EQUATE
COMPUTER PROGRAM, VERSION 2.0

Frank B. Baker
Laboratory of Experimental Design
Department of Educational Psychology
University of Wisconsin
Madison Wisconsin

The EQUATE program implements the test characteristic curve equating procedure due to Stocking and Lord(1983) for tests whose items have been scored on either a right wrong, graded or nominal basis. It uses the item parameter estimates of the common anchor items contained in the "from" instrument and the "to" instrument to compute the slope(A) and intercept (K) coefficients of the linear transformation of the ability metric. Given these coefficients, the item parameter estimates contained in the "from" file are converted to the metric of the "to" file. Optionally, the examinee ability estimates contained in a "from" file can be transformed into the "to" metric using the same coefficients. Both the transformed item and ability estimates are then stored in files named by the user in a standard format.

Under the dichotomous and graded response models, the slope and intercept equating coefficients are always computed under the assumption that the two instruments ICC's are modeled by a logistic ogive. When a test is specified as having a normal ogive metric, the numerical values of the item discrimination indices are multiplied by 1.702 in order to calculate the probability of correct response using the logistic ogive. There are four different combinations of metric for the parameters of the "from" and "to" instruments. When the "from" and "to" tests are in the same metric the transformed item discrimination parameters are reported in the common metric. When the "from" and "to" tests are in a different metric, the metric of the "to" test determines the metric of the transformed item discrimination parameters. If the "to" test has been specified as having a normal metric, and the "from test" is in a logistic metric, the transformed item discrimination parameter values are divided by 1.702 before they are saved in the transformed item parameter file. If the "to" test has been specified as having a logistic metric, the transformed item discrimination indices are in a logistic ogive metric. In all cases, only the values of the item discrimination indices are modified to reflect the ICC model. In the case of nominally scored items, the item response category item charactersitic curves (i.e. trace lines) are always in the logistic metric and the program will automatically specify this.

Under the nominal response model, the multivariate logistic function is assumed for the "from" and "to" tests and all parameters are in the slope-intercept form. The program accepts either the MULTILOG "unconstrained" parameters or pairs of slope and intercept parameters for each item response category.

The program computes the test characteristic curves for points along the ability scale rather than computing a true score for each examinee. In the nominal response case, it computes the probability of chosing each response category at each of the ability scale points.

The majority of the IRT parameter estimation programs solve the identification problem by standardizing the ability distribution to mean zero and unit variance. Consequently, the EQUATE computer program assumes that the "TO" test ability scale metric has a midpoint of zero and a unit of measurement of one. Thus, values of the equating coefficients are those that transform the metric of the "FROM" test results into this metric.

The program was written in *Professional FORTRAN* for the IBM PC series of microcomputers using subroutines taken from the book *Numerical Recipes* (Press,Flannery,Teukolsky, and Vetterling,1986) for the Broyden-Fletcher-Goldfarb-Shanno version of the Davidon-Fletcher-Powell method for the minimization of an

arbitrary function. The program runs in an interactive mode with the user answering a series of questions and providing the necessary information. The set of responses can also be put in a batch file and executed via the < feature of Dos 3.3 and higher.

To load the program into your computer do the following:

1. Go to the root directory and then type `MD EQUATE`
2. Go to the directory `CD\EQUATE`
3. Copy all of the provided disk into the directory via
   `COPY A:*.* C:\EQUATE`
   This will read the complete contents of the floppy disk in A into the EQUATE directory.

USING THE PROGRAM

The computer generated messages are indicated in this section by the capitalized lines. The program is executed by typing EQUATE and the following occurs:

`TYPE IN A TITLE FOR THE COMPUTER RUN`

You may enter a title that is up to 64 characters in length

`ENTER NUMBER OF ABILITY SCALE POINTS N=`

The number of points on the ability scale used to define the test characteristic curve is specified. The number of points must be <=100 (11 or 21 is a reasonable number). In the nominal response case, the number of points must be <=25.

`RESPONSE MODE DICHOTOMOUS, GRADED OR NOMINAL? D/G/N`

If the test items were scored right wrong, respond `D`

If the test items were scored on a graded basis respond `G`

If the test items were scored on a nominal basis, respond `N`

NOTE: Throughout the program, if a single letter response is called for, you may respond with either an upper or lower case letter.

If the items are dichotomously scored, the following interaction will take place:

`ENTER THE NUMBER OF PARAMETERS IN THE ICC MODEL`

The ICC model can have 1,2,or 3 parameters. The "from" and "to" instruments must have been analyzed under the same ICC model. If a one parameter model is specified, the program expects that the Rasch Model logistic metric (with a discrimination value of unity) underlies both the "from" and "to" item parameters.

