Chapter 2
Observed Score Equating Using the Random Groups Design

As was stressed in Chap. 1 the same specifications property is an essential property of equating, which means that the forms to be equated must be built to the same content and statistical specifications. We also stressed that the symmetry property is essential for any equating relationship. The focus of the present chapter is on methods that are designed to achieve the observed score equating property, along with the same specifications and symmetry properties. As was described in Chap. 1, these observed score equating methods are developed with the goal that, after equating, converted scores on two forms have at least some of the same score distribution characteristics in a population of examinees.

In this chapter, these methods are developed in the context of the random groups design. Of the designs discussed thus far, the assumptions required for the random groups design are the least severe and most readily achieved. Thus, very few sources of systematic error are present with the random groups design. Because of the minimal assumptions required with the random groups design, this design is ideal for use in presenting the basic statistical methods in observed score equating, which is the focus of the present chapter.

The definitions and properties of mean, linear, and equipercentile equating methods are described in this chapter. These methods are presented, initially, in terms of population parameters (e.g., population means and standard deviations) for a specific population of examinees. We also discuss the process of estimating equating relationships, which requires that statistics (e.g., sample means and standard deviations) be substituted in place of the parameters. The methods then are illustrated using a real data example. Following the presentation of the methods, issues in using scale scores are described and illustrated. We then briefly discuss equating using the single group design.

An important practical challenge in using the random groups design is to obtain large enough sample sizes so that random error (see Chap. 7 for a discussion of standard errors) is at an acceptable level (rules of thumb for appropriate sample sizes are given in Chap. 8. For the equipercentile equating method, in Chap. 3 we describe statistical smoothing methods that often are used to help reduce random error when conducting equipercentile equating using the random groups design.

M. J. Kolen and R. L. Brennan, *Test Equating, Scaling, and Linking*,
Statistics for Social and Behavioral Sciences, DOI: 10.1007/978-1-4939-0317-7_2,

For simplicity, the statistical methods in this chapter are developed using a testing situation in which tests consist of test items that are scored correct (1) or incorrect (0), and where the total score is the number of items answered correctly. Near the end of the chapter, a process for equating tests that are scored using other scoring schemes is described.

2.1 Mean Equating

In mean equating, Form X is considered to differ in difficulty from Form Y by a constant amount along the score scale. For example, under mean equating, if Form X is 2 points easier than Form Y for high-scoring examinees, it is also 2 points easier than Form Y for low-scoring examinees. Although a constant difference might be overly restrictive in many testing situations, mean equating is useful for illustrating some important equating concepts.

As was done in Chap. 1, define Form X as the new form, let X represent the random variable score on Form X, and let x represent a particular score on Form X (i.e., a realization of X); and define Form Y as the old form, let Y represent the random variable score on Form Y, and let y represent a particular score on Form Y (i.e., a realization of Y). Also, define $\mu(X)$ as the mean on Form X and $\mu(Y)$ as the mean on Form Y for a population of examinees. In mean equating, scores on the two forms that are an equal (signed) distance away from their respective means are set equal:

$$x - \mu(X) = y - \mu(Y). \tag{2.1}$$

Then solve for y and obtain

$$m_Y(x) = y = x - \mu(X) + \mu(Y). \tag{2.2}$$

In this equation, $m_Y(x)$ refers to a score x on Form X transformed to the scale of Form Y using mean equating.

As an illustration of how to apply this formula, consider the situation discussed in Chap. 1, in which the mean on Form X was 72 and the mean on Form Y was 77. Based on this example, Eq. (2.2) indicates that 5 points would need to be added to a Form X score to transform a score on Form X to the Form Y scale. That is,

$$m_Y(x) = x - 72 + 77 = x + 5.$$

For example, using mean equating, a score of 72 on Form X is considered to indicate the same level of achievement as a score of 77 ($77 = 72 + 5$) on Form Y. And, a score of 75 on Form X is considered to indicate the same level of achievement as a score of 80 on Form Y. Thus, mean equating involves the addition of a constant (which might be negative) to all raw scores on Form X to find equated scores on Form Y.

2.2 Linear Equating

Rather than considering the differences between two forms to be a constant, linear equating allows for the differences in difficulty between the two test forms to vary along the score scale. For example, linear equating allows Form X to be more difficult than Form Y for low-achieving examinees but less difficult for high-achieving examinees.

In linear equating, scores that are an equal (signed) distance from their means in standard deviation units are set equal. Thus, linear equating can be viewed as allowing for the scale units, as well as the means, of the two forms to differ. Define $\sigma(X)$ and $\sigma(Y)$ as the standard deviations of Form X and Form Y scores, respectively. The linear conversion is defined by setting standardized deviation scores (z-scores) on the two forms to be equal such that

$$\frac{x - \mu(X)}{\sigma(X)} = \frac{y - \mu(Y)}{\sigma(Y)}. \tag{2.3}$$

If the standard deviations for the two forms were equal, Eq. (2.3) could be simplified to equal the mean equating Eq. (2.2). Thus, if the standard deviations of the two forms are equal, then mean and linear equating produce the same result. Solving for y in Eq. (2.3),

$$l_Y(x) = y = \sigma(Y)\left[\frac{x - \mu(X)}{\sigma(X)}\right] + \mu(Y), \tag{2.4}$$

where $l_Y(x)$ is the linear conversion equation for converting observed scores on Form X to the scale of Form Y. By rearranging terms, an alternate expression for $l_Y(x)$ is

$$l_Y(x) = y = \frac{\sigma(Y)}{\sigma(X)}x + \left[\mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\mu(X)\right]. \tag{2.5}$$

This expression is a linear equation of the form *slope* $(x) +$ *intercept* with

$$\mathit{slope} = \frac{\sigma(Y)}{\sigma(X)}, \quad \text{and} \quad \mathit{intercept} = \mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\mu(X). \tag{2.6}$$

What if the standard deviations in the mean equating example were $\sigma(X) = 10$ and $\sigma(Y) = 9$? The slope is $9/10 = .9$, and the intercept is $77 - (9/10)72 = 12.2$. The resulting conversion equation is $l_Y(x) = .9x + 12.2$. What is $l_Y(x)$ if $x = 75$?

$$l_Y(75) = .9(75) + 12.2 = 79.7.$$

How about if $x = 77$ or $x = 85$?

$$l_Y(77) = .9(77) + 12.2 = 81.5, \text{ and}$$
$$l_Y(85) = .9(85) + 12.2 = 88.7.$$

These equated values illustrate that the difference in test form difficulty varies with score level. For example, the difference in difficulty between Form X and Form Y for a Form X score of 75 is 4.7(79.7 − 75), whereas the difference for a Form X score of 85 is 3.7(88.7 − 85).

2.3 Properties of Mean and Linear Equating

In general, what are the properties of the equated scores? From Chapter 1, **E** is the expectation operator. The mean of a variable is found by taking the expected value of that variable. Using Eq. (2.2), the mean converted score $m_Y(x)$, for mean equating is

$$\mathbf{E}[m_Y(X)] = \mathbf{E}[X - \mu(X) + \mu(Y)] = \mu(X) - \mu(X) + \mu(Y) = \mu(Y). \tag{2.7}$$

That is, for mean equating the mean of the Form X scores equated to the Form Y scale is equal to the mean of the Form Y scores. In the example described earlier, the mean of the equated Form X scores is 77 [recall that $m_Y(x) = x + 5$ and $\mu(X) = 72$], the same value as the mean of the Form Y scores. Note that standard deviations were not shown in Eq. (2.7). What would be the standard deviation of Form X scores converted using the mean equating Eq. (2.2)? Because the Form X scores are converted to Form Y by adding a constant, the standard deviation of the converted scores would be the same as the standard deviation of the scores prior to conversion. That is, under mean equating, $\sigma[m_Y(X)] = \sigma(X)$.

Using Eq. (2.5), the mean equated score for linear equating can be found as follows:

$$\begin{aligned}\mathbf{E}[l_Y(X)] &= \mathbf{E}\left[\frac{\sigma(Y)}{\sigma(X)}X + \mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\mu(X)\right]\\ &= \frac{\sigma(Y)}{\sigma(X)}\mathbf{E}(X) + \mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\mu(X)\\ &= \mu(Y),\end{aligned}\tag{2.8}$$

because $\mathbf{E}(X) = \mu(X)$.

The standard deviation of the equated scores is found by first substituting Eq. (2.5) for $l_Y(X)$ as follows:

$$\sigma[l_Y(X)] = \sigma\left[\frac{\sigma(Y)}{\sigma(X)}X + \mu(Y) - \frac{\sigma(Y)}{\sigma(X)}\mu(X)\right]$$

To continue, the standard deviation of a score plus a constant is equal to the standard deviation of the score. That is, $\sigma(X + \mathit{constant}) = \sigma(X)$. By recognizing in the linear equating equation that the terms to the right of the addition sign are a constant, the following holds:

$$\sigma[l_Y(X)] = \sigma\left[\frac{\sigma(Y)}{\sigma(X)}X\right].$$

Also note that the standard deviation of a score multiplied by a constant equals the standard deviation of the score multiplied by the constant. That is, $\sigma(\mathit{constant}\ X) = \mathit{constant}\ \sigma(X)$. Noting that the ratio of standard deviations in the large parentheses is also a constant that multiplies X,

$$\sigma[l_Y(X)] = \frac{\sigma(Y)}{\sigma(X)}\sigma(X) = \sigma(Y). \tag{2.9}$$

Therefore, the mean and standard deviation of the Form X scores equated to the Form Y scale are equal to the mean and standard deviation, respectively, of the Form Y scores. In the example described earlier for linear equating, the mean of the equated Form X scores is 77 and the standard deviation is 9; these are the same values as the mean and standard deviation of the Form Y scores.

Consider the equation for mean equating, Eq. (2.2), and the equation for linear equating (2.5). If either of the equations were solved for x, rather than for y, the equation for equating Form Y scores to the scale of Form X would result. These conversions would be symbolized by $m_X(y)$ and $l_X(y)$, respectively. Equating relationships are defined as being *symmetric* because the equation used to convert Form X scores to the Form Y scale is the inverse of the equation used to convert Form Y scores to the Form X scale.

