DOCUMENT RESUME

ED 401 319 TM 025 892

AUTHOR	Flowers, Claudia P.; And Others
TITLE	A Description and Demonstration of the Polytomous-DFIT Framework.
PUB DATE	Apr 96
NOTE	48p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Item Response Theory; *Mathematical Models; Simulation; *Test Bias; Test Length
IDENTIFIERS	Item Bias Detection; Parametric Analysis; *Polytomous Variables

ABSTRACT

N. S. Raju, W. J. van der Linden, and P. F. Fleer (in press) have proposed an item response theory-based, parametric procedure for the detection of differential item functioning (DIF)/differential test functioning (DTF) known as differential functioning of item and test (DFIT). DFIT can be used with dichotomous, polytomous, or multidimensional data. This study describes and provides a simulated demonstration of the polytomous-DFIT framework. Factors manipulated in the simulation were: (1) length of test (20 and 40 items); (2) focal group distribution; (3) number of DIF items; (4) direction of DIF; and (5) type of DIF. The DFIT framework was effective in identifying DTF and DIF in polytomously scored data for the conditions simulated. The preliminary findings provide promising results and indicate directions for future research. (Contains 3 figures, 6 tables, and 21 references.) (Author/SLD)

A Description and Demonstration of the Polytomous-DFIT Framework

Claudia P. Flowers

University of North Carolina at Charlotte

Chris Oshima

Georgia State University

Nambury Raju

Georgia Institute of Technology

Correspondence should be sent to:

Claudia Flowers
Dept of Teaching Specialties
UNC Charlotte
9201 University City Blvd
Charlotte, NC 28223

Running Head: Polytomous-DFIT

A Description and Demonstration of the Polytomous-DFIT Framework

ABSTRACT

Raju, van der Linden, and Fleer (in press) have proposed an IRT-based, parametric DIF/DTF procedure known as differential functioning of item and test (DFIT). DFIT can be used with dichotomous, polytomous, or multidimensional data. This study describes and provides a simulated demonstration of the polytomous-DFIT framework. Factors manipulated in the simulation were (a) length of test (20 and 40 item) (b) Focal Group distribution (c) number of DIF items (d) direction of DIF and (e) type of DIF. The preliminary findings provided promising results and indicated directions for future research.

Index terms and phrases: *Differential item functioning (DIF), Differential test functioning (DTF), Differential functioning of items and test (DFIT), IRT, Polytomous data, Unidiminsionality, Simulation*

A DESCRIPTION AND DEMONSTRATION OF THE POLYTOMOUS-DFIT FRAMEWORK

Differential Test Functioning (DTF) and Differential Item Functioning (DIF) research has focused primarily on dichotomously-scored items and test. With the increased use of polytomously-scored items and evidence of greater discrepancy in ethnic groups' performance using performance-based assessment (Dunbar, Koretz, & Hoover, 1991; Zwick, Donoghue, & Grima, 1993), there has been increased interest in polytomous DIF/DTF procedures. A new IRT-based, parametric procedure proposed by Raju, van der Linden, and Fleer (in press), known as differential functioning of item and test (DFIT), can be used with dichotomous, polytomous, or multidimensional data.

The DFIT framework has many useful features for test developers. First, it is the only parametric IRT-based, psychometric measure of differential functioning at both the test and item levels. When IRT is used to develop tests, IRT-based DIF/DTF procedures that use item parameter estimates, such as DFIT, maintain a common framework in test development. Second, DFIT has an index that does not assume that all items in the test, other than the one under study, are unbiased. Third, during the development phase DFIT provides an additional tool for determining the overall effect of eliminating an item from a test. Fourth, DFIT allows examining DIF/DTF in a mixed test format such as a combination of polytomous and dichotomous items. Finally, DFIT allows flexibility in examining potential bias in tests.

Raju et al. (in press) offered an empirical demonstration of DFIT using dichotomous data, and Oshima, Raju, and Flowers (1993) demonstrated the multidimensional case. This study describes and provides a simulated demonstration of the polytomous-DFIT framework.

Polytomous-DFIT

As with the dichotomous models, many polytomous models exist, such as Samejima's (1969) graded response model; Master's (1982) partial credit model; the rating scale model (Andrich 1978); the nominal response model (Bock, 1972); the generalized partial credit model (Muraki, 1992); and the free-response model (Samejima, 1972). Even though the DFIT framework can be used with any polytomous model, this study will use Samejima's graded response model to describe and demonstrate the polytomous-DFIT framework.

Samejima's graded response model (1969) assumes an ordered response; that is, the more steps successfully completed, the larger the category score. Higher category scores indicate a greater ability. In the graded response model, the probability of person s responding in or above category k to item i is:

$$P^{+}_{ik}(\theta) = \frac{\exp[Da_i(\theta_s - b_{ik})]}{1 + \exp[Da_i(\theta_s - b_{ik})]} \qquad (1)$$

where b_{ik} is the boundary or threshold between category k and k-1 associated with item i; a_i is the item slope or discrimination parameter; and θ_s is the ability parameter. This equation is similar to the two-parameter dichotomous model except that more than one function is needed per item. For each item the number of functions is one less than the number of categories. The item discrimination parameter, a, is constant across all categories in an item but varies across items in a test. This results in all category characteristic curves (CCC) having equal slopes for each category in an item which ensures no crossing of the curves. For each item, multiple difficulty parameters, b, are required. The number of b-parameters is one less than the number of categories.

To calculate the probability of responding in a particular category, the adjacent category is subtracted from the cumulative probability. This can be expressed as

$$P_{ik}(\theta) = P^{+}_{ik}(\theta) - P^{+}_{i,k+1}(\theta) \ . \qquad (2)$$

This function is often called the item category response function (ICRF). Because the first and last categories lack an adjacent category, Samejima (1969) defined $P^{+}_{i0}(\theta)$ and $P^{+}_{im}(\theta)$ as

$$P^{+}_{i0}(\theta) = 1 \qquad (3)$$

and

$$P^{+}_{im}(\theta) = 0 \;, \qquad (4)$$

where m equals the number of categories. The probability of responding in the first category for item i is

$$P_{i1}(\theta) = P^{+}_{i0}(\theta) - P^{+}_{i1}(\theta) = 1 - P^{+}_{i1}(\theta) \;. \qquad (5)$$

The probability of responding in the last category for item i is

$$P_{im}(\theta) = P^{+}_{i,m-1}(\theta) - P^{+}_{im}(\theta) = P^{+}_{i,m-1}(\theta) \;. \qquad (6)$$

The number of ICRFs per item is equal to the number of categories.

After the probability for responding in each category is estimated, a measure of the item expected score can be calculated. Raju et al. (in press) suggests that for polytomously-scored data an expected score (ES_{si}) for item i can be computed for examinee s as

$$ES_{si} = \sum_{k=1}^{m} P_{ik}(\theta_s) X_{ik} \qquad (7)$$

where X_{ik} is the score for category k; m is the number of categories; and P_{ik} is the probability of responding to category k (see Equation 2). This is referred to as the item true score function (ITSF). Summing the expected item scores across a test will result in the true test score function for each examinee as

$$T_s = \sum_{i=1}^{n} ES_{si} \tag{8}$$

where n is the number of items in the test. Once the true item and test scores are known then the DFIT for the polytomous framework is identical to the DFIT framework for the dichotomous case.

DFIT framework requires two item expected scores (ES) and two true test scores (T) to be calculated for each Focal Group examinee (i.e., the group of interest). If a single examinee is a member of the Focal Group (F), an expected score (see Equation 7) for an item, ES_{siF}, can be calculated. If the same examinee is treated as a member of the Reference Group (R) (i.e., comparison group), then an expected score, ES_{siR}, can be calculated as if examinee s were a member of the Reference Group. If the item is functioning differentially, the two expected scores would not be equal.

