Linear Equating Models for the Common-Item Nonequivalent-Populations Design

Michael J. Kolen and Robert L. Brennan
American College Testing Program

The Tucker and Levine equally reliable linear methods for test form equating in the common-item nonequivalent-populations design are formulated in a way that promotes understanding of the methods. The formulation emphasizes population notions and is used to draw attention to the practical differences between the methods. It is shown that the Levine method weights group differences more heavily than the Tucker method. A scheme for forming a synthetic population is suggested that is intended to facilitate interpretation of equating results. A procedure for displaying form and group differences is developed that also aids interpretation.

Test form equating of observed scores adjusts for small unintended differences in difficulty among multiple forms of a test for a specified population of examinees. Equating requires a design for collecting data and a method for equating forms. *Linear observed score methods* under the *common-item nonequivalent-populations design* are the focus of the present paper. In the common-item nonequivalent-populations design, two groups of examinees from different populations are each administered different test forms that have a subset of items in common. This design for equating tests is often used in practice. One of the reasons for its popularity is that only one form of a test needs to be administered on a given test date. A second reason is that in cases where the common items do not contribute to an examinee's score, this design may be used when the test items comprising the scored portion of the examination are disclosed to examinees.

Linear methods are most often used with this design. In linear equating, the linear transformation is estimated that leads to the transformed scores on one test form having the same mean and standard deviation as scores on another test form for an explicitly defined population of examinees. Linear methods are attractive because they involve only a simple linear transformation of raw to scaled scores.

Strong statistical assumptions are required when conducting equating using the common-item nonequivalent-groups design, because any examinee is administered only one of the two forms to be equated. The different methods for equating using this design can be distinguished by their statistical assumptions. The *Tucker* and *Levine equally reliable* methods (Angoff, 1971) are examined in the present paper. These methods are formulated in a way that is intended to promote better understanding and interpretation of equating results. This is done by using a common notation, emphasizing population notions, and drawing

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 11, No. 3, September 1987, pp. 263–277
© Copyright 1987 Applied Psychological Measurement Inc.
0146-6216/87/030263-15$2.00

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

attention to differences between the two methods. This is followed by a consideration of how to display group and form differences to aid in the interpretation of equating results.

Equating Models

Multiple test forms to be equated should be designed to be similar in content and statistical characteristics. For the common-item nonequivalent-populations design, a new form is equated to an old form by using a set of items that is common to both forms. The set of common items is constructed to be similar to each of the full-length forms in content balance and in the statistical characteristics of its items. Scores on the common items may contribute to the total scores on each form (an *internal* set of common items), or they may not contribute to the total scores (an *external* set of common items). The new form is administered to examinees who are considered to be from a somewhat different population than the examinees who were previously administered the old form.

Refer to the new test form as X, the old form as Y, and the set of common items as V. Examinees from Population 1 are administered X and V. Examinees from Population 2 were previously administered Y and V. For an internal set of common items, X and Y include scores on the common items. For an external set of common items, X and Y do not include scores on the common items. For example, consider an examinee who earned a score of 5 on the common items and a score of 20 on the items that were not in common. If the common items are external, then $X = 20$ and $V = 5$. If the common items are internal, then $X = 25$ and $V = 5$.

Observed score equating functions are necessarily defined for a single population of examinees. In the common-item nonequivalent-populations design, Populations 1 and 2 must be combined in some way to arrive at a single examinee population for defining the equating relationship. Braun and Holland (1982, p. 21) introduced the concept of a *synthetic population* to address this issue. They conceived of the synthetic population as consisting of two strata, Population 1 and Population 2. Populations 1 and 2 are proportionally weighted by w_1 and w_2 ($w_1 + w_2 = 1$; $w_1, w_2 \geqslant 0$) to arrive at the synthetic population. Statistics are computed for the synthetic population by first computing the statistics for each of the two strata separately and then forming weighted averages of these statistics using w_1 and w_2 as weights.

The linear equation for equating scores on X to the scale of Y is

$$\ell(x) = \frac{\sigma_s(Y)}{\sigma_s(X)}[x - \mu_s(X)] + \mu_s(Y) \quad , \tag{1}$$

where the subscript s indicates the synthetic population.

The synthetic population parameters in Equation 1 can be expressed in terms of Population 1 and Population 2 parameters as follows:

$$\mu_s(X) = w_1\mu_1(X) + w_2\mu_2(X) \quad , \tag{2}$$

$$\mu_s(Y) = w_1\mu_1(Y) + w_2\mu_2(Y) \quad , \tag{3}$$

$$\sigma_s^2(X) = w_1\sigma_1^2(X) + w_2\sigma_2^2(X) + w_1w_2[\mu_1(X) - \mu_2(X)]^2 \quad , \tag{4}$$

and

$$\sigma_s^2(Y) = w_1\sigma_1^2(Y) + w_2\sigma_2^2(Y) + w_1w_2[\mu_1(Y) - \mu_2(Y)]^2 \quad , \tag{5}$$

where the subscripts 1 and 2 refer to Populations 1 and 2, respectively. Also, the mean and variance of V for the synthetic population are expressed, respectively, as

$$\mu_s(V) = w_1\mu_1(V) + w_2\mu_2(V) \tag{6}$$

and

$$\sigma_s^2(V) = w_1\sigma_1^2(V) + w_2\sigma_2^2(V) + w_1w_2[\mu_1(V) - \mu_2(V)]^2 \quad . \tag{7}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

The following parameters presented in Equations 2 through 7 cannot be estimated directly: $\mu_1(Y)$, $\sigma_1(Y)$, $\mu_2(X)$, and $\sigma_2(X)$. This is true because, in the study design, Y is not administered to examinees from Population 1 and X is not administered to examinees from Population 2. The Tucker and Levine equally reliable methods require statistical assumptions that make these parameters expressible as functions of parameters that can be estimated directly. The required statistical assumptions distinguish the Tucker method from the Levine method.

Tucker Method

The Tucker method was originally described by Gulliksen (1950, pp. 299–301), who attributed it to Ledyard Tucker. Angoff (1971), Braun and Holland (1982), and Kolen (1985) have also presented descriptions of this method. To derive the Tucker method, assumptions are made about the similarity of the linear regressions of total score on common item score for Populations 1 and 2.