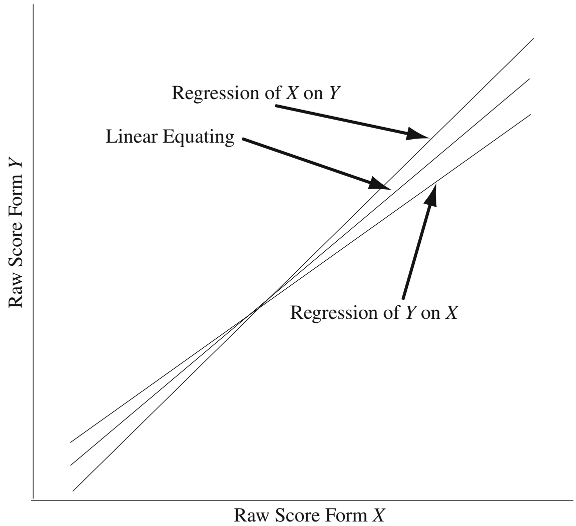
The equation for linear equating (2.5) is deceptively like a linear regression equation. The difference is that, for linear regression, the $\sigma(Y)/\sigma(X)$ terms are multiplied by the correlation between X and Y. However, a linear regression equation does not qualify as an equating function because the regression of X on Y is different from the regression of Y on X, unless the correlation coefficient is 1. For this reason, regression equations cannot, in general, be used as equating functions. The comparison between linear regression and linear equating is illustrated in Fig. 2.1. The regression Y on X is different from the regression of X on Y. Also note that there is only one linear equating relationship graphed in the figure. This relationship can be used to transform Form X scores to the Form Y scale, or to transform Form Y scores to the Form X scale.

2.4 Comparison of Mean and Linear Equating

Figure 2.2 illustrates the equating of Form X and Form Y using the hypothetical test forms already discussed. The equations for equating scores on Form X to the Form Y scale are plotted in this figure.

Also plotted in this figure are the results from the "identity equating." In the *identity equating*, a score on Form X is considered to be equivalent to the identical score on Form Y; for example, a 40 on Form X is considered to be equivalent to a 40

Fig. 2.1 Comparison of linear regression and linear equating



on Form Y. Identity equating would be the same as mean and linear equating if the two forms were identical in difficulty all along the score scale.

To find a Form Y equivalent of a Form X score using the graph, find the Form X value of interest on the horizontal axis, go up to the function, and then go over to the vertical axis to read off the Form Y equivalent.

How to find the Form Y equivalent of a Form X score of 72 is illustrated in the figure using the arrows. This equivalent is 77, using either mean or linear equating. The score 72 is the mean score on Form X. As indicated earlier, both mean and linear equating will produce the same result at the mean.

Now refer to the identity equating line in the figure, and note that the line for mean equating is parallel to the line for the identity equating. The lines for these two methods will always be parallel. As can be seen, the line for mean equating is uniformly 5 points vertically above the line for the identity equating, because Form Y is, on average, 5 points less difficult than Form X. Refer to the line for linear equating. This line is not parallel to the identity equating line. The linear equating line is further above the identity equating line at the low scores than at the high scores. This observation is consistent with the earlier discussion in which the difference in difficulty between Form X and Form Y was shown to be greater at the lower scores than at the higher scores.

Assume that the test in this example is scored number-correct. Number-correct scores for this 100-item test can range from 0 to 100. Figure 2.2 illustrates that equated scores from mean and linear equating can sometimes be out of the range of possible observed scores. The dotted lines at 0 on Form X and at 100 illustrate the boundaries of possible observed scores. For example, using linear equating, a score of 100 on Form X equates to a score of approximately 102 on Form Y. Also, using linear equating, a score of 0 on Form Y equates to a score of approximately −14 on

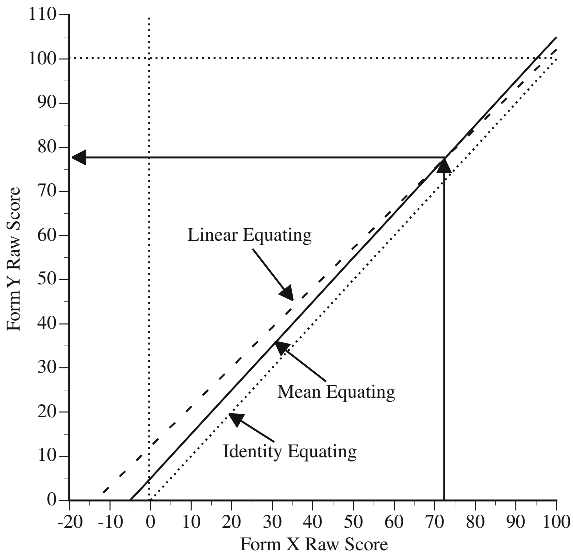


Fig. 2.2 Graph of mean and linear equating for a hypothetical 100-item test

Form X. There are a variety of ways to handle this problem. One way is to allow the top and bottom to "float." For example, the highest equated score might be allowed to exceed the highest raw score. An alternative is to truncate the conversion at the highest and lowest scores. In the example, truncation involves setting all converted scores greater than 100 equal to 100 and setting all converted scores less than 0 equal to 0. That is, all Form Y scores that equate to Form X scores below 0 would be set to 0 and all Form X scores that equate to Form Y scores above 100 would be set to 100. In practice, the decision about how to handle equated scores outside the range typically interacts with the score scale that is used for reporting scores. Sometimes this issue is effectively of no consequence, because no one achieves the extreme raw scores on Form X that equate to unobtainable scores on Form Y.

In summary, in mean equating the conversion is derived by setting the deviation scores on the two forms equal, whereas in linear equating the standardized deviation scores (z-scores) on the two forms are set equal. In mean equating, scores on Form X are adjusted by a constant amount that is equal to the difference between the Form Y and Form X means. In linear equating, scores on Form X are adjusted using a linear equation that allows for the forms to be differentially difficult along the score scale. In mean equating, the mean of the Form X scores equated to the Form Y scale is equal to the mean of the Form Y scores; whereas in linear equating, the standard deviation as well as the mean are equal. In general, mean equating is less complicated than linear equating, but linear equating provides for more flexibility in the conversion than does mean equating.

2.5 Equipercentile Equating

In equipercentile equating, a curve is used to describe form-to-form differences in difficulty, which makes equipercentile equating even more general than linear equating. Using equipercentile equating, for example, Form X could be more difficult than Form Y at high and low scores, but less difficult at the middle scores.

The equating function is an equipercentile equating function if the distribution of scores on Form X converted to the Form Y scale is equal to the distribution of scores on Form Y in the population. The equipercentile equating function is developed by identifying scores on Form X that have the same percentile ranks as scores on Form Y.

The definition of equipercentile equating developed by Braun and Holland (1982) is adapted for use here. Consider the following definitions of terms, some of which were presented previously:

X is a random variable representing a score on Form X, and x is a particular value (i.e., a realization) of X.

Y is a random variable representing a score on Form Y, and y is a particular value (i.e., a realization) of Y.

F is the cumulative distribution function of X in the population.

G is the cumulative distribution function of Y in the same population.

e_Y is a symmetric equating function used to convert scores on Form X to the Form Y scale.

G^* is the cumulative distribution function of e_Y in the same population. That is, G^* is the cumulative distribution function of scores on Form X converted to the Form Y scale.

The function e_Y is defined to be the equipercentile equating function in the population if

$$G^* = G. \tag{2.10}$$

That is, the function e_Y is the equipercentile equating function in the population if the cumulative distribution function of scores on Form X converted to the Form Y scale is equal to the cumulative distribution function of scores on Form Y.

Braun and Holland (1982) indicated that the following function is an equipercentile equating function when X and Y are continuous random variables:

$$e_Y(x) = G^{-1}[F(x)], \tag{2.11}$$

where G^{-1} is the inverse of the cumulative distribution function G.

As previously indicated, to be an equating function, e_Y must be symmetric. Define

e_X as a symmetric equating function used to convert scores on Form Y to the Form X scale, and

F^* as the cumulative distribution function of e_X in the population. That is, F^* is the cumulative distribution function of scores on Form Y converted to the Form X scale.

By the symmetry property,

$$e_X^{-1}(x) = e_Y(x) \text{ and } e_Y^{-1}(y) = e_X(y). \tag{2.12}$$

Also,

$$e_X(y) = F^{-1}[G(y)], \tag{2.13}$$

is the equipercentile equating function for converting Form Y scores to the Form X scale. In this equation, F^{-1} is the inverse of the cumulative distribution function F.

Following the definitions in Eqs. (2.10–2.13), an equipercentile equivalent for the population of examinees can be constructed in the following manner: For a given Form X score, find the percentage of examinees earning scores at or below that Form X score. Next, find the Form Y score that has the same percentage of examinees at or below it. These Form X and Form Y scores are considered to be equivalent. For example, suppose that 20 % of the examinees in the population earned a Form X score at or below 26 and 20 % of the examinees in the population earned a Form Y score at or below 27. Then a Form X score of 26 would be considered to represent the same level of achievement as a Form Y score of 27. Using equipercentile equating, a Form X score of 26 would be equated to a Form Y score of 27.

The preceding discussion was based on an assumption that test scores are continuous random variables. Typically, however, test scores are discrete. For example, number-correct scores take on only integer values. With discrete test scores, the definition of equipercentile equating is more complicated than the situation just described. Consider the following situation. Suppose that a test is scored number-correct and that the following is true of the population distributions:

1. 20 % of the examinees score at or below 26 on Form X.
2. 18 % of the examinees score at or below 27 on Form Y.
3. 23 % of the examinees score at or below 28 on Form Y.

What is the Form Y equipercentile equivalent of a Form X score of 26? No Form Y score exists that has precisely 20 % of the scores at or below it. Strictly speaking, no Form Y equivalent of a Form X score of 26 exists. Thus, the goal of equipercentile equating stated in Eq. (2.10) cannot be met strictly when test scores are discrete.

How can equipercentile equating be conducted when scores are discrete? A tradition exists in educational and psychological measurement to view discrete test scores as being continuous by using percentiles and percentile ranks. In this approach, an integer score of 28, for example, is considered to represent scores in the range 27.5–28.5. Examinees with scores of 28 are considered to be uniformly distributed in this range. The percentile rank of a score of 28 is defined as being the percentage of scores *below* 28. However, because only 1/2 of the examinees who score 28 are considered to be below 28 (the remainder being between 28 and 28.5), the percentile rank of 28 is the percentage of examinees who earned integer scores of 27 and below, plus 1/2 the percentage of examinees who earned an integer score of 28. Placing the preceding example in the context of percentile ranks, 18 % of the examinees earned

Table 2.1 Form X score distribution for a hypothetical four-item test

x	$f(x)$	$F(x)$	$P(x)$
0	.2	.2	10
1	.3	.5	35
2	.2	.7	60
3	.2	.9	80
4	.1	1.0	95

a Form Y score below 27.5 and 5 % (23–18 %) of the examinees earned a score between 27.5 and 28.5. So the percentile rank of a Form Y score of 28 would be 18 % + 1/2(5 %) = 20.5 %. In the terminology typically used, the percentile rank of an integer score is the percentile rank at the midpoint of the interval that contains that score.