The same reasoning can be applied at the test level. The true test score (see Equation 8), T_s, is calculated by summing the ES_{si} across all the items in the test. Two true test scores can be calculated for each Focal Group examinee: one true score for the examinee as a member of the Focal Group (T_{sF}) and one as if he or she were a member of the Reference Group (T_{sR}). The greater the difference between the two true scores, the greater the DTF. According to Raju et al. (in press), a measure of DTF at the examinee level may be defined as

$$D_s^2 = (T_{sF} - T_{sR})^2 . \tag{9}$$

DTF across examinees may be defined as

$$DTF = \epsilon (T_{sF} - T_{sR})^2 \tag{10}$$

where ϵ stands for expectation. If the expectation is taken over the Focal Group examinees, then

$$DTF = \underset{F}{\epsilon} (T_{sF} - T_{sR})^2 . \tag{11}$$

Using the definition in Equation 9, Equation 11 may be rewritten as

$$DTF = \underset{F}{\epsilon} D_s^2 \tag{12}$$

in which case θ could be integrated out of the function by

$$DTF = \int_{\theta} D_s^2 f_F(\theta)\, d\theta \tag{13}$$

where $f_F(\theta)$ is the density function of θ in the Focal Group. Then

$$DTF = \sigma_D^2 + (\mu_{TF} - \mu_{TR})^2 = \sigma_D^2 + \mu_D^2 \tag{14}$$

where μ_{TF} is the mean true score for the Focal Group examinees; μ_{TR} is the mean true score for the same examinees as if they were members of the Reference Group; and σ_D^2 is the variance of D.

Differential functioning at the item level can be derived from Equation 11. If

$$d_{si} = ES_{siF} - ES_{siR} \tag{15}$$

then

$$DTF = \epsilon[(\sum_{i=1}^{n} d_{si})^2] \tag{16}$$

where n is the number of items in a test. This can be rewritten as

$$DTF = \sum_{i=1}^{n} [Cov(d_i, D) + \mu_{d_i}\mu_D] \tag{17}$$

where $Cov(d_i,D)$ is the covariance of the difference in expected scores (d_i) and the difference in true scores (D), and μ_{di} and μ_D are the means of d_{is} and D_s, respectively. In this case DIF can be written as

$$DIF_i = \epsilon(d_i, D) = Cov(d_i, D) + \mu_{d_i}\mu_D. \tag{18}$$

Raju et al. (in press) refer to this DIF as compensatory DIF (C-DIF). If DIF in Equation 18 was expressed as C-DIF, then

$$DTF = \sum_{i=1}^{n} C\text{-}DIF_i \ . \tag{19}$$

The additive nature of DTF allows for possible cancellation at the test level. This occurs when one item displays DIF in favor of one group and another item displays DIF for the other group. This combination of DIF items will have a canceling effect on the overall DTF. The sum of the C-DIF indices reflects the net directionality. For practical applications, a test developer could examine the DTF, then determine which item needs to be eliminated based on its C-DIF value and its overall contribution to DTF.

Raju et al. (in press) proposed a second index, named NC-DIF, that assumes that all items other than the one under study are free from differential functioning. In the dichotomous case, NC-DIF is closely related to other existing DIF indices such as Lord's chi-square and the unsigned area. If all other items are DIF free, then $d_j = 0$ for all $j \neq i$ where i is the item being studied and Equation 18 can be rewritten as

$$NC\text{-}DIF_i = \sigma_{d_i}^{\ 2} + \mu_{d_i}^{\ 2}. \tag{20}$$

Raju et al. (in press) noted that items having significant NC-DIF do not necessarily have significant C-DIF in the sense of contributing significantly to DTF. For example, if one item favors the Reference Group and another item favors the Focal Group, significant NC-DIF occurs for both items even though the two C-DIF indices may not be significant because of their canceling effect at the test level. This will often lead to a greater number of significant NC-DIF items than C-DIF items.

In addition to cancellation at the test level, polytomously-scored items allow for potential cancellation at the item level within a person. Cancellation at the item level within a person is only possible using polytomously-scored items. Because each item has multiple categories in the polytomous case, which leads to multiple probabilities, there is a possibility that one category may cancel the effects in another category when computing d_i for a given examinee. For example, if the Focal Group-based P_{1i} is greater than the Reference Group-based P_{1i} but the Focal Group-based P_{2i} is less than the the Reference Group-based P_{2i}, a cancellation will occur, keeping d_i close to zero, thereby indicating no differential functioning at the item level within a person. Figure 1 provides a visual displays of DIF cancellation for a three category response item.

Insert Figure 1 about here

The degree of cancellation is dependent on several factors. First, location and shape of the Focal Group distribution, which is used to weight DFIT values, would determine which areas of the IRF is emphasized. In other words, if more of the Focal Group members were located in the area where the categories changed in direction of DIF, more cancellation would occur. Second, the a-parameter values, which determine the slope of the IRF, influences the difference between the probabilities for the Focal and Reference Groups. That is, all other things being equal, high a-parameter values tend to have smaller differences between the Focal and Reference Group probabilities. Figure 2 displays two nonuniform DIF items (with 3 categories) with a .5 difference between the a-parameters for the Focal and Reference Groups. The only difference between the figures is one nonuniform DIF item has greater a-parameter values that the other DIF item. Finally, the distance between the b-parameters for each category will determine the amount of overlap. All these factors can interact in different ways to create situations where there is more or less cancellation.

Insert Figure 2 about here

DFIT Significance Test

To help in the decision making, statistical significance testing can be performed. Assume that the difference (D) between the true scores is normally distributed with a mean of μ_D and a standard deviation of σ_D. A Z score for examinee s is

$$Z_s = \frac{D_s - \mu_D}{\sigma_D} \tag{21}$$

where Z_s^2 has a chi-square distribution with 1 degree of freedom. The sum of Z_s^2 across N examinees has a chi-square distribution with N degrees of freedom:

$$\chi_N^2 = \Sigma Z_s^2 = \frac{\Sigma (D_s - \mu_D)^2}{\sigma_D^2}. \tag{22}$$

If $\epsilon(DTF) = \mu_D^2 = 0$, then by substitution

$$\chi_N^2 = \frac{\Sigma D_s^2}{\sigma_D^2} = \frac{N(DTF)}{\sigma_D^2}. \tag{23}$$

If an unbiased estimator is substituted for σ_D^2 then

$$\chi_{N-1}^{2} = \frac{N(DTF)}{\hat{\sigma}_{D}^{2}}. \tag{24}$$

A significant chi-square value indicates that one or more items are functioning differentially. Raju et al. (in press) suggest removing items that contribute significantly to DTF until the chi-square value is no longer significant. According to Raju el al. (in press), items so deleted are designated as having significant C-DIF. Therefore, Raju et al. did not propose a separate significance test for C-DIF.

Raju et al. (in press) defined a similar chi-square test for NC-DIF. This test was shown to be overly sensitive for large sample sizes (Fleer, 1993). Fleer suggested empirically establishing a critical (cutoff) value for NC-DIF. This critical value was determined from a Monte Carlo study of non-DIF items.

Method

Data Simulation

A graded response model with five-response categories was used to generate the simulated data sets. Item parameters used in previous studies (Cohen & Kim, 1991; Fleer, 1993) were modified to accommodate the graded response model. The modified item parameters are contained in Tables 1 and 2.