The equating of graded response tests is based upon the following assumptions:

1. The maximum number of response categories per item is 9.
2. Items are weighted in monotonic manner, i.e. 1,2,3,4,5,6,7,8,9 or 9,8,7,6,5,4,3,2,1.

The equating of nominal response tests is based only on assumption1 1.

`ENTER THE NAME OF THE "FROM METRIC" ITEM PARAMETER FILE`

The name of the file containing the item parameter values (estimates) for the instrument whose metric is to be transformed is entered. The name typically is something like

`C:\BKR\TEST1.ITM`

If the tests are nominally scored, the following message will appear:

IS "FROM" FILE IN GENIRV OR MULTILOG FORM? G/M

GENIRV form: The parameters(estimates) are assumed to be in a(1),c(1),a(2),c(2),.....a(nc),c(nc) order, where a is the slope and c is the intercept. It is important to note that the Nominal response case uses the slope-intercept parameterization of an item rather than the discrimination-difficulty form. If you are using a pair of slope intercept parameters for each item response category, respond with G.

MULTILOG form: If you are reading in a MULTILOG .SAV file containing the "unconstrained" parameters respond with M. The program will automatically convert them to the slope intercept parameterization using the T deviation contrasts. See Example 9-10 in the MULTILOG VER 6.0 manual and pages 3-38,39 in particular. The EQUATE PROGRAM assumes that the default T matrix of deviation contrasts was used. If you use something else you will have to do the conversion before entering EQUATE and set the parameters in the slope intercept form and respond to the query with G.

ENTER FORMAT OF "FROM METRIC" FILE

A standard FORTRAN format statement, enclosed in parentheses, is to be provided. The program assumes the item parameters are in the following order:

Dichotomous response

1-parameter ICC model b

2-parameter ICC model a b

3-parameter ICC model a b c

Graded response

a,b(1),b(2),b(3),....b(nc) where the b(i)'s are the location parameters of the (NC-1) BOUNDARY CURVES. Note: The MULTILOG output files use 2 lines per item, but the first line contains non-essential information. Thus,the format for such a file in the graded response case is (/5F10.3) for items with 4 boundary curves.

Nominal response

GENIRV format: The format should characterize the slope and intercept parameters for a single item response category. For example: (7X,F8.4,1X,F8.4).

MULTILOG format: In the Nominal response case, the MULTILOG .SAV file containing the "unconstrained" parameters has a peculiar format. There are either one or two header lines for each item depending on how many response categories are used. There are "unconstrained" parameters called a,c each having NRC-1 elements, and the d parameters with NRC-2 elements and a single value of H. These are stored in lines each having eight E10.3 fields. Thus, your format statement must encompass 3*(NRC-1) elements. For example (/8E10.3/E10.3) would be the format for a nominally scored item with four response categories. You should print the .SAV file to make sure of the number of header lines and the number of reported estimates, some of which may have values of zero (d's and H) but must still be read by EQUATE.

ENTER NUMBER OF ITEMS IN THE "FROM" TEST

The number of items in the "from" instrument is entered as a integer number. The number of items must be <= 100. If nominally scored items are used, the number of items must be <=35.

If the tests are nominally scored, the next query will not appear and the metric is automatically set to logistic.

IS FROM METRIC LOGISTIC OR NORMAL? L/N

If the parameters were calibrated in the normal ogive metric(such as when LOGIST is used), respond N. If the parameters were calibrated using a logistic metric(such as when the LOG option is used in BILOG), respond L. If the one parameter, i.e. Rasch model, is used, you must respond with L.

ENTER THE NAME OF THE "TO METRIC" ITEM PARAMETER FILE

The name of the file containing the item parameter values (estimates) for the instrument whose metric is to be the basal metric is entered. The name typically is something like

C:\BKR\TEST2.DAT

If the tests are nominally scored, the following message will appear:

IS "TO" FILE IN GENIRV OR MULTILOG FORM? G/M

GENIRV form: The parameters(estimates) are assumed to be in a(1),c(1),a(2),c(2),.....a(nc),c(nc) order, where a is the slope and c is the intercept. It is important to note that the Nominal response case uses the slope-intercept parameterization of an item rather than the discrimination-difficulty form. If you are using a pair of slope intercept parameters for each item response category, respond with G.

MULTILOG form: If you are reading in a MULTILOG .SAV file containing the "unconstrained" parameters respond with M. The program will automatically convert them to the slope intercept prameterization using the T deviation contrasts. See Example 9-10 in the MULTILOG VER 6.0 manual and pages 3-38,39 in particular. The EQUATE PROGRAM assumes that the default T matrix of deviation contrasts was used. If you use something else you will have to do the conversion before entering EQUATE and set the parameters in the slope intercept form and respond to the query with G.

