An Illustration on the Quantile-Based Calculation of the Standard Error of Equating in Kernel Equating

Jorge González and **Gabriel Wallin**

1 Introduction

Equating methods rely on the comparison of score distributions using what is called an equating transformation function. Let F_X and G_Y be the score distributions of the random variables X and Y, corresponding to the test scores on two test forms X and Y, and defined on $\mathcal{X}$ and $\mathcal{Y}$, respectively. The equipercentile equating function $\varphi : \mathcal{X} \mapsto \mathcal{Y}$, computed as

$$\varphi(x) = G_Y^{-1}(F_X(x)), \tag{1}$$

maps the scores from one test form into the scale of the other (Braun & Holland, 1982; González & Wiberg, 2017). The equating transformation is a functional parameter that in practice is estimated using score data. Although various measures for the assessment of equating functions have been proposed (Wiberg & González, 2016), the uncertainty in the estimation of the equating transformation has mainly been measured by the standard error of equating (SEE),

$$\text{SEE}_Y(x) = \sqrt{\text{Var}(\hat{\varphi}(x))}. \tag{2}$$

J. González (✉)
Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile

Laboratorio Interdiciplinario de Estadística Social, LIES, Pontificia Universidad Católica de Chile, Santiago, Chile
e-mail: jorge.gonzalez@mat.uc.cl

G. Wallin
CNRS, Laboratoire J.A. Dieudonné, team Maasai, Université Côte d'Azur Inria, Nice, France
e-mail: gabriel.wallin@inria.fr

Different methods to calculate the SEE include exact formulas (see, e.g., Kolen and Brennan, 2014, Table 7.2); the Delta method (Lord, 1982), (Braun & Holland, 1982, p. 33), (Holland et al., 1989); and the bootstrap (Tsai et al., 2001). Another method proposed in Liou and Cheng (1995) and later extended by Liou et al. (1997) is based on the Bahadur's representation of sample quantiles (Bahadur, 1966; Ghosh, 1971). In this paper, this method will be refereed to as the *Quantile-Based SEE* (QB-SEE).

Liou and Cheng (1995) used the QB-SEE method and obtained results for traditional equipercentile equating under the single group (SG), the equivalent groups (EG), and the nonequivalent groups with anchor test (NEAT) designs. Later, Liou et al. (1997) extended this work to include the kernel equating transformation using Gaussian and Uniform kernels, considering only the NEAT design. These authors did however not make any comparison between the QB-SEE and the more traditionally used Delta method for the estimation of the SEE under the kernel equating framework. In this paper, we aim to fill this gap.

The paper is organized as follows. In Sect. 2 we briefly revise the kernel equating transformation and the way the SEE is calculated using the Delta method. Next, in Sect. 3 we introduce the QB-SEE method and give the details on how it can be used under the kernel equating framework. An illustration of the comparison between the QB-SEE and the delta method applied to the estimated kernel equating transformation is shown in Sect. 4. The paper ends in Sect. 5 summarizing the main results and discussing on future research.

2 Equating and the Standard Error of Equating

In this section we briefly review the basics of kernel equating and the way the SEE has been calculated within this framework. Next, we introduce the QB-SEE method and show how it adapts to be used in KE.

2.1 Kernel Equating

Kernel equating (Holland & Thayer, 1989; von Davier et al., 2004) is a semiparametric method used to estimate the equating function (González & von Davier, 2013). The score distributions are estimated using both kernel density estimation techniques (the nonparametric part) and maximum likelihood estimates of score probabilities (the parametric part).

Let $X(h_X)$ be a continuized version of the discrete score random variable X, defined as

$$X(h_X) = a_X(X + h_X V) + (1 - a_X)\mu_X,$$

where V is a continuous random variable with mean 0 and variance σ_V^2, $a_X^2 = \sigma_X^2/(\sigma_X^2 + \sigma_V^2 h_X^2)$, μ_X and σ_X^2 are the mean and variance of X, and h_X is a smoothing parameter. The estimated score distribution of $X(h_X)$ is obtained as

$$\hat{F}_{h_X}(x) = \sum_j \hat{r}_j K(\hat{R}_{jX}(x)),$$

where $r_j = \Pr(X = x_j)$ are score probabilities, typically modelled using log-linear models estimated by maximum likelihood, $\hat{R}_{jX}(x) = (x - \hat{a}_X x_j - (1 - \hat{a}_X)\hat{\mu}_X)/\hat{a}_X h_X$, and K is a kernel defined by the distribution of V. In this paper we will assume that $V \sim N(0, \sigma_V^2)$ so that $K = \Phi$, the standard normal (or Gaussian) distribution function.