Insert Tables 1 and 2 about here

Next, the item probabilities for five categories per item for a simulated examinee was generated using Equation 1. Recall that five categories result in four probabilities per item. In order to assign a score for each simulated examinee the following procedure was used. First, each simulated examinee was randomly assigned an ability parameter (θ) from a standard normal distribution. Using the item parameters in Tables 1 and 2 along with the randomly assigned ability parameter (θ), each simulated examinee has four probabilities per item. For example, using the item parameters for Item 1 in Table 1 and randomly assigning an ability parameter (θ) of 1.0, the following item probabilities (P^{+}_{sik}) are calculated for examinee *s* in category *k* on item *i*: P^{+}_{s11} = .932, P^{+}_{s12} = .817, P^{+}_{s13} = .592 and P^{+}_{s14} = .321. Next, for each simulated examinee a single random number (X) was sampled from a uniform distribution over the interval [0,1]. If the randomly sampled number was less than the calculated probability at the boundary category *k* but greater than the calculated probability at *k+1*, then the score assigned was the value of category k. This can be expressed as

$$P^{+}_{sik} > X_{si} > P^{+}_{si(k+1)} \quad (28)$$

where X_{si} is the single random number for examinee s on item i. In the example, if examinee s was assigned a single uniform random number of .853, then the simulated examinee is assigned a score of 1 because .853 is less than P^{+}_{s11} (.932) but greater than P^{+}_{s12} (.817). This example assumes that examinees can score either 0, 1, 2, 3, or 4.

Factors Manipulated

Two different ability distributions were simulated for the Focal Group. In the first condition the Focal and Reference Groups had equal ability distributions. That is, the ability parameter for each group was randomly selected from a N(0,1) distribution. This condition is referred to as the "no impact" condition. In the second condition, the Focal Group was sampled from a N(-1,1) distribution resulting in a lower ability level than that in the Reference Group. This condition is referred to as the "impact" condition.

Two test lengths, 20 and 40 items, were simulated in this study. Sample size and scoring options were constant in this study. One thousand examinees for each group, Focal and Reference, were simulated. This sample size ensures adequate precision for parameter estimations (Muraki & Bock, 1993) prior to DIF/DTF analyses. All items consisted of five scoring options (i.e., 0, 1, 2, 3, and 4). Each condition will be evaluated on five replications.

Four proportions of test-wide DIF (0%, 5%, 10%, and 20%) and two conditions of direction of DIF (unidirectional and balanced-bidirectional) were simulated. In the 20-item test, 0, 1, 2, 3, and 4 items were embedded with DIF. In the unidirectional conditions, all items favored the Reference Group. In the balanced-bidirectional conditions, items favoring the Reference Group were perfectly balanced with items favoring the Focal Group. In the 5% condition, which has one DIF item, the bidirectional condition could not simulated. In addition, items were generated to simulate uniform DIF (for which $a_{iR} = a_{iF}$ and $b_{iR} \neq b_{iF}$) and nonuniform DIF (for which $a_{iR} \neq a_{iF}$ either with $b_{iR} \neq b_{iF}$ or $b_{iR} = b_{iF}$). Only the 20% DIF condition contains nonuniform DIF items. In this condition, two nonuniform DIF and two uniform DIF items were embedded.

Similar conditions were simulated in the 40-item test. DIF was embedded in 0, 2, 4, and 8 items. Directional and balanced-bidirectional DIF was simulated using the same method as the 20-item test. Nonuniform DIF was embedded only in the 20% DIF condition. The true item parameters for the DIF items are contained in Tables 1 and 2. Figure 3 provides a visual display of the simulation design.

Insert Figure 3 about here

Parameter Estimations and Linking Method

Item and ability parameters were estimated using the computer program PARSCALE 2 (Muraki & Bock, 1993). The maximum marginal likelihood procedure and EM algorithm were used to estimate the item parameters. Default values were used for all estimations. Estimation of underlying abilities were made using Bayesian EAP procedure which incorporates normal priors.

The estimation of equating coefficients was made by means of Baker's modified test characteristic curve method as implemented by the EQUATE 2.0 computer program (Baker, 1993). In this study, all parameter estimates for the Reference Group were equated to the underlying metric of the Focal Group.

Several researchers (Lord, 1980; Drasgow, 1987; Candell & Drasgow, 1988; Lautenschlager & Park, 1988; Miller & Oshima, 1992) have shown that an iterative linking procedure improves identification of DIF items. To minimize error introduced by the equating procedure, a two-stage linking procedure was used in this study. After the initial linking with all test items, a DIF analysis was performed. If items were identified as displaying DIF, as indicated by an NC-DIF index that exceeded the critical value, the linking procedure was performed again without these

DIF items. Finally, all items were transformed using the linking coefficients obtained in the second iteration. A Fortran program written by Raju (1995) was used to calculate the DFIT indices.

Results

Before the DFIT procedure was applied to the simulated data, a recovery analysis was undertaken. Two indices were used to examine the item parameter recovery; a correlation coefficient (i.e., true parameters with estimated parameters) and RMSD. The recovery analyses results indicated an acceptable recovery of the underlying item parameters (i.e., high correlation coefficients and low RMSDs). None of the data sets had extreme results to warrant exclusion from the DTF/DIF analyses.

Establishing Critical Values

As mentioned previously, the chi-square value for NC-DIF was shown to be overly sensitive for large samples sizes. To protect against a Type I error, an empirical critical value was established for all DIF indices. Two thousand DIF-free items were simulated and DIF analyses were conducted. An alternative cutoff was established by finding the value at the 99th percentile. This resulted in an alternative cutoff value of .016.

Detection of DIF

Two indicators were calculated to determine the accuracy of DIF detection: true positive (TP) and false positive (FP). A true positive is an embedded DIF item with a DIF index value that exceeds the cutoff value; conversely, a false positive is a non-DIF item with a DIF index value that exceeds the criterion established for DIF. High true positive values (i.e., close to 1) and low false positive values (i.e., close to 0) are desirable for DTF/DIF indices.

Additional analyses were conducted using true item parameters to calculate C-DIF and NC-DIF. These analyses bypassed the PARSCALE estimations and linking procedure and are referred to as "True" conditions. "True" conditions consist of one analysis per condition as opposed to the "Estimated" conditions that consist of five replications per condition. The "True" conditions are reported first and used as the standard to which the "Estimated" conditions are compared.

Comparisons should not be made across conditions because of confounding factors. That is, not only does the number of DIF items change across conditions but the magnitude of DIF (a difference of 1.0 or .5 between the b-parameters) and the type of DIF (uniform and nonuniform) are not consistent across conditions. The discrepancy between the "True" and the "Estimated" conditions should be the focus for comparisons.

C-DIF Results

Items with significant C-DIF were identified by using a chi-square test (at the .01 level of significance) or a cutoff value of .016. Items were removed one at a time until a nonsignificant DTF or a value less than .016 was obtained. Items that were removed to achieve either of these criteria were classified as having significant C-DIF. Recall that C-DIF values are summed across the entire test. The balanced-bidirectional tests should not have any items identified as DIF because of C-DIF cancellation; therefore, true positives are relevant only in the 20 and 40-item unidirectional conditions (Conditions 1, 2, and 3). Tables 3 and 4 contain the results at the condition level and item level for DFIT analyses in terms of identifying C-DIF items.

Insert Tables 3 and 4 about here

C-DIF "True" conditions. For the 20-item conditions, all items with significant C-DFI were identified except in Condition 3. In Condition 3, .75 of the true C-DIF items were detected (see Table 3). Item level results indicated that all uniform DIF items and nonuniform DIF items with differences in the b-parameters were detected; whereas, the nonuniform DIF item with differences in only the a-parameters (Item 18) was not detected. No false positives were detected in any of the conditions.

Similar results were obtained in the 40-item conditions. Again, all significant C-DIF items were identified except in Condition 3. Again items with differences in b-parameters were detected but items with differences in only the a-parameter (Items 20 and 40) were not detected. No false positives were observed.

