DFIT8

DFIT8 for Windows
User's Manual

Differential Functioning of Items and Tests

T. C. Oshima, Steve Kushubar,
John C. Scott, and Nambury S. Raju

System Requirements

Microsoft Windows 2000/XP/Vista

7 Mb free disk space

2 Mb free RAM

For Further Information

Assessment Systems Corporation

2233 University Avenue, Suite 200

St. Paul, Minnesota 55114, U.S.A.

Voice: 651/647-9220

Fax: 651/647-0412

E-Mail: support@assess.com

World Wide Web page: www.assess.com

User's Manual for DFIT8: Differential Functioning of Items and Tests

First Edition, Windows Version (DFIT8.04)

Your DFIT8 License

Unless you have purchased multiple licenses for DFIT8, your license is a single-user license. You may install DFIT8 on up to two computers (e.g., a desktop and a laptop) so long as there is no possibility that the two copies will be used simultaneously.

Technical Assistance

If you need technical assistance using DFIT8, please visit the Support section of our Web site, www.assess.com. If the answer to your question is not posted, please email us at support@assess.com. Technical assistance for DFIT8 is provided for one year from the date you purchase your license. Please provide us with the invoice number for your license purchase when you request technical assistance.

Citation

Oshima, T. C., Kushubar, S., Scott, J.C. & Raju N.S. (2009). *DFIT8 for Window User's Manual: Differential functioning of items and tests.* St. Paul MN: Assessment Systems Corporation.

Contents

1 Introduction 1

Item Response Theory Models 1

Differential Functioning of Items and Tests 3

The DFIT Framework for DIF and DTF 4

Significance Tests in DFIT 6

References 7

2 Formatting Input Data for Analysis 8

Dichotomous IRT Calibration 8

Polytomous IRT Calibration 8

Linking 9

References 10

3 Definition File 11

Running DFIT 11

Dichotomous Run Definition File 11

Polytomous Graded Response Model Definition File 14

4 Output File 17

Viewing Output 17

Content of the Output File 17

Appendix A. Examples 18

1. A dichotomous example (Example 1) 18

2. A polytomous example (Example 2) 25

Appendix B. Troubleshooting 33

1 Introduction

Nambury S. Raju (1937 – 2005) developed a framework for assessing the differential functioning of items and tests (DFIT) based on item response theory (IRT). In DFIT, differential item functioning (DIF) and differential test functioning (DTF) are assessed by using item and score parameter estimates obtained from an IRT calibration. The framework can be applied to test data that are either dichotomous (Raju, van der Linden, & Fleer, 1995) or polytomous (Flowers, Oshima, & Raju, 1999). Although both unidimensional and multidimensional DFIT models have been developed (Oshima, Raju, & Flowers, 1997), this program implements only the unidimensional models.

Item Response Theory Models

For dichotomously scored items, the three-parameter logistic model describes the probability of an individual answering item *i* correctly as a function of ability (θ) as follows.

$$P_i(\theta_s) = c_i + (1 - c_i)\frac{\exp\left[Da_i(\theta_s - b_i)\right]}{1 + \exp\left[Da_i(\theta_s - b_i)\right]}. \qquad (1)$$

The item characteristics are represented by the a_i, b_i, and c_i parameters. The location parameter, b_i reflects the difficulty of the item. The discrimination parameter, a_i, relates to the steepness of the curve. The c_i parameter reflects the lower asymptote, or the probability that a person with extremely low θ would correctly answer the item. D is a scaling constant typically set at 1.702.

Polytomous IRT models require the estimation of multiple item response functions (IRF) representing the different response categories. Although various polytomous models have been developed in the IRT literature, the current version of DFIT8 uses Samejima's (1969) graded

response model (GRM) which is designed for ordered response categories. Theoretically, however, DFIT can be applied to different types of polytomous models.

According to GRM, for an item with *m* response categories, there will be *m*-1 boundary response functions (BRF). A BRF represents the probability of person *s* responding above response category *k* on item *i*,

$$P_{ik}^{*}(\theta_s) = \frac{\exp\left[\mathrm{D}a_i\left(\theta_s - b_{ik}\right)\right]}{1+\exp\left[\mathrm{D}a_i\left(\theta_s - b_{ik}\right)\right]}, \qquad (2)$$

where b_{ik} is a location parameter that designates the boundary between response categories *k* and *k*+1, and a_i is the item discrimination parameter.

The probability of responding in a particular response category can be computed from the difference between adjacent BRFs. This function is referred to as the category response function (CRF):

$$P_{ik}(\theta) = P_{i(k-1)}^{*}(\theta) - P_{ik}^{*}(\theta). \qquad (3)$$

Because the first and last response categories lack an adjacent boundary, Samejima (1969) defined $P_{i0}^{*}(\theta) = 1$, and $P_{im}^{*}(\theta) = 0$. There will be as many CRFs for an item as there are response categories.

The expected score of individual *s* on item *i*, $ES_{si}(\theta_s)$, can be defined as a weighted average of the category values, where the weights reflect the probability of the individual selecting each category (i.e, the CRFs),

$$ES_i(\theta_s) = \sum_{k=1}^{m} P_{ik}(\theta_s) X_{ik}, \qquad (4)$$

where X_{ik} is the value assigned to category *k* on item *i*. For a dichotomously scored item, the expected score is equal to the probability of answering an item correctly (Equation 1).

The total score on a test can be defined as the sum of the scores on the individual items. This total test score can also be modeled as a function of θ and the resulting curve is called the test response function (TRF). The TRF is defined as sum of expected score functions across *n* items,

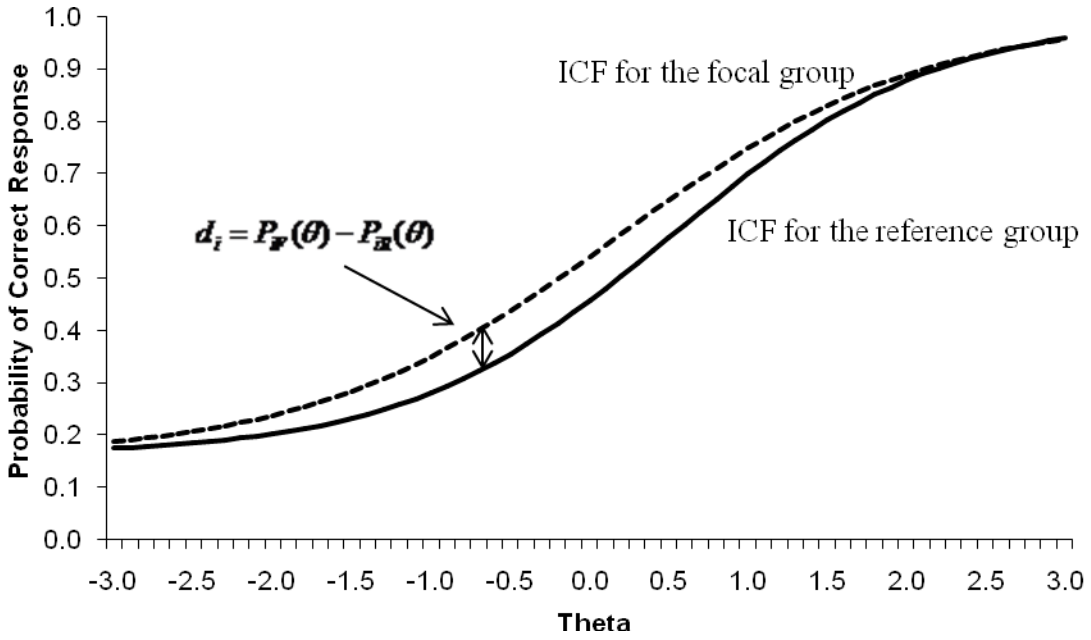
$$T(\theta_s) = \sum_{i=1}^{n} ES_i(\theta_s). \quad (5)$$

Differential Functioning of Items and Tests

The DFIT analysis starts with two sets of item parameter estimates (one from a reference group and another from a focal group) which are placed on the scale of the focal group along with the focal group's θ estimates. According to IRT, the item response functions (IRFs; also called item characteristic curves or item characteristic functions) are invariant over subgroups of examinees. Therefore, DIF is conceptualized by measuring the differences between the two IRFs .

Figure 1 graphically depicts the differences between a reference group IRF and a focal group IRF in the dichotomous model, by plotting probability (Y axis) against θ (X axis). The curve plotted is the IRF. According to IRT, the two curves should be invariant after the item parameter estimates from each group are placed on the common scale. Thus, any gap between the two IRFs indicates DIF. The gap can be measured in various ways. For example Raju (1988) developed a DIF index based on the area of the gap. In the DFIT framework, on the other hand, the gap is measured by the average squared distance of the probability difference at θ levels based on the focal group. This approach, compared to the area approach, offers substantial advantages as it can be easily expanded to different models (unidimensional or multidimensional models) in various scoring formats (dichotomous or polytomous data). The DFIT approach also makes several applications possible that are useful in DIF research (e.g., DTF, differential bundle functioning (Oshima, Raju, Flower, & Slinde, 1998)).

Figure 1. Two IRFs/ICFs



The DFIT Framework for DIF and DTF

The DFIT framework offers two types of DIF, non-compensatory DIF (NCDIF) and compensatory DIF (CDIF). NCDIF is defined as

$$NCDIF_i = E_F\left[d_i(\theta)^2\right], \tag{6}$$

where

$$d_i(\theta) = P_{iF}(\theta) - P_{iR}(\theta) \tag{7}$$

for the dichotomous model (see Figure 1), and

$$d_i(\theta) = ES_{iF}(\theta) - ES_{iR}(\theta) \tag{8}$$

for the polytomous model. E_F denotes the expectation taken over the θ distribution from the focal group.

