Standard Errors of Kernel Equating: Accounting for Bandwidth Estimation

Kseniia Marcq[1] (iD), and **Björn Andersson**[1] (iD)

Abstract
In standardized testing, equating is used to ensure comparability of test scores across multiple test administrations. One equipercentile observed-score equating method is kernel equating, where an essential step is to obtain continuous approximations to the discrete score distributions by applying a kernel with a smoothing bandwidth parameter. When estimating the bandwidth, additional variability is introduced which is currently not accounted for when calculating the standard errors of equating. This poses a threat to the accuracy of the standard errors of equating. In this study, the asymptotic variance of the bandwidth parameter estimator is derived and a modified method for calculating the standard error of equating that accounts for the bandwidth estimation variability is introduced for the equivalent groups design. A simulation study is used to verify the derivations and confirm the accuracy of the modified method across several sample sizes and test lengths as compared to the existing method and the Monte Carlo standard error of equating estimates. The results show that the modified standard errors of equating are accurate under the considered conditions. Furthermore, the modified and the existing methods produce similar results which suggest that the bandwidth variability impact on the standard error of equating is minimal.

Keywords
achievement testing, classical test theory, equating, item response theory, standard errors

Standardized testing is commonly used for assessing individual achievement and its results greatly influence high-stakes decisions ranging from university admissions to various industry certifications. Standardized testing generally requires alternate test forms to be administered on multiple occasions. As a consequence, the tests often differ in difficulty from one administration to another, which poses a challenge with respect to comparability and fairness of the resulting test scores. In order to address this challenge, a statistical procedure known as equating is employed with the

[1]University of Oslo, Oslo, Norway

Corresponding Author:
Björn Andersson, Centre for Educational Measurement, University of Oslo. Postboks 1161 Forskningsparken, Oslo 0318, Norway.
Email: bjorn.andersson@cemo.uio.no

paramount goal of adjusting the scores on the test forms so that they yield interchangeable results (Kolen & Brennan, 2014).

Observed-score equating is one of the fundamental methods used in test equating. Rooted in classical test theory, it is concerned with establishing the equivalence of the observed scores on two test forms and includes both linear and equipercentile equating functions (von Davier, 2011). In this study, we focus on an equipercentile observed-score equating method called kernel equating, which was initially introduced by Holland and Thayer (1989) and further developed by von Davier et al. (2004).

The conceptual framework of kernel equating follows that of equipercentile observed-score equating and posits a series of steps to obtain the equated scores: (1) pre-smoothing of the data to reduce sampling variability, (2) obtaining discrete score probability distributions, (3) obtaining continuous approximations to the discrete score distributions, (4) calculating the equating function, and (5) calculating the standard errors of equating (SEEs) (von Davier, 2011; von Davier et al., 2004). A feature that distinguishes kernel equating from other equipercentile methods is that the continuous approximations of the score probability distributions are achieved through kernels that utilize bandwidth parameters. The bandwidths allow the density functions to be as smooth as possible while retaining the properties of the original distributions. Estimating such parameters, however, introduces additional sampling variability. This variability is typically not accounted for when calculating the standard errors of kernel equating, and therefore constitutes a threat to their accuracy (Holland et al., 1989; von Davier et al., 2004).

Accurate estimation of the SEE is integral to making correct inferences and comparisons. When estimated incorrectly, it can lead to unjustified certainty. One previous study derived modified standard errors of kernel equating when using a variant of the Silverman's rule of thumb for bandwidth estimation (Andersson et al., 2014). The current study derives modified SEEs when using the more commonly applied approach to estimate the bandwidth by minimizing a penalty function. Such an approach is more generally appropriate and does not rely on a particular distributional assumption for the test scores. Thus, the objective of this article is to introduce a modified method of calculating the SEE which accounts for the additional variability stemming from the bandwidth estimation. The new approach is compared via simulations to the current method of calculating the SEE (Holland et al., 1989) across several sample sizes and test lengths.

We structure this article as follows. In the subsequent section, we give a brief background to the kernel method of equating and expand on the issue of bandwidth estimation and how it influences sampling variability. We also discuss how the standard errors of kernel equating are currently estimated. Next, the asymptotic variance of the bandwidth parameter estimator is derived and is incorporated in a modified method for calculating the SEE. This modified method is further verified and compared to the existing method in a simulation study. Lastly, the results are reported and discussed.

The Kernel Method of Test Equating

Data Collection Designs

An observed-score equating procedure consists of two fundamental components, namely, the data collection design and the equating method (von Davier et al., 2004). Hence, before we focus on the equating itself, it is essential to review, if only briefly, the common approaches to collecting the data. There are several data collection designs widely used in practice and they can roughly be divided into two categories: designs which use examinees from a common population taking both test forms and designs which use common items on the test forms (von Davier et al., 2004). The first category of data collection designs includes the equivalent groups, the single group, and the

counterbalanced designs, and the second category includes the common-item non-equivalent groups design. In this study, we focus on the equivalent groups design, where two independent random samples are drawn from a common population, P, and one group takes test form X, while the other takes test form Y. In the following, we will use X and Y to denote both the test forms and the random variable corresponding to the test scores from each of the test forms.

The choice of an appropriate data collection design is subject to considerations like the available sample size, time, and costs. The designs subsequently affect the equating procedure implying that some designs, such as the equivalent groups design, allow for a relatively straightforward comparison between the test forms. Other designs are much more complex, such as the common-item non-equivalent groups design. A more detailed account of the considerations and procedures involved in various data collection designs can be found in von Davier et al. (2004).

Kernel Equating

In the following, we adopt the notation of von Davier et al. (2004). Let the target population be T, the possible score values on the test form X be x_j for $j = 1, \ldots, J$; and let the possible score values on the test form Y be y_k for $k = 1, \ldots, K$. Thus, we define the score probabilities as

$$r_j = \text{Prob}\{X = x_j | T\}, \tag{1}$$

and

$$s_k = \text{Prob}\{Y = y_k | T\}. \tag{2}$$

Further, an equipercentile equating function is defined in terms of the cumulative distribution functions (CDFs) which are given by

$$F(x) = \text{Prob}(X \leq x) = \sum_{j, x_j \leq x} r_j, \tag{3}$$

and

$$G(y) = \text{Prob}(Y \leq y) = \sum_{k, y_k \leq y} s_k. \tag{4}$$

When the CDFs are continuous, we obtain the equipercentile equating function of X to Y from

$$y = \text{Equi}_Y(x) = G^{-1}(F(x)). \tag{5}$$

Strictly speaking, however, most score distributions are discrete, and their continuous approximations are required. Kernel equating addresses this problem by introducing a series of steps which can be applied to various data collection designs and which provides continuous CDFs. The steps of kernel equating are pre-smoothing, estimation of the score probabilities, continuous approximation to the discrete score distributions, equating, and calculating the SEE (von Davier et al., 2004). We now briefly review the first two steps and dedicate subsequent subsections to present the remaining steps in more detail as they pertain to the subject at hand.

Pre-Smoothing. In the pre-smoothing step, a parametric statistical model is fitted to the observed data. This can be done by fitting log-linear or item response theory (IRT) models to the data. The methods are described in detail in Andersson and Wiberg (2017), and Holland and Thayer (1987), and are not repeated here.