Holland and Thayer (1989) presented a statistical justification for using percentiles and percentile ranks. In their approach, they use what they refer to as a *continuization* process and a kernel smoothing process. Given a discrete integer-valued random variable X and a random variable U that is uniformly distributed over the range $-1/2$ to $+1/2$, they defined a new random variable, $X^* = X + U$.

This new random variable is continuous. The cumulative distribution function of this new random variable corresponds to the percentile rank function. The inverse of the cumulative distribution of this new function exists and is the percentile function. Holland and Thayer (1989) also generalized their approach to incorporate continuization processes that are based on distributions other than the uniform.

This approach was developed further by von Davier et al. (2004) and is discussed in more detail in Chap. 3. In the present chapter, the traditional approach to percentiles and percentile ranks is followed.

The equipercentile methods presented next assume that the observed scores on the tests to be equated are integer scores that range from zero through the number of items on the test, as would be true of tests scored number-correct. Generalizations to other scoring schemes are discussed as well.

2.5.1 Graphical Procedures

Equipercentile equating using graphical methods provides a conceptual framework for subsequent consideration of analytic methods. A hypothetical four-item test is used to illustrate the graphical process for equipercentile equating. Data for Form X are presented in Table 2.1.

In this table, x refers to test score and $f(x)$ to the proportion of examinees earning the score x. For example, the proportion of examinees earning a score of 0 is .20. $F(x)$ is the cumulative proportion at or below x. For example, the proportion of examinees scoring 3 or below is .9. $P(x)$ refers to the percentile rank, and for an

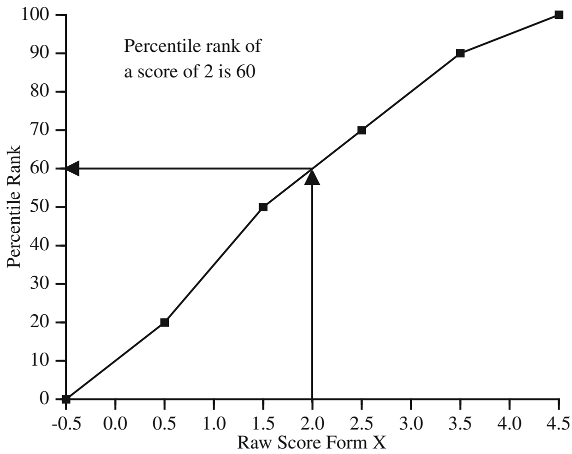


Fig. 2.3 Form X percentile ranks on a hypothetical four-item test

integer score it equals the percentage of examinees below x plus 1/2 the percentage of examinees at x—i.e., for integer score x, $P(x) = 100[F(x-1) + f(x)/2]$.

To be consistent with traditional definitions of percentile ranks, the percentile rank function is plotted as points at the upper limit of each score interval. For example, the percentile rank of a score of 3.5 is 90, which is 100 times the cumulative proportion at or below 3. Therefore, to plot the percentile ranks, plot the percentile ranks at each integer score plus .5. The percentile ranks at an integer score plus .5 can be found from Table 2.1 by taking the cumulative distribution function values, $F(x)$, at an integer and multiplying them by 100 to make them percentages. Figure 2.3 illustrates how to plot the percentile rank distribution for Form X.

A percentile rank of 0 is also plotted at a Form X score of $-.5$. The points are then connected with straight lines. An example is presented for finding the percentile rank of a Form X integer score of 2 using the arrows in Fig. 2.3. As can be seen, the percentile rank of a score of 2 is 60, which is the same result found in Table 2.1.

In Fig. 2.3, percentile ranks of scores between $-.5$ and 0.0 are greater than zero. These nonzero percentile ranks result from using the traditional definition of percentile ranks, in which scores of 0 are assumed to be uniformly distributed from $-.5$ to .5. Also, scores of 4 are considered to be uniformly distributed between 3.5 to 4.5, so that scores above 4 have percentile ranks less than 100. Under this conceptualization, the range of possible scores is treated as being between $-.5$ and the highest integer score $+.5$.

Data from Form Y also need to be used in the equating process. The data for Form Y are presented along with the Form X data in Table 2.2. In this table, y refers to Form Y scores, $g(y)$ to the proportion of examinees at each score, $G(y)$ to the proportion at or below each score, and $Q(y)$ to the percentile rank at each score. Percentile ranks for Form Y are plotted in the same manner as they were for Form X. To find the equipercentile equivalent of a particular score on Form X, find the Form Y score with the same percentile rank. Figure 2.4 illustrates this process for finding

Table 2.2 Form X and Form Y distributions for a hypothetical four-item test

y	$g(y)$	$G(y)$	$Q(Y)$	x	$f(x)$	$F(x)$	$P(x)$
0	.1	.1	5	0	.2	.2	10
1	.2	.3	20	1	.3	.5	35
2	.2	.5	40	2	.2	.7	60
3	.3	.8	65	3	.2	.9	80
4	.2	1.0	90	4	.1	1.0	95

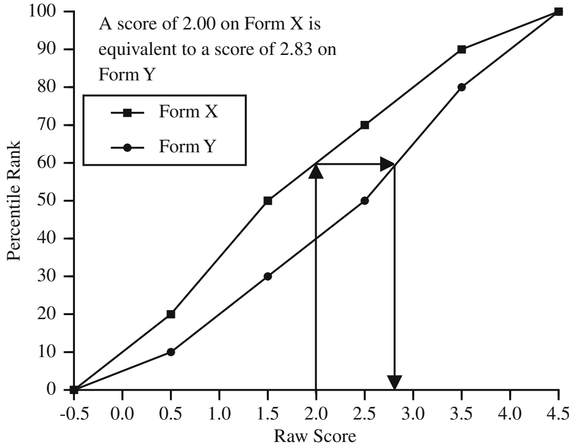


Fig. 2.4 Graphical equipercentile equating for a hypothetical four-item test

the equipercentile equivalent of a Form X score of 2. As indicated by the arrows, a Form X score of 2 has a percentile rank of 60. Following the arrows, it can be seen that the Form Y score of about 2.8 (actually 2.83) is equivalent to the Form X score of 2.

The equivalents can also be plotted. To construct such a graph, plot, as points, Form Y equivalents of Form X scores at each integer plus .5. Then plot Form X equivalents of Form Y scores at each integer plus .5. To handle scores below the lowest integer scores $+.5$, a point is plotted at the (x, y) pair $(-.5, -.5)$. The plotted points are then connected by straight lines. This process is illustrated for the example in Fig. 2.5. As indicated by the arrows in the figure, a Form X score of 2 is equivalent to a Form Y score of 2.8 (actually 2.83), which is consistent with the result found earlier. This plot of equivalents displays the Form Y equivalents of Form X scores.

In summary, the graphical process of finding equipercentile equivalents is as follows: Plot percentile ranks for each form on the same graph. To find a Form Y equivalent of a Form X score, start by finding the percentile rank of the Form X score. Then find the Form Y score that has that same percentile rank. Equivalents can be plotted in a graph that shows the equipercentile relationship between the two forms.

One issue that arises in equipercentile equating is how to handle situations in which no examinees earn a particular score. When this occurs, the score that corresponds

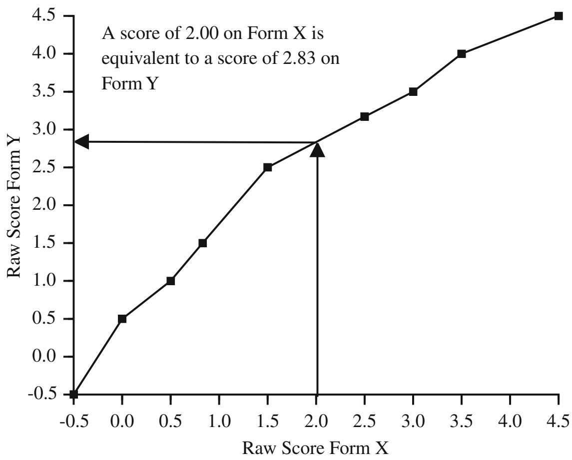


Fig. 2.5 Equipercentile equivalents for a hypothetical four-item test

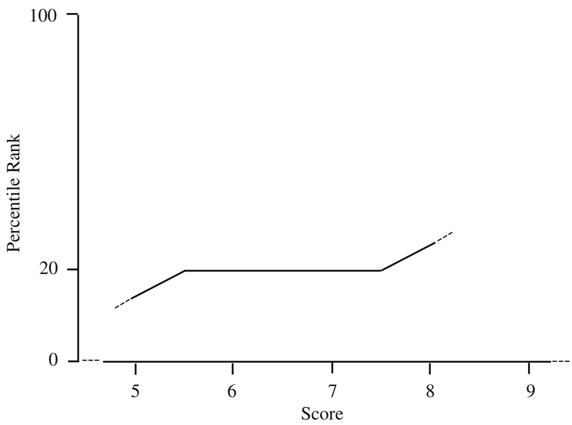


Fig. 2.6 Illustration of percentile ranks when no examinees earn a particular score

to a particular percentile rank might not be unique. Suppose for example that x has a percentile rank of 20. To find the equipercentile equivalent, the Form Y score that has a percentile rank of 20 needs to be found. Suppose, however, that there is no unique score on Form Y that has a percentile rank of 20, as illustrated in Fig. 2.6.

The percentile ranks shown in Fig. 2.6 could occur if no examinees earned scores of 6 and 7. In this case, the graph indicates that scores in the range 5.5 to 7.5 all have percentile ranks of 20. The choice of the Form Y score that has a percentile rank of 20 is arbitrary. In this situation, usually the middle score would be chosen. So, in the example the score with a percentile rank of 20 would be designated as 6.5. Choosing the middle score is arbitrary, technically, but doing so seems sensible.

2.5.2 Analytic Procedures

The graphical method discussed in the previous section is not likely to be viable for equating a large number of real forms in real time. In addition, equating using graphical procedures can be inaccurate. What is needed are formulas that provide more formal definitions of percentile ranks and equipercentile equivalents. The following discussion provides such formulas. The result of applying these formulas is to produce percentile ranks and equipercentile equivalents that are equal to those that would result using the graphical procedures.

