ACT Research Report Series 89-7

# Performance of Three Conditional DIF Statistics in Detecting Differential Item Functioning on Simulated Tests

Judith A. Spray

October 1989

ACT

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243

# PERFORMANCE OF THREE CONDITIONAL DIF STATISTICS IN DETECTING DIFFERENTIAL ITEM FUNCTIONING ON SIMULATED TESTS

Judith A. Spray

## ABSTRACT

Computer simulations were conducted to study the behavior of three conditional differential item functioning (DIF) statistics in the detection of true or asymptotic DIF. The statistics included the standardized difference in proportion-correct (STD), the Mantel-Haenszel common odds-ratio (MH) and the root mean weighted squared difference in proportion-correct (RMWSD). The simulated tests were based on actual administrations of the ACT Assessment to certain focal and base examinee populations. Sample sizes of examinees were varied while true DIF and test length remained fixed. Results of these simulations showed that the MH and STD statistics were preferred as DIF indicators for sample sizes greater than 250.

**PERFORMANCE OF THREE CONDITIONAL DIF STATISTICS IN DETECTING DIFFERENTIAL ITEM FUNCTIONING ON SIMULATED TESTS**

In the fall of 1988, several members of the American College Testing Program's Test Development Division conducted computer simulations to study the behavior of three conditional differential item functioning (DIF) statistics, in terms of DIF or item bias detection. The statistics selected for inclusion in this study were the standardized difference in proportion-correct (Dorans & Kulick, 1986), Mantel-Haenszel common odds-ratio (Holland & Thayer, 1986; Mantel & Haenszel, 1959), and the root mean weighted squared difference in proportion-correct (Dorans & Kulick, 1986).

Item bias statistics which condition on some examinee ability measure are thought to be better measures of DIF than those statistics that use the simple unconditional difference in proportion-correct values, sometimes referred to as impact. The unconditional impact does not take into account underlying differences in ability distributions between populations or groups of interest. The conditional procedures, on the other hand, reflect proportion-correct differences only between examinees with comparable ability in each population or group. These DIF statistics have been used by other testing programs and services to detect or flag test items on tests where DIF might be problematic. The statistics were defined as follows.

The populations or groups of interest were referred to as the focal ($F$) group and the base ($B$) group. Then $s$ indexed each observed score category of a $k$-item test, or $s = 0, 1, \ldots, k$. Then

$N_{F_s}$ = the number of examinees in the $F$ group at score $s$,

$N_{B_s}$ = the number of examinees in the $B$ group at score $s$,

$N_s$ = the total number of examinees in F and B at score s,

$\omega_{F_s} = N_{F_s} / \sum_{s=0}^{k} N_{F_s}$, the relative frequency of F at s,

$\omega_{B_s} = N_{B_s} / \sum_{s=0}^{k} N_{B_s}$, the relative frequency of B at s, and

$\omega_s = N_s / \sum_{s=0}^{k} N_s$, the total relative frequency of F and B at s.

Also $R_{F_s}$ and $R_{B_s}$ were the numbers of examinees (i.e., absolute frequency) in F and B respectively at s who answered the item correctly. The proportion-correct values for each group at s were given by

$$P_{F_s} = R_{F_s} / N_{F_s} \text{, and}$$

$$P_{B_s} = R_{B_s} / N_{B_s} \quad .$$

## The STD Statistic

The standardized difference in proportion-correct was defined as

$$STD = \sum_{s=0}^{k} \omega_{F_s} (P_{F_s} - P_{B_s}) \quad . \tag{1}$$

The signed difference, $(P_{F_s} - P_{B_s})$, was weighted by the relative frequency of F because $\omega_{F_s}$ provided the greatest weight to differences at those score levels most frequently observed in the focal group.

## The MH Statistic

If $W_{F_s}$ and $W_{B_s}$ were the absolute frequencies of incorrect responses to this item in F and B, respectively at s, and $N_s$ was the total number of responses at s, then the Mantel-Haenszel common odds-ratio was

$$MH = \frac{\sum_{s=0}^{k} R_{B_s} W_{F_s} / N_s}{\sum_{s=0}^{k} R_{F_s} W_{B_s} / N_s} .$$

If $Q_{F_s}$ and $Q_{B_s}$ were defined as $(1 - P_{F_s})$ and $(1 - P_{B_s})$ respectively, then this index also could be written as

$$MH = \frac{\sum_{s=0}^{k} P_{B_s} Q_{F_s} \frac{N_{B_s} \cdot N_{F_s}}{N_s}}{\sum_{s=0}^{k} P_{F_s} Q_{B_s} \cdot \frac{N_{B_s} \cdot N_{F_s}}{N_s}}$$

or even as a function of several relative weights or densities,

$$MH = \frac{\sum_{s=0}^{k} P_{B_s} Q_{F_s} \frac{\omega_{B_s} \cdot \omega_{F_s}}{\omega_s}}{\sum_{s=0}^{k} P_{F_s} Q_{B_s} \frac{\omega_{B_s} \cdot \omega_{F_s}}{\omega_s}} . \qquad (2)$$

The RMWSD Statistic

And finally, the root mean weighted squared difference in proportion-correct was defined as

$$\text{RMWSD} = \left\{ \sum_{s=0}^{k} \omega_{F_s} \left(P_{F_s} - P_{B_s}\right)^2 \right\}^{\frac{1}{2}} . \tag{3}$$

**Asymptotic DIF Indices**

These conditional item bias or DIF statistics are limited by the use of a manifest ability measure, the observed test score, which suffers from several drawbacks. First, the observed test score contains some measurement error which, according to the underlying model adopted to describe an examinee's true ability, is a function of the test's reliability or a function of this true but latent examinee ability. Because the observed test score was used in this study as the conditional ability measure and because this test score measure was discrete and finite, examinees from each group with different true but latent abilities could have been classified into the same score category, thereby violating to some degree the conditional approach to assessing DIF. A second problem is that, in practice, the observed test response data are based on available sample sizes from each group. These sample sizes may be quite small for some minority groups of interest whereas the latent ability distributions may be based on populations of very large (e.g., infinite) sizes. A third problem is that the observed test score consists of the sum of each individual's item scores. Thus if one or more of the test items is biased or favors one group over another, the observed test score may be biased as well.