Estimation of the Score Probabilities. Having estimated the score distributions with a pre-smoothing model, the score probabilities can be obtained using a linear or non-linear transformation which, following von Davier et al. (2004), we call the design function. The design function depends on the data collection design. For instance, consider the equivalent groups design and let $r = (r_1,\ldots,r_J)^t$ denote the column vector of the score probabilities of X and $s = (s_1,\ldots,s_K)^t$ denote the column vector of the score probabilities of Y. The design function (DF) is then a simple identity function, that is

$$\text{DF}(r,s) = \begin{pmatrix} \mathbf{I}_J & 0 \\ 0 & \mathbf{I}_K \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} r \\ s \end{pmatrix}, \tag{6}$$

where $\mathbf{I}_J$ and $\mathbf{I}_K$ are $J \times J$ and $K \times K$ identity matrices. Design functions for other data collection designs are given explicitly in von Davier et al. (2004).

Continuous Approximation and Equating

The third step in kernel equating, distinguishing it from other equipercentile methods, is how the continuous approximations to the discrete CDFs, $F_{h_X}(x)$ and $G_{h_Y}(y)$ to $F(x)$ and $G(y)$, are obtained. In kernel equating, this is achieved by applying a kernel with a smoothing bandwidth parameter (von Davier et al., 2004). There are different kernels available in the literature (Lee et al., 2008), but in this study we focus on the Gaussian kernel, which is most commonly used. Following the notation of von Davier et al. (2004), let $\Phi(\cdot)$ denote the CDF of the Gaussian distribution, and let h_X denote the bandwidth parameter. Then, the Gaussian kernel smoothing of the distribution of X has the CDF defined by

$$F_{h_X}(x) = \sum_j r_j \Phi\big(R_{jX}(x)\big), \tag{7}$$

where $R_{jX}(x)$ is given by

$$R_{jX}(x) = \frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}, \tag{8}$$

and a_X, μ_X, and σ_X^2 are functions of r, that is

$$\mu_X = \sum_j x_j r_j, \tag{9}$$

$$\sigma_X^2 = \sum_j (x_j - \mu_X)^2 r_j, \tag{10}$$

$$a_X = \sqrt{\frac{\sigma_X^2}{\sigma_X^2 + h_X^2}}. \tag{11}$$

$\widehat{G}_{h_Y}(y)$ is defined analogously.

It is evident from equations (7)–(11) that for the continuous approximation to be carried out, the bandwidth parameters h_X and h_Y have to be estimated. The primary goal of introducing such parameters is to make the density functions as smooth as possible while retaining the properties of the original distributions. Various methods of estimating the bandwidth parameter have been suggested in previous research (Andersson et al., 2014; von Davier et al., 2004). Of particular interest to this study

is a method described in Holland and Thayer (1989) and von Davier et al. (2004), which estimates the bandwidth parameter by minimizing a penalty function, named PEN_1 (h_X) in von Davier et al. (2004), with respect to the bandwidth. The penalty function itself is based on the squared distances between proportions and the density function and is given by

$$PEN_1(h_X) = \sum_j \left(r_j - f_{h_X}(x_j)\right)^2,$$ (12)

where $f_{h_X}(x_j)$ is the density function, found by differentiating Equation 7 with respect to x, that is

$$f_{h_X}(x) = \sum_j r_j \phi\left(R_{jX}(x)\right) \frac{1}{a_X h_X},$$ (13)

and $R_{jX}(x)$ is given in Equation 8. The density obtained as a result of estimating the bandwidth by minimizing PEN_1 is typically a good approximation of the discrete score distribution. Note that sometimes it can be beneficial to smooth the density function further, in which case an additional component named PEN_2 in von Davier et al. (2004) can be applied combined with PEN_1. However, because of the complexity in accounting for estimation variability with PEN_2, we focus exclusively on PEN_1 in the present study.

Once the continuous approximations are obtained, the equating function estimator for equating X to Y is given by

$$\widehat{e}_Y\left(x; \hat{\boldsymbol{r}}, \hat{\boldsymbol{s}}\right) = \widehat{G}_{h_Y}^{-1}\left(\widehat{F}_{h_X}(x; \hat{\boldsymbol{r}}); \hat{\boldsymbol{s}}\right).$$ (14)

The equating function for equating Y to X is analogous and found by substitution.

Standard Error of Kernel Equating

The SEE is the measure of random equating error or uncertainty which stems from the equating function being subject to estimation and thereby sampling variability. We largely base this subsection on the work of Holland et al. (1989), who derived the asymptotic standard error for the kernel method of equating using the standard delta method for computing large sample approximations to the sampling variances of functions of statistics. Before proceeding, we see it appropriate to briefly introduce the multivariate delta method (Rao, 1973).

Adopting the notation of Rao (1973), let the ($k \times 1$)-dimensional random vector $\sqrt{n}(T_{kn} - \theta_k)$ converge to a multivariate normal distribution with zero mean and covariance Σ, where T_{kn} is an estimator, θ_k is the true parameter vector, and n denotes the sample size. Let g denote a vector-valued function with components $g_1, \ldots, g_q$, such that all the entries of g are differentiable. Then, $\sqrt{n}(g(T_{kn}) - g(\theta_k))$ converges to a multivariate normal distribution with zero mean and covariance of $G\Sigma G'$, that is

$$\sqrt{n}(g(T_{kn}) - g(\theta_k)) \xrightarrow{d} N\left(0, G\Sigma G'\right),$$ (15)

where G is the ($q \times k$) Jacobian matrix of partial derivatives of g with respect to θ_k. In this study, the equivalent of T_{kn}, θ_k, and g are the estimator of the score probabilities, the true score probabilities, and the equating function, respectively. The delta method was employed by Holland et al. (1989), where they defined the SEE for equating X to Y by

$$\mathrm{SEE}_Y(x) = \sqrt{\mathrm{Var}(\hat{e}_Y(x; \hat{\boldsymbol{r}}, \hat{\boldsymbol{s}}))}. \tag{16}$$

The SEE for equating Y to X is defined analogously.

Treating the bandwidth parameters h_X and h_Y as constants, Holland et al. (1989) assert that all the uncertainty in the equating function comes from the estimation of the score probabilities $\boldsymbol{r}$ and $\boldsymbol{s}$. Hence, the variance of the equating function, and in turn the SEE, reflects the data collection design, the choice of the pre-smoothing technique used in the estimation of the population score probabilities, and the equating function itself.

Reiterating the notation used previously, let $\boldsymbol{r}$ and $\boldsymbol{s}$ define the vectors of the pre-smoothed score distributions. The calculation of the SEE per Holland et al. (1989) then requires two components: the vector ∂_{e_Y} with derivatives of the equating function e_Y with respect to $\boldsymbol{r}$ and $\boldsymbol{s}$, and the asymptotic covariance matrix $\Sigma_{(\hat{\boldsymbol{r}}, \hat{\boldsymbol{s}})}$. Using the delta method (Rao, 1973), the variance of the equating function $\hat{e}_Y$ can then be expressed as

$$\mathrm{Var}(\hat{e}_Y(x; \hat{\boldsymbol{r}}, \hat{\boldsymbol{s}})) = \partial_{e_Y} \Sigma_{(\hat{\boldsymbol{r}}, \hat{\boldsymbol{s}})} [\partial_{e_Y}]', \tag{17}$$

where $\boldsymbol{s}$ and $\Sigma_{(\hat{\boldsymbol{r}}, \hat{\boldsymbol{s}})}$ is the covariance matrix of the independently estimated score probabilities, given by

$$\Sigma_{(\hat{\boldsymbol{r}}, \hat{\boldsymbol{s}})} = \begin{bmatrix} \Sigma_{\hat{\boldsymbol{r}}} & 0 \\ 0 & \Sigma_{\hat{\boldsymbol{s}}} \end{bmatrix}. \tag{18}$$

The matrix $\Sigma_{(\hat{\boldsymbol{r}}, \hat{\boldsymbol{s}})}$ has dimensions $(J + K) \times (J + K)$ where J is the dimension of $\boldsymbol{r}$ and K is the dimension of $\boldsymbol{s}$ (von Davier et al., 2004). The calculation of $\Sigma_{(\hat{\boldsymbol{r}}, \hat{\boldsymbol{s}})}$ for different equating designs can be found in Andersson and Wiberg (2017), Holland et al. (1989) and von Davier et al. (2004).

