DOCUMENT RESUME

ED 126 134 TM 005 390

AUTHOR Brigman, S. Leellen; Bashaw, W. L.
TITLE Multiple Test Equating Using the Rasch Model.
PUB DATE [Apr 76]
NOTE 34p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Ability; Achievement Tests; Computer Programs; *Equated Scores; *Item Analysis; *Mathematical Models; Matrices; Measurement Techniques; Probability; Standard Error of Measurement; *Statistical Analysis
IDENTIFIERS *Rasch Model; *Test Equating

ABSTRACT

Procedures are presented for equating simultaneously several tests which have been calibrated by the Rasch Model. Three multiple test equating designs are described. A Full Matrix Design equates each test to all others. A Chain Design links tests sequentially. A Vector Design equates one test to each of the other tests. For each design, the Rasch model test equating constants were obtained for four reading vocabulary tests. The standard errors of the constants based on each design are also provided and the appropriate use of each design is discussed. (Author/DEP)

MULTIPLE TEST EQUATING USING THE RASCH MODEL

S. Leellen Brigman
Indiana University

W. L. Bashaw*
University of Georgia

American Educational Research Association
Annual Meeting
April 1976
San Francisco, California

Abstract

This paper presents procedures for equating simultaneously several tests which have been calibrated by the Rasch Model. Three multiple test equating designs are described. A Full Matrix Design equates each test to all others. A Chain Design links tests sequentially. A Vector Design equates one test to each of the other tests. For each design, the Rasch model test equating constants were obtained for four reading vocabulary tests. The standard errors of the constants based on each design are also provided and the appropriate use of each design is discussed.

MULTIPLE TEST EQUATING USING THE RASCH MODEL

S. Leellen Brigman
Indiana University

W. Louis Bashaw
University of Georgia

Angoff (1971) has stated that for test scores to be meaningful, the instruments of measurement must meet three requirements. First, an appropriate scale structure must be defined so that the scores may be communicated, i.e., the scaling process. The second requirement is that special norms or interpretive guides must be prepared for the user of the scores, i.e., the process of norming. The third requirement for a test score to be meaningful is that provisions be made for the maintenance and perpetuation of the scale on which the original test scores are reported, i.e., the process of equating or calibration.

A general definition of test equating is a psychometric process which converts the system of units of one test to the system of units of a second test such that the scores derived from the two tests after conversion will be directly equivalent. Two restrictions are implied by this definition: (1) the measures (tests or forms) must measure the same characteristic, and (2) the conversion must be unique, a transformation of the system

of units only, except for random errors associated with the unreliability of the data and the errors associated with the method used for determining the transformation. The second restriction indicates that the resulting conversion should be independent of the persons from whom the data were obtained to develop the conversion; and thus, the conversion should be freely applicable to all situations, i.e., sample-free test calibration.

Angoff (1971) has made one of the few definitive efforts to provide those in psychometrics with a discussion of equating versus calibration of tests, a discussion of the equipercentile and linear models for test equating and/or calibration, and sampling designs for the equating of two tests or the calibration of a test to a reference scale. Most practical applications of test equating have involved the equating of two tests or the calibration of one test to a reference scale. It is obvious why this is true when one reviews the complexities of the sampling design and the procedures associated with equipercentile and linear equating of multiple tests within one study as evidenced in the Educational Testing Service's Anchor Test Study: Final Report (1972).

A third model, the simple logistic model or Rasch model may also be used to equate or calibrate tests. The

Rasch model provides the researcher with a mathematical model that reduces the complexities of the sampling designs and equating or calibration procedures, especially when equating multiple tests in one study.

In 1969 Panchapakesan successfully applied the simple logistic model (Rasch model) to the problem of equating linked test forms and tests administered to matched samples. To equate scores on two tests, Panchapakesan estimated a constant which represented the difference in the origins of the scales of the two tests. This additive constant could be used to equate the scores on one test to the scores on the second test. The use of the Rasch model for equating multiple tests in one study was not attempted until 1973 in the federally funded Rasch Project (Rentz, Bashaw, Cartledge, and Brigman, 1975).