To define percentile ranks, let K_X represent the number of items on Form X of a test. Define X as a random variable representing test scores on Form X that can take on the integer values $0, 1, \ldots, K_X$. Define $f(x)$ as the discrete density function for $X = x$. That is,

$$\begin{aligned}
&f(x) \geq 0 \text{ for integer scores } x = 0, 1, \ldots, K_X;\\
&f(x) = 0 \text{ otherwise; and}\\
&\sum f(x) = 1.
\end{aligned}$$

Define $F(x)$ as the discrete cumulative distribution function. That is, $F(x)$ is the proportion of examinees in the population earning a score *at or below* x. Therefore,

$$\begin{aligned}
&0 \leq F(x) \leq 1 \text{ for } x = 0, 1, \ldots, K_X;\\
&F(x) = 0 \text{ for } x < 0; \text{ and}\\
&F(x) = 1 \text{ for } x > K_X.
\end{aligned}$$

Consider a possible noninteger value of x. Define x^* as that integer that is closest to x such that $x^* - .5 \leq x < x^* + .5$. For example, if $x = 5.7$, $x^* = 6$; if $x = 6.4$, $x^* = 6$; and if $x = 5.5$, $x^* = 6$. The percentile rank function for Form X is

$$\begin{aligned}
P(x) = {} & 100\{F(x^* - 1) + [x - (x^* - .5)][F(x^*) - F(x^* - 1)]\},\\
& \quad -.5 \leq x < K_X + .5,\\
= {} & 0, \qquad x < -.5,\\
= {} & 100, \qquad x \geq K_x + .5.
\end{aligned} \tag{2.14}$$

To illustrate how this equation functions, consider the following example based on the data in Table 2.1. Calculate the percentile rank for a score of $x = 1.3$, using Eq. (2.14):

$$\begin{aligned}
P(1.3) &= 100\{F(0) + [1.3 - (1 - .5)][F(1) - F(0)]\}\\
&= 100\{.2 + [.8][.5 - .2]\} = 100\{.2 + .24\} = 44.
\end{aligned}$$

In this case, $x^* = 1.0$, because 1 is the integer score that is closest to 1.3. The term $[F(1) - F(0)] = .5 - .2 = .3$ represents the proportion of examinees earning a score of 1. These scores are considered to range from .5 to 1.5. The term $[1.3 - (1 - .5)] = .8$ indicates that the score of 1.3 is, proportionally, .8 of the distance between .5 and 1.5. So, $[.8][.3] = .24$ represents the probability of scoring between .5 and 1.3. The probability of scoring below .5 is represented by $F(0) = .2$. Therefore, the percentile rank of a score of 1.3 equals 44.

The inverse of the percentile rank function, which often is referred to as the percentile function, is symbolized as P^{-1}. Two alternate percentile functions are given as follows. These functions produce the same result, unless some of the probabilities are zero. Given a percentile rank (e.g., the 10th percentile rank), this inverse function is used to find the score corresponding to that percentile rank. To find this function, solve Eq. (2.14) for x. Specifically, for a given percentile rank P^*, the percentile is

$$\begin{aligned} x_U(P^*) = P^{-1}[P^*] &= \frac{P^*/100 - F(x_U^* - 1)}{F(x_U^*) - F(x_U^* - 1)} + (x_U^* - .5), \quad 0 \le P^* < 100, \\ &= K_X + .5, \qquad\qquad P^* = 100. \end{aligned} \tag{2.15}$$

In Eq. (2.15), for $0 \le P^* < 100$, x_U^* is the *smallest* integer score with a cumulative percent $[100F(x)]$ that is *greater than* P^*. An alternate expression for the percentile is

$$\begin{aligned} x_L(P^*) = P^{-1}[P^*] &= \frac{P^*/100 - F(x_L^*)}{F(x_L^* + 1) - F(x_L^*)} + (x_L^* + .5), \quad 0 < P^* \le 100, \\ &= -.5, \qquad\qquad P^* = 0. \end{aligned} \tag{2.16}$$

In Eq. (2.16), for $0 < P^* \le 100$, x_L^* is the *largest* integer score with a cumulative percent $[100F(x)]$ that is *less than* P^*. If the $f(x)$ are nonzero at all score points $0, 1, \ldots, K_X$, then $x = x_U = x_L$, and either expression can be used. If some of the $f(x)$ are zero, then $x_U \ne x_L$ for at least some percentile ranks. In this case, the convention $x = (x_U + x_L)/2$ is used. This convention produces the same results as the one described in association with Fig. 2.6 using the graphical procedures. In most situations, it seems reasonable to assume that the $f(x)$ are all nonzero over the integer score range $0, 1, \ldots, K_X$. For this reason, and to simplify issues, when considering population distributions in the following discussion, only Eq. (2.15) is used with $x_U = x$. When considering estimates of population distributions, estimated probabilities of zero are often encountered (i.e., when no examinees in a sample earn a particular score).

As an example of how to use Eq. (2.15), find the score corresponding to a percentile rank of 62 using the inverse of the percentile rank function using the data in Table 2.1. In this case $x_U^* = 2$ because, in Table 2.1, it is the *smallest* integer score with $F(x)$ that is *greater than* .62. Then

$$\begin{aligned}
P^{-1}(62) &= \frac{62/100 - F(1)}{F(2) - F(1)} + (2 - .5) \\
&= \frac{.62 - .5}{.7 - .5} + (2 - .5) = .12/.20 + 1.5 = .60 + 1.5 = 2.1.
\end{aligned}$$

In equipercentile equating, the interest is in finding a score on Form Y that has the same percentile rank as a score on Form X. Referring to y as a score on Form Y, let K_Y refer to the number of items on Form Y, let $g(y)$ refer to the discrete density of y, let $G(y)$ refer to the discrete cumulative distribution of y, let $Q(y)$ refer to the percentile rank of y, and let Q^{-1} refer to the inverse of the percentile rank function for Form Y. Then the Form Y equipercentile equivalent of score x on Form X is

$$e_Y(x) = y = Q^{-1}[P(x)], \qquad -.5 \le x \le K_X + .5. \tag{2.17}$$

This equation indicates that, to find the equipercentile equivalent of score x on the scale of Form Y, first find the percentile rank of x in the Form X distribution. Then find the Form Y score that has that same percentile rank in the Form Y distribution. Equation (2.17) is symmetric. That is, to find the Form X equivalent of a Form Y score, Eq. (2.17) is solved for y, giving $e_X(y) = P^{-1}[Q(y)]$.

Analytically, to find $e_Y(x)$ given by Eq. (2.17), use the analog of Eq. (2.15) for the Form Y distribution. That is, use

$$\begin{aligned}
e_Y(x) &= Q^{-1}[P(x)] \\
&= \frac{P(x)/100 - G(y_U^* - 1)}{G(y_U^*) - G(y_U^* - 1)} + (y_U^* - .5), \qquad 0 \le P(x) < 100, \\
&= K_Y + .5, \qquad\qquad P(x) = 100.
\end{aligned} \tag{2.18}$$

[Note that, to use this equation when some Form Y scores have zero probabilities, it also is necessary to use y_L^* as described in the discussion following Eq. (2.16).] Refer to Table 2.2. As an example of finding equipercentile equivalents, find the Form Y equipercentile equivalent of a Form X score of 2. The percentile rank of a Form X score of 2 is $P(2) = 60$, as is shown in Table 2.2. To find the equipercentile equivalent, the Form Y score that has a percentile rank of 60 must be found. Because 3 is the score with the *smallest* $G(y)$ that is *greater than* .60, $y_U^* = 3$. Thus, using Eq. (2.18),

$$e_Y(x) = Q^{-1}[60] = \frac{60/100 - .5}{.8 - .5} + (3 - .5) = .1/.3 + 2.5 = 2.8333.$$

The raw score equipercentile equivalents that result typically are noninteger. Noninteger scores arise through the continuization process used to define percentiles and percentile ranks. Issues related to rounding to integers are considered later in the discussion of scale scores.

Table 2.3 Form Y equivalents of Form X scores for a hypothetical four-item test

x	$f(x)$	$e_Y(x)$
0	.2	.50
1	.3	1.75
2	.2	2.8333
3	.2	3.50
4	.1	4.25

2.5.3 Properties of Equated Scores in Equipercentile Equating

Conducting equipercentile equating using Eq. (2.18) always results in equated scores in the range $-.5 \leq e_Y(x) \leq K_Y + .5$. Thus, equipercentile equating has the desirable property that the equated scores will always be within the range of possible scores under the traditional conceptualization of percentiles and percentile ranks. The problem of having equated scores that are out of the range of possible scores which occur with mean and linear equating does not occur with equipercentile equating.

Ideally, in equipercentile equating the equated scores on Form X would have the same distribution as the scores on Form Y. As was previously indicated, if test scores were continuous, then these distributions would be the same. However, test scores are discrete. A continuization process involving percentiles and percentile ranks was used to render the problem mathematically tractable. However, when the results of equating are applied to discrete scores, the equated Form X score distribution will differ from the Form Y distribution.

Consider the following illustration. Using the hypothetical four-item test from Tables 2.2 and 2.3 provides the Form Y equivalents of scores resulting from the use of Eq. (2.18). The moments that result are shown in Table 2.4, where skewness and kurtosis are defined for Form X, respectively, as

$$sk(X) = \frac{\mathbf{E}[X - \mu(X)]^3}{[\sigma(X)]^3}, \text{ and} \tag{2.19}$$

$$ku(X) = \frac{\mathbf{E}[X - \mu(X)]^4}{[\sigma(X)]^4}. \tag{2.20}$$

Central moments for other variables are defined similarly. To arrive at the moments of the equated scores, $e_Y(x)$, in Table 2.4, the Form X scores were equated to Form Y scores. For example, as indicated in Table 2.3, the proportion of examinees earning an $e_Y(x)$ of .50 is .20.

Moments of these equated scores then were found. Ideally, the moments for $e_Y(x)$ in Table 2.4 would be equal to those for y. As can be seen, however, there are departures. These departures are a result of the discreteness of the scores. The departures in Table 2.4 are relatively large because the test is so short. Departures likely would

2.4 Moments for equating Form X and Form Y of a hypothetical four-item test

Score	μ	σ	sk	ku
y	2.3000	1.2689	−.2820	1.9728
x	1.7000	1.2689	.2820	1.9728
$e_Y(x)$	2.3167	1.2098	−.0972	1.8733

be considerably less with longer, more realistic tests. For tests of realistic lengths, not being able to achieve the equal distribution goal precisely often is more of a theoretical concern than a practical one.

The approach taken here is to compare moments of the equated scores to the moments of the Form Y scores as was just done. von Davier et al. (2004) introduced the *percent relative error* index to compare these moments. The percent relative error is computed by finding the difference between a particular moment for the equated scores and that same moment for the Form Y scores. This difference is then divided by the same moment for the Form Y scores.

2.6 Estimating Observed Score Equating Relationships

So far, the methods have been described using population parameters. In practice, sample statistics are all that are available, and these sample statistics are substituted for the parameters in the preceding equations.