Defining $s_k = \Pr(Y = y_k)$, and with similar expressions for $\hat{R}_{kY}(y)$, a_Y, and $Y(h_Y)$, the score distribution of the continuized Y scores, $\hat{G}_{h_Y}$, is obtained leading to calculate the kernel equating function as

$$\varphi(x, \hat{\mathbf{r}}, \hat{\mathbf{s}}) = \hat{G}_{h_Y}^{-1}(\hat{F}_{h_X}(x)),$$

where $\hat{\mathbf{r}} = (\hat{r}_1, \ldots, \hat{r}_J)^\top$ and $\hat{\mathbf{s}} = (\hat{s}_1, \ldots, \hat{s}_K)^\top$.

Because $\hat{\mathbf{r}}$ and $\hat{\mathbf{s}}$ are maximum likelihood estimates, the Delta method (e.g., Rao, 1973; Lehmann, 1999), described next, has been used to estimate the uncertainty on the estimation of φ.

2.2 SEE in Kernel Equating

The SEE in KE is based on the Delta method. The following theorem from von Davier et al. (2004) formalizes the result.

Theorem 1 (Delta method for the SEE in KE)

If $\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\mathbf{s}} \end{pmatrix} \overset{\cdot}{\sim} N\left(\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}, \Sigma\right)$, *then*

$$\varphi(x; \hat{\mathbf{r}}, \hat{\mathbf{s}}) \overset{\cdot}{\sim} N\left(\varphi(x; \mathbf{r}, \mathbf{s}), J_\varphi \Sigma J_\varphi^\top\right),$$

where

$$\Sigma = \begin{pmatrix} \Sigma_{\hat{\mathbf{r}}} & \Sigma_{\hat{\mathbf{r}}, \hat{\mathbf{s}}} \\ \Sigma_{\hat{\mathbf{r}}, \hat{\mathbf{s}}}^\top & \Sigma_{\hat{\mathbf{s}}} \end{pmatrix},$$

and

$$J_\varphi = \left(\frac{\partial \varphi}{\partial \mathbf{r}}, \frac{\partial \varphi}{\partial \mathbf{s}}\right).$$

When the score probabilities are obtained using maximum likelihood estimates from log-linear models and estimated for different equating designs using a design function, $DF(\hat{\mathbf{r}}, \hat{\mathbf{s}})$, von Davier et al. (2004) showed that the asymptotic variance obtained via the Delta method can be written as

$$Var(\hat{\varphi}) = ||J_\varphi J_{DF} C||^2$$

where J_φ is the Jacobian of the equating function, J_{DF} is the Jacobian matrix of the design function and C is a factor of the covariance matrix such that $\Sigma = CC^\top$. From this result, the SEE for the kernel equating function is defined as

$$\text{SEE}_Y(x) = ||J_\varphi J_{DF} C||, \tag{3}$$

which in this paper is denoted as $\text{SEE}_Y^A(x)$.

3   Quantile-Based Estimation of SEE

The QB-SEE method is based on the so called Bahadur's representation of sample quantiles. The main result is presented in Ghosh (1971) and reproduced here.