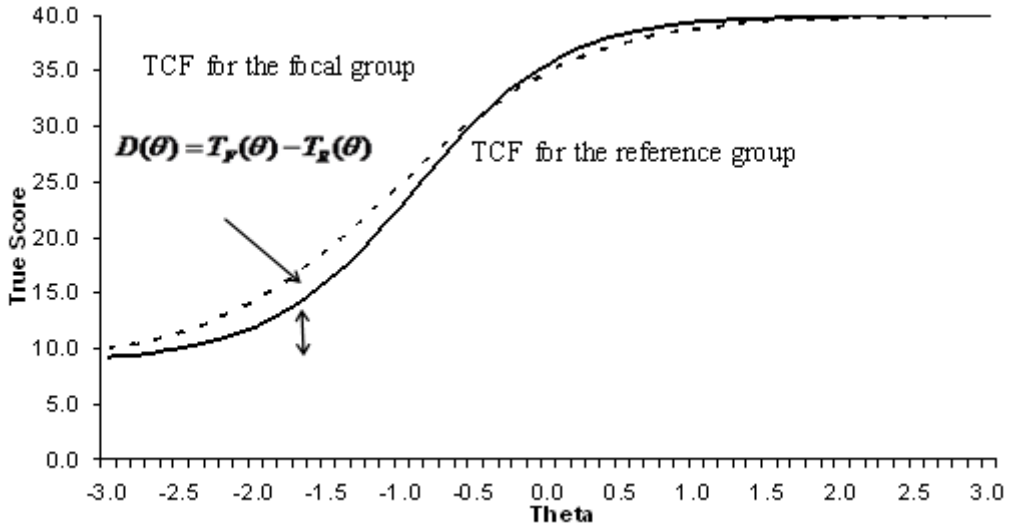
DTF is similarly defined but at the test level. As shown in Figure 2, the two curves compared are the TRFs and the distance is defined as,

$$D(\theta) = T_F(\theta) - T_R(\theta)\,. \tag{9}$$

DTF is then defined as the expected value of the squared difference between focal and reference groups, where the expectation is taken across the θ distribution from the focal group,

$$DTF = E_F\left[D(\theta)^2\right]. \tag{10}$$

Figure 2. Two TRFs/TCFs



Despite the mathematical similarity between NCDIF and DTF, the relationship between the two is not straightforward as NCDIF does not take the item covariances into account. CDIF, on the other hand, has a straightforward relationship with DTF,

$$DTF = \sum_{i=1}^{n} CDIF_i\,. \tag{11}$$

CDIF is defined as

$$CDIF_i = E_F\left(d_i D\right) = Cov\left(d_i, D\right) + \mu_{d_i}\mu_D\,, \tag{12}$$

where Cov stands for covariance.

NCDIF and CDIF both relate to DIF, but address different issues. Like most item-level DIF indices, NCDIF assumes that all items other than the studied item are DIF free, and can be used like any other DIF indices. CDIF, on the other hand, takes into account the pattern of DIF across items, and can be used to investigate the effect of removing the DIF items on DTF.

Significance Tests in DFIT

DFIT8 employs a recently developed significance test for NCDIF and DTF. The test is called the item parameter replication (IPR) method (Oshima, Raju & Nanda, 2006; Raju, Fortmann, Kim, Morris, Nering, & Oshima, in press) and provides a means of deriving cutoff values that are tailored to a particular data set. The IPR method begins with estimates of item parameters for the focal group and the sampling variances and covariances of these estimates. Based on these initial estimates, a large number of replications (typically 1,000 pairs) of item parameters are simulated. Then the cutoff value for each alpha level is determined by the empirical sampling distribution of NCDIF/DTF obtained under the null hypothesis that focal and reference groups have identical parameters. The IPR method produces cutoff values for each item.

The significance of CDIF is not tested directly. Instead, items with large CDIF are removed one by one until DTF reaches non-significance by a chi-square test. Those removed CDIF items are then considered significant.

References

More detailed information on DFIT can be found in the following references. Researchers new to DFIT might find the instructional module developed for the National Council on Measurement in Education (Oshima & Morris, 2008) useful.

Flowers, C. P., Oshima, T. C., & Raju, N. S. (1999). A description and demonstration of the polytomous-DFIT framework. *Applied Psychological Measurement, 23*, 309-326.

Oshima, T. C., & Morris, S. B. (2008). An NCME Instructional Module on Raju's Differential Functioning of Items and Tests (DFIT). *Educational Measurement: Issues and Practice, 27,* 43-50.

Oshima, T. C., Raju, N. S., & Flowers, C. P. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential functioning of items and tests. *Journal of Educational Measurement, 34,* 253-272.

Oshima, T. C., Raju, N. S., Flowers, C. P., & Slinde, J. (1998). Differential bundle functioning (DBF) using the DFIT framework: Procedures for identifying possible sources of differential functioning. *Applied Measurement in Education, 11,* 353-369.

Oshima, T. C., Raju, N. S., & Nanda (2006). A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework. *Journal of Educational Measurement, 43,* 1-17.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika, 53,* 495-502.

Raju, N. S., K. A. Fortmann, Kim, W., Morris, S. B., Nering, M., & Oshima, T. C. (in press). The item parameter replication method for detecting differential functioning in the DFIT framework. *Applied Psychological Measurement.*

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). An IRT-based internal measure of test bias with applications for differential item functioning. *Applied Psychological Measurement, 19,* 353-368.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph,* No. 17.

2

Formatting Input Data for Analysis

Dichotomous IRT Calibration

Any dichotomous IRT calibration program can be used to obtain item parameter estimates, their variance and covariance for each group (focal or reference). BILOG-MG3 (Zimowski, Muraki, Mislevy, & Bock, 2002), for example, provides an estimated covariance file (*.cov) which contains item parameter estimates, error variances and covariances. The *.cov file needs to be rearranged to be readily read into DFIT8. The newly created file should contain the following information in the specific order for each item for the focal group: b, a, c, V(b), V(a), V(c), Cov(b,a), Cov(b,c) and Cov(a,c), where V is the variance and Cov is the covariance. BILOG-MG3 also provides the ability file (*.sco) which contains the θ estimates. DFIT depends on the accurate estimation of item parameters, and accurate parameter estimation typically requires large samples of examinees. Therefore, DFIT is not recommended for application with small samples. For dichotomous IRT models, the required sample size for each group will depend on the number of parameters estimated: Conservative sample sizes (N) are $N > 200$ for the one-parameter model, $N > 500$ for the two-parameter model, and $N > 1{,}000$ for the three-parameter model (Crocker & Algina, 1986).

Polytomous IRT Calibration

Any polytomous IRT calibration program can be used to obtain item parameter and variance estimates. The covariance information may not be readily available, however. For example, PARSCALE (Muraki & Bock, 2003) does not provide the covariance information. A Fortran program (Polycov; Morris 2007) is available to calculate the item variances and covariances from Parscale output and to create the input data file for DFIT8. It is not recommended to assume that the covariances are equal to 0. The input data file to DFIT8 has to provide focal group information in a specific order. Using a 5-category item as an example, it will be a, b_1, b_2,

b_3, b_4, V(a), V(b_1), V(b_2), V(b_3), V(b_4), Cov(a,b_1), Cov(a,b_2), Cov(a,b_3), Cov(a,b_4), Cov(b_1,b_2), Cov(b_1,b_3), Cov(b_1,b_4), Cov(b_2,b_3), Cov(b_2,b_4), and Cov(b_3,b_4) in that order for each item for the focal group.

As with the dichotomous case, large sample sizes are required for polytomous DFIT. A sample size recommendation for accurate estimation for the polytomous models should be followed (e.g., $N > 500$ is recommended by Reise & Yu, 1990).

Linking

In the DFIT framework, the item parameter estimates from the reference group are placed on the scale of the focal group. Linking coefficients (a multiplicative coefficient and an additive coefficient) can be obtained by using linking programs such as IPLINK (Lee & Oshima, 1996) or PIE for dichotomous models, IRTEQ (Han, 2007) for either dichotomous or polytomous models, or POLYEQUATE for polytomous models. Iterative linking, or two-stage linking is recommended. Iterative linking is a process by which only items that are considered to be DIF-free are used as linking items. The steps for conducting two-stage linking are as follows:

1. Obtain linking coefficients using all items (the first stage),
2. Run DFIT to identify large DIF items,
3. Remove those large DIF items from the test,
4. Obtain revised linking coefficients (the second stage), and finally
5. Run DFIT for all items again using those revised linking coefficients.

References

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart & Winston.

Han, K. T. (2007). IRTEQ: Windows application that implements IRT scaling and equating [computer program]. Amherst, MA: University of Massachusetts Amherst, Center for Educational Assessment. Available for download at http://www.umass.edu/remp/software/irteq/

Lee, K., & Oshima, T. C. (1996). *IPLINK: Multidimensional and unidimensional item parameter linking in item response theory.* Applied Psychological Measurement, *20,* 230. Available for download at http://education.gsu.edu/eps/4493.html

Morris, B. S. (2008). *Polycov* [Computer program]. Chicago: Illinois Institute of Technology. Available for download at http://www.iit.edu/~morris/polycov/polycov.html

Muraki, E., & Bock, R. D. (2003). *PARSCALE 4: IRT Item Analysis and Test Scoring for Rating-Scale Data.* Scientific Software: Chicago, IL.

Reise, S. P. & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement, 27,* 133-144.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2002). *BILOG-MG3* [Computer software]. St.Paul, MN: Assessment Systems Corporation.

