviation of the labeled distribution, if in the continuization step of KE process, the following is defined:

$$a_{T_X} = \sqrt{\frac{\sigma^2_{T_X}}{\sigma^2_X + h^2_X}}\ ,\ a_{T_Y} = \sqrt{\frac{\sigma^2_{T_Y}}{\sigma^2_Y + h^2_Y}}\ ,\ a_{T_{A_P}} = \sqrt{\frac{\sigma^2_{T_{A_P}}}{\sigma^2_{A_P} + h^2_{A_P}}}\ ,\ \text{and}\ a_{T_{A_Q}} = \sqrt{\frac{\sigma^2_{T_{A_Q}}}{\sigma^2_{A_Q} + h^2_{A_Q}}},$$

and

$$\begin{aligned} X_T(h_X) &= a_{T_X}(X + h_X V) + (1 - a_{T_X})\mu_X \\ Y_T(h_Y) &= a_{T_Y}(Y + h_Y V) + (1 - a_{T_Y})\mu_Y \\ A_{P,T}(h_{A_P}) &= a_{T_{A_P}}(A_P + h_{A_P} V) + (1 - a_{T_{A_P}})\mu_{A_P} \\ A_{Q,T}(h_{A_Q}) &= a_{T_{A_Q}}(A_Q + h_{A_Q} V) + (1 - a_{T_{A_Q}})\mu_{A_Q}. \end{aligned}$$

Let

$$\begin{aligned} &F_{hX_T}(x) = \text{Prob}(X_T(h_X) \le x) \\ &G_{hY_T}(y) = \text{Prob}(Y_T(h_Y) \le y) \\ &H_{P,hA_{P,T}}(a) = \text{Prob}(A_{P,T}(h_{A_P}) \le a) \\ &H_{Q,hA_{Q,T}}(a) = \text{Prob}(A_{Q,T}(h_{A_Q}) \le a), \end{aligned}$$

and then for the equating function:

$$\hat{e}_{Y(T)}(x) \ = \ G^{-1}_{h_{Y_T}}(H_{Q,h_{A_{Q,T}}}(H^{-1}_{P,h_{A_{P,T}}}(F_{h_{X_T}}(x)))), \qquad (26)$$

with large bandwidths (preferably no less than 30 times of the related standard deviations, respectively), it follows that

$$\hat{e}_{Y(T)}(x) \ = \ \text{LTSE}(x). \qquad (27)$$

The $\hat{e}_{Y(T)}(x)'s$ are called CTSEE functions.

Example and Discussion

Any equating method can be decomposed as a linear part and a nonlinear part. The decomposition will be quite natural if using the KE framework. The original equating in KE (i.e., with quite small bandwidths) is a sum of the linear portion (i.e., with very large bandwidths) and the remainder, which can be written as:

$$e_Y(x) = \text{Lin}_{e_y}(x) + \text{Rsd}_{e_y}(x), \tag{28}$$

where $\text{Rsd}_{e_y}(x) = e_Y(x) - \text{Lin}_{e_y}(x)$, is the residue of the equating function minus its linear portion. There is a belief that if two equating methods are related but different, the majority of their differences lies in their linear portions. This idea led to the development of the hybrid Levine equipercentile equating (von Davier et al., 2006). So if one method is to replace another method under some conditions for data, the difference between these two methods is expected to be mainly in their linear portions. The following example, (29), checks to see if it is true for both CE and CTSEE.

The design and data sets are given in Chapter 10 of von Davier et al. (2004). Test X has 78 items with external anchor A_P, which has 35 items, and test Y has 78 items with anchor $A_Q = A_P$. Using the same notation as before,

$$\mu_X = 39.2515, \mu_Y = 32.6866, \mu_{A_P} = 17.0540, \mu_{A_Q} = 14.3864,$$
$$\sigma_X = 17.2252, \sigma_Y = 16.7271, \sigma_{A_P} = 8.3329, \sigma_{A_Q} = 8.2082,$$
$$Cov(X, A_P) = Cov(T_X, T_{A_P}) = \rho_{T_X, T_{A_P}} \sigma_{T_X} \sigma_{T_{A_P}} = 126.4198,$$

and

$$Cov(Y, A_Q) = Cov(T_Y, T_{A_Q}) = \rho_{T_Y, T_{A_Q}} \sigma_{T_Y} \sigma_{T_{A_Q}} = 120.0982, \tag{29}$$

where $\rho_{T_X, T_{Ap}}$ is the correlation coefficient of T_X and T_{A_P}. Next, calculate σ_{T_X}, and so on, using the formulas in Brennan (1990) with the assumptions of the classical congeneric model, in particular, that both $\rho_{T_X, T_{Ap}}$ and $\rho_{T_Y, T_{AQ}}$ = 1. Later, this paper will cover the case that at least one of them < 1.