In studying or comparing the performance of several DIF statistics in research settings, it was useful to define some measures of asymptotic DIF at the latent variable level based on the distributions of latent ability, so that these problems could be eliminated. Then, the performance capabilities of the item bias statistics could be compared to the true DIF indices which served as a true reference for DIF identification. The purpose of the present study was to compare the performances of three conditional DIF statistics under computer-simulated test-taking conditions in identifying test items with various degrees of DIF.

These true DIF indices are called asymptotic DIF indices. They were constructed as follows. Each of the DIF statistics, as given by equations (1), (2), and (3), was redefined by (a) replacing the manifest ability measure, $\underline{s}$, with a latent ability measure, $\theta$, (b) allowing $\theta$ to be open and continuous on the real line such that $-\infty < \theta < \infty$, (c) replacing the summation operators with integration operations, and (d) allowing the finite samples of the groups or populations to approach infinity. In addition the relative frequencies at the discrete scores, $\underline{s}$, were replaced by densities of $\theta$ in $\underline{F}$, $\underline{B}$ and the combined distribution. These arbitrary density functions of $\theta$ were denoted by $g_F(\theta)$, $g_B(\theta)$ and $g^*(\theta)$, respectively. The combined group density could be written as

$$g^*(\theta) = \alpha g_F(\theta) + (1 - \alpha) g_B(\theta),$$

where a mixing weight or proportion, $\alpha$, was defined as $0 \le \alpha \le 1$.

For the purposes of this study, the definition of asymptotic DIF was made by replacing the proportions-correct and -incorrect at each score category or, $P_{B_s}$, $Q_{F_s}$, $P_{F_s}$, and $Q_{B_s}$, with probability functions of the latent ability measure, $\theta$. It was assumed that the success probabilities, $P_B(\theta)$

and $P_F(\theta)$, were defined by the three-parameter logistic item response function with known item parameters for each group and for each item, or generally by

$$P(\theta) = c + \frac{(1 - c)}{1 + e^{-1.7a(\theta - b)}} \quad .$$

## The Asymptotic STD Index

The STD index was measured on the *p*-value or proportion-correct scale. It indicated, on average, how members of *F* differed from comparable members of *B*. The STD statistic for any item of a *k*-item test was computed in practice by equation (1). Negative values of STD indicated that the item favored *B* while positive values of STD indicated that the item favored *F*. Values of STD near zero indicated no DIF.

An asymptotic value of STD was defined by

$$\underset{\to\infty}{\text{STD}}(\theta) = \int_{-\infty}^{\infty} \left[ P_F(\theta) - P_B(\theta) \right] g_F(\theta) \, d\theta \quad . \qquad (4)$$

## The Asymptotic MH Index

The MH statistic could be interpreted as an odds-ratio. It indicated how much more (or less) likely it was that a comparable member of *B* answered the item correctly than a comparable member of *F*.

The MH defined in (2) was an estimate of the common odds-ratio across the *k* + 1 levels of the 2 x 2 tables (i.e., group by response). The MH statistic had a value at or near 1.0 if there was no item bias present between *F* and *B*. If the item favored *F*, MH was less than 1.0. If the item favored *B*, MH was greater than 1.0.

An asymptotic measure of the MH index was defined by

$$\underset{\rightarrow\infty}{\mathrm{MH}}(\theta) = \frac{\int_{-\infty}^{\infty} P_B(\theta)Q_F(\theta)\,\frac{g_B(\theta)g_F(\theta)}{g^*(\theta)}\,d\theta}{\int_{-\infty}^{\infty} P_F(\theta)Q_B(\theta)\,\frac{g_B(\theta)g_F(\theta)}{g^*(\theta)}\,d\theta} \quad . \tag{5}$$

The Asymptotic RMWSD Index

The RMWSD statistic given in (3) was a nondirectional DIF indicator, thought to be capable of detecting cases of nonuniform item bias or situations where DIF favored one group at one part of the ability range but favored the other group at another part of the ability range.

An asymptotic value of RMWSD was defined by

$$\underset{\rightarrow\infty}{\mathrm{RMWSD}}(\theta) = \left\{ \int_{-\infty}^{\infty} \left[P_F(\theta) - P_B(\theta)\right]^2 g_F(\theta)\, d\theta \right\}^{\frac{1}{2}} \quad . \tag{6}$$

## Obtaining Item Parameters

Three tests from previously administered forms of the ACT Assessment Program (AAP) were used to obtain item parameter estimates on several different comparison groups of examinees. The first test used was form number 29B of the Mathematics Usage test (Test A). The groups of examinees randomly selected from a national population of test-takers were 2000 white or Caucasian examinees who were designated as the base (*B*) group and 2000 black examinees who were designated as the focal (*F*) group. The second test was AAP form number 27F of the Social Studies Reading Test (Test B). The examinee groups were again randomly selected to yield 2000 white or Caucasian students (B) and 2000 black students (F). The third test was AAP form number 29C of the Mathematics Usage Test (Test C). For this test 2000 members of B were chosen from the national sample of males and 2000 members of F from the national sample of females.