C-DIF "Estimated" conditions. In the "Estimated" 20-item/no impact conditions there was a decrease for the true positives in Conditions 2 and 3 as compared to the "True" parameter conditions. In Condition 2, the true positive rate decreased from 1.00 to .90 and in Condition 3, the true positive rate dropped from .75 to .65 (see Table 3). In addition to nonuniform DIF not being detected, several of the uniform DIF items were not detected in either Condition 2 or Condition 3. Additionally, the false positive rates increased in Conditions 2 and 3. In Condition 2, the false positive rate increased slightly from .00 to .03. In Condition 3, the false positive rate had a larger increase from .00 to .18. This was due to two repetitions within this condition that identified 4 and 6 non-DIF items. The remaining three repetitions identified none or one false positive item.

For the 20-item/impact conditions, the results are identical to the 20-item/no impact conditions except for the false positive rate in Condition 3. A lower false positive rate (.03) was detected in the impact condition compared to the no impact condition (.18).

A similar trend was observed in the 40-item conditions. In the 40-item/no impact conditions, the true positive rates decreased in both Conditions 2 and 3. The true positive rate decreased from 1.00 to .80 and from .75 to .68 for Conditions 2 and 3, respectively. The item-level analyses revealed that all nonuniform and several uniform DIF were not detected. The false positive rates increased slightly in almost all conditions.

The 40-item/impact conditions had similar results to the 40-item/no impact conditions except for two instances. In Condition 2, the true positive rate decreased from .80 to .50. Due to such a substantial decrease in detection rate, an additional five repetitions were simulated. The results of the additional repetitions were similar to the finding in the 40-item/no impact condition. For the additional repetitions in this condition the true positive rate was .80 and the false positive rate was .03.

NC-DIF Results

True positives and false positives were determined by NC-DIF values that exceeded .016. Tables 5 and 6 contain the results of the true positives and false positives for NC-DIF. Recall that the "True" conditions bypass item parameter estimations and linking procedures and are used as a standard for evaluating the "Estimated" conditions.

Insert Tables 5 and 6 about here

<u>NC-DIF "True" conditions</u>. In the "True" 20-item conditions, the true positive rate was 1.0 except for Conditions 3 and 5 which had a true positive rate of .75 and .50, respectively. Analyses at the item level revealed that the DIF items not detected were Item 18 (Condition 3) and Items 3 and 4 (Condition 5). All of these items are nonuniform DIF items with differences in only the a-parameters. No false positive items were detected.

For the "True" 40-item conditions, all conditions had perfect true positive detection rates except Conditions 3 and 6. In Condition 3, the true positive detection rate was .88 and in Condition 6, the true positive rate was .75. In all conditions uniform DIF items were detected. In Condition 3, Item 20, a nonuniform DIF item, was detected whereas Item 40, another nonuniform DIF item, was not detected. The only difference between these items' characteristics was that Item 20 had a lower a-parameter (Reference Group = 1.00 and Focal Group = 0.50) as compared to Item 40 (Reference Group = 1.80 and Focal Group = 1.30). In Condition 6, two nonuniform DIF items were detected (Items 15 and 16) and two nonuniform DIF items were not detected (Items 5 and 6). Again, the discrimination parameters were lower for Items 15 and 16 than for Items 5 and 6. No false positives were detected.

NC-DIF "Estimated" conditions. The results of the "Estimated" conditions are similar to the "True" conditions. In the 20-item/no impact conditions, the results were identical to the "True" conditions except in Condition 1 where the false positive rate slightly increased from .00 to .01.

In the 20-item/impact case, Conditions 3 and 5 showed a slight increase in the true positive rates, from .75 and .50 to .80 and .55 , respectively.

In the 40-item/no impact condition, the "Estimated" conditions were similar to the "True" conditions. There was a slight decrease in true positive detection rate in Condition 6, from .75 to .70. There was also a slight increase in false positive rates in Conditions 3 and 6, from .00 to .01.

For the 40-item/impact case, the results were identical to the "True" condition except in Condition 6 where the true positive detection rate increased from .75 to .80. Additionally, the false positive rates in Conditions 1 and 2 increased slightly, from .00 to .01 for both conditions.

Conclusions

The DFIT framework was effective in identifying DTF and DIF in polytomously-scored data for the conditions simulated. Test length (20 and 40 items), Focal Group distribution (no impact and impact), number of DIF items (0%, 5%, 10%, and 20%), and direction of DIF (unidirectional and balanced-bidirectional) had little effect on the true positive and false positive detection rates across all conditions.

As expected, the type of DIF (uniform and nonuniform) affected the detection of DIF in the DFIT framework. Both indices, C-DIF and NC-DIF, successfully identified DIF items with differences in the b-parameters. However, nonuniform DIF items with higher a-parameters were not detect whereas lower a-parameter items were detected. As mentioned previously, the lower a-parameter items tend to result in greater differences between the Focal and Reference Groups.

Overall, C-DIF was not as stable as NC-DIF. This finding is similar to the findings of the unidimensional (Fleer, 1993) and multidimensional-dichotomous (Oshima, Raju, & Flowers, 1993) cases. In this study, C-DIF had two conditions that varied from what was expected (40-item/impact, Condition 2 and 20-item/no impact, Condition 3). When additional simulations were performed, the results were consistent with the theoretical expectations. A possible explanation for the occasional erratic detection rate is that the estimation and linking errors associated with the "Estimated" conditions accumulate across the entire test. The calculation of DTF involves summing the C-DIF values across the entire test which includes all the errors related to each item.

For example, a linking error would magnify the error in the same direction throughout the test. If the linking additive component was overestimated by .2, then .2 would be added to each item which are then summed across the entire test. NC-DIF, which had stable results across all conditions, is calculated from information related to one item; consequently, this leads to much more stable results.

Limitations

While this study supports the validity of the polytomous-DFIT framework, the results are specific to the conditions simulated. In this study, the method in which DIF was embedded (i.e., placing differences in each category) may be unrealistic and provide optimal conditions for detecting DIF/DTF. This high detection rate created a ceiling effect that limited the investigation of the influence of factors that were manipulated in this study. Ability group distribution and values of the a and b-parameters should have an influence in the detection of DIF/DTF. The efficacy of the DFIT framework should be researched in more conditions with other IRT models.

Future Research

The findings in this study are preliminary and encourage future research areas for DFIT. First, critical (cutoff) values for C-DIF and NC-DIF need to be investigated. In this study, the critical value was established by using an empirical method which was optimal for the detection of DIF/DTF specific to this study. A Type-I and Type-II error simulation study should be performed. For DFIT to have practical use, critical values at various alpha levels with different IRT models need to be established.

The reason for the occasional instability of C-DIF needs to be determined. C-DIF offers a unique method for assessing the overall effect of removing or adding an item to a test.

Finally, many conditions need to be experimentally manipulated. Sample size, amount of DIF, length of test, distribution of Focal Group, and many other conditions need to be systematically investigated. Additionally, the DFIT framework should be applied to tests with mixed item formats (i.e., dichotomous and polytomous items). These systematic investigations would help establish guidelines and limitations of the DFIT procedure.

Summary

The preliminary findings of the polytomous-DFIT framework provided promising results and indicated directions for future research. The DFIT procedure provides unique tools for examining and interpreting DIF and DTF. The value of the DFIT will ultimately be determined by its adaptability for use in the practical setting.

References

Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43, 561-573.

Baker, F. B. (1993). EQUATE2: Computer program for equating two metrics in item response theory [computer program]. Madison:University of Wisconsin, Laboratory of Experimental Design.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29-51.

Candell, G. L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. Applied Psychological Measure, 12,253-260.