The second component, ∂_{e_Y}, can be defined as follows

$$\partial_{e_Y} = \left[\frac{\partial e_Y}{\partial \boldsymbol{r}}, \frac{\partial e_Y}{\partial \boldsymbol{s}} \right]. \tag{19}$$

Recalling Equation 14, the derivatives needed to compute ∂_{e_Y} are defined in Holland et al. (1989) as

$$\frac{\partial e_Y}{\partial r_j} = \frac{1}{G'} \frac{\partial F(x, \boldsymbol{r})}{\partial r_j}, \tag{20}$$

$$\frac{\partial e_Y}{\partial s_k} = -\frac{1}{G'} \frac{\partial G(e_Y(x); \boldsymbol{s})}{\partial s_k}, \tag{21}$$

where $\frac{\partial e_Y}{\partial \boldsymbol{r}}$ is a row vector with dimensions $1 \times J$, $\frac{\partial e_Y}{\partial \boldsymbol{s}}$ is a row vector with dimensions $1 \times K$, and G' is the density evaluated at $e_Y(x)$, that is

$$G' = \frac{\partial G(e_Y(x); \boldsymbol{s})}{\partial y}, \tag{22}$$

and

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

where $\frac{\partial F(x; \boldsymbol{r})}{\partial x}$ is given in Equation 13, $R_{jX}(x; \boldsymbol{r})$ in Equation 8, and

$$M_{jX}(x; \boldsymbol{r}) = \frac{1}{2}(x - \mu_X)(1 - a_X^2)z_{jX}^2 + (1 - a_X)x_j, \tag{24}$$

where z_{jX} is defined as

$$z_{jX} = \frac{x_j - \mu_X}{\sigma_X}. \tag{25}$$

The derivatives of e_X are analogous to those above and can be computed by substitution.

At this point, it is important to emphasize that the method of SEE calculation described above treats the bandwidth parameters h_X and h_Y as fixed and not as functions of $\boldsymbol{r}$ and $\boldsymbol{s}$. Hence, the additional variability introduced by the bandwidth estimation is currently not accounted for in the calculation of the SEEs, and consequently poses a challenge with respect to their accuracy (Holland et al., 1989; von Davier et al., 2004).

Accounting for Bandwidth Estimation Variability in Kernel Equating

In this section, we first derive the bandwidth parameter estimator variance and then introduce a modified method for the calculation of the analytical SEE that accounts for bandwidth estimation variability.

Asymptotic Variance and Standard Error of the Bandwidth Parameter Estimator

Recalling the standard delta method restated in the previous section (Rao, 1973), it is important to note that the bandwidth parameter estimator is not defined explicitly but rather as an implicit function of other asymptotically normal variables. Therefore, we use a generalization of the delta method presented by Benichou and Gail (1989) which facilitates computing the asymptotic variance of the implicitly defined bandwidth parameter estimator. Following the notation of von Davier et al. (2004), h_X denotes the bandwidth parameter selected to minimize PEN$_1$ defined by Equation (12) and $\boldsymbol{r}$ denotes the vector of estimated score probabilities. Consider further that PEN$_1$ is a continuously differentiable function of the estimated score probabilities $\boldsymbol{r}$ in h_X, and the function is minimized so that $\frac{\partial \text{PEN}_1}{\partial h_X} = 0$. Applying the implicit function theorem (Rao, 1973), we can then define h_X as a function of $\boldsymbol{r}$ such that $h_X = g_{h_X}(\boldsymbol{r})$, and compute the partial derivatives of $g_{h_X}(\boldsymbol{r})$ with respect to $\boldsymbol{r}$ as

$$\frac{\partial g_{h_X}(\boldsymbol{r})}{\partial \boldsymbol{r}} = -\left(\frac{\partial^2 \text{PEN}_1}{\partial h_X^2}\right)^{-1} \frac{\partial^2 \text{PEN}_1}{\partial h_X \partial \boldsymbol{r}'}, \tag{26}$$

where $\frac{\partial^2 \text{PEN}_1}{\partial h_X^2}$ is a scalar second order partial derivative of PEN$_1$ with respect to h_X and $\frac{\partial^2 \text{PEN}_1}{\partial h_X \partial \boldsymbol{r}'}$ is a $1 \times J$ vector of second-order partial derivatives of PEN$_1$ with respect to $\boldsymbol{r}$. The $\frac{\partial^2 \text{PEN}_1}{\partial h_X^2}$ and $\frac{\partial^2 \text{PEN}_1}{\partial h_X \partial \boldsymbol{r}'}$ derivatives are unequivocally calculated using the chain rule and implicit differentiation. The equations, however, are lengthy, and we summarize them in the appendix.

Let $\Sigma_{\hat{\boldsymbol{r}}}$ denote the asymptotic covariance matrix of the estimated score probabilities $\boldsymbol{r}$ with dimensions $J \times J$ where J is the dimension of $\boldsymbol{r}$. By applying the delta method for implicit functions (Benichou & Gail, 1989), we can define the asymptotic variance of the bandwidth parameter estimator $\widehat{h}_X$ as

$$\text{Var}(\widehat{h}_X) = \frac{\partial g_{h_X}(\boldsymbol{r})}{\partial \boldsymbol{r}} \Sigma_{\hat{\boldsymbol{r}}} \left[\frac{\partial g_{h_X}(\boldsymbol{r})}{\partial \boldsymbol{r}}\right]', \tag{27}$$

and its standard error as

$$\mathrm{SE}\left(\widehat{h}_X\right) = \sqrt{\mathrm{Var}\left(\widehat{h}_X\right)}. \tag{28}$$

The variance and the standard error of $\widehat{h}_Y$ are analogous to those given for $\widehat{h}_X$ and can be computed by substituting X by Y and r by s.

Standard Error of Equating Accounting for Bandwidth Variability

To account for the bandwidth estimation variability when computing the SEEs, we apply the chain rule together with the delta method and obtain a modified expression for the SEEs (Cox, 1984). Treating $\widehat{h}_X$ and $\widehat{h}_Y$ as functions of the score probability estimators $\hat{r}$ and $\hat{s}$, we redefine Equation 14 by adding a term which accounts for the bandwidth estimation variability to obtain

$$
\begin{aligned}
\mathrm{Var}\left(\widehat{e}_Y\left(x;\,\hat{r},\hat{s},\widehat{h}_X(\hat{r}),\widehat{h}_Y(\hat{s})\right)\right) = &\ \frac{\partial e_Y(x)}{\partial(r,s)}\Sigma_{(\hat{r},\hat{s})}\left[\frac{\partial e_Y(x)}{\partial(r,s)}\right]' + \frac{\partial(h_X(r),h_Y(s))}{\partial(r,s)}\frac{\partial e_Y(x)}{\partial(h_X,h_Y)} \\
&\times \Sigma_{(\hat{r},\hat{s})}\left[\frac{\partial(h_X(r),h_Y(s))}{\partial(r,s)}\frac{\partial e_Y(x)}{\partial(h_X,h_Y)}\right]',
\end{aligned} \tag{29}
$$