The purpose of the present study was to develop and illustrate the procedure for obtaining Rasch model test equating constants for three multiple test equating designs. The three designs included in this research differed in the number of independent samples and the number of combinations of tests used to determine the Rasch model test equating constants. Each of the designs requires a different manipulation of the data to obtain the set of Rasch model test equating constants.

Rasch's Structural Model for Items of a Test

Georg Rasch is a Danish mathematician who has been instrumental in the development and investigation of mathematical foundations for "objective measurement" especially in the domains of educational and psychological testing. Rasch (1966a) has stated that "specific objectivity" exists when:

> The comparison of any two subjects can be carried out in such a way that no other parameters are involved than those of the two subjects (p. 104) and when.... any two stimuli can be compared indendently of all other parameters than those of the two stimuli. (p. 105)

In Rasch's 1960 book, Probabilistic Models for Some Intelligence and Attainment Tests, he presented a detailed discussion of three "models for measuring". Rasch (1961) stated that:

> Each model specifies a distribution function for the potential respons s of a given person to a given stimulus of a certain set of allied stimuli, and this distribution function depends upon a parameter characterizing the person and a parameter characterizing the stimulus. (p. 321)

An important property of these models when analyzing data is the ability to detach the person parameters from the stimulus parameters and visa versa.

In the field of educational and psychological measurement, "A Structural Model for Items of a Test"

has become known as the "Rasch model". In the development of the Rasch model three assumptions were made. Rasch (1966b) has listed the assumptions as follows:

> (a) To each situation in which a subject (s=1,2...n) has to answer an item (i=1, 2...m) there is a corresponding probability of a correct answer (X_{si}=1) which we shall write in the form
>
> $$Pr\ (X_{si} = 1) = \frac{\lambda_{si}}{1 + \lambda_{si}},\ (\lambda_{si} \geq 0).$$
>
> (b) The situation parameter λ_{si} is the product of two factors
>
> $$\lambda_{si} = \pi_s\ \omega_i$$
>
> where π_s pertains to the subject and ω_i to the item.
>
> (c) Given the values of the parameters, all answers are stochastically independent. (p. 50)

Rentz, Bashaw, Cartledge and Brigman (1975) have defined three "antecedent conditions" which are necessary for model fit when analyzing data with the Rasch model. These conditions are implications of the assumptions of the model. The first condition is that the item pool to be analyzed must be unidimensional. The second antecedent condition is that of equal item discrimination; the rate of increase in the probability of passing an item as the ability increases must be equal for all items. The third antecedent condition is that guessing must be

absent or minimal in the item responses to reduce the probability of passing an item by chance.

In a 1967 presentation to the Invitational Conference on Test Problems, Benjamin Wright operationalized and demonstrated the Rasch model's claims of objectivity. The two basic outcomes, or consequent conditions, of the Rasch model with which Wright dealt in his presentation were: (1) the calibration of test items independent of the sample of subjects, and (2) the measurement of a person on the latent trait independent of the particular items used.

Wright and Panchapakesan (1969) have described estimation techniques for the Rasch model parameters, π_s and ω_i; where ω_i, the item parameters, are invariant over different samples, and π_s, the person ability parameters, are invariant over samples and item sets. Associated with their work are computer programs which can be used to perform Rasch analyses of test data; one program is commonly referred to as the MESAMAX program written by Wright and Skirmont (1972) and employed in the present study. Another is called CALFIT, written by Wright and Mead (1975).

Equating Tests with the Rasch Model

To employ the Rasch model for equating tests, two general conditions must be met: (1) the tests to be equated must be parallel, and (2) the tests must provide an acceptable fit with the model. Equated or equivalent scores when using the Rasch model can be defined as scores on two tests which give rise to the same estimate of ability.

When Rasch model analysis of an n-item test is performed, n Rasch item easiness estimates are obtained on a log easiness scale with a mean of zero. The MESAMAX program provides easiness estimates that are positive for the easier items and negative for the harder items. Also, for the n-item test, n - 1 Rasch ability estimates will be obtained for the raw scores on a scale of log ability (ability estimates are not obtained for a raw score of zero or a maximum raw score of n). The ability estimates are positive for the higher scores and negative for the lower scores.