Item parameter estimates were obtained from the $N$ = 4000, $k$ = 40, Test A item responses via the joint maximum likelihood method from the LOGIST (Wingersky, Barton & Lord, 1982) computer program and from the $N$ = 4000, $k$ = 52, Test B and $N$ = 4000, $k$ = 40, Test C responses via the marginal maximum likelihood method of the BILOG (Mislevy & Bock, 1984) computer program. These item parameter estimates are listed in Tables 1, 2, and 3 for Tests A, B, and C, respectively.

Item parameter estimates from the three calibrations of each F group ($a_F$, $b_F$, $c_F$) were rescaled to the respective B group ($a_F^*$, $b_F^*$, $c_F^*$) by the family of simple linear transformations,

$$a_F^* = a_F/A,$$

$$b_F^* = b_F \bullet A + B$$

$$c_F^* = c_F$$

with

$$A = SD(b_B)/SD(b_F) \text{ and}$$
$$B = (\overline{b_B}) - A \bullet (\overline{b_F}) \quad .$$

## Asymptotic DIF Results

Using the item parameters given in Tables 1-3, the asymptotic DIF indices were calculated from the integral expressions given by equations (4), (5), and (6). The density functions for each group and each test were arbitrarily set as follows: (a) Test A, $\theta_F \sim N(-.5, 1.0)$, $\theta_B \sim N(.0, 1.0)$; (b) Test B, $\theta_F \sim N(-.5, 1.0)$, $\theta_B \sim N(.0, 1.0)$; (c) Test C, $\theta_F \sim N(.0, 1.0)$, $\theta_B \sim N(.0, 1.0)$. Two values of $\alpha$, the mixing parameter for the combined density function, $g^*(\theta)$, were used in the evaluation of the theoretical or asymptotic value of MH, as given in equation (5). These values of $\alpha$ were .50 and .09, which represented mixing ratios of 1:1 and 1:10, respectively. Note that varying $\alpha$ only affected the asymptotic value of MH, because only this index required the use of the combined or mixed density function of $\theta$. Tables 4, 5, and 6 give the values of these asymptotic DIF indices under the conditions described above.

The evaluation of these asymptotic DIF values revealed few problematic or "biased" test items. Using a completely arbitrary set of criteria, items were deemed to be biased if any one of the following conditions was met:

$$\text{(i)} \quad |\underset{\to\infty}{STD}(\theta)| \geq .10,$$

$$\text{(ii)} \quad \underset{\to\infty}{MH}(\theta) \geq 2.0, \text{ or}$$

$$\underset{\to\infty}{MH}(\theta) \leq .5,$$

$$\text{(iii)} \quad \underset{\to\infty}{RMWSD}(\theta) \geq .10 \quad .$$

The item statistics satisfying these criteria are in boldface in Tables 4, 5, and 6.

Results from Test A

Table 4 shows that four items from MA29B were identified as being biased by at least one of the asymptotic or asymptotic DIF indices. These included items 5, 7, 19, and 29. Figures 1-4 illustrate the comparison plots of the item characteristic curves (ICCs) between $P_B(\theta)$ and $P_F(\theta)$ with the known item parameters given in Table 1. Several types of item bias or DIF are represented by these figures. Each of the DIF indices was capable of detecting uniform bias or situations in which the direction of the bias remained constant throughout most of the ability range. The root mean weighted squared difference in proportion-correct additionally was supposed to be able to detect "cross-over" effects of nonuniform bias, as depicted in Figure 4 for item number 29. Therefore, an item identified as showing DIF through the RMWSD index was biased either uniformly, nonuniformly or both. The source of this bias was not revealed in the statistic, nor was the direction of the bias identified.

The indices also differed in the weighting functions or ability distributions used in their computation. Items 5 and 7 (Figures 1 and 2, respectively) yielded the greatest group differences at the lower portion of the ability range, where significant frequencies from the F group ability distribution gave more weight to these differences. Consequently, the standardized difference in proportion-correct and the root mean weighted squared difference in proportion-correct identified these items as biased. Item 29 was a classic nonuniform bias case where the ICC cross-over occurred near the center of the combined F and B groups. Subsequently, the $RMWSD_{\to\infty}(\theta)$ index identified the item as biased while the directional indices did not. (See Figure 4). Item 5 was labeled as a biased test item as long as the definition of item bias included the stipulation that bias can only occur "where the F ability is distributed." On the other hand, item 19 had the greatest ICC differences at the upper end of the ability scale where there were fewer F examinees. The results from Table 4 illustrate these facts. Test A showed quite a variety of the types of item bias or DIF that can occur.

## Results from Test B

Table 5 shows that five items from Test B were identified as biased test items according to at least one of the criteria defined previously. However, there were many items within the last reading passage of the test that were quite close to the .10 criterion for the $STD_{\to\infty}(\theta)$ and $RMWSD_{\to\infty}(\theta)$ indices, and all of the signs of the $STD_{\to\infty}(\theta)$ index indicated that the F group was disadvantaged. Because these items appeared in the last reading passage, it was hypothesized that a speededness effect was probably present. Figures 5-9 show the F and B comparisons of the ICCs of items 23, 34, 44, 47, and 50.

## Results from Test C

Table 6 shows that only three items (numbers 4, 18 and 28) from Test C were identified as biased according to the criteria. This situation was a bit different from the previous test results in that it was expected that the $STD_{\to\infty}(\theta)$ and $MH_{\to\infty}(\theta)$ results should have agreed more closely with each other due to the fact that the F and B ability distributions were the same. This was true, as the only item (item 4) to be identified as uniformly biased in favor of B, was so identified by both $STD_{\to\infty}(\theta)$ and $MH_{\to\infty}(\theta)$. The other index, $RMWSD_{\to\infty}(\theta)$, appears to have been much more liberal, identifying items not selected by either $STD_{\to\infty}(\theta)$ or $MH_{\to\infty}(\theta)$. And neither of the remaining two items exhibited the ICC cross-over that was observed in Test A (See Figures 11-12). Item 4 had some cross-over (see Figure 10) but this occurred at very low ability values, less than -2.50.