Assumptions. First, it is assumed that the linear regression function (slope and intercept) for the regression of X on V is the same for Populations 1 and 2. A similar assumption is made for Y and V. To state this more explicitly, let α represent a regression slope so that, for example,

$$\alpha_1(X|V) = \sigma_1(X,V)/\sigma_1^2(V) \tag{8}$$

is the slope for the linear regression of X on V for Population 1. Let β represent a regression intercept so that, for example,

$$\beta_1(X|V) = \mu_1(X) - \alpha_1(X|V)\mu_1(V) \tag{9}$$

is the intercept for the linear regression of X on V for Population 1. The Tucker method requires that

$$\alpha_1(X|V) = \alpha_2(X|V), \quad \alpha_1(Y|V) = \alpha_2(Y|V) \tag{10}$$

and

$$\beta_1(X|V) = \beta_2(X|V), \quad \beta_1(Y|V) = \beta_2(Y|V) \quad . \tag{11}$$

In Tucker equating, it is also assumed that the variance of X given V is the same for Populations 1 and 2. A similar assumption is made for Y given V. These assumptions are stated more explicitly as

$$\sigma_1^2(X)[1-\rho_1^2(X,V)] = \sigma_2^2(X)[1-\rho_2^2(X,V)] \tag{12}$$

and

$$\sigma_1^2(Y)[1-\rho_1^2(Y,V)] = \sigma_2^2(Y)[1-\rho_2^2(Y,V)] \quad , \tag{13}$$

where ρ^2 refers to a squared correlation. Sometimes stronger assumptions are used for deriving these equations, such as the assumption of homogeneity of variance used by Braun and Holland (1982), but the assumptions listed here are sufficient.

Intermediate results. These assumptions allow for the parameters that cannot be estimated directly to be expressed in terms of quantities that can be estimated directly. Given the Tucker assumptions, it can be shown that for Population 1,

$$\mu_1(Y) = \mu_2(Y) + \alpha_2(Y|V)[\mu_1(V) - \mu_2(V)] \quad , \tag{14}$$

$$\sigma_1^2(Y) = \sigma_2^2(Y) + \alpha_2^2(Y|V)[\sigma_1^2(V) - \sigma_2^2(V)] \quad , \tag{15}$$

and

$$\sigma_1(Y,V) = \sigma_2(Y,V)\,\frac{\sigma_1^2(V)}{\sigma_2^2(V)} \quad . \tag{16}$$

For Population 2,

$$\mu_2(X) = \mu_1(X) - \alpha_1(X|V)[\mu_1(V) - \mu_2(V)] \quad , \tag{17}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

$$\sigma_2^2(X) = \sigma_1^2(X) - \alpha_1^2(X|V)[\sigma_1^2(V) - \sigma_2^2(V)] \quad , \tag{18}$$

and

$$\sigma_2(X,V) = \sigma_1(X,V)\,\frac{\sigma_2^2(V)}{\sigma_1^2(V)} \quad . \tag{19}$$

Final results. To arrive at the Tucker equating equation, substitute Equations 14 through 19 in Equations 2 through 7. This gives

$$\mu_s(X) = \mu_1(X) - w_2\alpha_1(X|V)[\mu_1(V) - \mu_2(V)] \quad , \tag{20}$$

$$\mu_s(Y) = \mu_2(Y) + w_1\alpha_2(Y|V)[\mu_1(V) - \mu_2(V)] \quad , \tag{21}$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2\alpha_1^2(X|V)[\sigma_1^2(V) - \sigma_2^2(V)] + w_1 w_2 \alpha_1^2(X|V)[\mu_1(V) - \mu_2(V)]^2 \quad , \tag{22}$$

and

$$\sigma_s^2(Y) = \sigma_2^2(Y) + w_1\alpha_2^2(Y|V)[\sigma_1^2(V) - \sigma_2^2(V)] + w_1 w_2 \alpha_2^2(Y|V)[\mu_1(V) - \mu_2(V)]^2 \quad , \tag{23}$$

where all parameters to the right of the equal signs in Equations 20 through 23 are estimated directly using data from the study design. Equations 20 through 23 are inserted into Equation 1 to produce the Tucker linear equating function.

It can be shown that Equations 20 through 23 produce counterparts of the Tucker method equation described by Angoff (1971, p. 580), if weights are chosen proportional to sample size—that is, $w_1 = n_1/(n_1 + n_2)$ and $w_2 = n_2/(n_1 + n_2)$, where n_1 and n_2 are the sample sizes of examinees included in the equating study from Populations 1 and 2, respectively. Gulliksen (1950, pp. 299–301) presented a version of the Tucker method that differs from Angoff's version. The present equations will result in counterparts to Gulliksen's if w_1 and w_2 are set to 1 and 0, respectively. To weight the populations equally, w_1 and w_2 are each set to .5. The above equations are appropriate whether V is an internal or an external set of common items. If V is internal, then the score on V contributes to the score on X. If V is external, then the score on V does not contribute to the score on X.

Levine Equally Reliable Method

The Levine equally reliable method was originally developed by Levine (1955). Angoff (1971) has also presented a description of this method. Woodruff (1986) presented a derivation based on a congeneric test theory model. Unlike the Tucker method, which considers only observed scores, the derivation of the Levine equally reliable method (referred to as the Levine method in the following discussion) requires assumptions about true scores.

Assumptions. Define $\tilde{X}$, $\tilde{Y}$, and $\tilde{V}$ as true scores. The Levine method assumes that $\tilde{X}$ and $\tilde{V}$ as well as $\tilde{Y}$ and $\tilde{V}$ correlate perfectly for the two populations. That is,

$$\rho_1(\tilde{X},\tilde{V}) = \rho_2(\tilde{X},\tilde{V}) = \rho_1(\tilde{Y},\tilde{V}) = \rho_2(\tilde{Y},\tilde{V}) = 1.0 \quad . \tag{24}$$

Thus, for Levine equating it is assumed that the total test (X or Y) and the common items (V) are measuring the same thing in the sense that they have disattenuated correlations of 1.0.