Cohen, A. S., & Kim, S. H. (1991). A comparison of Load's chi-square and Raju's area measures on detection of DIF. Paper presented at the International Symposium on Measurement, University of Ottawa, Ontario.

Drasgow, F. (1987). Study of the measurement of bias of two standardized psychological tests. Journal of Applied Psychology, 72(1), 19-29.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4, 289-303.

Fleer, P. F. (1993). A Monte Carlo assessment of a new measure of item and test bias. Unpublished PhD dissertation, Illinois Institute of Technology.

Guilford, J. P. (1954). Pschometric methods (2nd Ed.). New York: McGraw-Hill.

Lautenschlager, G. J. & Park, D. G. (1988). IRT item bias detection procedures: Issues of model misspecification, robustness, and parameter linking. Applied Psychological Measurement, 12, 365-376.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Erlbaum, Hillsdale, New Jersey.

Miller, M. D., & Oshima, T. C. (1992). Effect of sample size, number of biased items, and magnitude of bias on a two-stage item bias estimation method. Applied Psychological Measurement, 16, 381-388.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. Applied Psychological Measurement,14(1), 59-71.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16, 159-176.

Muraki, E., & Bock, R.D. (1993). PARSCALE2: IRT based test scoring and item analysis for graded open-ended exercises and performance task [computer program]. Chicago: Scientific Software International.

Oshima, T. C., Raju, N., & Flowers, C. P. (1993). Evaluation of DTF and DIF in two-dimensional IRT. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Raju, N. (1995). DFITPU: A Fortran program for calculating DIF/DTF. [computer program]. Georgia Institute of Technology.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (in press). IRT-based internal measure of differential functioning of items and tests. Applied Psychological Measurement.

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

Samejima, F. (1972). A general model for free-response data. Psychometrika Monograph, No. 18.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. Journal of Educational Measurement, 30(3), 233-251.

For further information and/or reprint requests contact Claudia Flowers at UNC Charlotte, Department of Teaching Specialties, Charlotte, NC 28223 or e-mail CPFlower@UNCCVM.UNCC.EDU

Table 1

Item Parameters for Generating 20-Item Test Conditions

(a) Unidirectional DIF

	Reference					Focal Condition 1 (5% DIF)					Focal Condition 2 (10% DIF)					Focal Condition 3 (20% DIF)				
Item	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4
1	0.55	-1.80	-0.60	0.60	1.80															
2	0.73	-2.32	-1.12	0.08	1.28															
3	0.73	-1.80	-0.60	0.60	1.80	0.73	-0.80	0.40	1.60	2.80	0.73	-1.30	-0.10	1.10	2.30	0.73	-0.80	0.40	1.60	2.80
4	0.73	-1.80	-0.60	0.60	1.80															
5	0.73	-1.28	-0.08	1.12	2.32															
6	1.00	-2.78	-1.58	-0.38	0.82															
7	1.00	-2.32	-1.12	0.08	1.28															
8	1.00	-2.32	-1.12	0.08	1.28						1.00	-1.32	-0.12	1.08	2.28	0.50	-1.82	-0.62	0.58	1.78
9	1.00	-1.80	-0.60	0.60	1.80															
10	1.00	-1.80	-0.60	0.60	1.80															
11	1.00	-1.80	-0.60	0.60	1.80															
12	1.00	-1.80	-0.60	0.60	1.80															
13	1.00	-1.28	-0.08	1.12	2.32											1.00	-0.78	0.42	1.62	2.82
14	1.00	-1.28	-0.08	1.12	2.32															
15	1.00	-0.82	0.38	1.58	2.78															
16	1.36	-2.32	-1.12	0.08	1.28															
17	1.36	-1.80	-0.60	0.60	1.80															
18	1.36	-1.80	-0.60	0.60	1.80											0.86	-1.80	-0.60	0.60	1.80
19	1.36	-1.28	-0.08	1.12	2.32															
20	1.80	-1.80	-0.60	0.60	1.80															

(continued)

Table 1 (Continued)

Item Parameters for Generating 20-Item Test Conditions

(b) Balanced-Bidirectional DIF

	Reference					Focal Condition 4 (10% DIF)					Focal Condition 5 (20% DIF)				
Item	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4
1	0.55	-1.80	-0.60	0.60	1.80										
2	0.73	-2.32	-1.12	0.08	1.28										
3	0.73	-1.80	-0.60	0.60	1.80	0.73	-1.30	-0.10	1.10	2.30	1.23	-1.80	-0.60	0.60	1.80
4	0.73	-1.30	-0.10	1.10	2.30	0.73	-1.80	-0.60	0.60	1.80	----	----	----	----	----
*4	1.23	-1.80	-0.60	0.60	1.80	----	----	----	----	----	0.73	-1.80	-0.60	0.60	1.80
5	0.73	-1.80	-0.60	0.60	1.80										
6	0.73	-1.28	-0.08	1.12	2.32										
7	1.00	-2.78	-1.58	-0.38	0.82										
8	1.00	-2.32	-1.12	0.08	1.28										
9	1.00	-2.32	-1.12	0.08	1.28										
10	1.00	-2.07	-0.87	0.33	1.53						----	----	----	----	----
*10	1.00	-1.80	-0.60	0.60	1.80	----	----	----	----	----					
11	1.00	-1.80	-0.60	0.60	1.80										
12	1.00	-1.80	-0.60	0.60	1.80						----	----	----	----	----
*12	1.00	-1.28	-0.08	1.12	2.32	----	----	----	----	----	1.00	-0.78	0.42	1.62	2.82
13	1.00	-1.80	-0.60	0.60	1.80						----	----	----	----	----
*13	1.00	-0.78	0.42	1.62	2.82	----	----	----	----	----	1.00	-1.28	-0.08	1.12	2.32
14	1.00	-1.28	-0.08	1.12	2.32										
15	1.00	-1.28	-0.08	1.12	2.32						----	----	----	----	----
*15	1.00	-0.82	0.38	1.58	2.78	----	----	----	----	----					
16	1.00	-0.82	0.38	1.58	2.78						----	----	----	----	----
*16	1.00	-0.32	0.88	2.08	3.28	----	----	----	----	----					
17	1.36	-2.32	-1.12	0.08	1.28										
18	1.36	-1.80	-0.60	0.60	1.80										
19	1.36	-1.28	-0.08	1.12	2.32										
20	1.80	-1.80	-0.60	0.60	1.80										

Note. * indicate Reference Group parameters used in Condition 5 (20% DIF). Blanks indicate use of the same item parameters listed for the Reference Group.