ENTER FORMAT OF "TO METRIC" FILE

A standard FORTRAN format statement, enclosed in parentheses, is to be provided. The program assumes the item parameters are in the following order:

Dichotomous response

1-parameter ICC model b

2-parameter ICC model a b

3-parameter ICC model a b c

Graded response

a,b(1),b(2),b(3),....b(nc) where the b(i)'s are the location parameters of the (NC-1) BOUNDARY CURVES. Note: The MULTILOG output files use 2 lines per item, but the first line contains non-essential information. Thus,the format for such a file in the graded response case is (/5F10.3) for items with 4 boundary curves.

Nominal response

GENIRV format: The format should characterize the slope and intercept parameters for a single item response category. For example: (7X,F8.4,1X,F8.4).

MULTILOG format: In the Nominal response case, the MULTILOG .SAV file containing the "unconstrained" parameters has a peculiar format. There are either one or two header lines for each item depending on how many response categories are used. There are "unconstrained" parameters called a,c each having NRC-1 elements, and the d parameters with NRC-2 elements and a single value of H. These are stored in lines each having eight E10.3 fields. Thus, your format statement must encompass 3*(NRC-1) elements. For example (/8E10.3/E10.3) would be the format for a nominally scored item with four response categories. You should print the .SAV file to make sure of the number of header lines and the number of reported estimates, some of which may have values of zero (d's and H) but must still be read by EQUATE.

ENTER NUMBER OF ITEMS IN THE "TO" TEST

The number of items in the "from" instrument is entered as a integer number. The number of items must be <= 100. If nominally scored items are used, the number of items must be <=35.

NOTE: The number of items in the "FROM and TO" instruments can be different. However, the number of anchor items must be the same. If the tests are nominally scored, the next query will not appear and the metric is automatically set to logistic.

IS TO METRIC LOGISTIC OR NORMAL? L/N

The program computes all true scores using the logistic ogive. If the parameters were calibrated in the normal ogive metric(such as when LOGIST is used), respond N. If the parameters were calibrated using a logistic metric(such as when the LOG option is used in BILOG), respond L. If the one parameter, i.e. Rasch model, is used, you must respond with L.

NOTE: The specification of the FROM and TO metrics do not have to be the same, i.e. one could equate LOGIST normal results to BILOG logistic etc.

ENTER NAME OF FILE TO STORE TRANSFORMED ITEM PARAMETERS

The name of the file is typically something like
C:\BKR\TRANSIP.DAT.

The transformed item parameters will be stored in the a,b,c order using the following formats:

1-parameter ICC model (F12.6)
2-parameter ICC model (2F12.6)
3-parameter ICC model (3F12.6)
Graded response model (10F10.3)

Note: a single discrimination index will be reported for each item followed by the (NRC-1) boundary curve difficulties.

Nominal response model (2F12.6)

Pairs of slopes and intercepts will be reported by item response category.

TRANSFORM THETAS? Y/N

A Y response informs the program that ability estimates for a group of examinees who were administered the "FROM" instrument is to be transformed into the "TO" metric using the obtained transformation coefficients. A N response by passes this aspect of the program.

ENTER THE NAME OF THE "FROM" THETA FILE

The file name is typically something like
C:\BKR\THETA2.DAT

ENTER THE FORMAT OF "FROM" THETA FILE

A standard FORTRAN format, enclosed in parentheses, is used to specify the file design.

ENTER NUMBER OF EXAMINEES

Enter the number of examinees (as an integer number) whose ability scores are to be transformed.

NOTE: There is no limit on the number of examinees whose "FROM" ability estimates are to be transformed.

ENTER NAME OF FILE TO STORE TRANSFORMED THETAS

The file name is typically something like
C:\BKR\NEWTHET.DAT.

The transformed thetas will be stored in the file via a (F10.3) FORTRAN format statement. [At this point, the program will print a summary of the specifications to allow the user to verify them before proceeding.]

ARE THESE SPECIFICATIONS OK? Y/N

A Y response tells the program that the program has been set up properly and the analysis should proceed. A N response indicates that something is incorrect and the whole process described above must be repeated.