3

Definition File

Running DFIT

In order to perform a DFIT analysis, the user must create a DFIT run definition file (*.dft). The run definition file provides all the information necessary to read the necessary data files, perform the analysis and generate the output file.

Definition files are created by using the File → New menu option.

The user can create the definition file by using the Normal View or Raw View (choices under the View pull-down menu). In the Normal View, the Windows frames assist the user to fill out the necessary information. In the Raw View, the user can type the syntax directly in text mode.

After the definition file is saved, the user can run the program by clicking Run in the pull-down menu.

Dichotomous Run Definition File

Here is a sample dichotomous run definition file:

```
A DFIT example for 2-PL
   40    2 1000 1000  -35  1.7020  1.0299  0.0576
1111111111111111111111111111111111111111
(9F8.4)
(3F8.4)
(12x,F8.4)
Focal.PAR
Reference.PAR
Theta.THT
Output.TXT
```

Each line in the dichotomous run definition file is described below:

Line 1

Any title up to 80 characters long.

Line 2

The first number represents the number of items (I5 format – i.e., an integer field five characters wide). The maximum number of Items allowed is 100.

The second number represents the IRT model. This number should be 1, 2, or 3 (I5 format).

The third number represents the number of examinees or subjects in the focal group. The maximum number of examinees allowed is 5,000 (I5 format).

The fourth number is the number of replications for generating (using the IPR method) item-level and test/scale-level cut-off scores for assessing the significance of NCDIF and DTF indices (I5 format). The maximum is 99999.

The fifth number is the number for generating random item parameters in the Monte Carlo procedure and it MUST be a negative integer (I5 format). If the Windows interface is used to make the present file, the number for generating random item parameters is created by the program.

The sixth number is *D* (F8.4 format. i.e., a 8-character field with a decimal point and four significant digits after the decimal point), the IRT constant (see Equations 1 and 2). It must be set 1.0 if the IRT model is logistic or 1.7 if the IRT model is normal ogive.

The seventh and eighth numbers are the multiplicative and additive constants (F8.4 format for both). These are used for placing the reference group item parameters on the same metric as the focal group item parameters. In the case of the 1-paramter logistic or Rasch Model, the multiplicative constant must be set to 1.0. If the focal and reference group parameters are already on a common metric, the multiplicative and additive constants must be 1.0 and 0.0 respectively.

Line 3

A '1' tells the program to include the item in the DIF/DTF analysis. A '0' excludes the item from the analysis. For example, if there are 30 items in a test and you want to exclude Item 5 from the DIF/DTF analysis, then Line 3 should be as follows:

111101111111111111111111111111

If you want to include all 30 items in the DIF/DTF analysis, Line 3 should read as follows:

111111111111111111111111111111

Line 4

The format for reading the item parameters and the variance-covariance matrix as a vector [*b*, *a*, *c*, V(*b*), V(*a*), V(*c*), Cov(*b*,*a*), Cov(*b*,*c*) and Cov(*a*,*c*) in that order] for each item in the focal group. The specific format could vary from one application to the next. In the example above, 9F8.4 means nine fields of 8 characters each, including a decimal point and four significant digits after the decimal point. Formats must be enclosed in parentheses.

Line 5

The format for reading item parameters as a vector (*b*, *a*, and *c* in that order) for each item in the reference group. Information about the variances and covariances is not needed for the reference group. The specific item format could vary from one application to the next.

Line 6

The format for reading the θ values for the focal group. The specific θ format could vary from one application to the next. In the example above. 12x means skip the first 12 columns and begin reading in column 13.

Line 7

The file path to the focal group item data file. If no path is specified (i.e., just a file name, as in the above example) , the program will look for the files in the same folder in which the program exists

Line 8

The file path to the reference group item data file.

Line 9

The file path to the θ values file.

Line 10

The file path to the output file.

Polytomous Graded Response Model Definition File

Here is a sample polytomous run definition file:

```
Sample DFIT run for a 25-item scale with 5 categories per item
   25 1010 1000  -65  1.0000  1.0732 -0.2180
1111111111111111111111111
5555555555555555555555555
(8F10.5/8F10.5/8F10.5/8F10.5/8F10.5/8F10.5/6F10.5)
(8F10.5/8F10.5/8F10.5/8F10.5/8F10.5/8F10.5/6F10.5)
(16x,f8.4)
Focal.PAR
Reference.PAR
Theta.THT
Output.TXT
12345
12345
12345
12345
.
.
.
```

Line 1

Any title up to 80 characters long

Line 2

The first number represents the number of items (I5 format). The maximum number of items allowed is 100.

The second number represents the number of examinees or subjects in the focal group. The maximum number of examinees allowed is 5,000 (I5 format).

The third number is the number of replications for generating (via Monte Carlo procedures) item-level and test/scale-level cut-off scores for assessing the significance of NCDIF and DTF indices (I5 format).

The fourth number is the number for generating random item parameters in the Monte Carlo procedure and it MUST be a negative integer (I5 format). If the Windows interface is used to make the present file, the number for generating random item parameters is created by the program.

The fifth number is D (F8.4 format). It must be set to 1.0 if the IRT model is logistic or 1.7 if the IRT model is normal ogive.

The sixth and seventh numbers are the multiplicative and additive constants (F8.4 format for both). These are used for placing the reference group item parameters on the same scale as the focal group item parameters. In the case of one-paramter logistic or Rasch model, the multiplicative constant must be set to 1.0. If the focal and reference group parameters are already on a common metric, the multiplicative and additive constants must be 1.0 and 0.0, respectively.

Line 3

A '1' tells the program to include the item in the DIF/DTF analysis. A '0' excludes the item from the analysis. For example, if there 25 items in a test and you want to exclude Item 5 from the DIF/DTF analysis, then Line 3 should be as follows:

1111011111111111111111111

If all 25 items are to be included in the DIF/DTF analysis, Line 3 should read as follows:

1111111111111111111111111

Line 4

Each number represents the number of categories for a given item. The maximum number of categories allowed is 9. If you have a scale with 25 items with 5 categories per item, Line 4 should read:

5555555555555555555555555

Line 5

The format for reading item parameters and the variance-covariance matrix as a vector [For a 5-category item, it will be a, b_1, b_2, b_3, b_4, V(a), V(b_1), V(b_2), V(b_3), V(b_4), Cov(a,b_1), Cov(a,b_2), Cov(a,b_3), Cov(a,b_4), Cov(b_1,b_2), Cov(b_1,b_3), Cov(b_1,b_4), Cov(b_2,b_3), Cov(b_2,b_4), and Cov(b_3,b_4) in that order] for each item in the focal group. The specific format could vary from one application to the next.

Line 6

The format for reading item parameters and their variances as a vector [for an item with 5 categories: a, b_1, b_2, b_3, b_4, V(a), V(b_1), V(b_2), V(b_3) and V(b_4) in that order] for each item in the Reference Group. The specific format could vary from one application to the next.

Line 7

The format for reading the θ values for the focal group. The specific θ format could vary from one application to the next.

Line 8

The file path to the focal group item data file.

Line 9

The file path to the reference group item data file.

Line 10

The file path to the θ values file.

Line 11

The file path to the output file.

Line 12 and beyond: Each row indicates the initial score values associated with the categories for a given item. For example, if you have 25 items in a scale, you need 25 lines (Lines 12-36).

4

Output File

Viewing Output

The output file is created after a successful run of the definition file and presented on the screen. It can be also viewed later using File ⟶ Open in the program, or using any other programs that display text data.

Content of the Output File

Items 1 – 5 listed below are input information to verify that the input files are read into DFIT8 correctly. Items 6 – 10 present the results from the DFIT analysis. See Appendix A for examples.

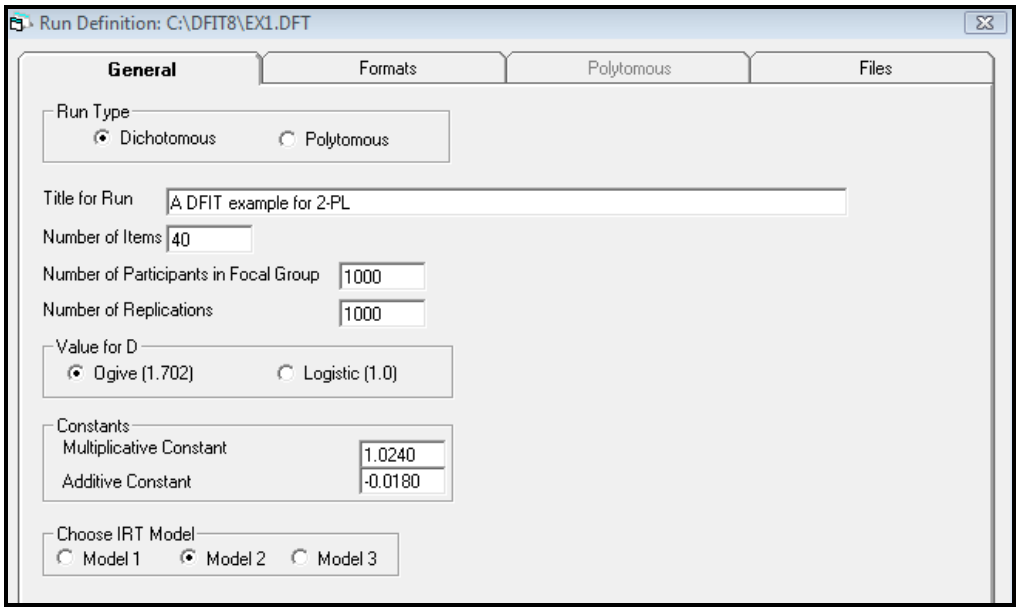
1. Title
2. Input variables
3. Focal group item parameters
4. Reference group (unequated) item parameters
5. Reference group (equated) item parameters
6. Monte Carlo generated item-level and test-level cut-offs (from the IPR method). Cutoff values for alpha = .001, .005, .01, 05, .10, and .50 are provided. DTF cutoff values are listed at the bottom.
7. Main output where CDIF, NCDIF, and NCDIF significance (*ns* indicates not significant) are reported. DIF category is for future use and it is blank.
8. Summary statistics for true scores
9. DTF related statistics, including the chi-square test
10. CDIF items to be removed to achieve non-significant DTF.