where $\frac{\partial e_Y(x)}{\partial(r,s)}$ is presented in equations (19)–(25) and $\Sigma_{(\hat{r},\hat{s})}$ in Equation 18. When evaluating these expressions in practice, the true parameters are replaced by the parameter estimates. The additional components are then a $(2 \times (J+K))$-matrix of partial derivatives of the bandwidth parameters as functions of the estimated score probabilities with respect to the estimated score probabilities, $\frac{\partial(h_X(r),h_Y(s))}{\partial(r,s)}$, calculated following Equation (26), that is

$$
\frac{\partial(h_X(r),h_Y(s))}{\partial(r,s)} = \begin{bmatrix} \left(-\left(\dfrac{\partial^2\mathrm{PEN}_1}{\partial h_X^2}\right)^{-1}\dfrac{\partial^2\mathrm{PEN}_1}{\partial h_X\partial r'}\right) & 0 \\[2em] 0 & \left(-\left(\dfrac{\partial^2\mathrm{PEN}_1}{\partial h_Y^2}\right)^{-1}\dfrac{\partial^2\mathrm{PEN}_1}{\partial h_Y\partial s'}\right) \end{bmatrix}, \tag{30}
$$

and $\frac{\partial e_Y(x)}{\partial(h_X,h_Y)}$, a $(2 \times J)$-matrix of first-order derivatives of the equating function with respect to the bandwidth parameters, h_X and h_Y, defined by

$$\frac{\partial e_Y}{\partial h_X} = \frac{1}{G'}\sum_j r_j\frac{\partial\Phi\left(R_{jX}(x)\right)}{\partial h_X}, \tag{31}$$

where

$$\frac{\partial\Phi\left(R_{jX}(x)\right)}{\partial h_X} = \sum_j r_j\phi\left(R_{jX}(x)\right)\frac{\partial R_{jX}(x)}{\partial h_X}, \tag{32}$$

and

$$\frac{\partial e_Y}{\partial h_Y} = -\frac{1}{G'}\sum_k r_k\frac{\partial\Phi\left(R_{kY}(y)\right)}{\partial h_Y}, \tag{33}$$

where

$$\frac{\partial \Phi(R_{kY}(y))}{\partial h_Y} = \sum_k r_k \phi(R_{kY}(y)) \frac{\partial R_{kY}(y)}{\partial h_Y}, \tag{34}$$

and G' is defined in Equation 22, with $R_{jX}(x)$ and $R_{kY}(y)$ given in Equation 8. Lastly, we define the SEE which accounts for bandwidth variability by

$$\text{SEE}_Y(x) = \sqrt{\text{Var}\left(\widehat{e}_Y\left(x; \widehat{\boldsymbol{r}}, \widehat{\boldsymbol{s}}, \widehat{h}_X(\widehat{\boldsymbol{r}}), \widehat{h}_Y(\widehat{\boldsymbol{s}})\right)\right)}. \tag{35}$$

Simulation Study

To confirm the accuracy of the presented derivations, we conducted a simulation study to evaluate the estimators of the standard error of the bandwidth parameter estimator and the modified SEEs. We evaluated the estimated standard errors with respect to the Monte Carlo standard errors and compared the modified standard errors to the original standard errors that do not account for the bandwidth estimation.

Simulation Design

Data for two test forms X and Y were simulated using the two-parameter logistic (2-PL) model within the framework of IRT (de Ayala, 2009), where test lengths of 20, 40, and 80 items were considered. The discrimination parameters for both test forms were selected from the U (0.5, 2)-distribution and the difficulty parameters for one test form were selected from the N (0.25, 1)-distribution and the other from the $N(-0.25, 1)$-distribution. These distributions were considered to mimic realistic item parameters found in standardized testing (National Center for Education Statistics, 2004).

The equivalent groups design was used in which two independent random samples of individuals are drawn from a single common population and where each random sample takes either of the test forms X and Y (von Davier et al., 2004). Dictated by the design, no differences in the latent distributions were present between the groups. The latent distributions were set to the standard normal distribution. The equivalent groups design was considered because of its simplicity. Relative to other data collection designs, it provided an opportunity for direct comparison of the results on the test forms X and Y without additional considerations or assumptions. The score distributions for the tests X and Y with 20, 40, and 80 items are provided in Figure 1. The means (SD) of the test score distributions with 20, 40, and 80 items were 10.35 (4.52), 21.40 (8.43), and 44.03 (16.06) for test X and 7.58 (4.24), 18.17 (8.22), and 35.95 (16.89) for test Y.

In order to systematically verify the accuracy of the modified method of calculating the SEE as well as to explore how well it performs in a variety of sample sizes, sample sizes 1000, 4000, and 16,000 were considered. The study was conducted using version 3.6.2 of the statistical software environment R (R Core Team, 2019), primarily employing the packages *kequate* (Andersson et al., 2013), *mirt* (Chalmers, 2012), and *numDeriv* (Gilbert & Varadhan, 2019), while also utilizing newly written code implementing the modified SEEs (available in the Supplementary material). In each simulation setting, we used 10,000 replications which enabled us to all but eliminate the simulation random error. The convergence rate for all simulation settings was 100%.
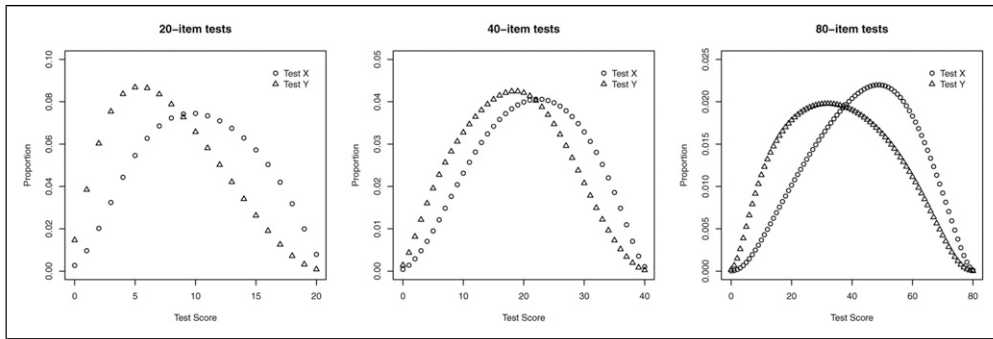
Figure 1. Score distributions with 20, 40, and 80 items, for each test *X* and *Y*.

The study followed the recommended kernel equating procedure (von Davier et al., 2004), albeit with a few adjustments to verify the derivations presented in this article. For each generated data set, the following steps were carried out:

(1) Pre-smoothing. The package *mirt* (Chalmers, 2012) was used to pre-smooth the irregularities of the raw data by estimating two separate 2-PL models to obtain item parameter estimates pertaining to each of the two groups and tests. The expectation-maximization (EM) algorithm was used for estimation (Bock & Aitkin, 1981), with the tolerance level 0.0001 and maximum number of iterations equal to 500. The asymptotic covariance matrix was estimated based on the method described in Oakes (1999).

(2) Estimation of score probabilities. Under the equivalent groups design, the score probabilities $\widehat{r}_j$ and $\widehat{s}_k$ were estimated based on the item parameter estimates and the assumed distribution of the latent variable (Andersson & Wiberg, 2017; von Davier et al., 2004).