One estimation problem that occurs in equipercentile equating is how to calculate the function P^{-1} when the frequency at some score points is zero. The conventions associated with Eqs. (2.15) and (2.16) for averaging the results is one procedure for producing a unique result. Another procedure is to add a very small relative frequency to each score, and then adjust the relative frequencies so they sum to one. If *adj* is taken as this small quantity, then the adjusted relative frequencies on Form Y are

$$\hat{g}_{adj}(y) = \frac{\hat{g}(y) + adj}{1 + (K_Y + 1) \cdot adj},$$

where $\hat{g}(y)$ is the relative frequency that was observed. For example, if $K_Y = 10$, $adj = 10^{-6}$, and $\hat{g}(2) = .02$, then

$$\hat{g}_{adj}(2) = \frac{.02 + 10^{-6}}{1 + (10 + 1) \cdot 10^{-6}} = .02000078.$$

A similar procedure could be used for Form X. The equating then can be done using the adjusted relative frequencies. Experience has shown that a value around $adj = 10^{-6}$ can be used without creating a serious bias in the equating. A third solution to the zero frequency problem is to use smoothing methods, which are the subject of Chap. 3.

Data for an example of an equating of Form X and Form Y of the original ACT Mathematics test are presented in Table 2.5. This test contains 40 multiple-choice items scored incorrect (0) or correct (1). Form X was administered to 4,329 examinees and Form Y to 4,152 examinees in a spiral administration, which resulted in random groups of examinees being administered Form X and Form Y. The sample sizes for the two forms differ, in part, because Form X always preceded Form Y in the distribution of booklets in each testing room. Thus, one more Form X than Form Y booklet was administered in some testing rooms. In the table, a "^" is used to indicate an estimate of a population parameter, and N_X and N_Y refer to sample sizes for the forms. Consider, for example, a score of 10 on Form Y. From Table 2.5, 194 examinees earned a score of 10, and 857 examinees earned a score of 10 or below; the proportion of examinees earning a score of 10 is .0467, the proportion of examinees at or below a score of 10 is .2064, and the estimated percentile rank of a score of 10 is 18.30.

Percentile ranks for Forms X and Y are plotted in Fig. 2.7. The percentile ranks are plotted for each score point plus .5. Form X appears to be somewhat easier than Form Y, because the Form X distribution is shifted to the right. The relative frequency distributions are shown in Fig. 2.8.

Both score distributions are positively skewed, and Form X again appears to be somewhat easier than Form Y. Estimates of central moments for Form X and Form Y are given in the upper portion of Table 2.6. Both forms have means, $\hat{\mu}$, less than 20 (which is 50 % of the 40 items), so it appears that the tests are somewhat difficult for these examinees. Form X is, on average, nearly 1 point easier than Form Y. Based on the standard deviations, $\hat{\sigma}$, the distribution for Form X is less variable than the distribution for Form Y. As indicated by the skewness values, $\widehat{sk}$ the distributions are positively skewed, where skewness for the population is defined in Eq. (2.19). Based on the kurtosis estimates, $\widehat{ku}$, the distributions have lower kurtosis than a normal distribution, which would have a kurtosis value of 3, where kurtosis for the population is defined in Eq. (2.20).

The conversions for mean, linear, and equipercentile equating are shown in Table 2.7 and are graphed in Fig. 2.9. The linear and equipercentile results were calculated using the RAGE-RGEQUATE computer program described in Appendix B, and are also described in Brennan et al. (2009, pp. 57–64). The moments for converted scores are shown in the bottom portion of Table 2.6. As expected, the mean converted scores for mean equating are the same as the mean for Form Y. For linear equating, the mean and standard deviation of the converted scores agree with the mean and standard deviation of Form Y. The first four moments of converted scores for equipercentile equating are very similar to those for Form Y. In Table 2.7, it can be seen that mean and linear equating produce results that are outside the range of possible raw scores. Because of the large number of values in Table 2.7 and the considerable similarity of equating functions in Fig. 2.9, differences between the functions are difficult to ascertain.

Table 2.5 Data for equating Form X and Form Y of the original ACT mathematics test

	Form Y					Form X				
Raw score	$N_Y\cdot\hat{g}(y)$	$N_Y\cdot\hat{G}(y)$	$\hat{g}(y)$	$\hat{G}(y)$	$\hat{Q}(y)$	$N_X\cdot\hat{f}(x)$	$N_X\cdot\hat{F}(x)$	$\hat{f}(x)$	$\hat{F}(x)$	$\hat{P}(x)$
0	0	0	.0000	.0000	.00	0	0	.0000	.0000	.00
1	1	1	.0002	.0002	.01	1	1	.0002	.0002	.01
2	3	4	.0007	.0010	.06	1	2	.0002	.0005	.03
3	13	17	.0031	.0041	.25	3	5	.0007	.0012	.08
4	42	59	.0101	.0142	.92	9	14	.0021	.0032	.22
5	59	118	.0142	.0284	2.13	18	32	.0042	.0074	.53
6	95	213	.0229	.0513	3.99	59	91	.0136	.0210	1.42
7	131	344	.0316	.0829	6.71	67	158	.0155	.0365	2.88
8	158	502	.0381	.1209	10.19	91	249	.0210	.0575	4.70
9	161	663	.0388	.1597	14.03	144	393	.0333	.0908	7.42
10	194	857	.0467	.2064	18.30	149	542	.0344	.1252	10.80
11	164	1021	.0395	.2459	22.62	192	734	.0444	.1696	14.74
12	166	1187	.0400	.2859	26.59	192	926	.0444	.2139	19.17
13	197	384	.0474	.3333	30.96	192	1118	.0444	.2583	23.61
14	177	561	.0426	.3760	35.46	201	1319	.0464	.3047	28.15
15	158	1719	.0381	.4140	39.50	204	1523	.0471	.3518	32.83
16	169	1888	.0407	.4547	43.44	217	1740	.0501	.4019	37.69
17	132	2020	.0318	.4865	47.06	181	1921	.0418	.4438	42.28
18	158	2178	.0381	.5246	50.55	184	2105	.0425	.4863	46.50
19	151	2329	.0364	.5609	54.28	170	2275	.0393	.5255	50.59
20	134	2463	.0323	.5932	57.71	201	2476	.0464	.5720	54.87
21	137	2600	.0330	.6262	60.97	147	2623	.0340	.6059	58.89
22	122	2722	.0294	.6556	64.09	163	2786	.0377	.6436	62.47
23	110	2832	.0265	.6821	66.88	147	2933	.0340	.6775	66.05
24	116	2948	.0279	.7100	69.61	140	3073	.0323	.7099	69.37
25	132	3080	.0318	.7418	72.59	147	3220	.0340	.7438	72.68
26	104	3184	.0250	.7669	75.43	126	3346	.0291	.7729	75.84
27	104	3288	.0250	.7919	77.94	113	3459	.0261	.7990	78.60
28	114	3402	.0275	.8194	80.56	100	3559	.0231	.8221	81.06
29	97	3499	.0234	.8427	83.10	106	3665	.0245	.8466	83.44
30	107	3606	.0258	.8685	85.56	107	3772	.0247	.8713	85.90
31	88	3694	.0212	.8897	87.91	91	3863	.0210	.8924	88.18
32	80	3774	.0193	.9090	89.93	83	3946	.0192	.9115	90.19
33	79	3853	.0190	.9280	91.85	73	4019	.0169	.9284	92.00
34	70	3923	.0169	.9448	93.64	72	4091	.0166	.9450	93.67
35	61	3984	.0147	.9595	95.22	75	4166	.0173	.9623	95.37
36	48	4032	.0116	.9711	96.53	50	4216	.0116	.9739	96.81
37	47	4079	.0113	.9824	97.68	37	4253	.0085	.9824	97.82
38	29	4108	.0070	.9894	98.59	38	4291	.0088	.9912	98.68
39	32	4140	.0077	.9971	99.33	23	4314	.0053	.9965	99.39
40	12	4152	.0029	1.0000	99.86	15	4329	.0035	1.000	99.83

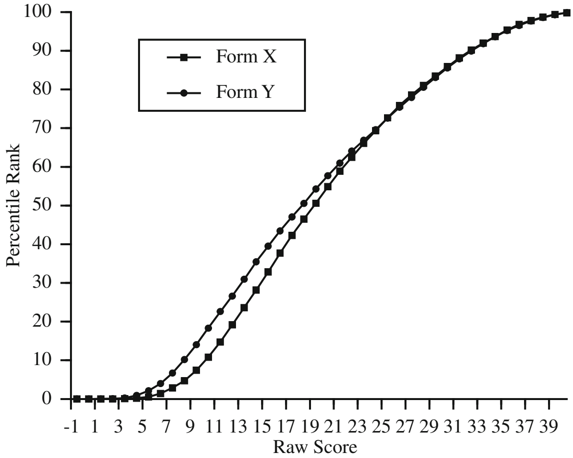


Fig. 2.7 Percentile ranks for equating Form X and Form Y of the original ACT Mathematics test

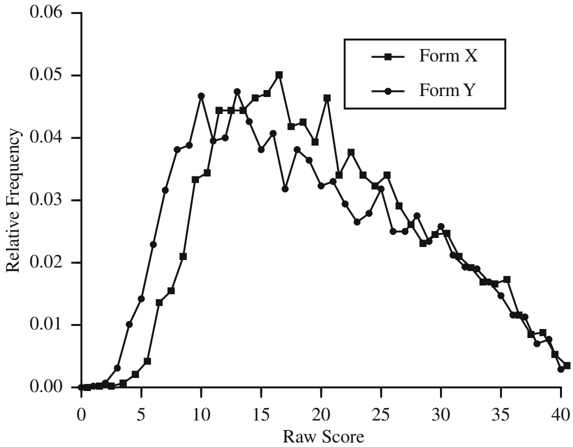


Fig. 2.8 Relative frequency distributions for Form X and Form Y of the original ACT Mathematics test

The use of considerably larger graph paper would help in such a comparison. Alternatively, difference-type plots can be used, as in Fig. 2.10. In this graph, the difference between the results for each method and the results for the identity equating are plotted. To find the Form Y equivalent of a Form X score, just add the vertical axis value to the horizontal axis value. For example, for equipercentile equating a Form X score of 10 has a vertical axis value of approximately -1.8. Thus, the Form Y equivalent of a Form X score of 10 is approximately $8.2 = 10 - 1.8$. This value is the same as the one indicated in Table 2.7 (8.1607), apart from error inherent in trying to read values from a graph.