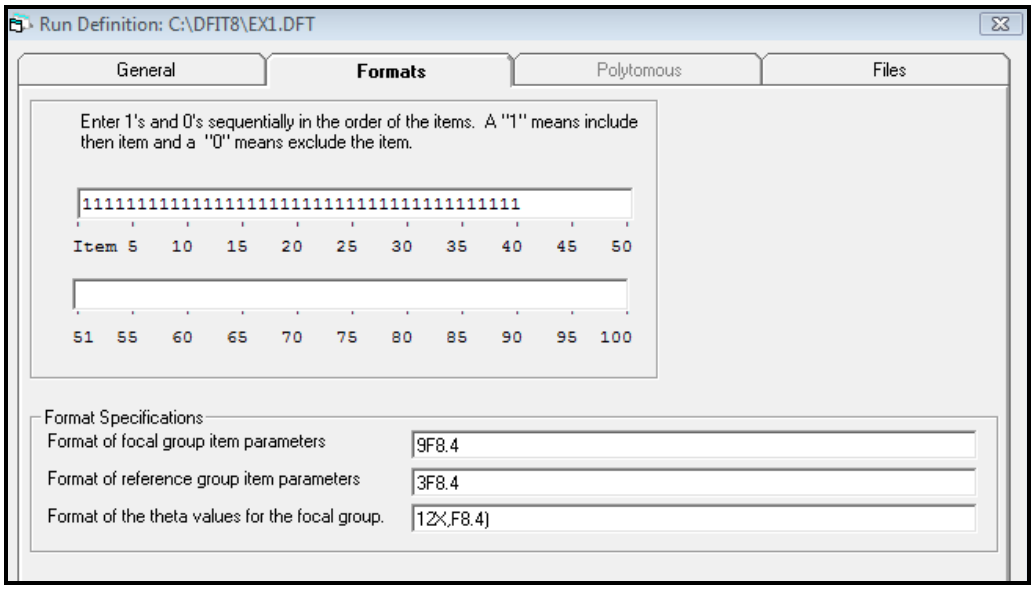
Appendix A. Examples

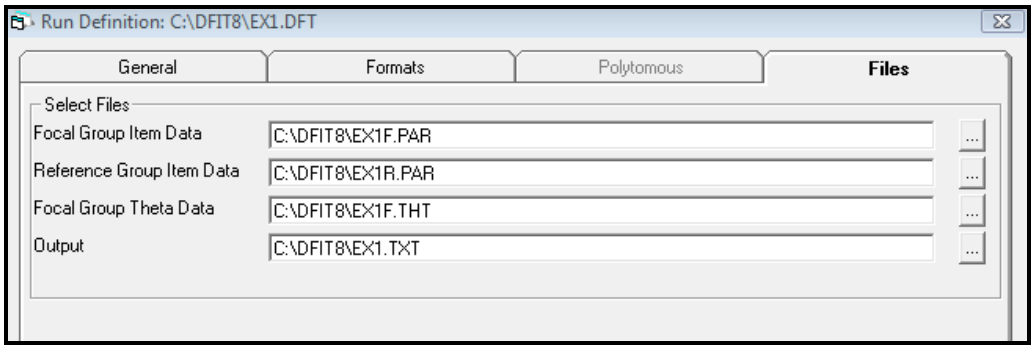
1. A dichotomous example (Example 1)

This is a 40-item simulated dichotomous test with 1,000 examinees in each group (reference and focal). DIF was embedded in Items 5, 10, 15, and 20. Items were calibrated separately for each group using BILOG-MG3 (2-parameter model). The item parameter estimates and their variance covariance estimates are listed in EX1R.PAR and EX1F.PAR for the reference group and focal group, respectively. In each file, there are five columns (b, a, V(b), V(a), and Cov(b,a) in that order). The linking coefficients were calculated using IPLINK (1.0240 and –0.0180 for the multiplicative and additive coefficients, respectively, after the second-stage linking). In the score file (EX1F.THT), the third column contains the θ estimates. For the IPR method, 1,000 replications were used to conduct the significance test.

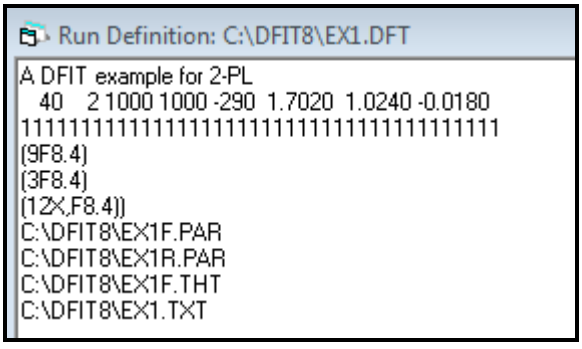
Definition File in the Normal View







Definition File in the Raw View



EX1F.PAR

File: C:\DFIT8\EX1F.PAR

```
 -.1268    .4919    .0068    .0020    .0005
 -.0039    .5873    .0049    .0023    .0001
-1.0306    .7138    .0072    .0037    .0032
-1.1965    .6637    .0099    .0035    .0041
  .9457    .6880    .0070    .0032   -.0029
 -.1201    .7076    .0037    .0029    .0004
 -.0792    .6191    .0045    .0027    .0001
```

EX1R.PAR

File: C:\DFIT8\EX1R.PAR

```
 -.1024    .5602    .0054    .0024    .0003
 -.1032    .5548    .0055    .0022    .0003
-1.0420    .7205    .0074    .0036    .0033
-1.1371    .7625    .0072    .0035    .0032
 -.0506    .7234    .0035    .0032    .0002
 -.0043    .6889    .0038    .0027    .0000
 -.0310    .7166    .0036    .0028    .0001
```

EX1F.THT

File: C:\DFIT8\EX1F.THT

```
   15    39  2.3506    .3860
   15    18  -.4157    .1574
   15    22   .3448    .2804
   15    34  1.4551    .3068
   15    12  -.4851    .1858
   15    23   .4353    .0897
   15    23   .4423    .0450
   15    29   .4844    .1841
   15    21   .4283    .1189
   15    29   .5705    .3101
   15    11  -.4836    .1823
   15    37  1.9399    .4328
   15     1 -2.4310    .4253
   15    23   .4369    .0816
   15    31  1.2512    .2587
   15    25   .4441    .0253
```

EX1.TXT (Selected output)

1. Title

2. Input variables

3. Focal group item parameters

File: C:\DFIT8\EX1.TXT

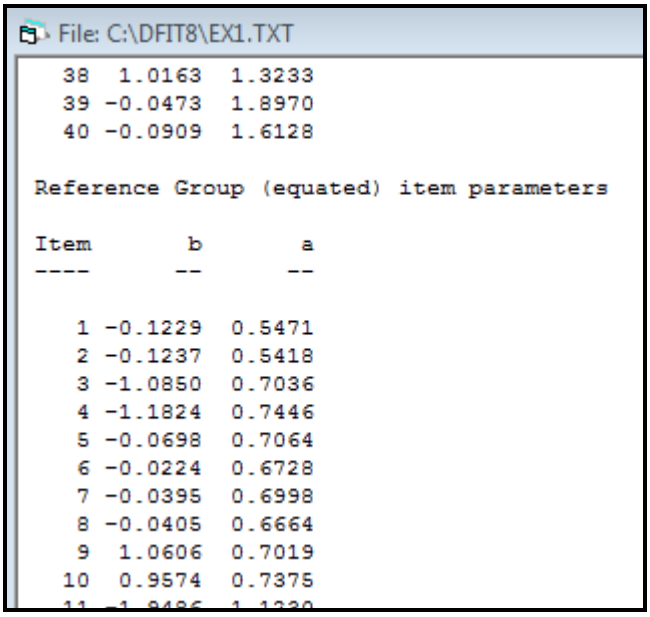
```
A DFIT example for 2-PL

 Number of Items:                          40
 Number of Item Parameters:                 2
 Number of Subjects:                     1000
 Number of Replications:                 1000
 Seed for Generating Random Numbers:     -290
 Logistic or Normal Ogive (D):         1.7020
 Multiplicative Constant:              1.0240
 Additive Constant:                   -0.0180

 Focal Group item parameters

 Item       b       a     V(b)     V(a)   C(b,a)
 ----       -       -     ----     ----   ------

    1 -0.1268  0.4919  0.0068  0.0020  0.0005
    2 -0.0039  0.5873  0.0049  0.0023  0.0001
    3 -1.0306  0.7138  0.0072  0.0037  0.0032
    4 -1.1965  0.6637  0.0099  0.0035  0.0041
    5  0.9457  0.6880  0.0070  0.0032 -0.0029
    6 -0.1201  0.7076  0.0037  0.0029  0.0004
    7 -0.0792  0.6191  0.0045  0.0027  0.0001
```