**Conditions for the Test Simulations**

The purpose of this study was to evaluate the performance of the DIF statistics as defined by equations (1), (2), and (3) under simulated testing conditions with the results from the evaluation of asymptotic DIF as obtained from equations (4), (5), and (6). Test simulation for each of the three tests (A, B, and C) was achieved in the following steps:

(1) Random samples of $\theta$ of size $N_F$ and $N_B$ were chosen from $g_F(\theta)$ and $g_B(\theta)$ in the mixing ratio of $\alpha$.

(2) Given the sampled value of $\theta$ and the item parameters of the appropriate test (i.e., A, B, or C), $P(\theta)$ was calculated by the three-parameter logistic item response function and compared against a uniform deviate, U, between zero and one. A response to that item, $y_{ij}$, was made according to the assignment,

$$y_{ij} = \begin{cases} 1, & .0 \le P(\theta) \le U \\ 0, & U < P(\theta) \le 1.0 \ . \end{cases}$$

(3) The item responses were then used to compute the DIF statistics

given by equations (1), (2), and (3).

(4) The DIF statistics were compared to the asymptotic DIF indices in two ways: (a) a rank correlation coefficient was computed between each statistic's value and the theoretical value for the *k* items on the test; (b) the statistic's "yes-no" performance was tallied as flagging or not flagging an item as biased according to a given, arbitrary criterion.

(5) All of these steps (1)-(4) were repeated or replicated 100 times.

The sample sizes and ratios between *F* and *B* included nine conditions, six for the $\alpha$ = .50 or 1:1 ratio and three for the $\alpha$ = .09 or 1:10 ratio. These were 2000:2000, 1000:1000, 500:500, 250:250, 100:100, 50:50, 200:2000, 100:1000, and 50:500.

## Test Simulation Results

The average rank correlations between the DIF estimates and their asymptotic values for the nine sample size ratios for tests A, B, and C, are given in Tables 7, 8, and 9, respectively. These tables show that the MH and STD estimates correlated fairly highly with their respective asymptotic counterparts for large (> 1000) samples and that these correlations gradually fell off as the samples became smaller. The RMWSD estimates, on the other hand, did not correlate highly with $\mathrm{RMWSD}_{\to\infty}(\theta)$, even for large (N = 2000) samples.

Tables 10 and 11 show the average DIF estimates and average standard deviations respectively, of the estimates across items for each test. The asymptotic DIF average across items is shown in the first row of Table 10. The RMWSD statistic overestimated $\underset{\to\infty}{\text{RMWSD}}(\theta)$, even for large samples, as seen in Table 10. The average RMWSD values were greater than the asymptotic values, increasing as the sample sizes decreased (i.e., the RMWSD estimator overestimated the asymptotic values even more). Variability also increased, as sample size decreased (see Table 11), which led to the eventual collapse of the relationship between the ranks of the estimates and the asymptotic value.

The MH estimator began to underestimate the asymptotic MH value for large sample sizes and then gradually progressed to the overestimation of $\underset{\to\infty}{\text{MH}}(\theta)$ as these sample sizes decreased. On the other hand, the STD estimator operated similarly to the RMWSD estimator in that it began to overestimate $\underset{\to\infty}{\text{STD}}(\theta)$ for large samples and continued to overestimate the asymptotic value as the sample sizes decreased. The magnitude of this overestimation increased as the sample size decreased. However, the STD statistic performed somewhat better than the RMWSD estimator, again as evidenced by the rank correlation coefficients in Tables 7, 8, and 9.

The variability of the DIF estimates was also of interest in these simulations. Table 12 summarizes the average standard deviations of the DIF estimates by test and by sample size. The variance of some of these estimates was certainly a function of their magnitude. The variance of the MH estimator increased as MH increased (i.e., as item bias increased in the direction of B) and decreased as MH decreased or approached zero (i.e., as item bias increased in the direction of F). However, overall these tests were unbiased in that they contained few biased items so this effect was probably minimized. Table 11 shows that variability increased as sample sizes decreased. This was true for all three DIF estimators and for all three tests.

The amount of increase in variability can be assessed (between adjacent sample sizes, $\underline{n}_1$ and $\underline{n}_2$, from two designs which provide some variability, $\sigma_1^2$ and $\sigma_2^2$ on some measure) by evaluating the relative efficiency of the two designs, where

$$\text{relative efficiency} = \frac{n_2\sigma_2^2}{n_1\sigma_1^2} .$$

Relative efficiency indicates the trade-off between using a smaller sample size (say, $\underline{n}_2 < \underline{n}_1$) in a design and observing an expected increase in variability, $\sigma_2^2 > \sigma_1^2$. If the variance increases by the same ratio as the decrease in sample size, then the relative efficiency between design 1 and design 2 would be 1.0, implying that no unusual relative loss in precision was observed by using a sample of $\underline{n}_2$ as opposed to $\underline{n}_1$. When the relative efficiency is greater than 1.0, it implies that the variability for design 2 has increased beyond the $\underline{n}_2$:$\underline{n}_1$ ratio and that the smaller sample, $\underline{n}_2$, is no longer as efficient in obtaining the $\sigma_2^2$:$\sigma_1^2$ variance ratio.

Table 12 gives the average relative efficiencies, averaged over items, between adjacent sample sizes for the $\alpha = .50$ mixing proportion for each DIF estimator. This table indicates that, up until samples of size 250, the relative efficiency of the designs remained around 1.0 for each DIF estimator. However, it deteriorated badly for MH as the sample dropped to 100 from 250 and even worse between 100 and 50 (i.e., the rate of deterioration accelerated for MH). The STD statistic also suffered a drop in relative efficiency in going from 250 to 100 (except for Test B) but then suffered no further loss in going from 100 to 50. The RMWSD estimator remained relatively efficient regardless of the sample sizes used.