Table 2.6 Moments for equating Form X and Form Y

Test Form	$\hat{\mu}$	$\hat{\sigma}$	$\widehat{sk}$	$\widehat{ku}$
Form Y	18.9798	8.9393	.3527	2.1464
Form X	19.8524	8.2116	.3753	2.3024
Form X equated to Form Y scale for various methods				
Mean	18.9798	8.2116	.3753	2.3024
Linear	18.9798	8.9393	.3753	2.3024
Equipercentile	18.9799	8.9352	.3545	2.1465

In Fig. 2.10, the horizontal line for the identity equating is at a vertical axis value of 0, which will always be the case with difference plots constructed in the manner of Fig. 2.10. The results for mean equating are displayed by a line that is parallel to, but nearly 1 point below, the line for the identity equating. The line for linear equating crosses the identity equating and mean equating lines. The equipercentile equating relationship appears to be definitely nonlinear. Referring to the equipercentile relationship, Form X appears to be nearly 2 points easier around a Form X score of 10, and the two forms appear to be similar in difficulty at scores in the range of 25 to 40.

The plot in Fig. 2.10 for equipercentile equating is somewhat irregular (bumpy). These irregularities are a result of random error in estimating the equivalents. Smoothing methods are introduced in Chap. 3, which lead to more regular plots and less random error.

2.7 Scale Scores

When equating is conducted in practice, raw scores typically are converted to scale scores. As described in Chap. 9, scale scores are constructed to facilitate score interpretation, often by incorporating normative or content information. For example, scale scores might be constructed to have a particular mean in a nationally representative group of examinees. The effects of equating on scale scores are crucial to the interpretation of equating results, because scale scores are the scores typically reported to examinees. A further discussion of methods for developing score scales is provided in Chap. 9. The use of scale scores in the equating context is described next.

2.7.1 Linear Conversions

The least complicated raw-to-scale score transformations that typically are used in practice are linear in form. For example, suppose that a national norming study was

Table 2.7 Raw-to-raw score conversion tables

Form X	Form Y equivalent using equating method		
Score	Mean	Linear	Equipercentile
0	−.8726	−2.6319	.0000
1	.1274	−1.5432	.9796
2	1.1274	−.4546	1.6462
3	2.1274	.6340	2.2856
4	3.1274	1.7226	2.8932
5	4.1274	2.8112	3.6205
6	5.1274	3.8998	4.4997
7	6.1274	4.9884	5.5148
8	7.1274	6.0771	6.3124
9	8.1274	7.1657	7.2242
10	9.1274	8.2543	8.1607
11	10.1274	9.3429	9.1827
12	11.1274	10.4315	10.1859
13	12.1274	11.5201	11.2513
14	13.1274	12.6088	12.3896
15	14.1274	13.6974	13.3929
16	15.1274	14.7860	14.5240
17	16.1274	15.8746	15.7169
18	17.1274	16.9632	16.8234
19	18.1274	18.0518	18.0092
20	19.1274	19.1405	19.1647
21	20.1274	20.2291	20.3676
22	21.1274	21.3177	21.4556
23	22.1274	22.4063	22.6871
24	23.1274	23.4949	23.9157
25	24.1274	24.5835	25.0292
26	25.1274	25.6722	26.1612
27	26.1274	26.7608	27.2633
28	27.1274	27.8494	28.1801
29	28.1274	28.9380	29.1424
30	29.1274	30.0266	30.1305
31	30.1274	31.1152	31.1297
32	31.1274	32.2039	32.1357
33	32.1274	33.2925	33.0781
34	33.1274	34.3811	34.0172
35	34.1274	35.4697	35.1016
36	35.1274	36.5583	36.2426
37	36.1274	37.6469	37.1248
38	37.1274	38.7355	38.1321
39	38.1274	39.8242	39.0807
40	39.1274	40.9128	39.9006

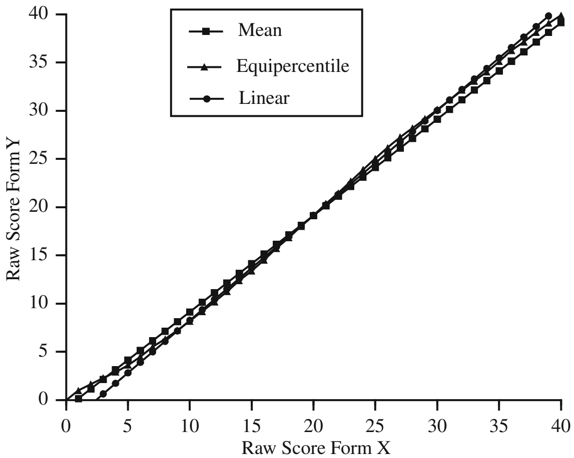


Fig. 2.9 Results for equating Form X and Form Y of the original ACT Mathematics test

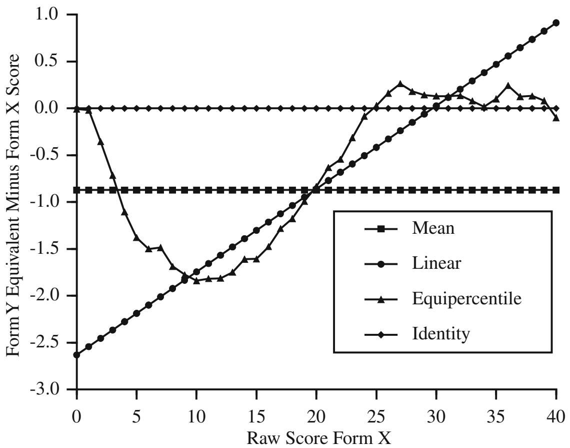


Fig. 2.10 Results expressed as differences for equating Form X and Form Y of the original ACT Mathematics test

conducted using Form Y of the 100-item test that was used earlier in this chapter to illustrate mean and linear equating. Assume that the mean raw score, $\mu(Y)$, was 70 and the standard deviation, $\sigma(Y)$, was 10 for the national norm group. Also assume that the mean scale score, $\mu(sc)$, was intended to be 20 and the standard deviation of the scale scores, $\sigma(sc)$, 5. Then the raw-to-scale score transformation (sc) for converting raw scores on the old form, Form Y, to scale scores is

$$sc(y) = \frac{\sigma(sc)}{\sigma(Y)} y + \left[\mu(sc) - \frac{\sigma(sc)}{\sigma(Y)} \mu(Y) \right]. \tag{2.21}$$

Substituting we have

$$\begin{aligned} sc(y) &= \frac{5}{10}y + \left[20 - \frac{5}{10}70\right] \\ &= .5y - 15. \end{aligned}$$

Now assume that scores on Form X are to be converted to scale scores based on the equating used in the earlier linear equating example. As was found earlier, the linear conversion equation for equating raw scores on Form X to raw scores on Form Y was $l_Y(x) = .9x + 12.2$. To find the raw-to-scale score transformation for Form X, substitute $l_Y(x)$ for y in the raw-to-scale score transformation for Form Y. This gives

$$\begin{aligned} sc[l_Y(x)] &= .5[l_Y(x)] - 15 \\ &= .5[.9x + 12.2] - 15 \\ &= .45x - 8.9. \end{aligned}$$

For example, a raw score of 74 on Form X converts to a scale score of $.45(74) - 8.9 = 24.4$. In this manner, raw-to-scale score conversions for all Form X raw scores can be found. When another new form is constructed and equated to Form X, a similar process can be used to find the scale score equivalents of scores on this new form.

2.7.2 Truncation of Linear Conversions

When linear transformations are used as scaling transformations, the score scale transformation often needs to be truncated at the upper and/or lower extremes. For example, the Form Y raw-to-scale score transformation, $sc(y) = .5y - 15$, produces scale scores below 1 for raw scores below 32. Suppose that scale scores are intended to be 1 or greater. The transformation for this form then would be as follows:

$$\begin{aligned} sc(y) &= .5y - 15, \quad & y \geq 32, \\ &= 1, & y < 32. \end{aligned}$$

Also, a raw score of 22 on Form X is equivalent to a raw score of $32 = .9(22) + 12.2$ on Form Y. So, the raw-to-scale score conversion for Form X is

$$\begin{aligned} sc[l_Y(x)] &= .45x - 8.9, \quad & x \geq 22, \\ &= 1, & x < 22. \end{aligned}$$

Truncation can also occur at the top end. For example, truncation would be needed at the top end for Form X but not for Form Y if the highest scale score was set to 35 on this 100-item test (the reader should verify this fact).

Scale scores are typically rounded to integers for reporting purposes. Define sc_{int}, as the scale score rounded to an integer. Then, for example, $sc_{int}[l_Y(x = 74)] = 24$, because a scale score of 24.4 rounds to a scale score of 24.

2.7.3 Nonlinear Conversions

Nonlinear raw-to-scale score transformations are often used in practice. Examples of nonlinear transformations include the following: normalized scales, grade equivalents, and scales constructed to stabilize measurement error variability (see Chap. 9). The use of nonlinear transformations complicates the process of converting raw scores to scale scores. The nonlinear function could be specified as a continuous function. However, when using discrete test scores (e.g., number-correct scores) the function is often defined at selected raw score values, and linear interpolation is used to compute scale score equivalents at other raw score values. The scheme for nonlinear raw-to-scale score transformations that is presented here is designed to be consistent with the definitions of equipercentile equating described earlier.

The first step in describing the process is to specify $sc(y)$, the raw-to-scale score function for Form Y. In the present approach, the conversions of Form Y raw scores to scale scores are specified at Form Y raw scores of $-.5$, $K_Y + .5$, and all integer score points through and including 0 to K_Y. The first two columns of Table 2.8 present an example. As can be seen, each integer raw score on Form Y has a scale score equivalent. For example, the scale score equivalent of a Form Y raw score of 24 is 22.3220. These equivalents resulted from an earlier equating of Form Y.

When Form X is equated to Form Y, the Form Y equivalents are typically noninteger. These noninteger equivalents need to be converted to scale scores, so a procedure is needed to find scale score equivalents of noninteger scores on Form Y. Linear interpolation is used in the present approach. For example, to find the scale score equivalent of a Form Y score of 24.5 in Table 2.8, find the scale score that is halfway between the scale score equivalents of Form Y raw scores of 24 (22.3220) and 25 (22.9178). The reader should verify that this value is 22.6199.

Note that scale score equivalents are provided in the table for raw scores of $-.5$ and 40.5. These values provide minimum and maximum scale scores when equipercentile equating is used. (As was indicated earlier, the minimum equated raw score in equipercentile equating is $-.5$ and the maximum is $K_Y + .5$.)