The zero point on the log easiness scale is an arbitrary origin. The origin is fixed in the computer program by setting the mean item easiness to zero. This zero point simultaneously fixes the zero point on the log ability scale. Thus, the zero point is arbitrary in the sense that it is defined by the set of items that

are analyzed. Equating can be considered as adjusting these arbitrary origins for sets of tests to a common origin.

There are two methods of obtaining Rasch model test equating constants. The first method, the item difficulty method, uses the Rasch model item parameter estimates as the initial values in the procedures for obtaining the constants in a test equating study. The second method, the ability method, uses the Rasch model ability parameter estimates as the initial values in the procedures for obtaining the constants.

For the item difficulty method, the two sets of item data for a pair of tests are pooled and calibrated as one test of $n_1 + n_2$ items. Thus, the $n_1 + n_2$ items are calibrated on a single scale of log easiness with a mean of zero. Since as in the following examples two tests are administered to the same group of subjects, any difference in the average item easiness estimates of the two tests represents the difference in the scale origins of the two tests. This difference in the scale origins is an additive constant that may be used to equate the two ability scales associated with the separate tests, i.e., it is a Rasch model equating constant.

For the ability method of Rasch model equating, each test is analyzed independently. Since the two tests as in the subsequent examples are administered to the same group of subjects, the average of the ability estimates of the two tests will be equal if the scale origins of the two tests are the same. To obtain an estimate of the difference in the scale origins, an average Rasch model ability estimate is calculated for each test. The difference in the two averages represents the difference in the origins of the scales of the two tests. As in the item difficulty method, the difference value is an additive constant that may be used to equate the two ability scales associated with the individual tests, i.e., it is a Rasch model equating constant.

The Designs and Procedures

Multiple test equating is defined as the process of simultaneous equating of more than two tests or the process of simultaneous calibration of more than one test to a reference scale. Multiple test equating designs may be seen as extensions of simple test equating designs and are defined by the present authors as the schemata for the administration and data collection that are required to equate more than two tests or calibrate more than one test to a reference scale in a single study.

The present research focused on three designs that may be employed in a multiple test equating study. The purpose and the desired product of the equating study dictate the choice of the design. The designs reflect the test data that must be collected and the steps in the equating procedures to estimate the Rasch model equating constants.

For the purpose of describing the three multiple test equating designs and the associated procedures for obtaining the Rasch model equating constants, a multiple test equating matrix was used. The symbols that are used in the following discussion are defined in Table 1 in the Appendix.

For k tests, a multiple test equating matrix is a k X k matrix. The elements, or cells, of the matrix, T_{ij}'s, represent all possible test pair combinations that could be administered to independent groups of subjects. For a cell in the multiple test equating matrix, the row index corresponds to the test that would be administered first to the group of subjects and the column index corresponds to the test that would be administered second. The diagonal cells in the multiple test equating matrix would represent two administrations of one test to a single group of subjects. Data for the

diagonal cells may or may not be collected in a study. The cells, or test pair combinations, below the diagonal of the multiple test equating matrix represent the counterbalanced testing orders of the test pair combinations above the diagonal of the matrix. If the researcher selects not to collect data for the diagonal cells of the multiple test equating matrix, the matrix consists of $k^2 - k$ cells or test pair combinations as seen in Figure 1.

In the Anchor Test Study and the Rasch Project, one of the tests to be equated was selected as an anchor, or base, test. For convenience of describing the different multiple test equating designs and their associated procedures for obtaining the final Rasch model test equating constants, the base test was always assigned to the first row and the first column of the multiple test equating matrix.

The Full Design for Multiple Test Equating

The Full Design for multiple test equating is defined as an equating design in which all test pair combinations in the multiple test equating matrix are administered. The data obtained on all test pairs are used to estimate the final Rasch model test equating constants. An illustration of the Full Design is the

the multiple test equating matrix (See Figure 1). The Full Design with test-parallel form combinations on the diagonal was used in the ATS and Rasch Project.