Theorem 2 (Ghosh, 1971)
Suppose that G is once differentiable at $\xi_p = G^{-1}(p)$ with $G'(\xi_p) > 0$. If $0 < p < 1$, then

$$\hat{\xi}_p = \xi_p + \frac{p - \hat{G}(\xi)}{G'(\xi)} + o_p(N^{-1/2}). \tag{4}$$

Liou and Cheng (1995) used this result to derive a formula for the SEE of the equipercentile equating transformation. After replacing p by F_X and checking regularity conditions, these authors took the variance in (4) to obtain

$$Var(\hat{G}^{-1}(F(x))) = \frac{1}{G'(\varphi)^2} \Big\{ \text{Var}(\hat{F}_X) + \text{Var}(\hat{G}_Y(\varphi)) - 2\text{Cov}(\hat{F}_X, \hat{G}_Y(\varphi)) \Big\}. \tag{5}$$

We call the square root of this expression the QB-SEE and denote it as $\text{SEE}_Y^B(x)$. In the next section we describe how the QB-SEE can be used to evaluate the kernel equating transformation under the NEAT design. For a critical review of the NEAT equating design see San Martín and González (2020).

3.1 Quantile-Based SEE in KE

The sample estimates of the score distributions can be replaced by kernel estimates in which case the QB-SEE formula becomes

$$\text{SEE}_Y^B(x) = \frac{1}{G'(\varphi)}\left\{\text{Var}(\hat{F}_{h_X}) + \text{Var}(\hat{G}_{h_Y}(\varphi)) - 2\text{Cov}(\hat{F}_{h_X}, \hat{G}_{h_Y}(\varphi))\right\}^{1/2}. \quad (6)$$

To derive the QB-SEE for the particular case of equating under the NEAT design, we introduce the following additional notation: $t_l = \Pr(A = a_l)$ are the marginal score probabilities for the anchor random variable A, and $r_{j|l}$ and $s_{k|l}$ are the conditional score probabilities of X and Y given A, respectively.

Following Liou et al. (1997), the variances and covariance terms in (6) can be obtained as

$$
\begin{aligned}
\text{Var}(\hat{F}_{h_X}) &= \text{Var}\left(\sum_j \hat{r}_j K(\hat{R}_{jX}(x))\right) \\
&\approx \sum_j \sum_{j'} K\big(R_{JX}(x)\big)K\big(R_{j'X}(x)\big)\text{Cov}\big[\hat{r}_j, \hat{r}_{j'}\big] \\
&= \sum_j K_{jX}^2\big(R_{JX}(x)\big)\text{Var}\big[\hat{r}_j\big] \\
&\quad + \sum_{j\neq j'}\sum K_{jX}\big(R_{JX}(x)\big)K_{j'X}\big(R_{JX}(x)\big)\text{Cov}\big[\hat{r}_j, \hat{r}_{j'}\big]
\end{aligned}
\quad , \quad (7)
$$

where

$$
\begin{aligned}
\text{Var}\big[\hat{r}_j\big] = \sum_l &\left\{\frac{\hat{r}_{j|l}[1 - \hat{r}_{j|l}]}{(n_X+1)\hat{h}(a_l)-1}\hat{h}^2(a_l) + \frac{\hat{h}(a_l)[1-\hat{h}(a_l)]}{n_X+n_Y}\hat{r}_{j|l}^2 \right. \\
&\left. + \frac{\hat{r}_{j|l}[1-\hat{r}_{j|l}]\hat{h}(a_l)[1-\hat{h}(a_l)]}{[(n_X+1)\hat{h}(a_l)-1](n_X+n_Y)}\right\} - \sum_{l\neq l'}\sum \frac{\hat{h}(a_l)\hat{h}(a_{l'})}{n_X+n_Y}\hat{r}_{j|a}\hat{r}_{j|a'}
\end{aligned}
\quad , \quad (8)
$$

and

$$
\begin{aligned}
\text{Cov}\big[\hat{r}_j, \hat{r}_{j'}\big] = \sum_a &\left\{\frac{\hat{h}(a_l)[1-\hat{h}(a_l)]}{n_X+n_Y}\hat{r}_{j|l}\hat{r}_{j'|l} - \frac{\hat{r}_{j|l}\hat{r}_{j'|l}}{(n_X+1)\hat{h}(a_l)-1}\hat{h}^2(a_l) \right. \\
&\left. - \frac{\hat{r}_{j|l}\hat{r}_{j'|l}\hat{h}(a_l)[1-\hat{h}(a_l)]}{[(n_X+1)\hat{h}(a_l)-1](n_X+n_Y)}\right\} - \sum_{l\neq l'}\sum \frac{\hat{h}(a_l)\hat{h}(a_{l'})}{n_X+n_Y}\hat{r}_{j|l}\hat{r}_{j'|l'}
\end{aligned}
\quad (9)
$$

Replacing terms accordingly, similar derivations lead to obtain the variance of G_Y.