In Levine equating, it is also assumed that the linear regression function of $\tilde{X}$ on $\tilde{V}$ is the same for Populations 1 and 2. A similar assumption is made for $\tilde{Y}$ on $\tilde{V}$. Note that the regression slope $\alpha_1(\tilde{X}|\tilde{V})$ is equal to $\rho_1(\tilde{X},\tilde{V})\sigma_1(\tilde{X})/\sigma_1(\tilde{V})$. Because $\rho_1(\tilde{X},\tilde{V}) = 1.0$ from Equation 24, $\alpha_1(\tilde{X}|\tilde{V}) = \sigma_1(\tilde{X})/\sigma_1(\tilde{V})$. Using similar reasoning it can be shown that the assumption of equal true score regression slopes can be stated as

$$\frac{\sigma_1(\tilde{X})}{\sigma_1(\tilde{V})} = \frac{\sigma_2(\tilde{X})}{\sigma_2(\tilde{V})} \tag{25}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

and

$$\frac{\sigma_1(\tilde{Y})}{\sigma_1(\tilde{V})} = \frac{\sigma_2(\tilde{Y})}{\sigma_2(\tilde{V})} \quad . \tag{26}$$

The true score regression intercepts also are assumed to be equal. In classical theory, for a given population the mean true score is equal to the mean observed score, for example, $\mu_1(\tilde{X}) = \mu_1(X)$; thus the equality of regression slopes assumption can be expressed as

$$\mu_1(X) - \frac{\sigma_1(\tilde{X})}{\sigma_1(\tilde{V})}\,\mu_1(V) = \mu_2(X) - \frac{\sigma_2(\tilde{X})}{\sigma_2(\tilde{V})}\,\mu_2(V) \tag{27}$$

and

$$\mu_1(Y) - \frac{\sigma_1(\tilde{Y})}{\sigma_1(\tilde{V})}\,\mu_1(V) = \mu_2(Y) - \frac{\sigma_2(\tilde{Y})}{\sigma_2(\tilde{V})}\,\mu_2(V) \quad . \tag{28}$$

In Levine equating it is also assumed that the measurement error variances are the same for the two populations. This set of assumptions can be expressed as

$$\sigma_1^2(X) - \sigma_1^2(\tilde{X}) = \sigma_2^2(X) - \sigma_2^2(\tilde{X}) \quad , \tag{29}$$

$$\sigma_1^2(Y) - \sigma_1^2(\tilde{Y}) = \sigma_2^2(Y) - \sigma_2^2(\tilde{Y}) \quad , \tag{30}$$

and

$$\sigma_1^2(V) - \sigma_1^2(\tilde{V}) = \sigma_2^2(V) - \sigma_2^2(\tilde{V}) \quad . \tag{31}$$

Intermediate results. By rearranging the terms in Equations 27 and 28 and using the relation in Equations 25 and 26, $\mu_2(X)$ and $\mu_1(Y)$ can be expressed as

$$\mu_2(X) = \mu_1(X) - \frac{\sigma_1(\tilde{X})}{\sigma_1(\tilde{V})}[\mu_1(V) - \mu_2(V)] \tag{32}$$

and

$$\mu_1(Y) = \mu_2(Y) + \frac{\sigma_2(\tilde{Y})}{\sigma_2(\tilde{V})}\,[\mu_1(V) - \mu_2(V)] \quad . \tag{33}$$

Note that expressions for $\sigma_1(\tilde{X})/\sigma_1(\tilde{V})$ and $\sigma_2(\tilde{Y})/\sigma_2(\tilde{V})$ are still needed for Equation 32 to be useful in practice.

To find an expression for $\sigma_2^2(X)$, first consider that from Equation 29,

$$\sigma_2^2(X) = \sigma_1^2(X) - \sigma_1^2(\tilde{X}) + \sigma_2^2(\tilde{X}) \quad . \tag{34}$$

From Equation 25, it can be shown that $\sigma_2^2(\tilde{X}) = \sigma_1^2(\tilde{X})\sigma_2^2(\tilde{V})/\sigma_1^2(\tilde{V})$. Substituting this quantity in Equation 34 gives

$$\sigma_2^2(X) = \sigma_1^2(X) - \sigma_1^2(\tilde{X}) + \sigma_1^2(\tilde{X})\,\frac{\sigma_2^2(\tilde{V})}{\sigma_1^2(\tilde{V})} = \sigma_1^2(X) - \frac{\sigma_1^2(\tilde{X})}{\sigma_1^2(\tilde{V})}\,[\sigma_1^2(\tilde{V}) - \sigma_2^2(\tilde{V})] \quad . \tag{35}$$

From Equation 31, it can be shown that $\sigma_1^2(\tilde{V}) - \sigma_2^2(\tilde{V}) = \sigma_1^2(V) - \sigma_2^2(V)$. Thus,

$$\sigma_2^2(X) = \sigma_1^2(X) - \frac{\sigma_1^2(\tilde{X})}{\sigma_1^2(\tilde{V})}\,[\sigma_1^2(V) - \sigma_2^2(V)] \quad . \tag{36}$$

By similar reasoning,

$$\sigma_1^2(Y) = \sigma_2^2(Y) + \frac{\sigma_2^2(\tilde{Y})}{\sigma_2^2(\tilde{V})}\,[\sigma_1^2(V) - \sigma_2^2(V)] \quad . \tag{37}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Final results. To use Equations 32, 33, 36, and 37, it is necessary to have representations of the ratios $\sigma_1(\tilde{X})/\sigma_1(\tilde{V})$ and $\sigma_2(\tilde{X})/\sigma_2(\tilde{V})$ that are expressible in terms of parameters that can be estimated directly from the design. One commonly used approach is to use error variance estimates derived by Angoff (1953). Angoff (1971) indicated that using this approach, when V is internal to X,

$$\frac{\sigma_1(\tilde{X})}{\sigma_1(\tilde{V})} = \frac{1}{\alpha_1(V|X)} \tag{38}$$

and

$$\frac{\sigma_2(\tilde{Y})}{\sigma_2(\tilde{V})} = \frac{1}{\alpha_2(V|X)} \quad , \tag{39}$$

where, for example, $\alpha_1(V|X) = \sigma_1(X,V)/\sigma_1^2(X)$.