To make the specification of conversion for Form Y to scale scores more explicit, let y_i refer to the i-th point that is tabled. For $-.5 \le y \le K_Y + .5$, define y_i^* as the tabled raw score that is the *largest* among the tabled scores that are *less than or equal* to y. In this case, the linearly interpolated raw-to-scale score transformation is defined as

Table 2.8 Raw-to-scale score conversion tables

Raw Score	Form Y scale Scores sc	Form Y sc_{int}	Form X Mean equating sc	Mean equating sc_{int}	Linear equating sc	Linear equating sc_{int}	Equipercentile sc	Equipercentile sc_{int}
−.5	.5000	1	.5000	1	.5000	1	.5000	1
0	.5000	1	.5000	1	.5000	1	.5000	1
1	.5000	1	.5000	1	.5000	1	.5000	1
2	.5000	1	.5000	1	.5000	1	.5000	1
3	.5000	1	.5000	1	.5000	1	.5000	1
4	.5000	1	.5000	1	.5000	1	.5000	1
5	.6900	1	.5242	1	.5000	1	.5000	1
6	1.6562	2	.8131	1	.5000	1	.5949	1
7	3.1082	3	1.8412	2	.6878	1	1.1874	1
8	4.6971	5	3.3106	3	1.7681	2	2.1098	2
9	6.1207	6	4.8784	5	3.3715	3	3.4645	3
10	7.4732	7	6.2930	6	5.0591	5	4.9258	5
11	8.9007	9	7.6550	8	6.5845	7	6.3678	6
12	10.3392	10	9.0839	9	8.0892	8	7.7386	8
13	11.6388	12	10.5047	11	9.6489	10	9.2622	9
14	12.8254	13	11.7899	12	11.1303	11	10.8456	11
15	14.0157	14	12.9770	13	12.4663	12	12.1050	12
16	15.2127	15	14.1682	14	13.7610	14	13.4491	13
17	16.3528	16	15.3579	15	15.0626	15	14.8738	15
18	17.3824	17	16.4839	16	16.3109	16	16.1515	16
19	18.3403	18	17.5044	18	17.4321	17	17.3912	17
20	19.2844	19	18.4606	18	18.4729	18	18.4958	18
21	20.1839	20	19.3990	19	19.4905	19	19.6151	20
22	20.9947	21	20.2872	20	20.4415	20	20.5533	21
23	21.7000	22	21.0845	21	21.2813	21	21.4793	21
24	22.3220	22	21.7792	22	22.0078	22	22.2695	22
25	22.9178	23	22.3979	22	22.6697	23	22.9353	23
26	23.5183	24	22.9943	23	23.3214	23	23.6171	24
27	24.1314	24	23.5964	24	23.9847	24	24.2949	24
28	24.7525	25	24.2105	24	24.6590	25	24.8496	25
29	25.2915	25	24.8212	25	25.2581	25	25.3538	25
30	25.7287	26	25.3472	25	25.7400	26	25.7841	26
31	26.1534	26	25.7828	26	26.2104	26	26.2176	26
32	26.6480	27	26.2164	26	26.7684	27	26.7281	27
33	27.2385	27	26.7232	27	27.4343	27	27,2908	27
34	27.9081	28	27.3238	27	28.2070	28	27.9216	28
35	28.6925	29	28.0080	28	29.1886	29	28.7998	29
36	29.7486	30	28.8270	29	30.5595	31	30.1009	30
37	31.2010	31	29.9336	30	32.1652	32	31.3869	31
38	32.6914	33	31.3908	31	33.7975	34	32.8900	33
39	34.1952	34	32.8830	33	35.2388	35	34.2974	34
40	35.4615	35	34.3565	34	36.5000	36	35.3356	35
40.5	36.5000	36	34.9897	35	36.5000	36	36.5000	36

$$\begin{aligned} sc(y) &= sc(y_i^*) + \frac{y - y_i^*}{y_{i+1}^* - y_i^*}[sc(y_{i+1}^*) - sc(y_i^*)], & -.5 \le y \le K_Y + .5, \\ &= sc(-.5), & y < -.5, \\ &= sc(K_Y + .5), & y > Ky + .5, \end{aligned} \tag{2.22}$$

where y_{i+1}^* is the *smallest* tabled raw score that is *greater than or equal to* y_i^*. Note that $sc(-.5)$ is the minimum scale score and that $sc(K_Y + .5)$ is the maximum scale score.

To illustrate how this equation works, refer again to Table 2.8. How would the scale score equivalent of a raw score of $y = 18.3$ be found using Eq. (2.22)? Note that $y_i^* = 18$, because this score is the *largest* tabled score that is *less than or equal to* y. Using Eq. (2.22),

$$\begin{aligned} sc(y) &= sc(18) + \frac{18.3 - 18}{19 - 18}[sc(19) - sc(18)] \\ &= 17.3824 + \frac{18.3 - 18}{19 - 18}[18.3403 - 17.3824] \\ &= 17.6698. \end{aligned}$$

To illustrate that Eq. (2.22) is a linear interpolation expression, note that the scale score equivalent of 18 is 17.3824. The scale score 18.3 is, proportionally, .3 of the way between 18 and 19. This .3 value is multiplied by the difference between the scale score equivalents at 19 (18.3403) and at 18 (17.3824). Then .3 times this difference is $.3[18.3403 - 17.3824] = .2874$. Adding .2874 to 17.3824 gives 17.6698.

Typically, the tabled scores used to apply Eq. (2.22) will be integer raw scores along with $-.5$ and $K_Y + .5$. Equation (2.22), however, allows for more general schemes. For example, scale score equivalents could be tabled at each half raw score, such as $-.5$, .0, .5, 1.0, etc.

In practice, integer scores, which are found by rounding $sc(y)$, are reported to examinees. The third column of the table provides these integer scale score equivalents for integer raw scores (sc_{int}). A raw score of $-.5$ was set equal to a scale score value of .5 and a raw score of 40.5 was set equal to a scale score value of 36.5. These values were chosen so that the minimum possible rounded scale score would be 1 and the maximum 36. In rounding, a convention is used where a scale score that precisely equals an integer score plus .5 rounds up to the next integer score. The exception to this convention is that the scale score is rounded down for the highest scale score, so that 36.5 rounds to 36.

To find the scale score equivalents of the Form X raw scores, the raw scores on Form X are first equated to raw scores on Form Y using Eq. (2.18). Then, substituting $e_Y(x)$ for y in Eq. (2.22),

$$sc[e_Y(x)] = sc(y_i^*) + \frac{e_Y(x) - y_i^*}{y_{i+1}^* - y_i^*}[sc(y_{i+1}^*) - sc(y_i^*)], -.5 \le e_Y(x) \le K_X + .5. \tag{2.23}$$

In this equation, y_i^* is defined as the *largest* tabled raw score that is *less than or equal to* $e_Y(x)$. This definition of y_i^* as well as the definition of y_{i+1}^* are consistent with their definitions in Eq. (2.22). The transformation is defined only for the range of Form X scores, $-.5 \leq x \leq K_X + .5$. There is no need to define this function outside this range, because $e_Y(x)$ is defined only in this range in Eq. (2.17). The minimum and maximum scale scores for Form X are identical to those for Form Y, which occur at $sc[e_Y(x = -.5)]$ and at $sc[e_Y(x = K_X + .5)]$, respectively.

As an example, Eq. (2.23) is used with the ACT Mathematics equating example. Suppose that the scale score equivalent of a Form X raw score of 24 is to be found using equipercentile equating. In Table 2.7, a Form X raw score of 24 is shown to be equivalent to a Form Y raw score of 23.9157. To apply Eq. (2.22), note that the largest Form Y raw score in Table 2.8 that is less than 23.9157 is 23. So, $y_i^* = 23$, and $y_{i+1}^* = 24$. From Table 2.8, $sc(23) = 21.7000$ and $sc(24) = 22.3220$. Applying Eq. (2.22),

$$\begin{aligned}
sc[e_Y(x = 24)] &= sc(23.9157) \\
&= sc(23) + \frac{23.9157 - 23}{24 - 23}[sc(24) - sc(23)] \\
&= 21.7000 + \frac{23.9157 - 23}{24 - 23}[22.3220 - 21.7000] \\
&= 22.2696.
\end{aligned}$$

For a Form X raw score of 24, this value agrees with the value using equipercentile equating in Table 2.8, apart from rounding. Rounding to an integer, $sc_{int}[e_Y(x = 24)] = 22$.

Mean and linear raw score equating results can be converted to nonlinear scale scores by substituting $m_Y(x)$ or $l_Y(x)$ for y in Eq. (2.22). The raw score equivalents from either the mean or linear methods might fall outside the range of possible Form Y scores. This problem is handled in Eq. (2.22) by truncating the scale scores. For example, if $l_Y(x) < -.5$, then $sc(y) = sc(-.5)$ by Eq. (2.22). The unrounded and rounded raw-to-scale score conversions for the mean and linear equating results are presented in Table 2.8.

Inspecting the central moments of scale scores can be useful in judging the accuracy of equating. Ideally, after equating, the scale score moments for converted Form X scores would be identical to those for Form Y. However, the moments typically are not identical, in part because the scores are discrete. If equating is successful, then the scale score moments for converted Form X scores should be very similar (say, agree, to at least one decimal place) to the scale score moments for Form Y. Should the Form X moments be compared to the rounded or unrounded Form Y moments? The answer is not entirely clear. However, the approach taken here is to compare the Form X moments to the Form Y unrounded moments. The rationale for this approach is that the unrounded transformation for Form Y most closely defines the score scale for the test, whereas rounding is used primarily to facilitate score interpretability. Following this logic, the use of Form Y unrounded moments for

Table 2.9 Scale score moments

Test Form	$\hat{\mu}_{sc}$	$\hat{\sigma}_{sc}$	$\widehat{sk}_{sc}$	$\widehat{ku}_{sc}$
Form Y				
unrounded	*16.5120*	*8.3812*	*−.1344*	*2.0557*
rounded	16.4875	8.3750	−.1025	2.0229
Form X equated to Form Y scale for various methods				
Mean				
unrounded	16.7319	7.6474	−.1868	2.1952
rounded	16.6925	7.5965	−.1678	2.2032
Linear				
unrounded	16.5875	8.3688	−.1168	2.1979
rounded	16.5082	8.3065	−.0776	2.1949
Equipercentile				
unrounded	16.5125	8.3725	−.1300	2.0515
rounded	16.4324	8.3973	−.1212	2.0294

comparison purposes should lead to greater score scale stability when, over time, many forms become involved in the equating process.

Moments are shown in Table 2.9 for Form Y and for Form X using mean, linear, and equipercentile equating. Moments are shown for the unrounded (sc) and rounded (sc_{int}) score transformations. Note that the process of rounding affects the moments for Form Y. Also, the Form X scale score mean for mean equating (both rounded and unrounded) is much larger than the unrounded scale score mean for Form Y. Presumably, the use of a nonlinear raw-to-scale score transformation for Form Y is responsible. When the raw-to-scale score conversion for Form Y is nonlinear, the mean scale score for Form X is typically not equal to the mean scale score for Form Y for mean and linear equating. Similarly, when the raw-to-scale score conversion for Form Y is nonlinear, the standard deviation of the Form X scale scores typically is not equal to the standard deviation of Form Y scale scores for linear equating.