4. Reference group (unequated) item parameters

File: C:\DFIT8\EX1.TXT

```
   39 -0.0923  1.7879  0.0011  0.0165  0.0002
   40 -0.0735  1.6198  0.0012  0.0133  0.0003

 Reference Group (unequated) item parameters

 Item       b       a
 ----      --      --

    1 -0.1024  0.5602
    2 -0.1032  0.5548
    3 -1.0420  0.7205
    4 -1.1371  0.7625
    5 -0.0506  0.7234
    6 -0.0043  0.6889
    7 -0.0210  0.7166
    8 -0.0220  0.6824
    9  1.0533  0.7187
   10  0.9525  0.7552
   11 -1.8854  1.1500
```

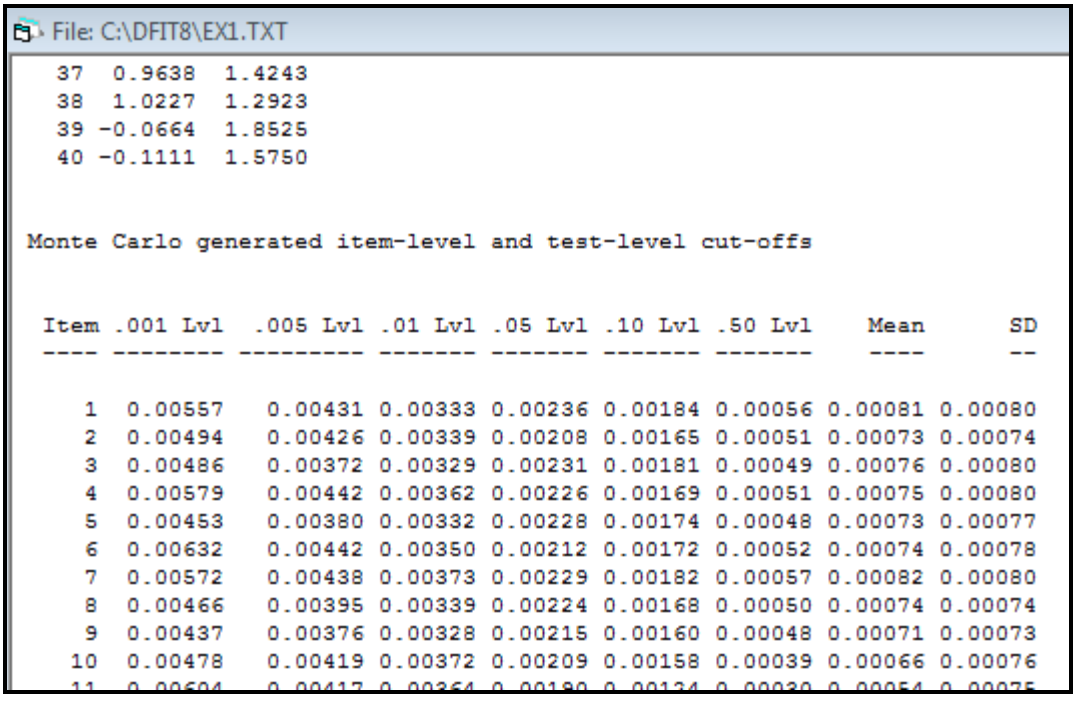
5. Reference group (equated) item parameters



File: C:\DFIT8\EX1.TXT

```
  38  1.0163  1.3233
  39 -0.0473  1.8970
  40 -0.0909  1.6128

Reference Group (equated) item parameters

Item       b       a
----      --      --

   1 -0.1229  0.5471
   2 -0.1237  0.5418
   3 -1.0850  0.7036
   4 -1.1824  0.7446
   5 -0.0698  0.7064
   6 -0.0224  0.6728
   7 -0.0395  0.6998
   8 -0.0405  0.6664
   9  1.0606  0.7019
  10  0.9574  0.7375
```

6. Monte Carlo generated item-level and test-level cut-offs (the IPR method)



File: C:\DFIT8\EX1.TXT

```
  37  0.9638  1.4243
  38  1.0227  1.2923
  39 -0.0664  1.8525
  40 -0.1111  1.5750


Monte Carlo generated item-level and test-level cut-offs


 Item .001 Lvl  .005 Lvl .01 Lvl .05 Lvl .10 Lvl .50 Lvl    Mean      SD
 ---- -------- --------- ------- ------- ------- -------    ----      --

    1  0.00557   0.00431 0.00333 0.00236 0.00184 0.00056 0.00081 0.00080
    2  0.00494   0.00426 0.00339 0.00208 0.00165 0.00051 0.00073 0.00074
    3  0.00486   0.00372 0.00329 0.00231 0.00181 0.00049 0.00076 0.00080
    4  0.00579   0.00442 0.00362 0.00226 0.00169 0.00051 0.00075 0.00080
    5  0.00453   0.00380 0.00332 0.00228 0.00174 0.00048 0.00073 0.00077
    6  0.00632   0.00442 0.00350 0.00212 0.00172 0.00052 0.00074 0.00078
    7  0.00572   0.00438 0.00373 0.00229 0.00182 0.00057 0.00082 0.00080
    8  0.00466   0.00395 0.00339 0.00224 0.00168 0.00050 0.00074 0.00074
    9  0.00437   0.00376 0.00328 0.00215 0.00160 0.00048 0.00071 0.00073
   10  0.00478   0.00419 0.00372 0.00209 0.00158 0.00039 0.00066 0.00076
```

7. Main output where CDIF, NCDIF, and NCDIF significance (*ns* indicates not-significant) are reported. DIF category is for future use and it is left blank.

File: C:\DFIT8\EX1.TXT

```
   40  0.00360   0.00335 0.00296 0.00182 0.00135 0.00043 0.00062 0.00061

  DTF  0.16970   0.12262 0.10372 0.06955 0.05664 0.02080 0.02692 0.02277


       Mean     SD    Mean    SD                                      DIF
Item    (d)     (d)   (ldl)   (ldl)  C(d,D)     CDIF      NCDIF  Sig. Category
----    ---     ---   -----  ------  ------    -----      -----  ---- --------

   1 -0.001  0.016   0.015   0.006  -0.001  0.00012   0.00026    ns
   2 -0.023  0.014   0.024   0.013   0.001  0.01735   0.00073    ns
   3 -0.009  0.006   0.009   0.006   0.000  0.00615   0.00011    ns
   4 -0.013  0.013   0.017   0.006   0.002  0.01024   0.00033    ns
   5 -0.221  0.068   0.221   0.068   0.007  0.15813   0.05342  .001
   6  0.023  0.010   0.023   0.010  -0.001 -0.01635   0.00062    ns
   7  0.008  0.020   0.019   0.010  -0.001 -0.00668   0.00047    ns
   8  0.000  0.018   0.017   0.006   0.001  0.00116   0.00034    ns
   9  0.009  0.007   0.011   0.004  -0.001 -0.00760   0.00014    ns
  10 -0.099  0.062   0.099   0.062  -0.003  0.06447   0.01366  .001
  11 -0.004  0.024   0.016   0.018   0.004  0.00615   0.00058    ns
  12  0.010  0.011   0.010   0.011   0.001 -0.00651   0.00023    ns
  13 -0.033  0.033   0.033   0.033   0.000  0.02292   0.00221   .05
  14 -0.015  0.008   0.015   0.008   0.001  0.01102   0.00027    ns
  15 -0.264  0.131   0.264   0.131   0.021  0.20153   0.08693  .001
  16  0.015  0.014   0.015   0.014   0.000 -0.01030   0.00042    ns
  17  0.021  0.009   0.021   0.009  -0.002 -0.01648   0.00054    ns
  18 -0.007  0.006   0.007   0.005   0.001  0.00528   0.00007    ns
  19  0.042  0.017   0.042   0.017  -0.003 -0.03189   0.00203    ns
  20 -0.136  0.052   0.136   0.052   0.009  0.10246   0.02130  .001
  21  0.001  0.006   0.006   0.001   0.000 -0.00074   0.00003    ns
  22 -0.029  0.027   0.033   0.022   0.004  0.02307   0.00155    ns
```

8. Summary statistics for true scores

File: C:\DFIT8\EX1.TXT

```
   40 -0.012  0.010   0.012   0.010   0.002  0.00983   0.00023    ns

                  True-F    True-R         D      CDIF       lDl
                  ------    ------         -      -----      ---

Mean              19.902    20.587  -0.68431   0.01275   0.68440
Variance          88.438    87.705   0.04172   0.00195   0.04160
Std. Dev.          9.404     9.365   0.20427   0.04411   0.20397

Correlation between focal & reference true scores:    0.99977
```