There was another way to assess the performance of these statistics. Based on the DIF criteria given previously on page 10, an item was identified as either biased or not biased and, in the case of MH or STD, if biased it was either biased in favor of F or B. Based on 100 simulations of the test, the number of times that each item was labeled as biased is listed in Tables 13-15 (Test A for MH, STD, and RMWSD), Tables 16-18 (Test B for MH, STD, and RMWSD) and Tables 19-21 (Test C for MH, STD, and RMWSD). For the directional statistics (MH and STD), the item could have been labeled as either biased in favor of F, in favor of B, or unbiased. The first number given in the columns of Tables 13, 14, 16, 17, 19, and 20 lists the number of times the item was found to be biased in favor of F while the second number gives the number of times the item was found to be biased in the B direction. The single number in each column in Tables 15, 18, and 21 gives the number of times that the RMWSD estimator classified an item as biased, regardless of direction.

The averages at the bottom of each column in Tables 13-21 represent the average number of biased test items on the test, averaged over 100 replications of simulated test administrations. These numbers should be compared to the number of actually biased items on the test, as defined by the evaluation of the asymptotic index and the criteria given on page 10. Recall that such "biased" test items are so indicated in boldface in Tables 4-6.

For example, according to Table 4, there was only one item on Test A that was biased in favor of F according to the MH index with $\alpha = .50$ (item number 19), while no items were actually biased in favor of B, according to the MH criterion. Therefore, the true or "asymptotic" numbers, 1/0, can be compared with those averages at the bottom of the first six columns in Table 13 in order to evaluate the performance of the MH estimator over 100 replications. In this example, the true 1/0 numbers were fairly well approximated by the

2000:2000, 1000:1000 and 500:500 sample sizes. For the $\alpha = .09$ or 1:10 mix, two items were classified as being biased, one in each direction, so that the averages in Table 13 for this mixing ratio could be compared against a 1/1 standard.

Table 14 shows the performance of the STD statistic in estimating $\underset{\to\infty}{\text{STD}}(\theta)$ on Test A, which identified two items as biased (see Table 4 again), or 1/1. The STD statistic held up fairly well through sample sizes of 2000:2000, 1000:1000, and 500:500. However, its performance for sample sizes under 100 was markedly poorer than the MH estimator in the *F* direction. For example for samples as small as 50:50, MH identified 6.6 items as biased for *F* while STD identified 18.4. However, the reverse was true in the *B* direction (5.9 for MH, compared with 1.9 for STD). In fact MH had a tendency to "even out" the bias classifications for *F* and *B* regardless of sample size, while STD tended to classify more items as biased in favor of *F* as the samples became smaller.

This same tendency occurred despite the test form. See Tables 16, 17 (Test B) and Tables 19, 20 (Test C). The RMWSD estimator grossly overestimated the number of biased items, regardless of test form, sample size or group composition (see Tables 15, 18 and 21).

In evaluating the performances of the MH and STD estimators, one should also consider and assess the bias identification process at each individual item level. For example for Test A, items 5, 7 and 19 were all problematic to some extent, regardless of the exact size of the asymptotic DIF parameter. For MH, these items were identified as problematic 46%, 34%, and 92% of the time, respectively, for samples as small as 250:250. These figures became 57%, 46%, and 76% for samples of 100:100.

The same "hit rates" for these three items for STD were 74%, 56%, and 52% for 250:250 and 56%, 73%, and 56% for 100:100. False positive rates were observed by looking at truly unbiased items (for example items 6, 12 and 20). The MH estimator classified these three items as biased (in either direction) 0%, 0%, and 0% for samples of 250:250 and 10%, 12%, and 13% for samples of 100:100. The STD statistic classified the same three items as biased 0%, 0%, and 2% of the time at 250:250 and 29%, 29%, and 21% at the smaller 100:100 ratio. As for the performance of the RMWSD estimator, its false positive error rates were so large, even for samples of 2000:2000, as to render it virtually useless for these sample sizes used in the simulations.

## Summary of Results

The results of three testing simulations, each test replicated 100 times, can be summarized as follows:

(1) MH tended to underestimate $\underset{\to\infty}{\underline{MH}}(\theta)$ slightly for samples of size 2000 but began to overestimate $\underset{\to\infty}{\underline{MH}}(\theta)$ as the sample sizes decreased;

(2) STD tended to overestimate $\underset{\to\infty}{\underline{STD}}(\theta)$ regardless of the sample size and this tendency increased, as did the magnitude of overestimation, as the sample size decreased;

(3) RMWSD overestimated $\underset{\to\infty}{\underline{RMWSD}}(\theta)$ even for the largest samples;

(4) MH and STD identified the truly biased test items at similar rates for samples greater than or equal to about 250;

(5) for smaller samples, the STD estimator tended to misclassify more items as biased in favor of <u>F</u> than did MH;

(6) RMWSD identified nearly all items as biased, even for samples as large as 500 to 1000;

(7) all DIF estimators exhibited increased variability as sample sizes decreased but STD and RMWSD showed more stable relative efficiency as the sample size of the experimental conditions dropped.