Using the numbers in (29) and the formula in Brennan (1990; with external A_P):

$$\sigma_{T_X} = \sqrt{\frac{Cov(X, A_P) + \sigma_X^2}{Cov(X, A_P) + \sigma_{A_P}^2} Cov(X, A_P)}$$
$$= \sqrt{\frac{126.4198 + 17.2252^2}{126.4198 + 8.3329^2} 126.4198} = \sqrt{273.1159} = 16.5262. \tag{30}$$

Similarly,

$$\sigma_{T_Y} = 16.0056, \quad \sigma_{T_{AP}} = 7.6497, \quad \sigma_{T_{AQ}} = 7.5035. \tag{31}$$

Substituting suitable numbers from (29) into (15) results in

$$\mathrm{Lin}_{\mathrm{CE}}(x) = 0.98584x - 0.57295. \tag{32}$$

Substituting the suitable numbers from (29) through (31) into (4) results in

$$\mathrm{LTSE}(x) = 0.98736x - 0.37874. \tag{33}$$

Hence,

$$\mathrm{LTSE}(x) - \mathrm{Lin}_{\mathrm{CE}}(x) = 0.00152x + 0.19421. \tag{34}$$

Using (28) and (34) and the fact that the linear part of CTSEE is LTSE, the difference between CTSEE and CE can be written as

$$\begin{aligned} \mathrm{CTSEE}(x) - \mathrm{CE}(x) &= (\mathrm{LTSE}(x) + \mathrm{Rsd}_{\mathrm{CTSEE}}(x)) - (\mathrm{Lin}_{\mathrm{CE}}(x) + \mathrm{Rsd}_{\mathrm{CE}}(x)) \\ &= \mathrm{LTSE}(x) - \mathrm{Lin}_{\mathrm{CE}}(x) + \mathrm{Rsd}_{\mathrm{CTSEE}}(x) - \mathrm{Rsd}_{\mathrm{CE}}(x) \\ &= 0.00152x + 0.19421 + \mathrm{Rsd}_{\mathrm{CTSEE}}(x) - \mathrm{Rsd}_{\mathrm{CE}}(x). \end{aligned} \tag{35}$$

Using the data above, CTSEE(x) and CE(x), both with bandwidths = 0.5, and LTSE(x) and $\mathrm{Lin}_{\mathrm{CE}}(x)$, both with bandwidths = 500, were computed with KE Software, which is currently under development at ETS. Both LTSE(x) - $\mathrm{Lin}_{\mathrm{CE}}(x)$ and $\mathrm{Rsd}_{\mathrm{CTSEE}}(x)$ - $\mathrm{Rsd}_{\mathrm{CE}}(x)$ are plotted in Figure 1.

It is obvious that LTSE(x) - $\mathrm{Lin}_{\mathrm{CE}}(x)$ has some bias, while $\mathrm{Rsd}_{\mathrm{CTSEE}}(x)$ - $\mathrm{Rsd}_{\mathrm{CE}}(x)$ has almost none. Simple computations show that Mean(LTSE(x) - $\mathrm{Lin}_{\mathrm{CE}}(x)$) = 0.254, and Mean($\mathrm{Rsd}_{\mathrm{CTSEE}}(x)$ - $\mathrm{Rsd}_{\mathrm{CE}}(x)$) = 0.059. The data sets are highly correlated (both correlation coefficients are in the range of 0.87–0.88), and LTSE(x) is quite close to $\mathrm{Lin}_{\mathrm{CE}}(x)$. Otherwise, Mean(LTSE($x$) - $\mathrm{Lin}_{\mathrm{CE}}(x)$) would be much bigger. Both LTSE(x) and $\mathrm{Lin}_{\mathrm{CE}}(x)$ as computed by the software agree with (33) and (32), respectively, up to four decimal places.

Computing σ_{T_X}, and so on, is impossible without additional assumptions. By using the classical congeneric model, which assumes that the true scores between the main test and its anchor are perfectly correlated, it is purely technical to compute the values of the terms in both

(30) and (31). In general, making such an assumption will contradict the purpose of CTSEE. However, in practice, for most equating cases, the correlation coefficient of two related true scores > 0.99, which is demonstrated in Figure 2.

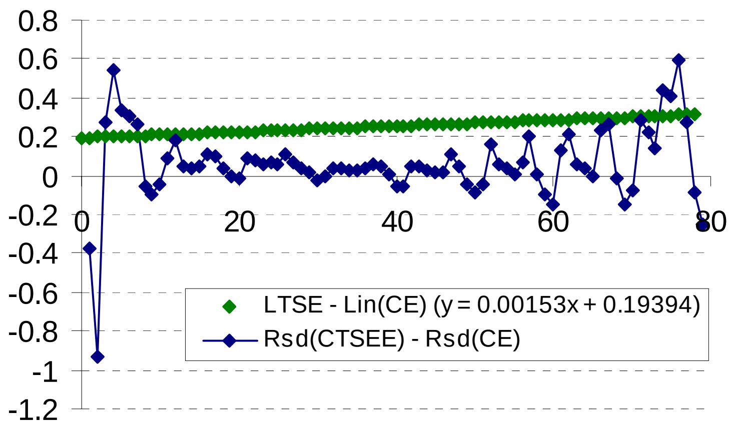

***Figure 1.* The linear difference and the residue (Rsd) difference between chained true score equipercentile equating (CTSEE) and chained equating (CE).**