(3) Continuous approximation. By adapting code from the package *kequate* (Andersson et al., 2013), continuous approximations to the discrete distributions were obtained by applying a Gaussian kernel with an optimal bandwidth parameter. Optimal bandwidth parameters $\widehat{h}_X$ and $\widehat{h}_Y$ were obtained by minimizing the first part of the penalty function, PEN_1 (von Davier et al., 2004). When optimizing the penalty function, a golden section search with successive parabolic interpolation (Brent, 1973) using the default tolerance of 1.50×10^{-8} was used.

The analytical derivations for the bandwidth parameter estimator variance were paramount to the study. Hence, upon obtaining the optimal bandwidths, the average standard errors of the bandwidth parameters were computed following the equations introduced in the previous section, and their accuracy was assessed using the Monte Carlo standard error (MCSE) as the criterion. When calculating the asymptotic variance of the bandwidth parameter estimator, the bandwidth parameters h_X and h_Y were replaced with the estimated parameters $\widehat{h}_X$ and $\widehat{h}_Y$, and the asymptotic covariance matrices of the estimated score probabilities $\Sigma_{\widehat{r}}$ and $\Sigma_{\widehat{s}}$ were calculated based on the implementation in *kequate* (Andersson et al., 2013).

(4) Equating. Upon obtaining continuous CDFs, an equipercentile equating function was applied to equate the test forms *X* and *Y*.

(5) Calculating the SEE. The average analytical SEEs were computed using the original method for calculating the SEE without accounting for the bandwidth variability (Holland et al., 1989), and the modified method of calculating the SEE accounting for the bandwidth variability. The Monte Carlo SEEs (MCSEE) were used as a criterion for comparing the accuracy of the modified and the original methods of the SEE calculation.

Furthermore, two measures were used to assess the performance and the accuracy of the modified method as compared to the original method. We computed the absolute differences of the means of the SEEs calculated with the original and the modified methods. Additionally, the average coverage probabilities were considered which explored the average proportion of time that the 95% confidence intervals calculated employing the original and the modified methods contained the true values of the equated results. The confidence intervals were estimated with $\widehat{e}_Y \pm z_{0.975} \times \widehat{\mathrm{SE}}(\widehat{e}_Y)$, with $z_{0.975}$ indicating the 0.975 quantile of the standard normal distribution.

The analytical derivations used in computing the bandwidth estimator variance and standard errors, as well as the SEE, were verified numerically using the R package *numDeriv* (Gilbert & Varadhan, 2019). The R code is available for review in the Supplementary material.

Simulation Results

The study largely depended on the accuracy of the asymptotic variance and standard error of the bandwidth parameter estimator derivations. The results of the simulation for the standard errors of the bandwidth parameter estimators $\widehat{h}_X$ and $\widehat{h}_Y$ given in Table 1 confirmed that the derivations were correct, and the asymptotic standard errors of the bandwidth parameter estimator (ASE) were accurate as witnessed by the comparison to the MCSE. As can be expected for asymptotic variance approximation (Ferguson, 1996), the differences between the ASEs and the MCSEs were larger in smaller sample sizes.

Subsequently incorporating the bandwidth estimation variability into computing the modified SEEs, Table 2 presents two performance measures used to compare the accuracy of the original standard errors of equating (ASEE) and the modified asymptotic standard errors of equating (ASEE_mod). These measures are the absolute aggregate differences between the SEEs for two pairs, ASEE - MCSEE and ASEE_mod - MCSEE, and the average coverage for both the original and the modified methods. From the average differences, it was evident that when compared to the MCSEE estimates, the modified asymptotic SEEs which take bandwidth variability into account were accurate for all sample sizes and test lengths. Furthermore, the modified asymptotic SEEs in most cases appeared to be nearly identical to those not accounting for bandwidth variability, suggesting that the bandwidth estimation influence on the SEEs was minimal. This finding was

Table 1. Asymptotic Standard Errors (ASE) and Monte Carlo Standard Errors (MCSE) for the Bandwidth Parameters h_X and h_Y With Sample Sizes N and Test Lengths Of 20, 40, and 80 Items.

N	h_X Parameter		h_Y Parameter	
	ASE	MCSE	ASE	MCSE
20 items				
1000	0.0035	0.0035	0.0035	0.0035
4000	0.0018	0.0017	0.0018	0.0017
16000	0.0009	0.0009	0.0009	0.0009
40 items				
1000	0.0034	0.0034	0.0037	0.0037
4000	0.0017	0.0017	0.0019	0.0019
16000	0.0009	0.0009	0.0009	0.0009
80 items				
1000	0.0033	0.0033	0.0036	0.0036
4000	0.0016	0.0016	0.0018	0.0018
16000	0.0008	0.0008	0.0009	0.0009

Table 2. Absolute Average Differences for the Original Asymptotic Standard Errors of Equating (ASEE) and the Modified Asymptotic Standard Errors of Equating (ASEE$_{mod}$) to the Monte Carlo Standard Errors of Equating (MCSEE) and Average Coverage of 95% Confidence Intervals Based on the ASEE and the ASEE$_{mod}$.

	Average Differences		Average Coverage	
N	ASEE-MCSEE	ASEE$_{mod}$-MCSEE	ASEE	ASEE$_{mod}$
20 items				
1000	0.0012	0.0016	94.97	95.03
4000	0.0002	0.0004	95.07	95.13
16000	0.0004	0.0004	94.87	94.93
40 items				
1000	0.0021	0.0019	94.97	94.99
4000	0.0012	0.0011	94.82	94.84
16000	0.0011	0.0011	94.66	94.69
80 items				
1000	0.0038	0.0041	94.87	94.90
4000	0.0050	0.0054	95.39	95.42
16000	0.0004	0.0006	94.90	94.93

further supported by the average coverage for both the original and the modified methods. Although the modified method performed better in most settings, the differences in coverage were small.

Discussion

The kernel method of equating is an equipercentile equating method in which number-correct scores are transformed into percentile rank scores from test form X to the scale of test form Y, and the scores from the two test forms with the same percentile rank are considered to be equivalent (von Davier et al., 2004). However, in order to obtain those equivalent scores, continuous approximations to the discrete score distributions are needed. To satisfy this requirement, kernel equating uses a Gaussian kernel with a smoothing bandwidth parameter that determines the characteristics of the continuous approximations to the raw discrete distributions (von Davier et al., 2004). The most commonly used method for bandwidth estimation is minimizing a penalty function with respect to the bandwidth parameter (von Davier et al., 2004). The bandwidth, in turn, is influenced by the estimated score probabilities and therefore is subject to variability. This variability, however, is not currently accounted for when calculating the SEE (Holland et al., 1989), challenging its accuracy and, ultimately, the fairness of the equated results.

The present study explored the issue of the additional variability stemming from the bandwidth estimation and its impact on the SEE. Building on the existing methodology of Holland et al. (1989) and von Davier et al. (2004), we derived the asymptotic variance of the bandwidth parameter estimator using the delta method for implicit functions (Benichou & Gail, 1989) and incorporated those derivations to expand the existing formulas for calculating the SEE (Holland et al., 1989). Thus, we have introduced SEEs that account for bandwidth estimation variability. A simulation study with 18 data sets generated for a wide range of sample sizes and test lengths was used to illustrate the results of the modified method as compared to the current method of the SEE calculation (Holland et al., 1989) and the MCSEEs.

The results offered several observations which are valuable to the testing industry. Firstly, the newly introduced SEE were accurate and close to the MCSEE estimates for all sample sizes and

test lengths, suggesting that the method is suitable for practical use. Secondly, using the MCSEE as a criterion, the results of the study indicate that the original (Holland et al., 1989) and the modified SEEs produce similar results, suggesting that the bandwidth estimation impact on the SEE is minimal.