BEST COPY AVAILABLE

Table 2

Item Parameters for Generating 40-Item Tests

(a) Unidirectional DIF

	Reference					Focal Condition 1 (5% DIF)					Focal Condition 2 (10% DIF)					Focal Condition 3 (20% DIF)				
Item	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4
1	0.55	-1.80	-0.60	0.60	1.80															
2	0.55	-1.80	-0.60	0.60	1.80															
3	0.73	-2.32	-1.12	0.08	1.28															
4	0.73	-2.32	-1.12	0.08	1.28															
5	0.73	-1.80	-0.60	0.60	1.80	0.73	-0.80	0.40	1.60	2.80	0.73	-0.80	0.40	1.60	2.80	0.73	-0.80	0.40	1.60	2.80
6	0.73	-1.80	-0.60	0.60	1.80															
7	0.73	-1.80	-0.60	0.60	1.80															
8	0.73	-1.80	-0.60	0.60	1.80															
9	0.73	-1.28	-0.08	1.12	2.32															
10	0.73	-1.28	-0.08	1.12	2.32	0.73	-0.28	0.92	2.12	3.32	0.73	-0.78	0.42	1.62	2.82	0.73	-0.78	0.42	1.62	2.82
11	1.00	-2.78	-1.58	-0.38	0.82															
12	1.00	-2.78	-1.58	-0.38	0.82															
13	1.00	-2.32	-1.12	0.08	1.28															
14	1.00	-2.32	-1.12	0.08	1.28															
15	1.00	-2.32	-1.12	0.08	1.28						1.00	-1.32	-0.12	1.08	2.28	0.50	-1.82	-0.62	0.58	1.78
16	1.00	-2.32	-1.12	0.08	1.28															
17	1.00	-1.80	-0.60	0.60	1.80															
18	1.00	-1.80	-0.60	0.60	1.80															
19	1.00	-1.80	-0.60	0.60	1.80															
20	1.00	-1.80	-0.60	0.60	1.80						1.00	-1.30	-0.10	1.10	2.30	0.50	-1.80	-0.60	0.60	1.80
21	1.00	-1.80	-0.60	0.60	1.80															
22	1.00	-1.80	-0.60	0.60	1.80															
23	1.00	-1.80	-0.60	0.60	1.80															
24	1.00	-1.80	-0.60	0.60	1.80															
25	1.00	-1.28	-0.08	1.12	2.32											1.00	-0.28	0.92	2.12	3.32
26	1.00	-1.28	-0.08	1.12	2.32															
27	1.00	-1.28	-0.08	1.12	2.32															
28	1.00	-1.28	-0.08	1.12	2.32															
29	1.00	-0.82	0.38	1.58	2.78															
30	1.00	-0.82	0.38	1.58	2.78											1.00	-0.32	0.88	2.08	3.28
31	1.36	-2.32	-1.12	0.08	1.28															
32	1.36	-2.32	-1.12	0.08	1.28															
33	1.36	-1.80	-0.60	0.60	1.80															
34	1.36	-1.80	-0.60	0.60	1.80															
35	1.36	-1.80	-0.60	0.60	1.80											0.86	-1.30	-0.10	1.10	2.30
36	1.36	-1.80	-0.60	0.60	1.80															
37	1.36	-1.28	-0.08	1.12	2.32															
38	1.36	-1.28	-0.08	1.12	2.32															
39	1.80	-1.80	-0.60	0.60	1.80															
40	1.80	-1.80	-0.60	0.60	1.80											1.30	-1.80	-0.60	0.60	1.80

(continued)

Table 2 (Continued)
(b) Balanced-Bidirectional DIF

	Reference					Focal Condition 4 (5% DIF)					Focal Condition 5 (10% DIF)					Focal Condition 6 (20% DIF)				
Item	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4
1	0.55	-1.80	-0.60	0.60	1.80															
2	0.55	-1.80	-0.60	0.60	1.80															
3	0.73	-2.32	-1.12	0.08	1.28															
4	0.73	-2.32	-1.12	0.08	1.28															
5	0.73	-2.30	-1.10	0.10	1.30	0.73	-1.30	-0.10	1.10	2.30	0.73	-1.30	-0.10	1.10	2.30	----	----	----	----	----
*5	0.73	-1.80	-0.60	0.60	1.80	----	----	----	----	----	----	----	----	----	----	1.23	-1.80	-0.60	0.60	1.80
6	0.73	-1.30	-0.10	1.10	2.30	0.73	-2.30	-1.10	0.10	1.30	0.73	-2.30	-1.10	0.10	1.30	----	----	----	----	----
*6	1.23	-1.80	-0.60	0.60	1.80	----	----	----	----	----	----	----	----	----	----	0.73	-1.80	-0.60	0.60	1.80
7	0.73	-1.80	-0.60	0.60	1.80															
8	0.73	-1.80	-0.60	0.60	1.80															
9	0.73	-1.28	-0.08	1.12	2.32															
10	0.73	-1.28	-0.08	1.12	2.32															
11	1.00	-2.78	-1.58	-0.38	0.82															
12	1.00	-2.78	-1.58	-0.38	0.82															
13	1.00	-2.32	-1.12	0.08	1.28															
14	1.00	-2.32	-1.12	0.08	1.28															
15	1.00	-2.57	-1.37	-0.17	1.03						1.00	-2.07	-0.87	0.33	1.53	----	----	----	----	----
*15	1.00	-2.32	-1.12	0.08	1.28	----	----	----	----	----	----	----	----	----	----	0.50	-2.32	-1.12	0.08	1.28
16	1.00	-2.07	-0.87	0.33	1.53						1.00	-2.57	-1.37	-0.17	1.03	----	----	----	----	----
*16	0.50	-2.32	-1.12	0.08	1.28	----	----	----	----	----	----	----	----	----	----	1.00	-2.32	-1.12	0.08	1.28
17	1.00	-1.80	-0.60	0.60	1.80															
18	1.00	-1.80	-0.60	0.60	1.80															
19	1.00	-1.80	-0.60	0.60	1.80															
20	1.00	-1.80	-0.60	0.60	1.80															
21	1.00	-1.80	-0.60	0.60	1.80															
22	1.00	-1.80	-0.60	0.60	1.80															
23	1.00	-1.80	-0.60	0.60	1.80															
24	1.00	-1.80	-0.60	0.60	1.80															
25	1.00	-1.28	-0.08	1.12	2.32											----	----	----	----	----
25	1.00	-1.78	-0.58	0.62	1.82	----	----	----	----	----	----	----	----	----	----	1.00	-0.78	0.42	1.62	2.82
26	1.00	-1.28	-0.08	1.12	2.32											----	----	----	----	----
*26	1.00	-0.78	0.42	1.62	2.82	----	----	----	----	----	----	----	----	----	----	1.00	-1.78	-0.58	0.62	1.82
27	1.00	-1.28	-0.08	1.12	2.32															
28	1.00	-1.28	-0.08	1.12	2.32															
29	1.00	-0.82	0.38	1.58	2.78											1.00	-0.32	0.88	2.08	3.28
30	1.00	-0.82	0.38	1.58	2.78											----	----	----	----	----
*30	1.00	-0.32	0.88	2.08	3.28	----	----	----	----	----	----	----	----	----	----	1.00	-0.82	0.38	1.58	2.78
31	1.36	-2.32	-1.12	0.08	1.28															
32	1.36	-2.32	-1.12	0.08	1.28															
33	1.36	-1.80	-0.60	0.60	1.80															
34	1.36	-1.80	-0.60	0.60	1.80															
35	1.36	-1.80	-0.60	0.60	1.80															
36	1.36	-1.80	-0.60	0.60	1.80															
37	1.36	-1.28	-0.08	1.12	2.32															
38	1.36	-1.28	-0.08	1.12	2.32															
39	1.80	-1.80	-0.60	0.60	1.80															
40	1.80	-1.80	-0.60	0.60	1.80															

Note. * indicate Reference Group parameters used in Condition 5 (20% DIF). Blanks indicate use of the same item parameters listed for the Reference Group.