9. DTF related statistics including the chi-square test

10. CDIF items to be removed to achieve non-significant DTF.

File: C:\DFIT8\EX1.TXT

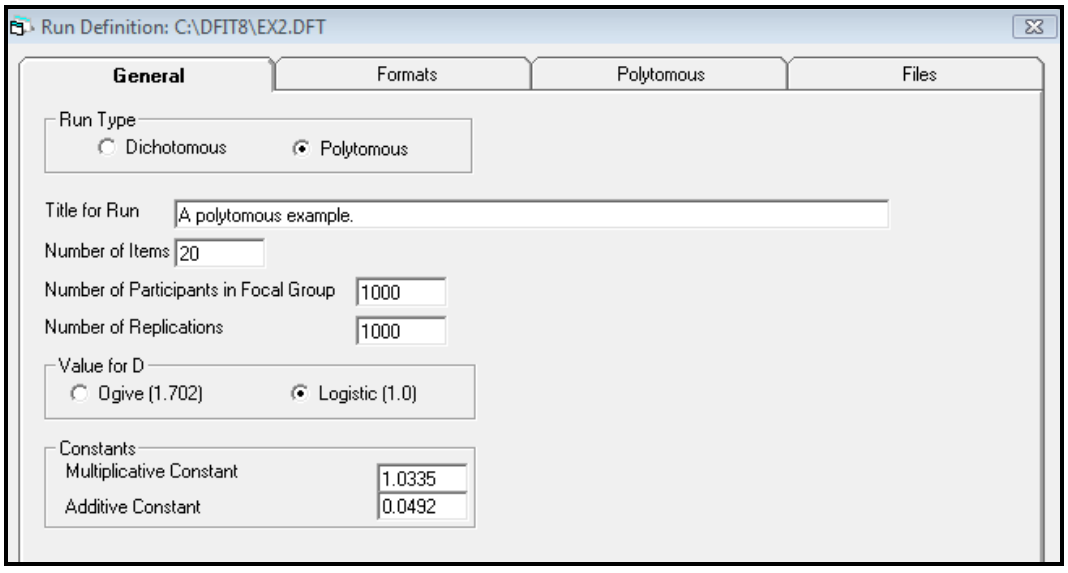
```
                     ------    ------         -      -----       ---

 Mean               19.902    20.587  -0.68431   0.01275   0.68440
 Variance           88.438    87.705   0.04172   0.00195   0.04160
 Std. Dev.           9.404     9.365   0.20427   0.04411   0.20397

 Correlation between focal & reference true scores:    0.99977

 Differential test functioning (DTF):  0.51001
                  Square-root of DTF:  0.71415
                          SD OF D**2:  0.25496

                    Chi-square value: 12223.32
                                Prob:   0.0000
   Degrees of freedom for chi-square:      999



                          Item Deletion Procedure A

                                                 DTF        Mean      Mean
Run      Item removed        DTF     Sig.     Category       (D)      (1D1)
---      ------------        ---     ----     --------     ------   --------

  1          None         0.51001    .001                -0.68431   0.68440

  2            15         0.19391    .001                -0.42094   0.42184

  3             5         0.04562      ns                -0.19983   0.20327
```

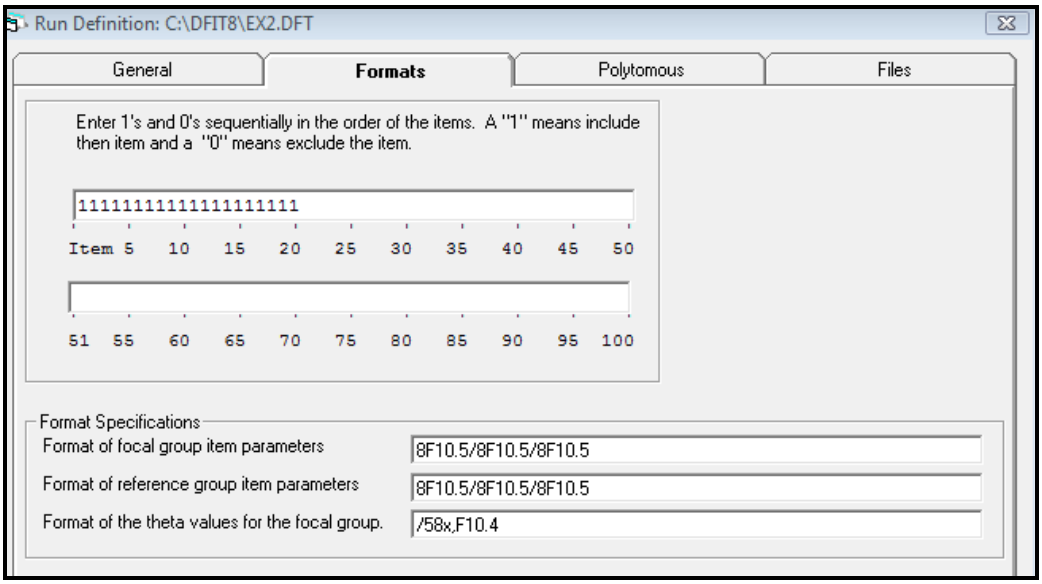
Results indicate that Items 5, 10, 15, and 20 show NCDIF at the .001 level. Item 13 shows NCDIF at the .05 level. The DTF of .51 is significant at the .001 level. After removing two significant CDIF items (Items 15 and 5), DTF is no longer significant.

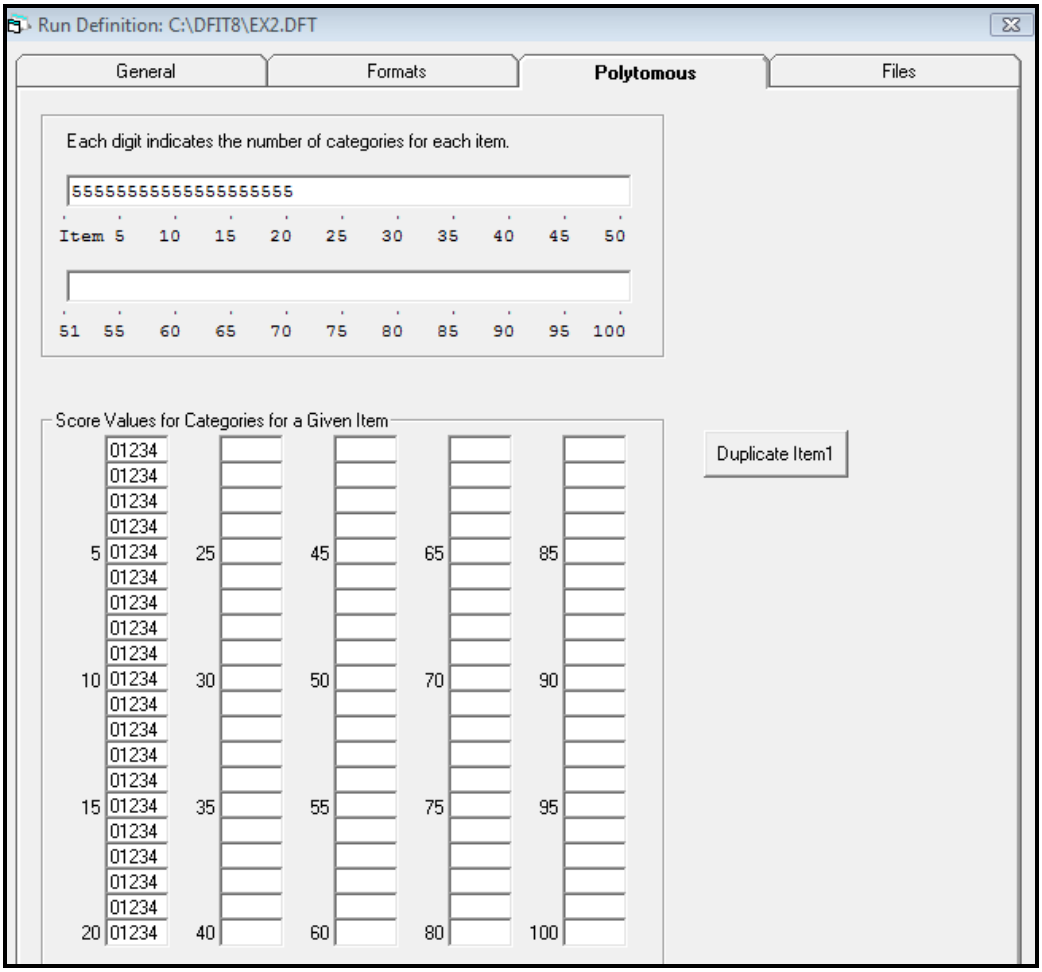
2. A polytomous example (Example 2)

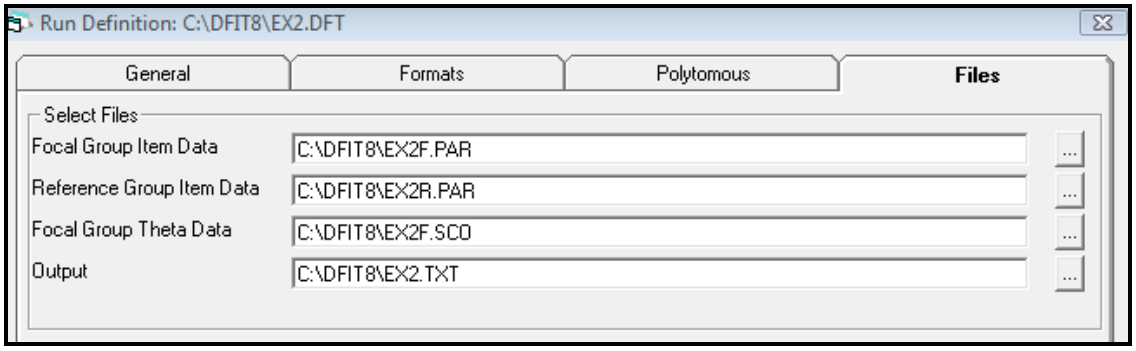
This is a 20-item simulated polytomous test (5-category responses scored 0, 1, 2, 3, 4) with 1,000 examinees in each group (reference and focal). DIF was embedded in Items 3, 8, 13, and 18. Items were calibrated separately for each group using PARSCALE (Muraki & Bock, 2003). The item parameter estimates and their variance covariance estimates are listed in EX2R.PAR and EX2F.PAR for the reference group and focal group, respectively. Those *.PAR files were created by POLYCOV which read the PARSCALE parameter files, added covariances, and arranged the information as required for DFIT input. In each file, there are 20 variables [a, b_1, b_2, b_3, b_4, V(a), V(b_1), V(b_2), V(b_3), V(b_4), Cov(a,b_1), Cov(a,b_2), Cov(a,b_3), Cov(a,b_4), Cov(b_1,b_2), Cov(b_1,b_3), Cov(b_1,b_4), Cov(b_2,b_3), Cov(b_2,b_4), and Cov(b_3,b_4) in that order]. The linking coefficients were calculated using EQUATE (1.0335 and 0.0492 for the multiplicative and additive coefficients, respectively, after the second-stage linking). In the score file (EX2F.SCO), the θ can be found at (/58x,F10.4). For the IPR method, 1000 replications were used to conduct the significance test.