The presented results apply directly to any pre-smoothing method, provided that the asymptotic covariance matrix of the score probabilities has been defined for such a method. However, in this study we only utilized IRT as the pre-smoothing method and the results may be different if instead using, for example, log-linear models. However, previous research has indicated that the SEEs are fairly accurate even when not accounting for the bandwidth estimation with the penalty function, and so we do not anticipate that the results will differ substantially when using pre-smoothing with log-linear models instead of IRT models.

The method for accounting for bandwidth estimation that we used in the present study can be generalized to additional kernels and equating designs by modifying the presented results to account for the different expressions of the equating function and score probabilities with such approaches. It is furthermore possible to utilize the delta method for implicit functions with other bandwidth estimation methods provided that those specify a function that is minimized which fulfills the properties required for the implicit function theorem and the delta method. One approach which does not fulfill these requirements is the method based on PEN_2, since the function PEN_2 is not differentiable and can have multiple local minima.

It is important to note that in this study, we derived the modified asymptotic SEE for two test forms in the setting of the equivalent groups data collection design. It can be the case that the bandwidth estimation influence on the SEE is greater for other data collection and equating designs. It would, therefore, be beneficial for future theoretical and empirical studies to focus on determining the bandwidth estimation impact on the SEE in these additional scenarios.

As a final note, we believe that it is theoretically more sound to use a method which successfully accounts for all sources of variability, however negligible those may be. Introducing the modifications to the formulas for the SEE calculation akin to those explored in this study can improve the accuracy of the standard errors of equating, and consequently, facilitate fairness and comparability of the equated results.

Appendix

Computation of the Penalty Function

In order to compute Equation 26, two partial derivatives of the PEN_1 (h_X) function need to be defined, the second-order partial derivative of PEN_1 (h_X) with respect to h_X and the second-order partial derivative of PEN_1 (h_X) with respect to h_X and r. Hence, we first need to compute the partial derivative of PEN_1 (h_X) with respect to h_X.

Recalling equations (8)–(13), we define the first partial derivative of PEN_1 (h_X) with respect to h_X as

$$
\begin{aligned}
\frac{\partial PEN_1}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\sum_j \left(r_j - f_{h_X}(x_j)^2\right)\right] \\
&= -2\sum_j \left(\left(r_j - f_{h_X}(x_j)\right)\frac{\partial f_{h_X}(x_j)}{\partial h_X}\right).
\end{aligned}
\tag{36}
$$

We then need to calculate $\frac{\partial f_{h_X}(x_j)}{\partial h_X}$ as

$$
\begin{aligned}
\frac{\partial f_{h_X}(x_j)}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\sum_j r_j \phi(R_{jX}(x))\frac{1}{a_X h_X}\right] \\
&= \sum_j r_j\left(\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\frac{1}{a_X h_X} + \phi(R_{jX}(x))\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right),
\end{aligned}
\tag{37}
$$

where $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ is

$$
\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X} = -\phi(R_{jX}(x))\frac{\partial R_{jX}(x)}{\partial h_X}R_{jX}(x),
\tag{38}
$$

with

$$
\begin{aligned}
\frac{\partial R_{jX}(x)}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}\right] \\
&= \frac{\partial a_X}{\partial h_X}(\mu_X - x_j)\frac{1}{a_X h_X} + (x - a_X x_j - (1 - a_X)\mu_X)\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right].
\end{aligned}
\tag{39}
$$

The remaining components needed for computing the first partial derivative of PEN_1 (h_X) with respect to h_X are then $\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]$ and $\frac{\partial a_X}{\partial h_X}$. Thus we calculate

$$
\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right] = -a_X^{-2}\frac{\partial a_X}{\partial h_X}\frac{1}{h_X} - \frac{1}{a_X h_X^2},
\tag{40}
$$

$$
\begin{aligned}
\frac{\partial a_X}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\frac{\sigma_X}{\sqrt{\sigma_X^2 + h_X^2}}\right] = -\frac{\sigma_X}{2(h_X^2 + \sigma_X^2)^{\frac{3}{2}}}\frac{\partial h_X^2}{\partial h_X} + \frac{\partial \sigma_X^2}{\partial h_X} \\
&= -\frac{\sigma_X h_X}{(h_X^2 + \sigma_X^2)^{\frac{3}{2}}}.
\end{aligned}
\tag{41}
$$

Using equations (36)–(41), we can compute the second partial derivative of PEN_1 (h_X) with respect to h_X as

$$
\begin{aligned}
\frac{\partial^2 \text{PEN}_1}{\partial h_X^2} &= \frac{\partial}{\partial h_X}\left[-2\sum_j\left((r_j - f_{h_X}(x_j))\frac{\partial f_{h_X}(x_j)}{\partial h_X}\right)\right] \\
&= -2\sum_j\left(\frac{\partial^2 f_{h_X}(x_j)}{\partial h_X^2}(r_j - f_{h_X}(x_j)) - \left[\frac{\partial f_{h_X}(x_j)}{\partial h_X}\right]^2\right),
\end{aligned}
\tag{42}
$$

where $\frac{\partial f_{h_X}(x_j)}{\partial h_X}$ is defined in Equation 37 and $\frac{\partial^2 f_{h_X}(x_j)}{\partial h_X^2}$ is given by

$$\frac{\partial^2 f_{h_X}(x_j)}{\partial h_X^2} \quad = \frac{\partial}{\partial h_X}\left[\sum_j r_j \left(\frac{\partial\left[\phi\left(R_{jX}(x)\right)\right]}{\partial h_X}\frac{1}{a_X h_X} + \phi\left(R_{jX}(x)\right)\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right)\right]$$

$$= \sum_j r_j\left(\frac{\partial^2\left[\phi\left(R_{jX}(x)\right)\right]}{\partial h_X^2}\frac{1}{a_X h_X} + \frac{\partial\left[\phi\left(R_{jX}(x)\right)\right]}{\partial h_X}\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right) \quad (43)$$

$$+ \sum_j r_j\left(\frac{\partial\left[\phi\left(R_{jX}(x)\right)\right]}{\partial h_X}\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right] + \phi\left(R_{jX}(x)\right)\frac{\partial^2}{\partial h_X^2}\left[\frac{1}{a_X h_X}\right]\right).$$

Recall that $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ is given in Equation 38 and $\frac{\partial R_{jX}(x)}{\partial h_X}$ in Equation 39. Hence, we define $\frac{\partial^2[\phi(R_{jX}(x))]}{\partial h_X^2}$ and $\frac{\partial^2 R_{jX}(x)}{\partial h_X^2}$ as

$$\frac{\partial^2\left[\phi\left(R_{jX}(x)\right)\right]}{\partial h_X^2} \quad = \frac{\partial}{\partial h_X}\left[-\phi\left(R_{jX}(x)\right)\frac{\partial R_{jX}(x)}{\partial h_X}R_{jX}(x)\right]$$

$$= -\frac{\partial\left[\phi\left(R_{jX}(x)\right)\right]}{\partial h_X}\frac{\partial R_{jX}(x)}{\partial h_X}R_{jX}(x) - \phi\left(R_{jX}(x)\right)\frac{\partial^2 R_{jX}(x)}{\partial h_X^2}R_{jX}(x) \quad (44)$$

$$- \phi\left(R_{jX}(x)\right)\left[\frac{\partial\left[R_{jX}(x)\right]}{\partial h_X}\right)\right]^2,$$

$$\frac{\partial^2 R_{jX}(x)}{\partial h_X^2} \quad = \frac{\partial}{\partial h_X}\left[\frac{\partial a_X}{\partial h_X}\left(\mu_X - x_j\right)\frac{1}{a_X h_X} + \left(x - a_X x_j - (1-a_X)\mu_X\right)\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right]$$

$$= \left(\mu_X - x_j\right)\left(\frac{\partial^2 a_X}{\partial h_X^2}\frac{1}{a_X h_X} + \frac{\partial a_X}{\partial h_X}\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right) \quad (45)$$

$$+ \left(\mu_X - x_j\right)\frac{\partial a_X}{\partial h_X}\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right] + \left(x - a_X x_j - (1-a_X)\mu_X\right)\frac{\partial^2}{\partial h_X^2}\left[\frac{1}{a_X h_X}\right].$$