By substituting Equations 38 and 39 into Equations 32, 33, 36, and 37 and then using these results in Equations 2 through 5, the equations for the Levine method using an *internal* anchor are as follows:

$$\mu_s(X) = \mu_1(X) - w_2\frac{1}{\alpha_1(V|X)}\,[\mu_1(V) - \mu_2(V)] \quad , \tag{40}$$

$$\mu_s(Y) = \mu_2(Y) + w_1\frac{1}{\alpha_2(V|Y)}\,[\mu_1(V) - \mu_2(V)] \quad , \tag{41}$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2\frac{1}{\alpha_1^2(V|X)}\,[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\frac{1}{\alpha_1^2(V|X)}\,[\mu_1(V) - \mu_2(V)]^2 \quad , \tag{42}$$

and

$$\sigma_s^2(Y) = \sigma_2^2(Y) + w_1\frac{1}{\alpha_2^2(V|Y)}\,[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\frac{1}{\alpha_2^2(V|Y)}\,[\mu_1(V) - \mu_2(V)]^2 \quad . \tag{43}$$

These are then substituted into Equation 1.

When V is external, Angoff (1971) indicated that

$$\frac{\sigma_1(\tilde{X})}{\sigma_1(\tilde{V})} = \frac{\sigma_1^2(X) + \sigma_1(X,V)}{\sigma_1^2(V) + \sigma_1(X,V)} \tag{44}$$

and

$$\frac{\sigma_2(\tilde{Y})}{\sigma_2(\tilde{V})} = \frac{\sigma_2^2(Y) + \sigma_2(Y,V)}{\sigma_2^2(V) + \sigma_2(Y,V)} \quad . \tag{45}$$

By substituting Equations 44 and 45 into Equations 32, 33, 36, and 37 and then these into Equations 2 through 5, the equations for the Levine method using an *external* anchor are as follows:

$$\mu_s(X) = \mu_1(X) - w_2\frac{\sigma_1^2(X) + \sigma_1(X,V)}{\sigma_1^2(V) + \sigma_1(X,V)}\,[\mu_1(V) - \mu_2(V)] \quad , \tag{46}$$

$$\mu_s(Y) = \mu_2(Y) + w_1\frac{\sigma_2^2(Y) + \sigma_2(Y,V)}{\sigma_2^2(V) + \sigma_2(Y,V)}\,[\mu_1(V) - \mu_2(V)] \quad , \tag{47}$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2\frac{[\sigma_1^2(X) + \sigma_1(X,V)]^2}{[\sigma_1^2(V) + \sigma_1(X,V)]^2}[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\frac{[\sigma_1^2(X) + \sigma_1(X,V)]^2}{[\sigma_1^2(V) + \sigma_1(X,V)]^2}[\mu_1(V) - \mu_2(V)]^2 \quad , \tag{48}$$

and

$$\sigma_s^2(Y) = \sigma_2^2(Y) + w_1\frac{[\sigma_2^2(Y) + \sigma_2(Y,V)]^2}{[\sigma_2^2(V) + \sigma_2(Y,V)]^2}[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\frac{[\sigma_2^2(Y) + \sigma_2(Y,V)]^2}{[\sigma_2^2(V) + \sigma_2(Y,V)]^2}[\mu_1(V) - \mu_2(V)]^2 \quad . \tag{49}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

These results are then substituted into Equation 1. If w_1 and w_2 are chosen proportional to sample size, then the Levine equally reliable equation presented by Levine (1955) and Angoff (1971) will result.

Data based on an actual equating are presented in Table 1 and some results for these data are presented in Table 2. These data and results are used in subsequent discussions.

Defining the Synthetic Group

The definition of an observed score equating relationship requires that a particular examinee group be specified. Because examinees from different groups take the forms to be equated in the nonequivalent-populations design, the equating relationship is defined for a synthetic group that is a combination of the groups taking the forms. As stated earlier, the synthetic group or population is conceived of as containing two strata. Examinees administered the new form are considered to be a random sample from Stratum 1, and examinees administered the old form are considered to be a random sample from Stratum 2. Weights w_1 and w_2 are used to weight the strata in defining the synthetic group. This process is illustrated in Figure 1. Traditionally, the weights have been chosen to be proportional to the sample size of examinees from each population. That is, $w_1 = n_1/(n_1 + n_2)$ and $w_2 = 1 - w_1$. Sometimes the weights are chosen to be equal, where $w_1 = w_2 = .5$. In general, different choices of weights lead to different equating relationships.

From a practical perspective, the synthetic group that leads to the most direct score interpretation is preferable. When a new form is administered and scored, the focus of score interpretation typically is on the group that just took the new form, and the interpretation given to the resulting scores depends very little on the particular form chosen to be the old form. The old form and old group are viewed merely as vehicles for placing new form scores on the scale used to report scores. Because the equal weighting ($w_1 = w_2 = .5$) and proportional weighting [$w_1 = n_1/(n_1 + n_2)$, $w_2 = 1 - w_1$] schemes incorporate the old group in the synthetic group, either of these schemes leads to a synthetic population that is not central to score interpretation. For this reason, the weights $w_1 = 1$ and $w_2 = 0$ may be preferable because, based on these weights, the synthetic group is defined as the new group. Equating based on $w_1 = 1$ and $w_2 = 0$ allows a direct comparison of how the new group performed on the new form to how the new group would have performed had it been administered the old form.

As an illustration, consider the equating data shown in Table 1. Equating results based on these data for the Tucker method are shown in Table 2. In the first row of Table 2, the mean of 95.70 is the observed score mean for the new group of examinees who took the new form. The value of 98.69 in this table is the mean, based on Tucker assumptions, that the new group would have earned had they taken the old form. Thus, the mean for the new group on the new form is 95.70, under the Tucker assumptions the mean for the new group on the old form is 98.69, and for the new group the new form is 2.99 points

Table 1
Numerical Illustration for Internal Equating Section*

New Form--Population 1	Old Form--Population 2
$\mu_1(X) = 95.7$	$\mu_2(Y) = 96.8$
$\sigma_1(X) = 13.4$	$\sigma_2(Y) = 13.4$
$\mu_1(V) = 23.2$	$\mu_2(V) = 22.5$
$\sigma_1(V) = 4.0$	$\sigma_2(V) = 4.3$
$\alpha_1(X\|V) = 2.9$	$\alpha_2(Y\|V) = 2.7$

*Based on a test with 125 items and 30 common items.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 2
Equating Results Based On Numerical Data In Table 1*

Method	$\mu_s(X)$	$\mu_s(Y)$	$\sigma_s(X)$	$\sigma_s(Y)$	SLP	INT	$\ell(100)$	$\ell(80)$
Tucker								
$w_1=1, w_2=0$	95.70	98.69	13.40	12.70	.948	7.96	102.77	83.80
$w_1=.5, w_2=.5$	94.69	97.75	13.82	13.09	.947	8.07	102.78	83.84
Levine								
$w_1=1, w_2=0$	95.70	99.32	13.40	12.14	.906	12.63	103.21	85.09
$w_1=.5, w_2=.5$	94.35	98.06	14.14	12.85	.908	12.36	103.19	85.03

*Values in this table are rounded to two decimal places.