To estimate the final Rasch model test equating constants for the Full Design, the researcher may use either the item difficulty method (item easiness estimates) or the ability method (person-ability estimates) to obtain the initial estimates of the difference in the scale origins of a test pair in a cell of the design. MESAMAX analyses are performed for each of the cells in the Full Design. For each test pair, the average of the test that was administered second in the test pair is subtracted from the average of the test that was administered first in the test pair. These differences in the averages are the Rasch model cell equating constants, denoted by c_{ij}, where i corresponds to the index of the test that was administered first in the test pair and j corresponds to the index of the test that was administered second in the test pair. The next step in the equating process is to organize the c_{ij}'s into their appropriate cells in the multiple test equating matrix with zeros inserted in the empty diagonal cells of the matrix.

To obtain a single Rasch model equating constant for each test in the Full Design, the sets of cell

equating constants are first combined to yield marginal equating means. These means are obtained by summing the c_{ij}'s in each row in the matrix and dividing by k to obtain the row marginal means, $c_{i.}$'s, and by summing the c_{ij}'s in each column in the matrix and dividing by k to obtain the column marginal means, $c_{.j}$'s. The two marginal means for a test correspond to the order of test administration with the row marginal mean reflecting the effects of the test when it was administered first in a test pair and the corresponding column marginal mean reflecting the effects of the test when it was administered second in a test pair.

The next step in the procedure is to combine the marginal means for a test in the Full Design. This is done by averaging the row marginal mean with its corresponding column marginal mean for each test. To do this, the signs of the column marginal means are reversed. The average of the row and column marginal mean, $(c_{i.} + c_{.j})/2$, are the preliminary Rasch model equating constants, $\bar{c}_i$'s, for the Full Design.

The last step in the process for obtaining Rasch model equating constants for the Full Design is to adjust the preliminary equating constants for the k tests to the base test. To do this, the preliminary equating

constant of the base test is subtracted from each of the k preliminary equating constants for the tests in the study. This will yield k Rasch model test equating constants, C_i's, that may be used to equate the ability estimates of the base test to the ability estimates that would have been obtained on the scale of ability that is associated with any other test in the study.

Four standardized reading vocabulary tests appropriate for fifth grade students were selected to illustrate the procedure associated with each design in this study. These four tests were taken from the following reading achievement batteries: (1) California Achievement Tests (1970)--Reading, Form A, Level 3 (CAT A3); (2) Iowa Test of Basic Skills (1970), Form S, Level 11 (ITBS 5, 11); (3) Metropolitan Reading Tests (1970), Form F, Intermediate Level (MAT FI); and (4) SRA Achievement Series (1971), Form E, Blue Edition (SRA EB). The CAT A3 was selected as the base test in all of the present illustrations. Using the data collected on these four tests in the Anchor Test Study, random samples of 500 subjects were drawn from each test pair cell in the multiple test equating matrix. Rasch analyses were performed for each test pair. Using the item difficulty method, an estimate of the mean item easiness was obtained for the set of items in each test.

Table 2 shows the cell equating constants and the intermediate values used to obtain the four final Rasch Model test equating constants. This illustrates the procedures for equating multiple tests in a Full Design. These equating constants are values that would be added to the ability estimate corresponding to a raw score that was obtained on the base test to determine the equivalent ability (equated score) on a second test.

Insert Table 2 about here

The Chain Design for Multiple Test Equating

The Chain Design for multiple test equating is defined as an equating design in which adjacent test pair combinations in the multiple test equating matrix are linked. The k tests in the multiple test equating study are numbered from 1 to k with the base test beginning the series. For the Chain Design, the cells in the multiple test equating matrix that are used in the equating study are T_{12}, T_{23}, T_{34},...$T_{k-1,k}$ and the cells of the counterbalanced test orders of these test pair combinations located below the diagonal in the multiple test equating matrix. For k tests, the Chain Design

requires that data for $k + k - 2$ test pair combinations be administered and used to estimate the Rasch model test equating constants. Figure 2 illustrates the Chain Design in the multiple test equating matrix for k tests.

Insert Figure 2 about here

To estimate the final Rasch model test equating constants for the Chain Design, the researcher may use either the item difficulty method or the ability method to obtain the initial estimates of the difference in the scale origins of a test pair in a cell in the design. MESAMAX analyses are performed for each of the $k + k - 2$ cells in the Chain Design. For each test pair, the average of the test that was administered second in the test pair is subtracted from the average of the test that was administered first in the test pair. These differences in the averages are Rasch model cell equating constants, denoted by c_{ij}, where i corresponds to the index of the test that was administered first in the test pair and j corresponds to the index of the test that was administered second in the test pair. The next step in the equating process is to organize the c_{ij}'s

into their appropriate cells in the multiple test equating matrix.