SPECIFY THE ANCHOR ITEM IDs IN THE "FROM" INSTRUMENT
ENTER LIST OF ANCHOR ITEMS SEPARATE WITH COMMAS TERMINATE WITH COLON

The program implements a scheme for indicating which items in an instrument are those to be used in computing the transformation coefficients. Anchor Item ID's can be entered as a single digit or as a range of numbers, such as
1,2,3,5-12,17,33:

Each specification must be separated by a comma and the last specification must be followed by a :

FOR G OR N CASES ENTER CATEGORY INFORMATION

GRADED response model. The program provides for varying numbers of response categories as well as two directions (low- high and high-low) of weight assignment. For each item in the test you will be asked for the number of response categories and the category number with the highest weight which must be an extreme category.

ITEM	N CATEGORIES	CATEGORY WITH HIGH WEIGHT
1	4	1
2	2	2
3	4	1
4	3	3
5	4	4

For example you should type 4 1 for the first item above with a space between the two numbers.

NOMINAL response model. The program provides for varying number of response categories per item. You will be asked for the number of categories on an item by item basis for the "FROM" test.

ITEM	N CATEGORIES
1	3
2	2
3	3
4	4

You respond with the number of response categories in each item, such as

3
2

SPECIFY THE ANCHOR ITEM IDs IN THE "TO" INSTRUMENT

ENTER LIST OF ANCHOR ITEMS SEPARATE WITH COMMAS TERMINATE WITH COLON

Note: The ID's of the anchor items in the "from" and "to" instruments can be different, but the number of anchor items in the two instruments must be the same. If they are not, the following error message will be printed and the program terminates:

NUMBER OF ANCHOR ITEMS NOT EQUAL "TO"= "FROM" =

Where the blanks will contain the numbers.

FOR G OR N CASES ENTER CATEGORY INFORMATION

GRADED response model. The program provides for varying numbers of response categories as well as two directions (low- high and high-low) of weight assignment. For each item in the test you will be asked for the number of response categories and the category number with the highest weight which must be an extreme category.

ITEM	N CATEGORIES	CATEGORY WITH HIGH WEIGHT
1	4	1
2	2	2
3	4	1
4	3	3
5	4	4

For example you should type 4 1 for the first item above with a space between the two numbers.

NOMINAL response model. The program provides for varying number of response categories per item. You will be asked for the number of categories on an item by item basis for the "FROM" test.

ITEM	N CATEGORIES
1	3
2	2
3	3
4	4

You respond with the number of response categories in each item, such as

4
3

Only in the case of the nominal response model, a provision has been made for identifying how the anchor items appearing in different relative positions in the the two tests are paired. Such a capability is NOT needed in the dichotomous and graded cases as the test characterisitic curve calculation does not care about the order in which the items enter the calculations.

The computer will print the identification number of an anchor item in the "FROM" test and ask you for the identification number of the corresponding item in the "TO" test.

"FROM" ITEM X IS MATCHED WITH "TO" ITEM

Where X is the identification number of an anchor item in the "FROM" test. You will respond with the corresponding identification number. If the number entered is not in the list of anchor items for the "TO" test, the following message will appear:

ITEM XXX NOT ON "TO" ANCHOR ITEM LIST"

Where XXX is the number you responded with to the previous query. The computer then checks to see if both items have the same number of response categories. If not the following message is printed:

```
NUMBER OF CATEGORIES DO NOT MATCH "FROM"= X "TO"= Y
```

Where X is the number of categories in the "FROM" item and Y is the number of categories in the "TO"item.

NOTE: If either of these error messages are printed, the program will return to the first message and ask you for a new corresponding "TO" item identification number. If you can not get things to match in three tries, the computer program will stop.

ONLY IN THE CASE OF A NOMINALLY SCORED TEST, the computer will ask for the following:

```
ENTER THE INITIAL VALUE FOR A
```

You should type in an intial value for the A (slope) equating coefficient as sx.xxxx, for example 1.5000

```
ENTER THE INITIAL VALUE FOR K
```

You should type in an initial value for the K (intercept) equating coefficient as sx.xxxx, for example -0.5000

The nominally scored test equating is quite sensitive to the values of the initial estimators, you should have a reasonably good idea of what values are appropriate.

While the transformation coefficients are being computed, the following message will appear on the screen:

```
COMPUTING TRANSFORMATION COEFFICIENTS
     ITERATION = 1
     ITERATION = 2
     etc.
```

RUNNING EXAMPLES ON THE DISK

The equate program can be run interactively from the keyboard, however, it is more efficient to use an editor to create the input run stream and store it as an ASCII i.e. text, file.