Definition File in the Normal View

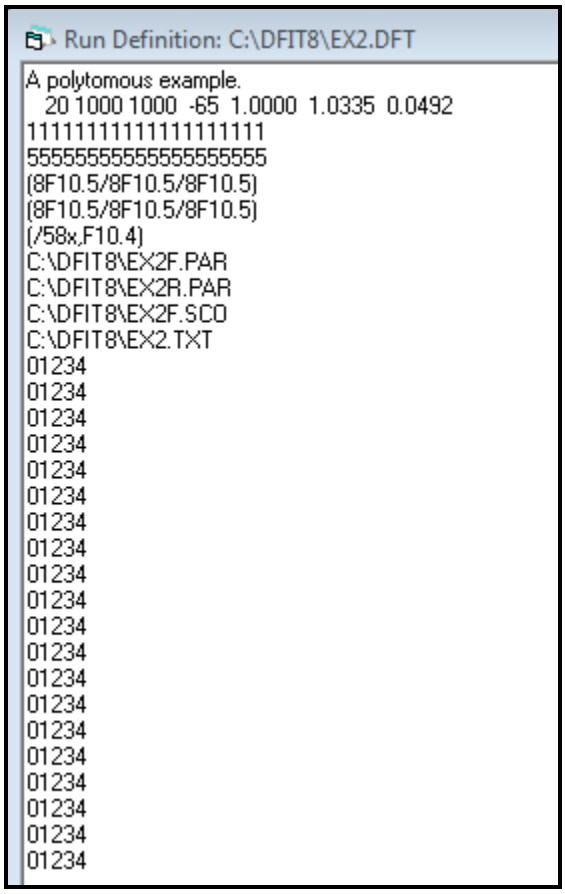



Run Definition: C:\DFIT8\EX2.DFT
General
Formats
Polytomous
Files
Enter 1's and 0's sequentially in the order of the items. A "1" means include then item and a "0" means exclude the item.
11111111111111111111
Item 5 10 15 20 25 30 35 40 45 50
51 55 60 65 70 75 80 85 90 95 100
Format Specifications
Format of focal group item parameters
8F10.5/8F10.5/8F10.5
Format of reference group item parameters
8F10.5/8F10.5/8F10.5
Format of the theta values for the focal group.
/58x,F10.4


Run Definition: C:\DFIT8\EX2.DFT
General
Formats
Polytomous
Files
Each digit indicates the number of categories for each item.
55555555555555555555
Item 5 10 15 20 25 30 35 40 45 50
51 55 60 65 70 75 80 85 90 95 100
Score Values for Categories for a Given Item
01234
Duplicate Item1
5 25 45 65 85
10 30 50 70 90
15 35 55 75 95
20 40 60 80 100



Definition File in the Raw View



EX2F.PAR

Run Definition: C:\DFIT8\EX2F.par

```
 0.55189 -1.77910 -0.69025  0.70902  1.85337  0.00128  0.01638  0.00683
 0.00690  0.01742  0.00327  0.00124 -0.00127 -0.00341  0.00711 -0.00126
-0.00731  0.00159 -0.00134  0.00729
 0.69564 -2.37868 -1.16347  0.12672  1.50311  0.00147  0.01806  0.00643
 0.00361  0.00884  0.00358  0.00167 -0.00020 -0.00226  0.00711  0.00069
-0.00490  0.00154 -0.00166  0.00214
 0.69400 -0.84214  0.37447  1.60908  2.74631  0.00155  0.00540  0.00384
 0.00953  0.02529  0.00140 -0.00051 -0.00246 -0.00444  0.00135 -0.00120
-0.00325  0.00289  0.00300  0.01122
 0.65985 -1.95057 -0.63769  0.64576  2.03350  0.00140  0.01376  0.00482
 0.00484  0.01479  0.00301  0.00096 -0.00097 -0.00316  0.00457 -0.00092
-0.00610  0.00126 -0.00101  0.00465
 0.64001 -1.35960 -0.11236  1.22009  2.54974  0.00140  0.00908  0.00414
 0.00779  0.02360  0.00221  0.00020 -0.00191 -0.00418  0.00265 -0.00181
```

EX2R.PAR

Run Definition: C:\DFIT8\EX2R.par

```
 0.50930 -1.88945 -0.72778  0.53368  1.87103  0.00123  0.02054  0.00808
 0.00716  0.02031  0.00372  0.00140 -0.00103 -0.00369  0.00884 -0.00056
-0.00943  0.00242 -0.00182  0.00727
 0.68177 -2.52672 -1.14592  0.13990  1.41861  0.00145  0.02098  0.00657
 0.00373  0.00855  0.00389  0.00168 -0.00023 -0.00218  0.00749  0.00060
-0.00520  0.00159 -0.00152  0.00234
 0.64920 -2.09475 -0.69691  0.69139  2.01866  0.00139  0.01584  0.00515
 0.00511  0.01499  0.00328  0.00105 -0.00104 -0.00317  0.00511 -0.00133
-0.00681  0.00110 -0.00125  0.00505
 0.67255 -2.00294 -0.66186  0.59783  1.92057  0.00142  0.01392  0.00474
 0.00455  0.01310  0.00304  0.00097 -0.00088 -0.00293  0.00451 -0.00078
-0.00564  0.00125 -0.00091  0.00419
 0.66343 -1.36376 -0.07894  1.18815  2.51284  0.00143  0.00857  0.00388
 0.00713  0.02176  0.00215  0.00015 -0.00180 -0.00399  0.00232 -0.00169
```

EX2F.SCO

Run Definition: C:\DFIT8\EX2F.SCO

```
5          |    1 GROUP 01    1.00
 1       1 |  1.00    1.80  1.00    -1.1592    0.2500
5          |    2 GROUP 01    1.00
 1       1 |  1.00    1.75  1.00    -1.2274    0.2507
5          |    3 GROUP 01    1.00
 1       1 |  1.00    3.35  1.00     0.5091    0.2462
5          |    4 GROUP 01    1.00
 1       1 |  1.00    1.55  1.00    -1.6724    0.2597
5          |    5 GROUP 01    1.00
 1       1 |  1.00    4.00  1.00     1.3623    0.2508
5          |    6 GROUP 01    1.00
 1       1 |  1.00    2.60  1.00    -0.4071    0.2449
5          |    7 GROUP 01    1.00
 1       1 |  1.00    2.05  1.00    -1.0901    0.2506
```

EX2.TXT (Selected output)