For equipercentile equating, the unrounded moments for Form X are similar to the unrounded moments for Form Y. The rounding process results in the mean of Form X being somewhat low. Is there anything that can be done to raise the mean of the rounded scores? Refer to Table 2.8. In this table, a raw score of 23 converts to an unrounded scale score of 21.4793 and a rounded scale score of 21. If the unrounded converted score had been only .0207 points higher, then the rounded converted score would have been 22. This observation suggests that the rounded conversion might be adjusted to make the moments more similar. Consider adjusting the conversion so that a raw score of 23 converts to a scale score of 22 (instead of 21) and a raw score of 16 converts to a scale score of 14 (instead of 13). The moments for the adjusted conversion are as follows: $\hat{\mu}_{sc} = 16.5165$, $\hat{\sigma}_{sc} = 8.3998$, $\widehat{sk}_{sc} = -.1445$, and $\widehat{ku}_{sc} = 2.0347$. Overall, the moments of the adjusted conversion seem closer to the moments of the original unrounded conversion. For this reason, the adjusted conversion might be used in practice.

Should the rounded conversions actually be adjusted in practice? To the extent that moments for the Form X rounded scale scores are made more similar to the unrounded scale score moments for Form Y, adjusting the conversions would seem advantageous. However, adjusting the conversions might lead to greater differences between the cumulative distributions of scale scores for Form X and Form Y at some scale score points. That is, adjusted conversions lead to less similar percentile ranks of scale scores across the two forms. In addition, adjusted conversions affect the scores of individual examinees.

Because adjusting can lead to less similar scale score distributions, and because it adds a subjective element into the equating process, we typically take a conservative approach to adjusting conversions. A rule of thumb that we often follow is to consider adjusting the conversions only if the moments are closer after adjusting than before adjusting, and the unrounded conversion is within .1 point of rounding to the next higher or lower value (e.g., 21.4793 in the example is within .1 point of rounding to 22). Smoothing methods are considered in Chap. 3, which might eliminate the need to consider subjective adjustments.

In the examples, scale score equivalents of integer raw scores were specified and linear interpolation was used between the integer scores. If more precision is desired, scale score equivalents of fractional raw scores could be specified. The procedures associated with Eqs. (2.22) and (2.23) are expressed in sufficient generality to handle this additional precision. Procedures using nonlinear interpolation also could be developed, although linear interpolation is likely sufficient for practical purposes.

When score scales are established, the highest and lowest possible scale scores are often fixed at particular values. For example, the ACT score scale is said to range from 1 to 36. The approach taken here to scaling when using nonlinear conversions is to fix the ends of the score scale at specific points. Over time, if forms become easier or more difficult, the end points could be adjusted. However, such adjustments would require careful judgment. An alternative procedure involves leaving enough room at the top and bottom of the score scale to handle these problems. For example, suppose that the rounded score scale for an original form is to have a high score of 36 for the first form developed. However, there is a desire to allow scale scores on subsequent forms to go as high as 40 if the forms become more difficult. For the initial Form Y, a scale score of 36 could be assigned to a raw score equal to K_Y and a scale score of 40.5 could be assigned to a raw score equal to $K_Y + .5$. If subsequent forms are more difficult than Form Y, the procedures described here could lead to scale scores as high as 40.5. Of course, alternate interpolation rules could lead to different properties. Rules for nonlinear scaling and equating also might be developed that would allow the highest and lowest scores to float without limit. The approach taken here is to provide a set of equations to be used for nonlinear equating and scaling that can adequately handle, in a consistent manner, many of the situations we have encountered in practice.

One practical problem sometimes occurs when the highest possible raw score does not equate to the highest possible scale score. For the ACT, for example, the highest possible raw score is assigned a scale score value of 36, regardless of the results of the equating. For the SAT (Donlon 1984, p. 19), the highest possible raw score is assigned a scale score value of 800, and other converted scores are sometimes adjusted, as well.

2.8 Equating Using Single Group Designs

If practice, fatigue, and other order effects do not have an effect on scores, then the statistical process for mean, linear, and equipercentile equating using the single group design (without counterbalancing) is essentially the same as with the random groups design. However, order typically has an affect, and for this reason the single group design (without counterbalancing) is not recommended.

When the single group design with counterbalancing is used, the following four equatings can be conducted:

1. Equate Form X and Form Y using the random groups design for examinees who were administered Form X first and Form Y first.
2. Equate Form X and Form Y using the random groups design for examinees who were administered Form X second and Form Y second.
3. Equate Form X and Form Y using the single group design for examinees who were administered Form X first and Form Y second.
4. Equate Form X and Form Y using the single group design for examinees who were administered Form X second and Form Y first.

Compare equatings 1 and 2. Standard errors of equating described in Chap. 7 can be used as a baseline for comparing the equatings. If the equatings give different results, apart from sampling error, then Forms X and Y are differentially affected by appearing second. In this case, only the first equating should be used. Note that the first equating is a random groups equating, so it is unaffected by order. The problem with using the first equating only is that the sample size might be quite small. However, when differential order effects occur, then equating 1 might be the only equating that would not be biased.

If equatings 1 and 2 give the same results, apart from sampling error, then Forms X and Y are similarly affected by appearing second. In this case, all of the data can be used. One possibility would be to pool all of the Form X data and all of the Form Y data, and equate the pooled distributions. Angoff (1971) and Petersen et al. (1989) provided procedures for linear equating. von Davier et al. (2004) described a systematic scheme that is based on statistical tests using log-linear models for equipercentile equating under the single group counterbalanced design.

2.9 Equating Using Alternate Scoring Schemes

The presentation of equipercentile equating and scale scores assumed that the tests to be equated are scored number-correct, with the observed scores ranging from 0 to the number of items. Although this type of scoring scheme is the one that is used most often with educational tests, alternative scoring procedures are becoming much more popular, and the procedures described previously can be generalized to other scoring schemes. For example, whenever raw scores are integer scores that range from 0 to a positive integer value, the procedures can be used directly by defining K as the maximum score on a form, rather than as the number of items on a form as has been done.

Some scoring schemes might produce discrete scores that are not necessarily integers. For example, when tests are scored using a correction for guessing, a fractional score point often is subtracted from the total score whenever an item is answered incorrectly. In this case, raw scores are not integers. However, the discrete score points that can possibly occur are specifiable and equally spaced. One way to conduct equating in this situation is to transform the raw scores. The lowest possible raw score is transformed to a score of 0, the next lowest raw score is transformed to a score of 1, and so on through K, which is defined as the transformed value of the highest possible raw score. The procedures described in this chapter then can be applied and the scores transformed back to their original units.

Equipercentile equating also can be conducted when the scores are considered to be continuous, which might be the case when equating forms of a computerized adaptive test. In many ways, equating in this situation is more straightforward than with discrete scores, because the definitional problems associated with continuization do not need to be considered. Still, difficulties might arise in trying to define score equivalents in portions of the score scale where few examinees earn scores. In addition, even if the range of scores is potentially infinite, the range of scores for which equipercentile equivalents are to be found needs to be considered.

2.10 Preview of What Follows

In this chapter, we described many of the issues associated with observed score equating using the random groups design, including defining methods, describing their properties, and estimating the relationships. We also discussed the relationships between equating and score scales. One of the major relevant issues not addressed in this chapter is the use of smoothing methods to reduce random error in estimating equipercentile equivalents. Smoothing is the topic of Chap. 3. Also, as we show in Chaps. 4 and 5, the implementation of observed score equating methods becomes much more complicated when the groups administered the two forms are not randomly equivalent. Observed score methods associated with IRT are described in Chap. 6. Estimating random error in observed score equating is discussed in detail in Chap. 7, and practical issues are discussed in Chap. 8. Scaling and linking are discussed in Chaps. 9 and 10.

Table 2.10 Score distributions for exercise 2.4

x	$f(x)$	$F(x)$	$P(x)$	y	$g(y)$	$G(y)$	$Q(y)$
0	.00			0	.00		
1	.01			1	.02		
2	.02			2	.05		
3	.03			3	.10		
4	.04			4	.20		
5	.10			5	.25		
6	.20			6	.20		
7	.25			7	.10		
8	.20			8	.05		
9	.10			9	.02		
10	.05			10	.01		

Table 2.11 Equated scores for exercise 2.4

x	$m_Y(x)$	$l_Y(x)$	$e_Y(x)$
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

2.11 Exercises

2.1. From Table 2.2 find $P(2.7)$, $P(.2)$, $P^{-1}(25)$, $P^{-1}(97)$.

2.2. From Table 2.2, find the linear and mean conversion equation for converting scores on Form X to the Form Y scale.

2.3. Find the mean and standard deviation of the Form X scores converted to the Form Y scale for the equipercentile equivalents shown in Table 2.3.

2.4. Fill in Tables 2.10 and 2.11.

2.5. If the standard deviations on Form X and Y are equal, which methods, if any, among mean, linear, and equipercentile will produce the same results? Why?

2.6. Suppose that a raw score of 20 on Form W was found to be equivalent to a raw score of 23.15 on Form X. What would be the scale score equivalent of a Form W raw score of 20 using the Form X equipercentile conversion shown in Table 2.8?

2.7. Suppose that the linear raw-to-scale score conversion equation for Form Y was $sc(y) = 1.1y + 10$. Also suppose that the linear equating of Form X to Form Y was $l_Y(x) = .8x + 1.2$. What is the linear conversion of Form X scores to scale scores?

2.8. In general, how would the shape of the distribution of Form X raw scores equated to the Form Y raw scale compare to the shape of the original Form X raw score distribution using mean, linear, and equipercentile equating?

References

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

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

Brennan, R. L., Wang, T., Kim, S., & Seol, J. (2009). *Equating recipes*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, University of Iowa.

Donlon, T. (Ed.). (1984). *The College Board technical handbook for the scholastic aptitude test and achievement tests*. New York: College Entrance Examination Board.

Holland, P. W., & Thayer, D. T. (1989). *The kernel method of equating score distributions (Technical Report 89–84)*. Princeton, NJ: Educational Testing Service.

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

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

http://www.springer.com/978-1-4939-0316-0

Test Equating, Scaling, and Linking
Methods and Practices
Kolen, M.J.; Brennan, R.L.
2014, XXVI, 566 p. 68 illus., Hardcover
ISBN: 978-1-4939-0316-0