## Conclusions

Of the three conditional DIF statistics studied in this series of simulations, only two, the MH and STD indices, are viable, useful candidates in the detection of DIF. The RMWSD statistic showed an unacceptable Type I error rate or tendency to classify items as DIF items when, in fact, they were not. The high Type I error rate for RMWSD was evident even for large sample sizes (e.g., 2000:2000). For sample sizes greater than or equal to 250, both the MH and STD statistics performed reasonably well in detecting asymptotic DIF and minimizing error rates. An unexpected finding in these simulations was the tendency for MH to yield unbiased results for moderate sample sizes of examinees (e.g., 500 to 250) but to show underestimates of $\underset{\to\infty}{\underline{MH}}(\theta)$ for large sample sizes and overestimates of $\underset{\to\infty}{\underline{MH}}(\theta)$ for small sample sizes of examinees. In contrast, the STD statistic consistently overestimated $\underset{\to\infty}{\underline{STD}}(\theta)$. These results were true for a fixed (i.e., finite) test length. This study did not pursue the question of consistency as both sample sizes of examinees <u>and</u> test items increased simultaneously. Thus, for whatever statistical reasons, MH produces a "nice unbiased-like" result for moderate sample sizes and moderate test lengths. This may yield enough of an advantage to suggest its use over that of STD.

## REFERENCES

Doran, N. J., & Kulick, E. M. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the scholastic aptitude test. Journal of Educational Measurement, 23, 355-368.

Holland, P. W., & Thayer, D. T. (1986). Differential item functioning and the Mantel-Haenszel procedure. Program Statistics Research Technical Report No. 86-69. Princeton, NJ: Educational Testing Service.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719-748.

Mislevy, R. J., & Bock, R. D. (1984). BILOG, maximum likelihood item analysis and test scoring: Logistic model. Mooresville, IN: Scientific Software, Inc.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). LOGIST user's guide. Princeton, NJ: Educational Testing Service.

TABLE 1

Item parameter estimates for the Focal (black) and Base (white) groups: Test A

| | a | | b | | c | |
|---|---|---|---|---|---|---|
| Item # | F | B | F | B | F | B |
| 1 | .447 | .397 | -1.818 | -2.237 | .056 | .139 |
| 2 | 2.037 | 1.495 | - .623 | - .565 | .113 | .096 |
| 3 | .431 | .537 | -1.342 | -1.116 | .056 | .139 |
| 4 | .717 | .644 | - .160 | - .250 | .056 | .139 |
| 5 | 1.077 | .458 | .364 | .945 | .296 | .139 |
| 6 | 1.032 | 1.145 | - .527 | - .502 | .099 | .088 |
| 7 | .928 | 1.261 | - .798 | - .469 | .056 | .023 |
| 8 | .480 | .574 | - .906 | - .435 | .056 | .139 |
| 9 | .821 | .857 | .070 | - .103 | .081 | .066 |
| 10 | 1.218 | .732 | .437 | - .278 | .357 | .139 |
| 11 | .606 | .584 | - .072 | - .154 | .056 | .139 |
| 12 | 1.631 | .964 | .075 | - .178 | .201 | .063 |
| 13 | .251 | .315 | - .421 | .234 | .056 | .139 |
| 14 | 1.197 | .920 | .376 | .241 | .112 | .061 |
| 15 | .853 | 1.290 | - .181 | - .104 | .000 | .057 |
| 16 | 1.375 | 1.074 | .121 | - .187 | .142 | .053 |
| 17 | 1.874 | 1.471 | - .049 | .067 | .142 | .093 |
| 18 | 2.165 | 1.336 | - .077 | .095 | .083 | .091 |
| 19 | 1.146 | 1.036 | .721 | 1.330 | .025 | .009 |
| 20 | .950 | 1.346 | - .042 | .092 | .000 | .071 |
| 21 | 1.218 | .839 | .211 | .161 | .134 | .004 |
| 22 | .491 | .555 | - .923 | - .572 | .056 | .139 |
| 23 | 2.165 | 1.382 | .509 | .495 | .166 | .190 |
| 24 | 1.538 | 1.065 | .680 | .700 | .177 | .209 |
| 25 | 1.470 | 1.206 | - .189 | .003 | .009 | .000 |
| 26 | 1.986 | 1.445 | .606 | .670 | .234 | .254 |
| 27 | .771 | .742 | .813 | .898 | .163 | .207 |
| 28 | 1.190 | .700 | .522 | .624 | .118 | .021 |
| 29 | 1.675 | .615 | .453 | .225 | .216 | .139 |
| 30 | 1.014 | .457 | 1.213 | 1.110 | .242 | .103 |
| 31 | 1.044 | .700 | .539 | .639 | .108 | .034 |
| 32 | .650 | 1.440 | 2.226 | 1.495 | .219 | .310 |
| 33 | 1.396 | 1.311 | .648 | .757 | .124 | .126 |
| 34 | 1.254 | 1.390 | .509 | .725 | .109 | .118 |
| 35 | 1.393 | .940 | .745 | .801 | .118 | .111 |
| 36 | 1.758 | 1.290 | 1.184 | 1.170 | .134 | .147 |
| 37 | 1.309 | 1.694 | .839 | .925 | .056 | .102 |
| 38 | 2.085 | 1.243 | .919 | .916 | .192 | .209 |
| 39 | 2.165 | 1.223 | .946 | 1.140 | .171 | .173 |
| 40 | 1.791 | 1.459 | 1.142 | 1.288 | .160 | .231 |