To obtain a single Rasch model equating constant for each test in the Chain Design, the two cell equating constants for a test pair combination, c_{ij} and c_{ji}, are combined to obtain a preliminary equating constant for each of the k tests in the Chain Design. To do this, the signs of the test pair cell equating constants for the cells above the diagonal of the multiple test equating matrix are reversed. The average of the two cell equating constants for a test pair in the Chain Design, $(c_{ij} + c_{ji})/2$, are the preliminary equating constants, $\bar{c}_i$'s, for the Chain Design.

The last step in the process for obtaining Rasch model equating constants for each of the tests in the Chain Design adjusted to the base test, consists of adding together all of the preliminary equating constants for the tests that link a particular test to the base test. Thus, the final Rasch model test equating constant for test i, C_i, is the sum of $\bar{c}_i$, $\bar{c}_{(i-1)}$, $\bar{c}_{(i-2)}, \ldots \bar{c}_1$. This will yield k Rasch model test equating constants, C_i's, that may be used to equate the ability estimates of the base test to the ability estimates that would have been obtained on the scale of ability that is associated with any other test in the study.

Table 3 presents the cell equating constants for the four tests in a Chain Design. Each of the constants may be added to the ability estimates of the base test to determine the equivalent abilities on the equated tests.

Insert Table 3 about here

The Vector Design for Multiple Test Equating

The Vector Design for multiple test equating is defined as an equating design in which all tests in the equating study are administered in combination with the base test. In the multiple test equating matrix, the base test appears in the test pair combinations of the first row and the first column of the matrix. For the Vector Design for k tests, the cells in the multiple test equating matrix that are used in the equating study are T_{12}, T_{13}, T_{14}, . . . T_{1k} and the cells of the counterbalanced test orders of these test pair combinations. For k tests, the Vector Design requires that data for k + k - 2 test pair combinations be used to estimate the Rasch model test equating constants. Figure 3 illustrates the Chain Design in the multiple test equating matrix for k tests.

Insert Figure 3 about here

To estimate the final Rasch model test equating constants for the Vector Design, the researcher may use either the item difficulty method or the ability method to obtain the initial estimates of the difference in the scale origins of a test pair in a cell in the design. MESAMAX analyses are performed for each of the $k + k - 2$ cells in the Vector Design. For each test pair, the average for the test that was administered second is subtracted from the average of the test that was administered first in the test pair. These differences in the averages are the Rasch model cell equating constants, denoted by c_{ij}'s, where i corresponds to the index of the test that was administered first in the test pair and j corresponds to the index of the test that was administered second in the test pair. The next step in the equating process is to organize the c_{ij}'s into their appropriate cells in the multiple test equating matrix.

To obtain a single Rasch model equating constant for a test in the Vector Design, say test i, the two cell equating constants for the test pair combination, c_{1i} and c_{i1}, are combined to obtain the final Rasch model

test equating constant for test i. This is done for each test in the study. To combine the c_{ij}'s for a test pair, the signs of the test pair cell equating constants for the cells in the row vector of the design are reversed. The average of the two cell equating constants for a test pair in the Vector Design, $(c_{1i} + c_{i1})/2$, are the final Rasch model test equating constants, C_i's, that may be used to equate the ability estimates of the base test to the ability estimates that would have been obtained on the scale of ability associated with any other test in the study.

Table 4 presents the data on the four reading vocabulary tests in a Vector Design. As in the previous two multiple test equating designs, the final equating constant for a particular test is the value to be added to the Rasch ability estimate based on the ability scale of the base test (CAT A3) to obtain the equivalent estimate on the ability scale of the particular test.

Insert Table 4 about here

Rasch Model Equating Errors

According to Angoff (1971), Donlon and Angoff (1971) and Rentz, et al. (1975), the major source of error in test equating is the unreliability of the test data, i.e., the standard error of measurement. An analysis of the estimated error in equating in the ATS found that the errors of equating would be a trival factor relative to the error of measurement.