Consider further that $\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]$ is defined in Equation 40 and $\frac{\partial a_X}{\partial h_X}$ in Equation 41. We can then observe that $\frac{\partial^2}{\partial h_X^2}\left[\frac{1}{a_X h_X}\right]$ can be computed as

$$\frac{\partial^2}{\partial h_X^2}\left[\frac{1}{a_X h_X}\right] = \frac{1}{h_X a_X^2}\left(\frac{2}{a_X}\left[\frac{\partial a_X}{\partial h_X}\right]^2 - \frac{\partial^2 a_X}{\partial h_X^2} + \frac{\partial a_X}{\partial h_X}\frac{1}{h_X}\right)$$

$$+ \left(\frac{\partial a_X}{\partial h_X}\frac{1}{a_X^2 h_X^2} + \frac{2}{a_X h_X^3}\right), \quad (46)$$

and

$$\frac{\partial^2 a_X}{\partial h_X^2} = \frac{\partial}{\partial h_X}\left[-\frac{\sigma_X h_X}{\left(h_X^2 + \sigma_X^2\right)^{\frac{3}{2}}}\right] = \frac{-\sigma_X}{\left(h_X^2 + \sigma_X^2\right)^{\frac{3}{2}}} - \frac{3h_X^2}{\left(h_X^2 + \sigma_X^2\right)^{\frac{5}{2}}}. \quad (47)$$

Lastly, we can compute the second partial derivative of $\text{PEN}_1\left(h_X\right)$ with respect to $\boldsymbol{r}$ as follows

$$\frac{\partial^2 \text{PEN}_1}{\partial h_X \partial r_i} = \frac{\partial}{\partial r_i}\left[-2\sum_j\left(\left(r_j - f_{h_X}(x_j)\right)\frac{\partial f_{h_X}(x_j)}{\partial h_X}\right)\right]$$

$$= -2\sum_j\left[\left(\frac{\partial r_j}{\partial h_X \partial r_i} - \frac{\partial f_{h_X}(x_j)}{\partial r_i}\right)\frac{\partial f_{h_X}(x_j)}{\partial h_X} + \left(r_j - f_{h_X}(x_j)\right)\frac{\partial^2 f_{h_X}(x_j)}{\partial h_X \partial r_i}\right], \tag{48}$$

where $\frac{\partial r_j}{\partial h_X \partial r_i} = 1$ if $i = j$ and $\frac{\partial r_j}{\partial h_X \partial r_i} = 0$ if $i \neq j$. Note that $\frac{\partial f_{h_X}(x_j)}{\partial h_X}$ is given in Equation 37. Then, the components needed for computing Equation 48 are $\frac{\partial f_{h_X}(x_j)}{\partial r_i}$ and $\frac{\partial^2 f_{h_X}(x_j)}{\partial h_X \partial r_i}$. We define $\frac{\partial f_{h_X}(x_j)}{\partial r_i}$ as

$$\frac{\partial f_{h_X}(x)}{\partial r_i} = \frac{\partial}{\partial r_i}\left[\sum_j r_j \phi\left(R_{jX}(x)\right)\frac{1}{a_X h_X}\right]$$

$$= \frac{1}{h_X}\left[\phi\left(R_{jX}(x)\right)\frac{1}{a_X} - \frac{\partial R_{jX}(x)}{\partial r_i}\sum_j\left(r_j\phi\left(R_{jX}(x)\right)R_{jX}(x)\right)\frac{1}{a_X}\right] \tag{49}$$

$$+ \frac{1}{h_X}\left[\frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\sum_j\left(r_j\phi\left(R_{jX}(x)\right)\right)\right],$$

where $\frac{\partial R_{jX}}{\partial r_i}$ and $\frac{\partial}{\partial r}\left[\frac{1}{a_X}\right]$ are given in Holland et al. (1989) as

$$\frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right] = -\frac{1}{2}a_X\frac{h_X^2}{\sigma_X^2}\frac{x_i^2 - \mu_X^2}{\sigma_X^2}, \tag{50}$$

and

$$\frac{\partial R_{jX}}{\partial r_i} = \left(-\frac{1}{a_X h_X}\right)\left[\frac{1}{2}(x - \mu_X)\left(1 - a_X^2\right)\left(\frac{x_i^2 - \mu_X^2}{\sigma_X^2}\right) + (1 - a_X)x_i\right]. \tag{51}$$

We further define $\frac{\partial^2 f_{h_X}(x_j)}{\partial h_X \partial r_i}$ as

$$\frac{\partial^2 f_{h_X}(x_j)}{\partial h_X \partial r_i} = \frac{\partial}{\partial h_X}\left[\frac{\partial f_{h_X}(x_j)}{\partial r_i}\right]. \tag{52}$$

Given Equation 49 is a lengthy expression, we further simplify the notation such that

$$\frac{\partial}{\partial h_X}\left[\frac{\partial f_{h_X}(x_j)}{\partial r_i}\right] = \frac{\partial}{\partial h_X}\left[\frac{1}{h_X}\right] \times P + \frac{1}{h_X} \times \frac{\partial P}{\partial h_X}$$

$$= -\frac{1}{h_X^2} \times P + \frac{1}{h_X} \times \frac{\partial P}{\partial h_X}, \tag{53}$$

where

$$P = \frac{\partial\left[\phi\left(R_{jX}(x)\right)\right]}{\partial h_X}\frac{1}{a_X} - \frac{\partial R_{jX}(x)}{\partial r_i}\sum_j\left(r_j\phi\left(R_{jX}(x)\right)R_{jX}(x)\right)\frac{1}{a_X}$$

$$+ \frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\sum_j\left(r_j\phi\left(R_{jX}(x)\right)\right). \tag{54}$$

Noting the three components in Equation 54, $\frac{\partial P}{\partial h_X}$ can then be presented as follows

$$\frac{\partial P}{\partial h_X} = \frac{\partial P1}{\partial h_X} - \frac{\partial P2}{\partial h_X} + \frac{\partial P3}{\partial h_X}. \tag{55}$$