Example of equating a dichotomously scored test

The file called DICHTEST.RUN, for equating two dichotomously scored tests, is provided on the disk. Assuming the EQUATE program is on the hard disk in the EQUATE directory, to execute the program in "batch mode" do the following:

1. CD:\EQUATE

2. Type EQUATE <DICHTEST.RUN

The program will compute the two transformation coefficients. The amount of time these calculations take depends upon the number of theta points and the number of anchor items. For large numbers, the program can run for several minutes. There will be a pause between the printing of the item parameter values and the ability values. Upon completion, the computer output presented below be printed.

```
EXAMPLE RUN OF EQUATE PROGRAM, DICHOTOMOUS RESPONSE CASE
  NUMBER OF ABILITY SCALE POINTS=   11
  DICHOTOMOUS RESPONSE MODEL
  ICC MODEL HAS 3 PARAMETERS

FROM METRIC ITEM PARAMETER FILE NAME IS
  C:\EQUATE\MYFILE2.DAT

  FILE FORMAT IS (3F12.6)
  NUMBER OF ITEMS IN "FROM" TEST IS  21
  FROM TEST IS IN LOGISTIC OGIVE METRIC

TO METRIC ITEM PARAMETER FILE NAME IS
  C:\EQUATE\MYFILE1.DAT
  FILE FORMAT IS (3F12.6)
  NUMBER OF ITEMS IN "TO" TEST IS  21
  TO TEST IS IN LOGISTIC OGIVE METRIC

TRANSFORMED ITEM PARAMETER FILE NAME IS
  C:\EQUATE\TRIP.DAT

 TRANSFORMED PARAMETERS WILL BE IN
   LOGISTIC OGIVE METRIC

NAME OF THE "FROM" THETA FILE IS
  C:\EQUATE\THETA2.DAT

     FILE FORMAT IS (F5.2)
```

```
  NUMBER OF EXAMINEES IN "FROM" THETA FILE IS 600

NAME OF FILE FOR TRANSFORMED THETAS IS
  C:\EQUATE\TRTHA.DAT

ANCHOR ITEM IDs FOR "FROM" INSTRUMENT ARE
1-21:

ANCHOR ITEM IDs FOR "TO" INSTRUMENT ARE
1-21:
 INITIAL VALUE FOR A=    3.3333  INITIAL VALUE FOR K=   -1.6666
 FUNCTION AT INITIAL VALUES= .000000
 NUMBER OF ITERATIONS PERFORMED = 5
 METRIC TRANSFORMATION COEFFICIENTS ARE
 A=   3.3333  K=   -1.6666
 FUNCTION VALUE IS = .00000



 SUMMARY STATISTICS FOR TRANSFORMED ITEMS
     MEAN B=    0.182 VARIANCE B=    0.915  STD DEV B=    0.956
     MEAN A=    1.027 VARIANCE A=    0.111  STD DEV A=    0.333
     MEAN C=    0.186 VARIANCE C=    0.006  STD DEV C=    0.079


 SUMMARY STATISTICS OF TRANSFORMED ABILITIES
              MEAN =    -1.666621
          VARIANCE=    70.833725
 STANDARD DEVIATION=     8.416277


   LABORATORY OF EXPERIMENTAL DESIGN
 DEPARTMENT OF EDUCATIONAL PSYCHOLOGY
      UNIVERSITY OF WISCONSIN
```

The batch run for this example, contained in file element
DICHTEST.RUN,
is as follows:

```
EXAMPLE RUN OF EQUATE PROGRAM, DICHOTOMOUS RESPONSE CASE
11
D
3
C:\EQUATE\MYFILE2.DAT
(3F12.6)
21
L
C:\EQUATE\MYFILE1.DAT
(3F12.6)
21
L
C:\EQUATE\TRIP.DAT
Y
C:\EQUATE\THETA2.DAT
(F5.2)
17
C:\EQUATE\TRTHA.DAT
Y
1-21:
1-21:
```

REFERENCES

Al-Dosary,I.M. (1983) Vertical Equating Using the Mean-Sigma and The Characteristic Curve Transformation Methods. Unpublished PhD dissertation, University of Wisconsin.

Baker, F. B. (1992) Equating tests under the graded response model. *Applied Psychological Measurement, 16*, 87-96.

Baker, F. B. (1993) Equating tests under the nominal response model. *Applied Psychological Measurement* .

Baker, F.B., Al-Karni, Al-Dosary, I.M. (1991) EQUATE:A computer program for the test characteristic curve method of IRT equating. *Applied Psychological Measurement, 15*,78.

Press,W.H., Flannery,B.P, Teukolsky,S.A.,& Vetterling,W.T. (1986) *Numerical Recipes*. Cambridge University Press, Cambridge England,pp818

Stocking, M.L. & Lord,F.M., (1983) Developing a Common Metric In Item Response Theory. *Applied Psychological Measurement, 7*, 201-210.