Rentz, et al. (1975), have identified three sources or error that are associated with test equating with the Rasch model: (1) the error of measurement, (2) the error of the equating constants, and (3) the "assignment" error. For Rasch calibrated tests, the standard error of measurement appears to be approximately 0.2 or more log ability units or 2.5 to 3.5 raw score units for typical length tests.

The second source of error in Rasch model test equating is associated with the equating constants. Depending on the equating design, the data manipulation procedures, and the number of tests, the estimates of the various errors of the equating constants are based on the standard errors of item easiness estimates and the formula for the addition of uncorrelated variance, $V(\Sigma a_i x_i) = \Sigma a_i^2 \cdot V(x_i)$.

The MESAMAX computer program provides the standard errors associated with each item easiness estimate. The first step was to obtain the average variance error of the items in a test. This yielded an average error for each test in a test pair combination. The next step was to add the two variance errors for a test pair to obtain a variance error associated with the cell equating constant. To obtain the variance errors associated with the final equating constants, the variance error for each cell was combined in the same order as were the cell equating constants. Finally, the standard error for each final equating constant was obtained by taking the square root of the variance error associated with the final equating constants. The standard errors associated with the final Rasch model test equating constants for each design are presented in Table 5.

Insert Table 5 about here

The errors associated with the equating constants were minor in comparison to other types of errors. For the Rasch Project the variance error of equating constants was in the order of 0.02 log ability units, or

approximately ten to fifteen percent of the standard error of measurement. This agreed with the estimates of the standard errors of equating provided by Donlon and Angoff (1971) for the SAT.

The third source of error, the assignment error, is associated with the assignment of equivalent raw scores on two tests. If a common log ability scale was used in reporting equivalent scores, there would be no assignment error.

Thus, the major source of error in test equating is still the error of measurement of the raw data. The second largest source of error in the Rasch Project was the assignment error which could be eliminated by "calibrating" all tests on a single log ability reference scale as opposed to raw-score-to-raw-score equating. Of the three sources of error in the Rasch model equating, the errors associated with the equating constants are minimal. Estimation procedures for the variance errors of the equating constants are presented in the Rasch Project final report.

Discussion and Summary

The purpose of this study was to develop and illustrate the procedures for obtaining Rasch model test equating constants for three multiple test equating designs. A multiple test equating matrix was defined and three multiple test equating designs, Full, Chain, and Vector, were described in terms of the matrix. The procedures for obtaining Rasch model test equating constants were delineated in a general form for each design. The Full Design closely parallels the design that was employed in the ATS and Rasch Project. This design included all possible combinations of two tests and the counterbalanced testing order combinations. Obviously this design would be the preferred design for equating a set of k tests since it contains the maximum information in terms of test pair combinations. But the Full Design requires large numbers of subjects and research funds as the number of tests included in the study increases. Also, all of the tests included in the Full Design must be appropriate for a common level or age group.

The Chain Design and Vector Design provide alternatives for the researcher who is limited in his resources. The Chain Design is particularly useful in a setting

where the tests in the equating study are sequential in the levels or age groups for which they are appropriate, e.g., the Sequential Tests of Educational Progress (STEP). The Vector Design is useful in a setting similar to the Full Design where limited resources are available and the tests in the equating study are appropriate for a common level or age group.

The investigation of the use of the Rasch model in the area of test equating began with Panchapakesan in 1969 and reached a major point with the Rasch Project in 1975. The simplicity of the Rasch model is characterized by the fact that only a single value, a test equating constant, is required to adjust the scale of one test to the scale of a second test. The present study has provided the general procedures and examples for obtaining the Rasch model test equating constants for three multiple test equating designs. These procedures are adaptable to any number of tests that might be included in a multiple test equating study.