Finally, the covariance term is calculated as

$$
\begin{aligned}
\mathrm{Cov}(\hat{F}_{h_X}, \hat{G}_{h_Y}) \approx \sum_j \sum_k K(R_{jX}) K(R_{kY}) \Bigg\{ & \sum_l \frac{\hat{h}(a_l)[1 - \hat{h}(a_l)]}{n_X + n_Y} \hat{r}_{j|l} \hat{s}_{k|l} \\
& - \sum_{l \neq l'} \frac{\hat{h}(l)\hat{h}(l')}{n_X + n_Y} \hat{r}_{j|l} \hat{s}_{k|l'} \Bigg\}
\end{aligned}
. \quad (10)
$$

In all previous equations either sample estimates or presmoothed score probabilities can be used as weights for the kernel. In the next section, the former case is considered for illustration and to compare the SEE for KE as calculated using the traditional delta method with the QB-SEE method

4   Illustration

4.1   Data

We use data described in Kolen and Brennan (2014). The data set consists of two 36-items test forms. Form X was administered to 1,655 examinees and form Y was administered to 1,638 examinees. Also, 12 out of the 36 items are common between both test forms (items 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, and 36). The data come with the distribution of the CIPE software which is freely available at https://education.uiowa.edu/centers/center-advanced-studies-measurement-and-assessment/computer-programs and is also available in the *equate* (Albano, 2016) and *SNSequate* (González, 2014) R packages.

4.2   Analyses

To investigate how $\mathrm{SEE}_Y^B(x)$ is related to $\mathrm{SEE}_Y^\Delta(x)$, we compared the SEE for KE under the NEAT-PSE design calculated using the Delta, the QB-SEE, and Bootstrap methods.

The SEE based on the Delta method is calculated using Equation (3), which is implemented in *SNSequate* and appears as one of the output values for a call to the `ker.eq()` function.

The QB-SEE is calculated using Equation (6). The variances and covariance components of the numerator on the right hand side are obtained using Equations 7–10, whereas the denominator corresponds to the derivative of G evaluated in the equated score, which in this case correspond to a Gaussian kernel.

The bootstrap SEE implements the procedure described in Kolen and Brennan (2014, Chap. 7) to compute the SEE using 500 replications.
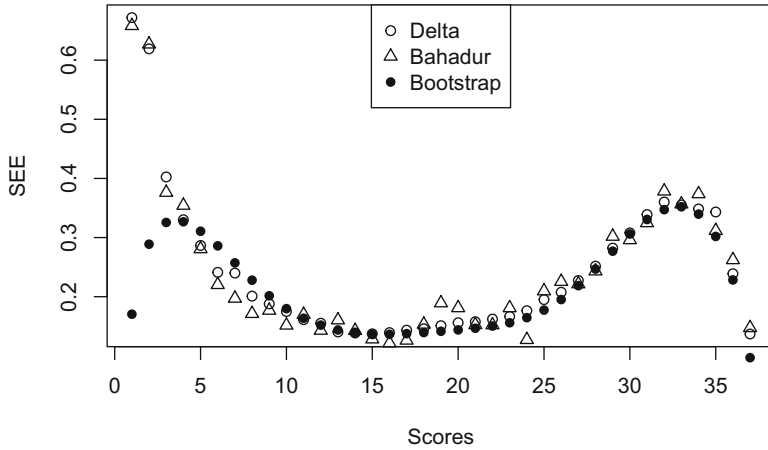
Fig. 1 SEE for the three compared methods

All the analyses were carried out using the R software (R Core Team, 2020) and the code is available from the authors upon request.