TABLE 2

Item parameter estimates for the Focal (black) and Base (white) groups: Test B

| | a | | b | | c | |
|---|---|---|---|---|---|---|
| Item # | F | B | F | B | F | B |
| 1 | .809 | .898 | - .086 | .202 | .155 | .218 |
| 2 | .639 | .680 | -1.300 | -1.184 | .298 | .190 |
| 3 | .844 | .505 | - .406 | -1.041 | .392 | .178 |
| 4 | .856 | .474 | 2.954 | 2.821 | .411 | .429 |
| 5 | .887 | .771 | .711 | 1.015 | .180 | .191 |
| 6 | .952 | .618 | .442 | .524 | .241 | .236 |
| 7 | 1.174 | .785 | 1.222 | 1.168 | .352 | .365 |
| 8 | .676 | .465 | .255 | .247 | .176 | .121 |
| 9 | 1.664 | 1.171 | 1.406 | 1.490 | .417 | .366 |
| 10 | .978 | .719 | - .332 | -.891 | .311 | .149 |
| 11 | 1.114 | .911 | - .177 | -.449 | .324 | .197 |
| 12 | .475 | .677 | .611 | .791 | .346 | .334 |
| 13 | 1.761 | 1.262 | - .164 | -.299 | .239 | .186 |
| 14 | .920 | .755 | - .126 | -.524 | .300 | .158 |
| 15 | .475 | .420 | - .035 | -.253 | .220 | .171 |
| 16 | .879 | .750 | .614 | .272 | .147 | .086 |
| 17 | .688 | .639 | - .588 | -.960 | .301 | .231 |
| 18 | 1.066 | 1.065 | .429 | .440 | .320 | .371 |
| 19 | .721 | .967 | -1.783 | -1.541 | .232 | .231 |
| 20 | .911 | .850 | -1.094 | -1.141 | .314 | .286 |
| 21 | .738 | .640 | - .371 | -.626 | .375 | .299 |
| 22 | .672 | .730 | .477 | .674 | .259 | .290 |
| 23 | .998 | .977 | - .251 | .508 | .164 | .223 |
| 24 | .795 | .607 | .369 | .389 | .180 | .220 |
| 25 | 1.044 | .852 | 1.039 | 1.233 | .217 | .198 |
| 26 | .901 | .716 | .216 | .249 | .106 | .079 |
| 27 | 1.349 | 1.242 | 1.521 | 1.589 | .352 | .407 |
| 28 | .743 | .760 | - .229 | .292 | .165 | .226 |
| 29 | .571 | .571 | -1.355 | -1.323 | .196 | .174 |
| 30 | .485 | .355 | .284 | .123 | .176 | .236 |
| 31 | .746 | .597 | .231 | .671 | .135 | .120 |
| 32 | .531 | .634 | .928 | 1.032 | .216 | .290 |
| 33 | .569 | .478 | - .108 | .107 | .129 | .194 |
| 34 | .938 | .766 | .354 | -.078 | .158 | .207 |
| 35 | .890 | .745 | .510 | .391 | .260 | .230 |
| 36 | .701 | .432 | .930 | 1.018 | .157 | .095 |
| 37 | .444 | .415 | 1.733 | 1.833 | .089 | .156 |
| 38 | .604 | .562 | .264 | .492 | .177 | .244 |
| 39 | .729 | .586 | .580 | .795 | .113 | .136 |
| 40 | .689 | .495 | .992 | 1.548 | .064 | .117 |
| 41 | .619 | .705 | 1.770 | 1.780 | .139 | .180 |

(table continues)

| Item # | a F | a B | b F | b B | c F | c B |
|---|---|---|---|---|---|---|
| 42 | 1.146 | .838 | 1.559 | 1.861 | .123 | .154 |
| 43 | .754 | .777 | .703 | 1.169 | .162 | .237 |
| 44 | .638 | .477 | - .225 | - .559 | .156 | .228 |
| 45 | .839 | .776 | .127 | .126 | .174 | .298 |
| 46 | .800 | .763 | 1.436 | 1.666 | .160 | .275 |
| 47 | .687 | .759 | 1.088 | 1.397 | .182 | .341 |
| 48 | .968 | .967 | .893 | 1.156 | .117 | .235 |
| 49 | .944 | .700 | .858 | .853 | .010 | .156 |
| 50 | .862 | .485 | .872 | .708 | .181 | .226 |
| 51 | .821 | .701 | 1.316 | 1.344 | .112 | .206 |
| 52 | .737 | 1.007 | 2.263 | 2.214 | .079 | .112 |

## TABLE 3

**Item parameter estimates for the Focal (females) and Base (males) groups: Test C**

| Item # | a F | a B | b F | b B | c F | c B |
|---|---|---|---|---|---|---|
| 1 | .759 | .723 | -1.192 | - .974 | .128 | .170 |
| 2 | .628 | .609 | .026 | - .075 | .198 | .118 |
| 3 | .975 | .897 | - .861 | - .685 | .099 | .077 |
| 4 | .755 | .844 | - .420 | -1.085 | .214 | .158 |
| 5 | .878 | .813 | - .335 | - .270 | .302 | .212 |
| 6 | .872 | .894 | .014 | - .139 | .221 | .224 |
| 7 | 1.108 | 1.228 | .121 | - .131 | .185 | .226 |
| 8 | .750 | .832 | - .454 | - .311 | .077 | .101 |
| 9 | .502 | .761 | .665 | .238 | .203 | .236 |
| 10 | .861 | 1.024 | .134 | - .162 | .218 | .160 |
| 11 | 1.276 | 1.173 | .223 | .247 | .212 | .197 |
| 12 | .881 | 1.030 | - .383 | .029 | .166 | .252 |
| 13 | .617 | .708 | .023 | - .290 | .151 | .155 |
| 14 | 1.315 | 1.386 | - .322 | - .106 | .070 | .117 |
| 15 | 1.322 | 1.264 | .210 | - .010 | .126 | .152 |
| 16 | .824 | .804 | .021 | .031 | .188 | .123 |
| 17 | .887 | .965 | .018 | .206 | .098 | .105 |
| 18 | 1.039 | 1.199 | - .086 | .272 | .096 | .113 |
| 19 | 1.010 | 1.073 | .373 | .402 | .110 | .129 |
| 20 | 1.305 | 1.060 | - .088 | .083 | .098 | .083 |
| 21 | .629 | .935 | .449 | .402 | .102 | .146 |
| 22 | .958 | 1.122 | .475 | .894 | .145 | .194 |
| 23 | 1.044 | 1.595 | .604 | .680 | .214 | .263 |
| 24 | 1.062 | 1.187 | .579 | .577 | .092 | .087 |
| 25 | 1.118 | 1.214 | .105 | .299 | .058 | .061 |
| 26 | 1.429 | 1.512 | .526 | .666 | .209 | .234 |
| 27 | .743 | .666 | 1.692 | 1.646 | .249 | .222 |
| 28 | .791 | .957 | 1.150 | .672 | .132 | .167 |
| 29 | 1.167 | .911 | .845 | .678 | .147 | .129 |
| 30 | .805 | .934 | .605 | .670 | .079 | .060 |
| 31 | .922 | 1.153 | 1.199 | .997 | .155 | .205 |
| 32 | .791 | 1.126 | 1.679 | 1.127 | .140 | .186 |
| 33 | 1.056 | 1.483 | 1.428 | 1.378 | .167 | .166 |
| 34 | .598 | .790 | 1.833 | 1.902 | .112 | .146 |
| 35 | 1.057 | 1.114 | .796 | .935 | .122 | .142 |
| 36 | 1.470 | 1.913 | .945 | 1.109 | .155 | .193 |
| 37 | 1.477 | 1.383 | 1.303 | 1.109 | .208 | .200 |
| 38 | 1.134 | 1.063 | .908 | 1.028 | .142 | .148 |
| 39 | 1.307 | 1.125 | 1.279 | 1.197 | .084 | .096 |
| 40 | 1.034 | .612 | 1.781 | 1.412 | .159 | .098 |