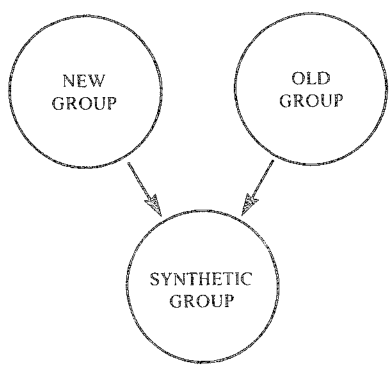
more difficult than the old form. Also, the equating line for the new group has a slope of .948 and an intercept of 7.96. All of these statements are made with reference to the new group, which is the group of interest at the time the new form is administered.

Based on the second row of Table 2, similar statements could be made for a synthetic population that contains 50% old-group examinees and 50% new-group examinees. However, from a practical perspective this synthetic group is seldom the group of interest at the time equating is being conducted.

Practical issues in defining the synthetic population also can be addressed from the perspective of a linking plan such as that shown in Figure 2. In this figure, Form 3 and subsequent forms are equated using two links. Typically, the results from the two links are averaged. Form 6 is equated to Forms 4 and 5. For traditionally defined synthetic groups, the Form 6 to Form 5 equating is based on a synthetic group containing examinees who took Form 5 and examinees who took Form 6. The Form 6 to Form 4 equating is based on a synthetic group containing examinees who took Form 4 and examinees who took Form 6. Thus, in this case, two different synthetic groups are used to equate Form 6. In effect, this means that the "synthetic" group for the averaged equating relationship is some ambiguous combination of the two component synthetic groups. However, if the new group is weighted 1 and the old group 0, both equatings are based on the examinee group that took Form 6. Again, the weights of 1 and 0 seem to have the potential to produce equatings that are more directly interpretable than traditional weights.

Refer again to Table 2. In this table, the slope and intercept of the equating functions do not appear to depend very much on the weighting scheme chosen. This observation is consistent with results obtained

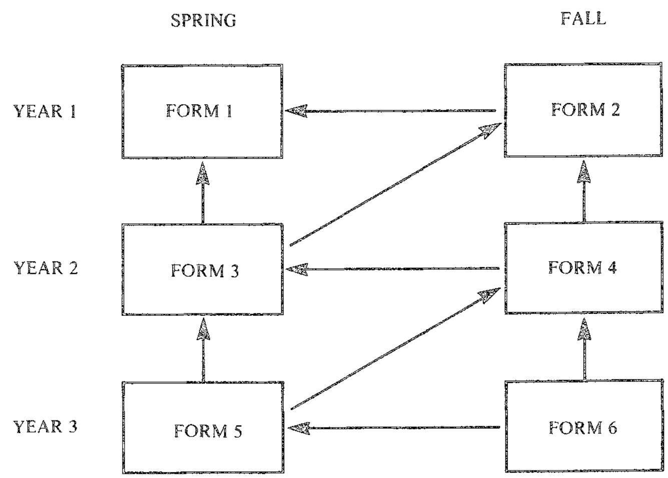
Figure 1
Synthetic Group in Common-Item Equating With Nonequivalent Groups



$w_1 + w_2 = 1$ and $w_1, w_2 \geq 0$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Figure 2
Simplified Linking Pattern for Common-Item Equating With Two Links



Notes: (1) Score scale and test specifications established on Form 1.
(2) Form 2 equated to score scale using one link.
(3) Form 3 and subsequent forms equated to the score scale using two links.

by the present authors in many different equatings. Thus, the choice of a synthetic group is evidently more of an interpretational issue than one that substantially affects the outcome of the equating process. However, the synthetic group formed using $w_1 = 1$ and $w_2 = 0$ does appear to lead to more readily interpretable equating results. It also greatly simplifies the equations and calculations. By using $w_1 = 1$ and $w_2 = 0$, all terms in this paper that contain w_2 as a multiplier become zero.

Comparison of the Tucker and Levine Equally Reliable Methods

The Tucker method requires that the regression of observed total score on common item score be identical for the two groups of examinees. The Levine method requires that the true scores on the *two tests* correlate 1.0. For these reasons, the Tucker method is often said to be appropriate when groups are more similar and tests less similar, and the Levine method is often said to be more appropriate when tests are more similar and groups less similar. However, research findings do not provide unambiguous evidence for these interpretations.

Both the Tucker and Levine methods described here are often said to be appropriate only if the tests to be equated are equally reliable. However, equal reliability was not assumed in the present derivation of either of these methods. Suppose that the assumptions for either one of the two equating methods were met. Then the scores on Y and the scores on X converted to the Y scale (ℓ in Equation 1) will have the same mean and standard deviation in the synthetic population, even if X and Y were unequally reliable. For unequally reliable forms, it might be argued that this property of equal observed means and standard deviations is less desirable than some other property. Such an argument has not been made clearly in the literature, however. (Methods that are typically referred to as methods for unequally reliable tests do not lead to the aforementioned property of equal observed score standard deviations when test forms are unequally reliable.) For tests that are designed to be as similar as possible in content and statistical

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

characteristics and to be equal in length, the Tucker and Levine methods described here should not be negatively affected by the small differences in reliability between Forms X and Y that are likely to exist.

Table 3 summarizes the equating equations presented earlier, and is designed to facilitate comparison across methods. The γ_1 and γ_2 terms listed in this table are all that distinguish the methods from each other. For example, if the γ_1 and γ_2 expressions shown for the Tucker method are used in the expressions for $\mu_s(X)$, $\mu_s(Y)$, $\sigma_s^2(X)$, and $\sigma_s^2(Y)$, then the Tucker method equations shown earlier will result.