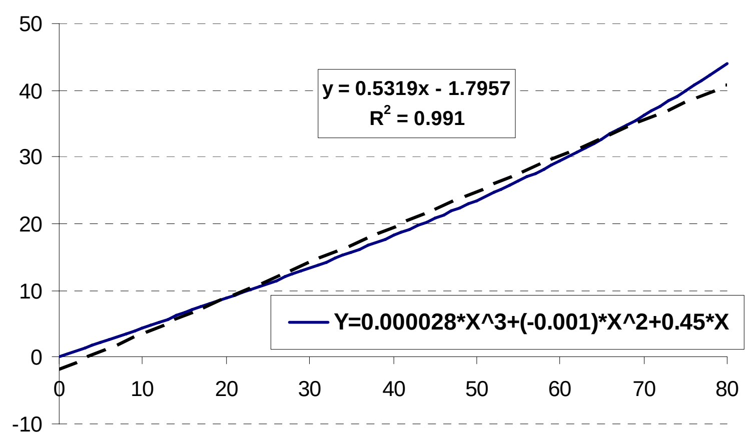

***Figure 2.* Computing the correlation coefficient for a curvilinear function.**

Both X and Y are assumed to be true scores, with Y a function of X, as shown in Figure 2, which may represent a general pattern of an equating function that is not linear. By fitting a line to the curve, ρ_{T_X,T_Y} = R = $\sqrt{0.991}$ = 0.9959.

For the case that the correlation coefficient of true scores is smaller than 0.99, an iterating process will be used. First, (30) is used to compute σ_{T_X}, and so on. Secondly, a true score equipercentile equating from T_X to T_{A_P} will be constructed, using previously computed σ_{T_X} and $\sigma_{T_{A_P}}$, resulting in a function from T_X to T_{A_P}. Hence $\rho_{T_X,T_{Ap}}$ can be computed. Then σ_{T_X} is computed again, but replacing $Cov(X, A_P)$ with $\dfrac{Cov(X, A_P)}{\rho_{T_X,T_{Ap}}}$ in (30). This process continues until $\rho_{T_X,T_{Ap}}$ converges. However, whether $\rho_{T_X,T_{Ap}}$ will converge or under what conditions it will converge are not the topics in this paper.

What happens to $X_T(h_X)$ when h → 0? It is apparent that a_{T_X} becomes ρ_X defined in (1), so

$$\lim_{h_X \to 0} X_T(h_X) = \lim_{h_X \to 0}(a_{T_X}(X + h_X V) + (1 - a_{T_X})\ \mu_X) = \rho_X X + (1 - \rho_X)\ \mu_X, \tag{36}$$

which is the squeezing process to replace the original score distribution by its true scores, proposed by Brennan and Lee (2006) and used by Wang and Brennan (2007) on the anchor marginal distributions for their modified PSE method.

Conclusion

CTSEE extends LTSE so that equating on true scores is not a linear function, just as chained equipercentile equating extends chained linear equating on observed scores. Under the KE framework, CTSEE and LTSE can be connected naturally by varying the values of the bandwidths. Additional computations for CTSEE are needed and can be done with the help of the classical congeneric model, although sometimes adjustments are needed.

Just like LTSE, CTSEE only equates true scores. This makes it less practical than other curvilinear equating methods. However, this new approach opens the field. More studies on this method will reveal properties so far unknown to researchers and practitioners, leading to the improvements of this method, as well as the developments of new methods.

References

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

Angoff, W. H. (1982). *Summary and derivation of equating methods used at ETS*. In P. W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 55—69). New York: Academic Press.

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

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

Brennan, R. L., & Lee, W. (2006). *Correcting for bias in single-administration decision consistency indexes* (CASMA Research Rep. No. 18). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, University of Iowa.

Dorans, N. J. (1990). Equating methods and sampling designs. *Applied Measurement in Education, 3*, 3–17.

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

Holland, P. W., & Thayer, D. T. (1989). *The kernel method of equating score distributions* (ETS Research Rep. No. RR-89-07). Princeton, NJ: ETS.

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

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

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison Wesley.

Moses, T., & Holland, P. W. (2006). *Notes on a general framework for observed score equating*. Unpublished manuscript.

von Davier, A. A., Fournier-Zajac, S., & Holland, P. W. (2006, April). *An equipercentile version of the Levine linear observed-score equating function using the methods of kernel* *equating*. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco.

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

Wang, T., & Brennan, R. L. (2007, April). *A modified frequency estimation equating method for the common-item non-equivalent groups design*. Paper presented at the annual meeting of the National Council of Measurement in Education, Chicago.

Notes

[1] The work that led to this paper was collaborative in every respect and the order of authorship is alphabetical.