1. Title

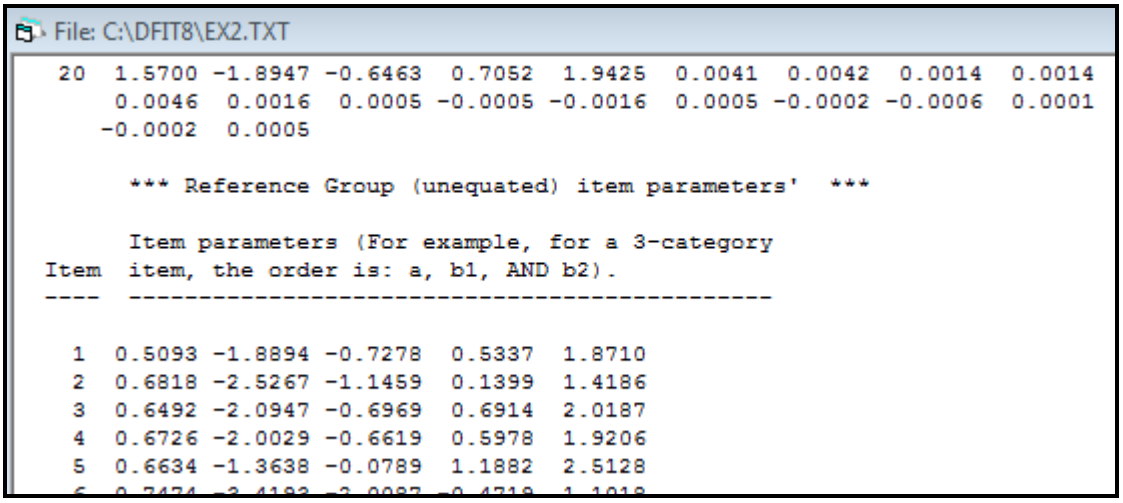
2. Input variables

3. Focal group item parameters

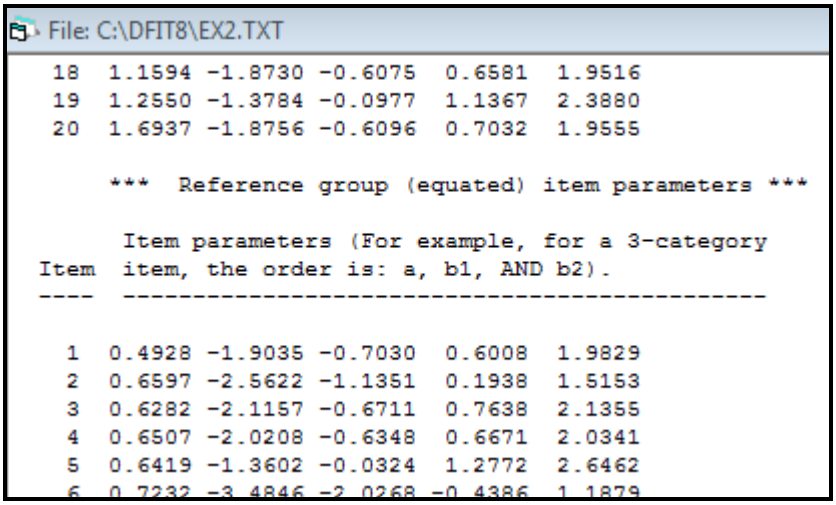
File: C:\DFIT8\EX2.TXT

```
A polytomous example.

 Number of Items:                          20
 Number of Subjects:                     1000
 Number of Replications:                 1000
 Seed for Generating Random Numbers:      -65
 Logistic or Normal Ogive (D):         1.0000
 Multiplicative Constant:              1.0335
 Additive Constant:                    0.0492

          ***** Focal group item parameters *****

       Item parameters, followed by variances and
       then by covariences. For example, for a
       3-category item, the order is: a, b1, b2, V(a),
 Item  V(b1), V(b2), Cov(a,b1), Cov(a,b2), and Cov(b1,b2).
 ----  ------------------------------------------------

   1  0.5519 -1.7791 -0.6902  0.7090  1.8534  0.0013  0.0164  0.0068  0.0069
      0.0174  0.0033  0.0012 -0.0013 -0.0034  0.0071 -0.0013 -0.0073  0.0016
     -0.0013  0.0073
   2  0.6956 -2.3787 -1.1635  0.1267  1.5031  0.0015  0.0181  0.0064  0.0036
      0.0088  0.0036  0.0017 -0.0002 -0.0023  0.0071  0.0007 -0.0049  0.0015
     -0.0017  0.0021
   3  0.6940 -0.8421  0.3745  1.6091  2.7463  0.0016  0.0054  0.0038  0.0095
      0.0253  0.0014 -0.0005 -0.0025 -0.0044  0.0014 -0.0012 -0.0033  0.0029
      0.0030  0.0112
   4  0.6599 -1.9506 -0.6377  0.6458  2.0335  0.0014  0.0138  0.0048  0.0048
      0.0148  0.0030  0.0010 -0.0010 -0.0032  0.0046 -0.0009 -0.0061  0.0013
```

4. Reference group (unequated) item parameters



5. Reference group (equated) item parameters



6. Monte Carlo generated item-level and test-level cut-offs (the IPR method)

File: C:\DFIT8\EX2.TXT

```
  18  1.1218 -1.8865 -0.5786  0.7294  2.0662
  19  1.2144 -1.3754 -0.0517  1.2240  2.5172
  20  1.6388 -1.8892 -0.5808  0.7760  2.0702


Monte Carlo generated item-level and test-level cut-offs


 Item .001 Lvl  .005 Lvl .01 Lvl .05 Lvl .10 Lvl .50 Lvl    Mean      SD
 ---- -------- --------- ------- ------- ------- -------    ----      --

    1  0.02208   0.01778 0.01376 0.00911 0.00699 0.00215 0.00312 0.00314
    2  0.01452   0.01180 0.01048 0.00736 0.00566 0.00170 0.00243 0.00237
    3  0.01847   0.01425 0.01266 0.00835 0.00615 0.00163 0.00266 0.00290
    4  0.01755   0.01377 0.01235 0.00692 0.00535 0.00161 0.00237 0.00250
    5  0.01855   0.01377 0.01210 0.00765 0.00563 0.00177 0.00254 0.00259
    6  0.01766   0.01093 0.01018 0.00643 0.00483 0.00140 0.00213 0.00226
```

7. Main output where CDIF, NCDIF, and NCDIF significance (*ns* indicates not-significant) are reported. DIF category is for future use and it is left blank.

File: C:\DFIT8\EX2.TXT

```
   20  0.00722   0.00571 0.00515 0.00330 0.00251 0.00084 0.00115 0.00110

  DTF  0.21806   0.18477 0.16902 0.12598 0.09857 0.03036 0.04245 0.04007


       Mean     SD    Mean    SD                                      DIF
Item    (d)     (d)   (ldl)   (ldl)  C(d,D)     CDIF      NCDIF  Sig. Category
----    ---     ---   -----   ------ ------    -----      -----  ---- --------

   1 -0.014  0.040   0.035   0.024  -0.012  0.00036   0.00181    ns
   2  0.006  0.022   0.020   0.012  -0.007 -0.01266   0.00054    ns
   3 -0.499  0.048   0.499   0.048  -0.008  0.44618   0.25178  .001
   4 -0.004  0.007   0.006   0.005  -0.002  0.00164   0.00006    ns
   5  0.028  0.004   0.028   0.004  -0.001 -0.02663   0.00080    ns
   6  0.006  0.029   0.025   0.016  -0.009 -0.01446   0.00090    ns
   7  0.046  0.024   0.047   0.024  -0.007 -0.04972   0.00276    ns
   8 -0.378  0.181   0.383   0.171   0.056  0.39962   0.17609  .001
   9 -0.004  0.041   0.035   0.022  -0.013 -0.00879   0.00169    ns
  10 -0.039  0.058   0.057   0.040  -0.018  0.01710   0.00490    ns
  11  0.014  0.018   0.019   0.012  -0.005 -0.01812   0.00050    ns
  12  0.034  0.043   0.046   0.030  -0.013 -0.04439   0.00302    ns
  13 -0.239  0.159   0.246   0.149   0.050  0.26715   0.08269  .001
  14  0.050  0.094   0.090   0.056   0.029 -0.01639   0.01124   .01
  15 -0.025  0.145   0.119   0.088   0.045  0.06740   0.02178  .001
  16  0.022  0.010   0.023   0.009   0.002 -0.01812   0.00059    ns
  17  0.023  0.048   0.043   0.031  -0.015 -0.03592   0.00281    ns
  18  0.010  0.076   0.063   0.045   0.024  0.01451   0.00592    ns
  19  0.007  0.024   0.022   0.012   0.007  0.00051   0.00063    ns
  20  0.047  0.010   0.047   0.010  -0.003 -0.04567   0.00234    ns
```

8. Summary statistics for true scores

File: C:\DFIT8\EX2.TXT

```
   16  0.022  0.010   0.023   0.009   0.002 -0.01812   0.00059    ns
   17  0.023  0.048   0.043   0.031  -0.015 -0.03592   0.00281    ns
   18  0.010  0.076   0.063   0.045   0.024  0.01451   0.00592    ns
   19  0.007  0.024   0.022   0.012   0.007  0.00051   0.00063    ns
   20  0.047  0.010   0.047   0.010  -0.003 -0.04567   0.00234    ns

                  True-F    True-R         D      CDIF       lDl
                  ------    ------         -     -----       ---

Mean               38.522    39.431  -0.90892   0.04618   0.90892
Variance          120.708   127.663   0.09744   0.02012   0.09744
Std. Dev.          10.987    11.299   0.31216   0.14185   0.31216

Correlation between focal & reference true scores:    1.00000
```

9. DTF related statistics including the chi-square test

10. CDF items to be removed to achieve non-significant DTF.

File: C:\DFIT8\EX2.TXT

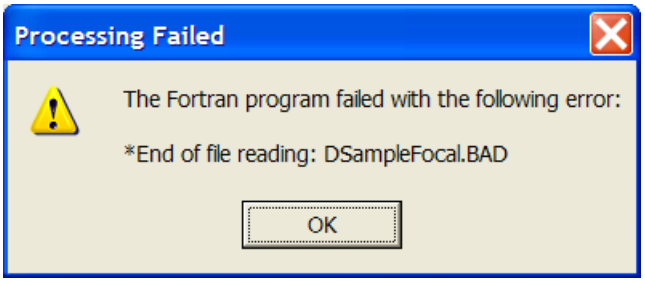
```
                  ------    ------         -     -----       ---

Mean               38.522    39.431  -0.90892   0.04618   0.90892
Variance          120.708   127.663   0.09744   0.02012   0.09744
Std. Dev.          10.987    11.299   0.31216   0.14185   0.31216

Correlation between focal & reference true scores:    1.00000

Differential test functioning (DTF):  0.92359
                 Square-root of DTF:  0.96103
                         SD OF D**2:  0.57450

                   Chi-square value:  9478.26
                               Prob:   0.0000
  Degrees of freedom for chi-square:      999



                        Item Deletion Procedure A

                                                DTF        Mean      Mean
Run      Item removed        DTF     Sig.     Category       (D)      (lDl)
---      ------------        ---     ----     --------    ------  --------

  1         None         0.92359     .001                -0.90892   0.90892

  2            3         0.28300     .001                -0.40944   0.43305

  3            8         0.03145       ns                -0.03106   0.12802
```

Results indicate that Items 3, 8, 13, and 15 show NCDIF at the .001 level. Item 14 shows NCDIF at the .01 level. The DTF of .92 is significant at the .001 level. After removing two significant CDIF items (Items 3 and 8), DTF is no longer significant.

Appendix B. Troubleshooting

Most problems that occur when running DFIT are a result of incorrect or poorly formatted data. If you open the file in Normal View, when you start the run DFIT will perform validation of the definition file.

The most likely error is an end of file error. The figure below shows a typical end of file error.



This can be caused by several different errors in the data files:

1. Insufficient data in the data file
2. Incorrect format specification in the run definition file
3. Incorrect number of items or examinees specified on line 2 of the run definition file