For the most typical cases encountered in equating, γ_1 and γ_2 for the Tucker method are less than γ_1 and γ_2 for the Levine method. It is shown in the Appendix that this property holds for internal equating sections if $\rho_1(X,V)$ and $\rho_2(Y,V)$ are nonzero, and that it holds for external equating sections if $\rho_1(X,V)$ and $\rho_2(Y,V)$ are positive. In practical terms, this implies that the differences between groups of examinees have a greater influence on Levine equating than on Tucker equating, because the γs are multipliers for the group differences in means and variances shown in Table 3.

The effect of γ is illustrated in Table 4, based on the data shown in Table 1. In Table 4, $w_1 = 1$

Table 3
Summary of Equating Equations

General Equations for Common Item Linear Equating with Nonequivalent Populations

$$\ell(x) = \frac{\sigma_s(Y)}{\sigma_s(X)}\left[x - \mu_s(X)\right] + \mu_s(Y) \quad , \quad \text{where}$$

$$\mu_s(X) = \mu_1(X) - w_2\gamma_1\left[\mu_1(V) - \mu_2(V)\right]$$

$$\mu_s(Y) = \mu_2(Y) + w_1\gamma_2\left[\mu_1(V) - \mu_2(V)\right]$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2\gamma_1^2\left[\sigma_1^2(V) - \sigma_2^2(V)\right] + w_1w_2\gamma_1^2\left[\mu_1(V) - \mu_2(V)\right]^2$$

$$\sigma_s^2(Y) = \sigma_2^2(Y) + w_1\gamma_2^2\left[\sigma_1^2(V) - \sigma_2^2(V)\right] + w_1w_2\gamma_2^2\left[\mu_1(V) - \mu_2(V)\right]^2$$

Tucker Method

$$\gamma_1 = \alpha_1(X|V) = \sigma_1(X,V)/\sigma_1^2(V)$$

$$\gamma_2 = \alpha_2(Y|V) = \sigma_2(Y,V)/\sigma_2^2(V)$$

Levine Method for Inclusive Equating Section (Using Angoff Error Estimates)

$$\gamma_1 = 1/\alpha_1(V|X) = \sigma_1^2(X)/\sigma_1(X,V)$$

$$\gamma_2 = 1/\alpha_2(V|Y) = \sigma_2^2(Y)/\sigma_2(Y,V)$$

Levine Method for Exclusive Equating Section (Using Angoff Error Estimates)

$$\gamma_1 = \left[\sigma_1^2(X) + \sigma_1(X,V)\right]/\left[\sigma_1^2(V) + \sigma_1(X,V)\right]$$

$$\gamma_2 = \left[\sigma_2^2(Y) + \sigma_2(Y,V)\right]/\left[\sigma_2^2(V) + \sigma_2(Y,V)\right]$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

and $w_2 = 0$, and $\mu_2(X)$ is calculated using the Tucker and Levine methods. The γ_1 factor can be viewed as a factor that expands the difference between $\mu_1(V)$ and $\mu_2(V)$ to the Form X scale. Based on these values of γ, the difference between the groups on the Form X scale is 2.03 points for the Tucker method and 2.71 for the Levine method. As indicated earlier, this sort of difference will nearly always be larger for the Levine method than for the Tucker method.

Decomposing Observed Differences in Means and Variances

After an equating has been performed a question that is frequently asked, implicitly or explicitly, is "Did form differences or group differences contribute more to the observed mean differences?" This imprecisely stated question seems to be aimed at decomposing the difference $\mu_1(X) - \mu_2(Y)$ into two parts, one part associated with form differences and the other part associated with population differences. A similar question may be asked about the observed difference in variances, $\sigma_1^2(X) - \sigma_2^2(Y)$. These questions seem rather unambiguous and amenable to a simple answer. However, there are several perspectives that can be used in addressing these questions; the different perspectives can give different answers, and the answers are not always as simple or unambiguous as the questions may suggest. The approach taken below seems sensible, but it gives results that are more complicated than might be desired, except when $w_1 = 1$ and $w_2 = 0$.

Table 4
Illustration of the Effects of γ On Means In Linear Equating

Parameters from Table 1:	$\mu_1(X) = 95.70$, $\mu_1(V) = 23.20$, $\mu_2(V) = 22.50$
Assume:	$w_1 = 1$, $w_2 = 0$
Based on the equation	
for $\mu_s(X)$ in Table 3:	$\mu_2(X) = \mu_1(X) - \gamma_1[\mu_1(V) - \mu_2(V)]$
For the Tucker Method:	$\gamma_1 = 2.90$, so
	$\mu_2(X) = 95.70 - 2.90\ (23.20 - 22.50)$
	$= 95.70 - 2.90\ (.70)$
	$= 95.70 - 2.03$
	$= 93.67$
For the Levine Method:	$\gamma_1 = 3.87$, so
	$\mu_2(X) = 95.70 - 3.87\ (23.20 - 22.50)$
	$= 95.70 - 3.87\ (.70)$
	$= 95.70 - 2.71$
	$= 92.99$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Decomposing $\mu_1(X) - \mu_2(Y)$

Begin with a simple tautology:

$$\mu_1(X) - \mu_2(Y) = \{\mu_s(X) - \mu_s(Y)\} + \{[\mu_1(X) - \mu_s(X)] - [\mu_2(Y) - \mu_s(Y)]\} \quad . \tag{50}$$

Note that $\mu_s(X) - \mu_s(Y)$ is the difference in means for the two forms for the synthetic population; this difference can be termed the "form difference factor." The remaining terms are referred to as the "population difference factor."

After some algebra, it can be shown that, for linear equating methods in general,

$$\begin{aligned} \mu_1(X) - \mu_2(Y) = \; & w_1\{\mu_1(X) - \mu_1(Y)\} && \{\text{Form difference for Population 1}\} \\ & + w_2\{\mu_2(X) - \mu_2(Y)\} && \{\text{Form difference for Population 2}\} \\ & + w_2\{\mu_1(X) - \mu_2(X)\} && \{\text{Population difference on X scale}\} \\ & + w_1\{\mu_1(Y) - \mu_2(Y)\} && \{\text{Population difference on Y scale}\} \quad , \end{aligned} \tag{51}$$

where the verbal descriptions in braces describe the terms *in braces* (i.e., excluding the w_1 and w_2 weights).