$\frac{\partial P1}{\partial h_X}$ is given by

$$\begin{aligned}
\frac{\partial P1}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\phi\big(R_{jX}(x)\big)\frac{1}{a_X}\right] \\
&= \frac{\partial\big[\phi\big(R_{jX}(x)\big)\big]}{\partial h_X}\frac{1}{a_X} + \phi\big(R_{jX}(x)\big)\frac{a_X h_X}{\sigma_X^2},
\end{aligned} \tag{56}$$

where $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ is given in Equation 38. $\frac{\partial P2}{\partial h_X}$ is defined as

$$\begin{aligned}
\frac{\partial P2}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\frac{\partial R_{jX}(x)}{\partial r_i}\sum_j\big(r_j\phi\big(R_{jX}(x)R_{jX}(x)\big)\big)\frac{1}{a_X}\right] \\
&= \frac{\partial^2 R_{jX}(x)}{\partial h_X \partial r_i}\sum_j\big(r_j\phi\big(R_{jX}(x)\big)R_{jX}(x)\big)\frac{1}{a_X} \\
&\quad + \frac{\partial R_{jX}(x)}{\partial r_i}\sum_j\left(r_j\frac{\partial\big[\phi\big(R_{jX}(x)\big)\big]}{\partial h_X}R_{jX}(x) + r_j\phi\big(R_{jX}(x)\big)\frac{\partial R_{jX}(x)}{\partial h_X}\right)\frac{1}{a_X} \\
&\quad + \frac{\partial R_{jX}(x)}{\partial r_i}\sum_j\big(r_j\phi\big(R_{jX}(x)R_{jX}(x)\big)\big)\frac{a_X h_X}{\sigma_X^2},
\end{aligned} \tag{57}$$

where $\frac{\partial R_{jX}(x)}{\partial r_i}$ is defined in Equation 51, $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ in Equation 38, and $\frac{\partial^2 R_{jX}(x)}{\partial h_X \partial r_i}$ is given by

$$\begin{aligned}
\frac{\partial^2 R_{jX}(x)}{\partial h_X \partial r_i} &= \frac{\partial}{\partial h_X}\left[\left(-\frac{1}{a_X h_X}\right)\left(\frac{1}{2}(x-\mu_X)\big(1-a_X^2\big)\left(\frac{x^2-\mu_X^2}{\sigma_X^2}\right)+(1-a_X)x\right)\right] \\
&= \left(\frac{1}{a_X h_X^2}-\frac{a_X}{\sigma_X^2}\right)\left(\frac{1}{2}(x-\mu_X)\big(1-a_X^2\big)\left(\frac{x_i^2-\mu_X^2}{\sigma_X^2}\right)+(1-a_X)x\right) \\
&\quad + \left(-\frac{1}{a_X h_X}\right)\left(\left(-(x-\mu_X)\left(\frac{x^2-\mu_X^2}{\sigma_X^2}\right)a_X\frac{\partial a_X}{\partial h_X}\right)+x\frac{\partial a_X}{\partial h_X}\right).
\end{aligned} \tag{58}$$

It remains to calculate $\frac{\partial P3}{\partial h_X}$ as follows

$$\begin{aligned}
\frac{\partial P3}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\sum_j\big(r_j\phi\big(R_{jX}(x)\big)\big)\right] \\
&= \frac{\partial^2}{\partial h_X \partial r_i}\left[\frac{1}{a}\right]\sum_j\big(r_j\phi\big(R_{jX}(x)\big)\big) + \frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\frac{\partial}{\partial h_X}\left[\sum_j\big(r_j\phi\big(R_{jX}(x)\big)\big)\right] \\
&= -\left(\left(\frac{x_i^2-\mu_X^2}{\sigma_X^2}\right)\frac{1}{2\sigma_X^2}\right)\left(\frac{\partial a_X}{\partial h_X}h_X^2 + 2a_X h_X\right)\sum_j\big(r_j\phi\big(R_{jX}(x)\big)\big) \\
&\quad + \frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\sum_j\left(r_j\frac{\partial\big[\phi\big(R_{jX}(x)\big)\big]}{\partial h_X}\right),
\end{aligned} \tag{59}$$

where $\frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]$ is given in Equation 50, $\frac{\partial a_X}{\partial h_X}$ in Equation 41, and $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ in Equation 38.

The partial derivatives of the $\text{PEN}_1 (h_Y)$ with respect to h_Y, $\frac{\partial \text{PEN}_1}{\partial h_Y}$, $\frac{\partial^2 \text{PEN}_1}{\partial h_Y}$ and $\frac{\partial^2 \text{PEN}_1}{\partial h_Y \partial s_t}$, are computed analogously.

Declaration of Conflicting Interests

Funding

ORCID iDs

Kseniia Marcq https://orcid.org/0000-0001-8215-0667
Björn Andersson https://orcid.org/0000-0002-9007-2440

References

Andersson, B., Bränberg, K., & Wiberg, M. (2013). Performing the kernel method of test equating with the package kequate. *Journal of Statistical Software*, *55*(6), 1–25. https://doi.org/10.18637/jss.v055.i06

Andersson, B., & von Davier, A. A., (2014). Improving the bandwidth selection in kernel equating. *Journal of Educational Measurement*, *51*(3), 223–238. https://doi.org/10.1111/jedm.12044

Andersson, B., & Wiberg, M. (2017). Item response theory observed-score kernel equating. *Psychometrika*, *82*(1), 48–66. https://doi.org/10.1007/s11336-016-9528-7

Benichou, J., & Gail, M. H. (1989). A delta method for implicitly defined random variables. *The American Statistician*, *43*(1), 41–44. https://doi.org/10.1080/00031305.1989.10475608

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*(4), 443–459. https://doi.org/10.1007/BF02293801

Brent, R. (1973). *Algorithms for minimization without derivatives*. Prentice-Hall.

Chalmers, R. P. (2012). Mirt : A multidimensional item response theory package for the R Environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Cox, C. (1984). An elementary introduction to maximum likelihood estimation for multinomial models: Birch's theorem and the delta method. *The American Statistician*, *38*(4), 283–287. https://doi.org/10.1080/00031305.1984.10483226

de Ayala, R. J (2009). *The theory and practice of item response theory*. Guilford Press.

Ferguson, T. S (1996). *A course in large sample theory*. Routledge.

Gilbert, P., & Varadhan, R (2019). Accurate numerical derivatives [Computer software manual]. https://CRAN.R-project.org/package=numDeriv (R package version 2016.8-1.1)

Holland, P. W., King, B. F., & Thayer, D. T. (1989). The standard error of equating for the kernel method of equating score distributions. *ETS Research Report Series*, *1989*(1), i–52. https://doi.org/10.1002/j.2330-8516.1989.tb00332.x

Holland, P. W., & Thayer, D. T. (1987). Notes on the use of log-linear models for fitting discrete probability distributions. *ETS Research Report Series*, *1987*(2), i–40. https://doi.org/10.1002/j.2330-8516.1987.tb00235.x

Holland, P. W., & Thayer, D. T. (1989). The kernel method of equating score distributions. *ETS Research Report Series*, *1989*(1), i–45. https://doi.org/10.1002/j.2330-8516.1989.tb00333.x

Kolen, M. J., & Brennan, R. L (2014). Test equating, scaling, and linking: methods and practices (3rd ed.). Springer.

Lee, Y.-H., & von Davier, A. A. (2008). Comparing alternative kernels for the kernel method of test equating: Gaussian, logistic, and uniform kernels. *ETS Research Report Series*, *2008*(1), i–26. https://doi.org/10.1002/j.2333-8504.2008.tb02098.x

National Center for Education Statistics. (2004). NAEP technical documentation - mathematics assessment IRT parameters. https://nces.ed.gov/nationsreportcard/tdw/analysis/scaling_irt_math.aspx

Oakes, D. (1999). Direct calculation of the information matrix via the EM algorithm. Journal of the Royal Statistical Society. Series B. *Statistical Methodology*, *61*(2), 479–482. https://doi.org/10.1111/1467-9868.00188

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. https://www.R-project.org/

Rao, C. R (1973). *Linear statistical inference and its applications* (2nd ed.). Wiley.

von Davier, A. A. (2011). *Statistical models for test equating, scaling*. Springer. https://doi.org/10.1007/978-0-387-98138-3

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