4.3 Results

The SEE obtained for the three compared methods are shown graphically in Fig. 1. Except for some score values in the lower range of the score scale, it can be seen that all the methods yielded similar estimations of SEE for the analyzed data.

The results for all SEE methods reflect that there are few test-takers in the tails of the score scale, as illustrated by the increased values of the SEE. The results also suggest that the QB-SEE and the Delta method produce very similar results, although leaning on different asymptotic results. Given that both the QB-SEE and Delta method SEE deviate from the bootstrap SEE in the lower tail, the results also indicate that they might be better approximations when the number of test-takers is large, which is expected given that they both are large-sample approximations.

5   Discussion

In this paper we have revisited the result of Bahadur on the asymptotic representation of sample quantiles and its use in the derivation of what we call the QB-SEE method of estimating the standard error of equating. The method was applied for kernel equating transformations under the NEAT design and compared to the more traditional Delta method of obtaining SEE. Results from a numerical illustration

shown that the QB-SEEs are very similar to the SEEs obtained using the Delta method, for the analyzed data set.

An advantage of the QB-SEE method is that it allows to separate sources of uncertainty influencing the SEE, as it can be grasped from (5). A comprehensive simulation study to assess how these variances and covariance terms vary according to different conditions is planned for future research. Another advantage of this method is that, in comparison to the Delta method, it does not rely on normality. This could open room for other models and methods for presmoothing that do not necessarily resort on the normality of parameter estimates, as it is the case of log-linear models estimated using maximum likelihood.

Future work include other methods to estimate the variance-covariance components in the SEE formulas and the evaluation of other kernels and other equating designs.

References

Albano, A. D. (2016). Equate: An R package for observed-score linking and equating. *Journal of Statistical Software, 74*(8), 1–36

Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics, 37*(3), 577–580

Braun, H., & Holland, P. (1982). Observed-score test equating: A mathematical analysis of some ets equating procedures. In P. Holland, & D. Rubin (Eds.), *Test equating* (vol 1, pp. 9–49). New York: Academic Press

Ghosh, J. K. (1971). A new proof of the bahadur representation of quantiles and an application. *The Annals of Mathematical Statistics, 42*(6), 1957–1961

González, J. (2014). SNSequate: Standard and nonstandard statistical models and methods for test equating. *Journal of Statistical Software, 59*(7), 1–30

González, J., & von Davier, M. (2013). Statistical models and inference for the true equating transformation in the context of local equating. *Journal of Educational Measurement, 50*(3), 315–320

González, J., & Wiberg, M. (2017). *Applying test equating methods using R*. New York: Springer

Holland, P., King, B., & Thayer, D. (1989). The standard error of equating for the kernel method. Technical Report 89–83, Educational Testing Service

Holland, P., & Thayer, D. (1989). The kernel method of equating score distributions. Technical report, Princeton: Educational Testing Service

Kolen, M., & Brennan, R. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York: Springer

Lehmann, E. L. (1999). *Elements of large-sample theory*. New York: Springer

Liou, M., & Cheng, P. E. (1995). Asymptotic standard error of equipercentile equating. *Journal of Educational and Behavioral Statistics, 20*(3), 259–286

Liou, M., Cheng, P. E., & Johnson, E. G. (1997). Standard error of the kernel equating methods under the common-item design. *Applied Psychological Measurement, 21*(4), 349–369

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

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, Vienna, Austria

Rao, C. R. (1973). *Linear statistical inference and applications*. New York: Wiley

San Martín, E., & González, J. (2020). A Critical View on the NEAT Equating Design: Statistical Modelling and Identifiability Problems. *Manuscript submitted for publication*

Tsai, T., Hanson, B., Kolen, M., & Forsyth, R. (2001). A comparison of bootstrap standard errors of IRT equating methods for the common-item nonequivalent groups design. *Applied Measurement in Education, 14*(1), 17–30

von Davier, A. A., Holland, P. & Thayer, D. (2004). *The Kernel method of test equating*. New York: Springer

Wiberg, M., & González, J. (2016). Statistical assessment of estimated transformations in observed-score equating. *Journal of Educational Measurement, 53*(1), 106–125