In effect, Equation 51 decomposes both the form and population difference factors in Equation 50 into two parts. Hence, Equation 51 illustrates that, in general, the form difference factor $\mu_s(X) - \mu_s(Y)$ confounds form differences for the two underlying populations. Similarly, the population difference factor confounds population differences on the X and Y scales.

For the specific cases of the Tucker and Levine procedures, it can be shown that, in terms of directly estimable parameters (e.g., those in the numerical example in Table 1), Equation 51 can be expressed as

$$\begin{aligned} \mu_1(X) - \mu_2(Y) = \; & w_1\{\mu_1(X) - \mu_2(Y) - \gamma_2[\mu_1(V) - \mu_2(V)]\} && \{\text{Form difference for Population 1}\} \\ & + w_2\{\mu_1(X) - \mu_2(Y) - \gamma_1[\mu_1(V) - \mu_2(V)]\} && \{\text{Form difference for Population 2}\} \\ & + w_2\{\gamma_1[\mu_1(V) - \mu_2(V)]\} && \{\text{Population difference on X scale}\} \\ & + w_1\{\gamma_2[\mu_1(V) - \mu_2(V)]\} && \{\text{Population difference on Y scale}\} \quad . \end{aligned} \tag{52}$$

In effect, this equation states that the form difference for Population 1 is obtained by subtracting the population difference on the Y scale from $\mu_1(X) - \mu_2(Y)$. Similarly, the form difference for Population 2 is $\mu_1(X) - \mu_2(Y)$ minus the population difference on the X scale.

It seems that when most people ask about the contribution of form and population differences to $\mu_1(X) - \mu_2(Y)$, they expect an answer in two parts, as in Equation 50. However, Equation 50 is susceptible to misunderstanding unless it is interpreted in the sense of the terms in Equations 51 or 52. This potential problem in interpretation disappears when $w_1 = 1$ and $w_2 = 0$ because then, for linear equating procedures in general,

$$\begin{aligned} \mu_1(X) - \mu_2(Y) = \; & \{\mu_1(X) - \mu_1(Y)\} && \{\text{Form difference for Population 1}\} \\ & + \{\mu_1(Y) - \mu_2(Y)\} && \{\text{Population difference on Y scale}\} \quad , \end{aligned} \tag{53}$$

and for the Tucker and Levine procedures in particular,

$$\begin{aligned} \mu_1(X) - \mu_2(Y) = \; & \{\mu_1(X) - \mu_2(Y) - \gamma_2[\mu_1(V) - \mu_2(V)]\} && \{\text{Form difference for Population 1}\} \\ & + \{\gamma_2[\mu_1(V) - \mu_2(V)]\} && \{\text{Population difference on Y scale}\} \quad . \end{aligned} \tag{54}$$

Decomposing $\sigma_1^2(X) - \sigma_2^2(Y)$

A similar approach can be taken to decomposing the difference in the variances of the total scores on Forms X and Y. Begin with the tautology

$$\begin{aligned} \sigma_1^2(X) - \sigma_2^2(Y) = \; & \{\sigma_s^2(X) - \sigma_s^2(Y)\} && \{\text{Form difference factor}\} \\ & + \{[\sigma_1^2(X) - \sigma_s^2(X)] - [\sigma_2^2(Y) - \sigma_s^2(Y)]\} && \{\text{Population difference factor}\} \quad . \end{aligned} \tag{55}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Given Equation 55, for linear equating procedures in general, it can be shown that

$$\begin{aligned}\sigma_1^2(X)-\sigma_2^2(Y) &= w_1\{\sigma_1^2(X)-\sigma_1^2(Y)\} && \{\text{Form difference for Population 1}\}\\ &+ w_2\{\sigma_2^2(X)-\sigma_2^2(Y)\} && \{\text{Form difference for Population 2}\}\\ &+ w_2\{\sigma_1^2(X)-\sigma_2^2(X)\} && \{\text{Population difference on X scale}\}\\ &+ w_1\{\sigma_1^2(Y)-\sigma_2^2(Y)\} && \{\text{Population difference on Y scale}\}\quad , \end{aligned} \tag{56}$$

and for the Tucker and Levine procedures in particular

$$\begin{aligned}\sigma_1^2(X)-\sigma_2^2(Y) &= w_1\{\sigma_1^2(X)-\sigma_2^2(Y)-\gamma_2^2[\sigma_1^2(V)-\sigma_2^2(V)]\} && \{\text{Form difference for Population 1}\}\\ &+ w_2\{\sigma_1^2(X)-\sigma_2^2(Y)-\gamma_1^2[\sigma_1^2(V)-\sigma_2^2(V)]\} && \{\text{Form difference for Population 2}\}\\ &+ w_2\{\gamma_1^2[\sigma_1^2(V)-\sigma_2^2(V)]\} && \{\text{Population difference on X scale}\}\\ &+ w_1\{\gamma_2^2[\sigma_1^2(V)-\sigma_2^2(V)]\} && \{\text{Population difference on Y scale}\}\quad . \end{aligned} \tag{57}$$

Note, however, that the weighted form differences for Populations 1 and 2 (first two terms in Equations 56 and 57) do not sum to the form difference factor $[\sigma_s^2(X) - \sigma_s^2(Y)]$ in Equation 55. Similarly, the weighted population difference factors on the X and Y scales (last two terms in Equations 56 and 57) do not sum to the population difference factor in Equation 55. The seeming inconsistency results from the fact that the form difference factor involves the term

$$w_1w_2\{[\mu_1(X)-\mu_2(X)]^2-[\mu_1(Y)-\mu_2(Y)]^2\} \tag{58}$$

and the population difference factor involves the negative of this term.

In short, the contribution of form and population differences to $\sigma_1^2(X) - \sigma_2^2(Y)$ is not quite as simple as Equation 55 appears to suggest. However, equations and interpretations are considerably simplified when $w_1 = 1$ and $w_2 = 0$, because then the second and third terms in Equations 56 and 57 are zero, the form difference factor in Equation 55 does indeed equal the form difference for Population 1, and the population difference factor in Equation 55 does indeed equal the population difference on the Y scale.