To illustrate the procedures for obtaining the Rasch model test equating constants, multiple test equatings of four tests were performed. Random samples of 500 cases of test pair data for each cell in the multiple test equating matrix were drawn from the data collected for the ATS and Rasch Project. A set of four Rasch model

test equating constants were obtained for each of the designs in the present study. Since each design has its own utility, no statistical comparisons of the sets of constants across designs were appropriate

Angoff (1971), Donlon and Angoff (1971) and Rentz, et al. (1975) have pointed out that the major source of error in test equating is the usual standard error of measurement due to the unreliability of the individual tests. However, the error associated with the equating constants must be examined. Adapting the procedures developed by Rentz, et al. (1975) to the three designs in the present study, standard errors of the equating constants were obtained for each design with 500 cases of data in each cell. Table 5 presented the standard errors of the equating constants. Using the procedure for obtaining crude estimates provided by Rentz, et al. (1975), the standard errors of the equating constants for the Full Design with four tests was .0106. The standard errors reported for the Full Design are all within .0006 of the crude estimate. For the Chain Design the crude estimates for the standard errors were .0173 for the first linked test equating constant (ITBS 5,11), .0245 for the second linked test equating constant (MAT FI), and .0300 for the third linked test equating

constant (SRA EB). All of the crude estimates of the standard errors for the constants in the Chain Design were within .0008 of the reported values for the standard errors of the equating constants for the Chain Design in Table 5. For the Vector Design the crude estimate of the standard error of the equating constants was .0173 and all ~~of~~ the reported values are within .0011 of the crude estimate. For the Chain and Vector Designs, the standard error of the equating constant for the base test is zero since no estimates are directly obtained for the equating constant for the base test.

From an examination of the standard errors of the equating constants reported in Table 5, it is obvious that the standard errors are smaller for the constants based on the Full Design. The standard errors of the equating constants for the Vector Design are only slightly larger than those for the Full Design. However, the standard errors of the equating constants based on the Chain Design increase as the number of links in the chain between the base test and the test to be equated increase. Employing the $n\sigma_e^2$ defined by Donlon and Angoff (1971) as the increase in the equating error and using the standard error of the equating constant for ITBS 5,11 of .0179, the expected standard error of the second link would be .0245 and for the third link would be .0310. The standard

error value obtained for MAT FI was .0240 and for SRA EB was .0290. Considering the concern generated by the continuing drop in the national norms on the SAT in recent years and the number of links in the equating of each new form to the original 1941 normative or reference form of the test, a strong possible explanation for the drop may be tied to the equating error in each new form of the test.

The present research has provided those interested in test equating the procedures for obtaining Rasch model test equating constants for three multiple test equating designs. The next step in the application of the Rasch model to the test equating domain is in the area of test calibration of multiple tests to a common reference scale. Calibration would seem a preferable process to equating if for no other reason than the potential of reducing the errors associated with raw-score-to-raw-score equating, i.e., the assignment error, and the errors associated with the equating constants.

Table 1

Symbols and Definitions

k	the number of tests to be equated
n_i	the number of items on test i
T_{ij}	test pair (i,j) with test i administered first
c_{ij}	Rasch model cell equating constant for equating test i to test j, with test i administered first
$c_{i.}$	ith row marginal mean of the c_{ij}'s
$c_{.j}$	jth column marginal mean of the c_{ij}'s
$\bar{c}_i$	Rasch model preliminary equating constant for test i
C_i	Final Rasch model equating constant for test i

Tests	1	2	3	.	.	.	k-1	k
1		T_{12}	T_{13}	.	.	.	$T_{1,1-1}$	T_{1k}
2	T_{21}		T_{23}	.	.	.	$T_{2,k-1}$	T_{2k}
3	T_{31}	T_{32}		.	.	.	$T_{3,k-1}$	T_{3k}
.	.	.	.		.	.	.	.
.	.	.	.	.		.	.	.
.	.	.	.	.	.		.	.
k-1	$T_{k-1,1}$	$T_{k-1,2}$	$T_{k-1,3}$	.	.	.		$T_{k-1,k}$
k	T_{k1}	T_{k2}	T_{k3}	.	.	.	$T_{k,k-1}$	