Table 3

C-DIF Results : The True Positive (TP) and the False Positive (FP) Proportions of DIF Identification

	No Impact				Impact			
	C-DIF True		C-DIF Estimated		C-DIF True		C-DIF Estimated	
	TP	FP	TP	FP	TP	FP	TP	FP
				20-Item Test				
Null Condition (0 DIF Items)	--	.00	--	.00	--	.00	--	.00
Unidirectional								
Condition 1 (1 DIF Items)	1.00	.00	1.00	.00	1.00	.00	1.00	.00
Condition 2 (2 DIF Items)	1.00	.00	.90	.03	1.00	.00	.90	.03
Condition 3 (4 DIF Items)	.75	.00	.65	.18	.75	.00	.65	.03
Balanced-Bidirectional								
Condition 4 (2 DIF Items)	--	.00	--	.00	--	.00	--	.01
Condition 5 (4 DIF Items)	--	.00	--	.02	--	.00	--	.01
				40-Item Test				
Null Condition (0 DIF Items)	--	.00	--	.00	--	.00	--	.00
Unidirectional								
Condition 1 (2 DIF Items)	1.00	.00	1.00	.01	1.00	.00	1.00	.02
Condition 2 (4 DIF Items)	1.00	.00	.80	.01	1.00	.00	.50	.02
Condition 3 (8 DIF Items)	.75	.00	.68	.01	.75	.00	.68	.01
Balanced-Bidirectional								
Condition 4 (2 DIF Items)	--	.00	--	.01	--	.00	--	.01
Condition 5 (4 DIF Items)	--	.00	--	.00	--	.00	--	.01
Condition 6 (8 DIF Items)	--	.00	--	.03	--	.00	--	.03

Note. True NC-DIF condition is based on one analysis. All other figures are based on 5 replications.

Table 4

C-DIF True Positive Proportions at the Item Level

Item	Difference in Item Parameters: a	bs	No Impact: True NC-DIF	No Impact: Est NC-DIF	Impact: True NC-DIF	Impact: Est NC-DIF
20-Item Test						
Unidirectional Conditions						
Condition 1						
3	0.0	+1.0	1.0	1.0	1.0	1.0
Condition 2						
3	0.0	+0.5	1.0	.8	1.0	.8
8	0.0	+1.0	1.0	1.0	1.0	1.0
Condition 3						
3	0.0	+1.0	1.0	.8	1.0	1.0
8	-0.5	+0.5	1.0	.8	1.0	.6
13	0.0	+0.5	1.0	.8	1.0	1.0
18	-0.5	0.0	.0	.2	.0	.0
40-Item Test						
Unidirectional Conditions						
Condition 1						
5	0.0	+1.0	1.0	1.0	1.0	1.0
10	0.0	+1.0	1.0	1.0	1.0	1.0
Condition 2						
5	0.0	+1.0	1.0	.8	1.0	.4
10	0.0	+0.5	1.0	.6	1.0	.4
15	0.0	+1.0	1.0	1.0	1.0	.8
20	0.0	+0.5	1.0	.8	1.0	.4
Condition 3						
5	0.0	+1.0	1.0	1.0	1.0	1.0
10	0.0	+0.5	1.0	1.0	1.0	1.0
15	-0.5	+0.5	1.0	.8	1.0	.2
20	-0.5	0.0	.0	.0	.0	.0
25	0.0	+0.5	1.0	1.0	1.0	1.0
30	0.0	+0.5	1.0	.8	1.0	1.0
35	-0.5	+0.5	1.0	1.0	1.0	1.0
40	-0.5	0.0	.0	.0	.0	.0

Table 5

NC-DIF Results : The True Positive (TP) and the False Positive (FP) Proportions

	No Impact				Impact			
	NC-DIF True		NC-DIF Estimated		NC-DIF True		NC-DIF Estimated	
	TP	FP	TP	FP	TP	FP	TP	FP
20-Item Test								
Null Condition (0 DIF Items)	--	.00	--	.00	--	.00	--	.00
Unidirectional								
Condition 1 (1 DIF Items)	1.00	.00	1.00	.01	1.00	.00	1.00	.01
Condition 2 (2 DIF Items)	1.00	.00	1.00	.00	1.00	.00	1.00	.00
Condition 3 (4 DIF Items)	.75	.00	.75	.00	.75	.00	.80	.00
Balanced-Bidirectional								
Condition 4 (2 DIF Items)	1.00	.00	1.00	.00	1.00	.00	1.00	.00
Condition 5 (4 DIF Items)	.50	.00	.50	.00	.50	.00	.55	.00
40-Item Test								
Null Condition (0 DIF Items)	--	.00	--	.00	--	.00	--	.00
Unidirectional								
Condition 1 (2 DIF Items)	1.00	.00	1.00	.00	1.00	.00	1.00	.01
Condition 2 (4 DIF Items)	1.00	.00	1.00	.00	1.00	.00	1.00	.01
Condition 3 (8 DIF Items)	.88	.00	.88	.01	.88	.00	.88	.00
Balanced-Bidirectional								
Condition 4 (2 DIF Items)	1.00	.00	1.00	.00	1.00	.00	1.00	.00
Condition 5 (4 DIF Items)	1.00	.00	1.00	.00	1.00	.00	1.00	.00
Condition 6 (8 DIF Items)	.75	.00	.70	.01	.75	.00	.80	.00

Note. True NC-DIF condition is based on one analysis. All other figures are based on 5 replications.

Table 6

NC-DIF True Positive Proportions at the Item Level

(a) 20-Item Test

Item	Difference in Item Parameters: a	bs	No Impact: True NC-DIF	No Impact: Est NC-DIF	Impact: True NC-DIF	Impact: Est NC-DIF
Unidirectional Conditions						
Condition 1						
3	0.0	+1.0	1.0	1.0	1.0	1.0
Condition 2						
3	0.0	+0.5	1.0	1.0	1.0	1.0
8	0.0	+1.0	1.0	1.0	1.0	1.0
Condition 3						
3	0.0	+1.0	1.0	1.0	1.0	1.0
8	-0.5	+0.5	1.0	1.0	1.0	1.0
13	0.0	+0.5	1.0	1.0	1.0	1.0
18	-0.5	0.0	.0	.0	1.0	.2
Balanced-Bidirectional Conditions						
Condition 4						
3	0.0	+0.5	1.0	1.0	1.0	1.0
4	0.0	-0.5	1.0	1.0	1.0	1.0
Condition 5						
3	+0.5	0.0	.0	.0	1.0	.0
4	-0.5	0.0	.0	.0	1.0	.0
12	0.0	+0.5	1.0	1.0	1.0	1.0
13	0.0	-0.5	1.0	1.0	1.0	1.0

Table 6 (Continued)

NC-DIF True Positive Proportions at the Item Level

(b) 40-Item Test

Item	Difference in Item Parameters: a	Difference in Item Parameters: bs	No Impact: True NC-DIF	No Impact: Est NC-DIF	Impact: True NC-DIF	Impact: Est NC-DIF
Unidirectional Conditions						
Condition 1						
5	0.0	+1.0	1.0	1.0	1.0	1.0
10	0.0	+1.0	1.0	1.0	1.0	1.0
Condition 2						
5	0.0	+1.0	1.0	1.0	1.0	1.0
10	0.0	+0.5	1.0	1.0	1.0	1.0
15	0.0	+1.0	1.0	1.0	1.0	1.0
20	0.0	+0.5	1.0	1.0	1.0	1.0
Condition 3						
5	0.0	+1.0	1.0	1.0	1.0	1.0
10	0.0	+0.5	1.0	1.0	1.0	1.0
15	-0.5	+0.5	1.0	1.0	1.0	1.0
20	-0.5	0.0	1.0	1.0	1.0	1.0
25	0.0	+0.5	1.0	1.0	1.0	1.0
30	0.0	+0.5	1.0	1.0	1.0	1.0
35	-0.5	+0.5	1.0	1.0	1.0	1.0
40	-0.5	0.0	.0	.0	.0	.0
Balanced-Bidirectional Conditions						
Condition 4						
5	0.0	+1.0	1.0	1.0	1.0	1.0
6	0.0	-1.0	1.0	1.0	1.0	1.0
Condition 5						
5	0.0	+1.0	1.0	1.0	1.0	1.0
6	0.0	-0.0	1.0	1.0	1.0	1.0
15	0.0	+0.5	1.0	1.0	1.0	1.0
16	0.0	-0.5	1.0	1.0	1.0	1.0
Condition 6						
5	+0.5	0.0	.0	.0	.0	.2
6	-0.5	0.0	.0	0	.0	.2
15	-0.5	0.0	1.0	1.0	1.0	1.0
16	+0.5	0.0	1.0	1.0	1.0	1.0
25	0.0	+1.0	1.0	1.0	1.0	1.0
26	0.0	-1.0	1.0	1.0	1.0	1.0
29	0.0	+0.5	1.0	1.0	1.0	1.0
30	0.0	-0.5	1.0	1.0	1.0	1.0