Example

Consider the illustrative data in Table 1 where $\mu_1(X) - \mu_2(Y) = 95.7 - 96.8 = -1.10$, and assume that Tucker equating is performed with $w_1 = w_2 = .5$. Using Equation 52,

$$\begin{aligned}\mu_1(X)-\mu_2(Y) &= .5\,[-1.10-2.7(23.2-22.5)]\\ &+ .5\,[-1.10-2.9(23.2-22.5)]\\ &+ .5\,(2.9)(23.2-22.5)\\ &+ .5\,(2.7)(23.2-22.5)\\ &= (-2.99-3.13+2.03+1.89)/2 = -1.10\quad . \end{aligned} \tag{59}$$

In terms of Equation 50,

$$\mu_1(X)-\mu_2(Y) = -3.06+1.96 = -1.10\quad . \tag{60}$$

Roughly speaking, then, the new Form X is about 3 points more difficult than the old Form Y, and the new Population 1 is about 2 points more able than the old Population 2. However, this statement must be understood in the sense of the numerical results obtained using Equation 52.

For the same case considered above, the difference in observed variances is $\sigma_1^2(X) - \sigma_2^2(Y) = (13.4)^2 - (13.4)^2 = 0$, but this zero difference does not mean that the form and population differences are also zero. In fact, in this case, using Equation 57 with the Tucker procedure,

$$\begin{aligned}\sigma_1^2(X)-\sigma_2^2(Y) &= .5\{0-[(2.7)^2][(4.0)^2-(4.3)^2]\}\\ &+ .5\{0-[(2.9)^2][(4.0)^2-(4.3)^2]\}\\ &+ .5[(2.9)^2][(4.0)^2-(4.3)^2]\\ &+ .5[(2.7)^2][(4.0)^2-(4.3)^2]\\ &= (+18.15+20.94-20.94-18.15)/2\quad . \end{aligned} \tag{61}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

If it is assumed that $w_1 = 1$ and $w_2 = 0$, then the decompositions using Tucker equating with the illustrative data in Table 1 are different, but considerably simpler. Specifically, using Equation 54,

$$\begin{aligned}\mu_1(X) - \mu_2(Y) &= [-1.10 - 2.7(23.2 - 22.5)] + 2.7(23.2 - 22.5) \\ &= -2.99 + 1.89 = -1.10 \quad .\end{aligned} \tag{62}$$

Using Equation 57, the difference $\sigma_1^2(X) - \sigma_2^2(Y) = (13.4)^2 - (13.4)^2 = 0$ is decomposed as follows:

$$\begin{aligned}\sigma_1^2(X) - \sigma_2^2(Y) &= \{0 - [(2.7)^2][(4.0)^2 - (4.3)^2]\} + [(2.7)^2][(4.0)^2 - (4.3)^2] \\ &= 0 + 18.15 - 18.15 = 0 \quad .\end{aligned} \tag{63}$$

Conclusions

The demonstration that the Levine method weights examinee group differences to a greater extent than the Tucker method should aid in the interpretation of differences between the results when the methods are used with real data. The scheme developed above for decomposing means and variances into form and group differences yields rather complicated equations. Still, these decompositions should be useful in practice. The explicit consideration of the synthetic population in this paper facilitates the comparison of equating methods and the decomposition of means and variances. This illustrates the importance of the synthetic population concept for interpreting the results of linear equating in the common-item nonequivalent-populations design.

Appendix

This appendix demonstrates some relationships between the γ factors for the Tucker and Levine methods. For internal common items, assume that $0 < \rho_1^2(X,V) < 1$. Also assume a subscript of 1 on all slopes, correlations, and variances. Let γ_T refer to the γ_1 for the Tucker method, and let γ_L refer to the γ_1 for the Levine method with an internal set of common items. Thus,

$$\gamma_T = \alpha(X|V) = \frac{\rho(X,V)\sigma(X)}{\sigma(V)} \tag{A1}$$

and

$$\gamma_L = \frac{1}{\alpha(V|X)} = \frac{\sigma(X)}{\rho(X,V)\sigma(V)} \quad . \tag{A2}$$

From this it can be shown that $\gamma_T/\rho(X,V) = \gamma_L\rho(X,V)$, which implies that $\gamma_T = \rho^2(X,V)\gamma_L$. Thus, for all $0 < \rho^2(X,V) < 1$, $\gamma_T < \gamma_L$. When $\rho^2(X,V) = 1$, $\gamma_T = \gamma_L$. When $\rho(X,V) = 0$, γ_L is undefined because $\rho(X,V)$ is in the denominator of γ_L.

For external equating sections it can be shown that

$$\gamma_T = \frac{\rho(X,V)\sigma(X)}{\sigma(V)} \tag{A3}$$

and

$$\gamma_L = \frac{[\sigma^2(X) + \rho(X,V)\sigma(X)\sigma(V)]}{[\sigma^2(V) + \rho(X,V)\sigma(X)\sigma(V)]} \quad . \tag{A4}$$

By finding a common denominator for γ_T and γ_L it can be shown that

$$\gamma_L - \gamma_T = \frac{\sigma^2(X)[1 - \rho^2(X,V)]}{\sigma(V)[\sigma(V) + \rho(X,V)\sigma(X)]} \quad . \tag{A5}$$

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Assuming that $\sigma(X)$ and $\sigma(V)$ are positive, then for any possible value of $\rho(X,V)$, $-1 \leq \rho(X,V) \leq 1$, the numerator is ≥ 0. The denominator is positive if $\rho(X,V) > -[\sigma(V)/\sigma(X)]$. Thus, when the correlation between X and V is positive, γ for the Levine method is greater than γ for the Tucker method. This property holds even for some negative values of $\rho(X,V)$.

References

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

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 508–600). Washington DC: American Council on Education. (Reprinted by Educational Testing Service, Princeton NJ, 1984.)

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.

Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.

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

Levine, R. S. (1955). *Equating the score scales of alternate forms administered to samples of different ability* (ETS Research Bulletin No. 23). Princeton NJ: Educational Testing Service.

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

Author's Address

Send requests for reprints or further information to Michael J. Kolen, ACT, P.O. Box 168, Iowa City IA 52243, U.S.A.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/