Figure 1. A Multiple Test Equating Matrix for K Tests

Table 2

Cell Equating Constants for the Full Design

Tests	CAT A3	ITBS 5,11	MAT FI	SRA EB	$c_{i.}$
CAT A3	0	.673	.420	-.150	.2352
ITBS 5,11	-.585	0	-.113	-.496	-.2985
MAT FI	-.458	.226	0	-.363	-.1488
SRA EB	.007	.559	.422	0	.2470
$c_{.j}$	-.2590	.3645	.1834	-.2523	
$\bar{c}_i$	.2473	-.3315	-.1655	.2496	
C_i	0	-.579	.413	.002	

Tests	1	2	3	4	.	.	.	k-1	k
1		T_{12}							
2	T_{21}		T_{23}						
3		T_{32}		T_{34}					
4			T_{43}						
.									
.									
.									
k-1									$T_{k-1,k}$
k								$T_{k,k-1}$	

Figure 2. A Chain Design for Multiple Test Equating for K Tests

Table 3

Cell Equating Constants in the Chain Design

Tests	CAT A3	ITBS 5,11	MAT FI	SRA EB
CAT A3		.673		
ITBS 5,11	-.585		-.113	
MAT FI		.226		-.363
SRA EB			.422	
$\bar{c}_i$	0	-.629	.1695	.3925
C_i	0	-.629	-.460	-.067

Tests	1	2	3	.	.	.	k
1		T_{12}	T_{13}	.	.	.	T_{1k}
2	T_{21}						
3	T_{31}						
.	.						
.	.						
.	.						
k	T_{k1}						

Figure 3. A Vector Design for Multiple Test Equating for K Tests

Table 4

Cell Equating Constants for the Vector Design

Tests	CAT A3	ITBS 5,11	MAT FI	SRA EB
CAT A3		.673	.420	-.150
ITBS 5,11	-.584			
MAT FI	-.458			
SRA EB	.007			
C_i	0	-.629	-.439	-.079

Table 5

Standard Errors of the Equating Constants for each Multiple Test Equating Design

Design	Standard Error of The Equating Constant			
	CAT A3	ITBS 5,11	MAT FI	SRA EB
Full	.0111	.0108	.0108	.0109
Chain	.0000[a]	.0179	.0240	.0292
Vector	.0000[a]	.0179	.0184	.0182

[a]Since no direct operations were performed to estimate the Rasch model test equating constant for the base test in the Chain and Vector Designs, the standard error of the equating constant for the base test is .0000.

REFERENCES

Anchor Test Study: Final report (Project Report). Berkeley, Calif.: Educational Testing Service, 1972.

Angoff, W. H. Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement: Second edition. Washington, D.C.: American Council on Education, 1971.

Donlon, T. F., & Angoff, W. H. The Scholastic Aptitude Test. In W. H. Angoff (Ed.), The College Board Admission Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests. Princeton: Educational Testing Service, 1971.

Panchapakesan, N. The simple logistic model and mental measurement. Unpublished doctoral dissertation, University of Chicago, 1969.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960.

Rasch, G. On general laws and the meaning of measurement in psychology. Proceedings for the Fourth Berkeley Symposium on Mathematical Statistics. Berkeley: University of California Press, 1961, 4, 321-334.

Rasch, G. An individualistic approach to item analysis. In P. F. Lazarfield, & N. W. Henry (Eds.), Reading in mathematical social science. Chicago: Science Research Associates, 1966. (a)

Rasch, G. An item analysis which takes individual difference into account. British Journal of Mathematical and Statistical Psychology, 1966, 19(1), 49-57. (b)

Rentz, R. R., Bashaw, W. L., Cartledge, C. G., & Brigman, S. L. Equating reading tests with the Rasch Model. (2 vols.). Athens, Ga.: Educational Research Laboratory, 1975.

Wright, B. D. Sample free test calibration and person free measurement. Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1967, 85-101.

Wright, B. D., & Mead, R. J. CALFIT, Computer Program, 1975. (Available from Benjamin D. Wright, Statistical Laboratory, Department of Education, University of Chicago, Chicago, Ill. 60615).

Wright, B. D., & Panchapakesan, N. A procedure for sample-free item analysis. Educational and Psychological Measurement, 1969, 29, 23-48.

Wright, B. D., & Skirmont, R. MESAMAX, Computer Program, 1972. (Available from Ronald Skirmont, CSIP, Inc., 1525 E. 53rd Street, Chicago, Ill. 60615).