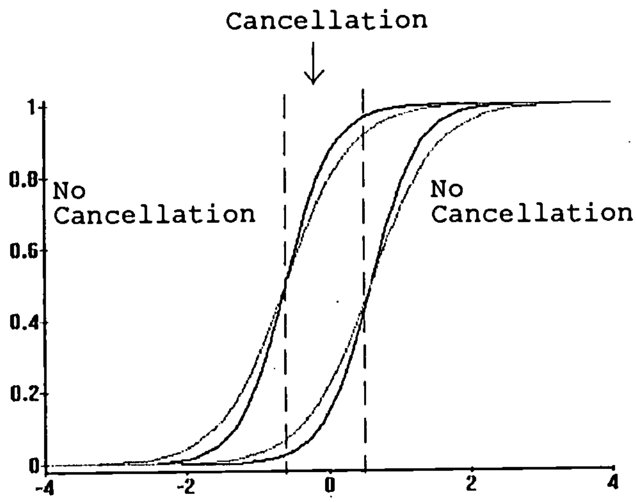
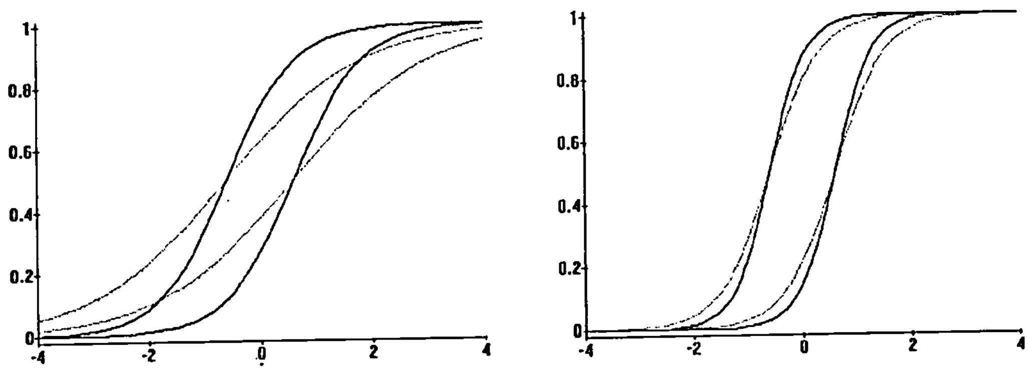


Figure 1. Cancellation Within an Examinee's True Item Score for a Three-Category Nonuniform DIF Item.



Low Discrimination
(a = .5 and 1.0)

High Discrimination
(a = 1.3 and 1.8)

Figure 2. Difference between Focal and Reference Groups with High and Low Discrimination Parameters.

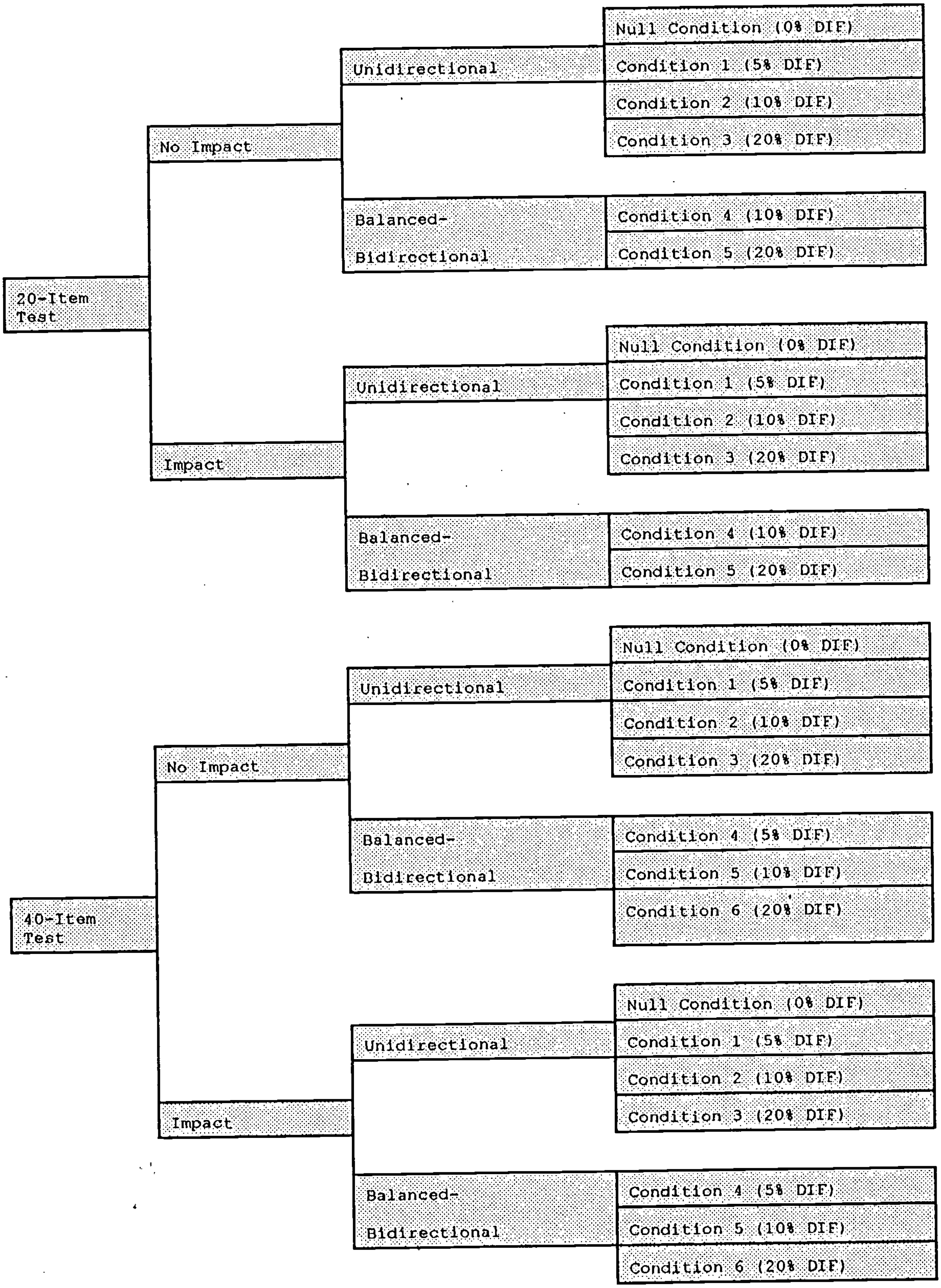


Figure 3. Simulation Design.

AERA April 8-12, 1996

TM025892

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A DESCRIPTION AND DEMONSTRATION OF THE POLYTOMOUS - DFIT FRAMEWORK

Author(s): FLOWERS, C.P., OSHIMA, T.C., RAJU, N.

Corporate Source:

Publication Date: Ap 8, 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[✓] **Sample sticker to be affixed to document**

Check here

Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here

Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: [signature]	Position: ASSISTANT PROFESSOR
Printed Name: CLAUDIA FLOWERS	Organization: THE UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE
Address: UNC CHARLOTTE 9201 UNIVERSITY CITY BLVD CHARLOTTE, NC 28223-0001	Telephone Number: (704) 547-4545
	Date: 5-28-96