**TABLE 4**

**Asymptotic DIF for Test A**

| Item # | $\text{STD}_{\to\infty}(\theta)$ | $\text{MH}_{\to\infty}(\theta)$ $\alpha = .50$ | $\text{MH}_{\to\infty}(\theta)$ $\alpha = .09$ | $\text{RMWSD}_{\to\infty}(\theta)$ |
|---|---|---|---|---|
| 1 | -.054 | 1.341 | 1.365 | .063 |
| 2 | .028 | .777 | .808 | .048 |
| 3 | -.015 | 1.106 | 1.085 | .022 |
| 4 | -.070 | 1.349 | 1.388 | .077 |
| 5 | -.153 | 1.995 | 2.011 | .171 |
| 6 | .012 | .952 | .936 | .023 |
| 7 | .102 | .572 | .543 | .126 |
| 8 | .038 | .845 | .840 | .042 |
| 9 | -.027 | 1.177 | 1.159 | .034 |
| 10 | -.015 | 1.142 | 1.089 | .080 |
| 11 | -.065 | 1.324 | 1.348 | .068 |
| 12 | .003 | 1.003 | .993 | .059 |
| 13 | .029 | .893 | .886 | .031 |
| 14 | -.014 | 1.095 | 1.091 | .034 |
| 15 | .004 | 1.016 | .980 | .050 |
| 16 | -.034 | 1.260 | 1.227 | .064 |
| 17 | .055 | .686 | .697 | .062 |
| 18 | .025 | .775 | .829 | .081 |
| 19 | .095 | .358 | .363 | .126 |
| 20 | .009 | .950 | .942 | .041 |
| 21 | .050 | .763 | .754 | .058 |
| 22 | .018 | .921 | .920 | .021 |
| 23 | -.045 | 1.272 | 1.293 | .065 |
| 24 | -.049 | 1.271 | 1.296 | .062 |
| 25 | .055 | .645 | .662 | .072 |
| 26 | -.022 | 1.098 | 1.115 | .039 |
| 27 | -.027 | 1.115 | 1.134 | .031 |
| 28 | .037 | .793 | .802 | .058 |
| 29 | -.060 | 1.301 | 1.338 | .106 |
| 30 | .004 | 1.001 | .987 | .044 |
| 31 | .030 | .827 | .836 | .046 |
| 32 | -.071 | 1.407 | 1.403 | .079 |
| 33 | .013 | .904 | .920 | .023 |
| 34 | .041 | .744 | .763 | .056 |
| 35 | -.020 | 1.108 | 1.129 | .042 |
| 36 | -.031 | 1.242 | 1.238 | .039 |
| 37 | -.008 | 1.018 | 1.058 | .033 |
| 38 | -.043 | 1.266 | 1.266 | .059 |
| 39 | -.009 | 1.034 | 1.055 | .048 |
| 40 | -.062 | 1.380 | 1.423 | .070 |
| True Total (F/B) | 1/1 | 1/0 | 1/1 | 4 |

**TABLE 5**

**Asymptotic DIF for Test B**

| Item # | $STD(\theta)_{\to\infty}$ | $MH(\theta)_{\to\infty}$ $\alpha = .50$ | $MH(\theta)_{\to\infty}$ $\alpha = .09$ | $RMWSD(\theta)_{\to\infty}$ |
|---|---|---|---|---|
| 1 | .022 | .876 | .893 | .037 |
| 2 | .049 | .758 | .746 | .059 |
| 3 | .011 | .928 | .944 | .030 |
| 4 | -.055 | 1.275 | 1.255 | .059 |
| 5 | .018 | .883 | .908 | .037 |
| 6 | -.022 | 1.069 | 1.099 | .047 |
| 7 | -.046 | 1.227 | 1.219 | .052 |
| 8 | .002 | .960 | .983 | .032 |
| 9 | .039 | .857 | .852 | .041 |
| 10 | -.029 | 1.178 | 1.167 | .048 |
| 11 | .011 | .973 | .955 | .036 |
| 12 | .061 | .781 | .774 | .065 |
| 13 | -.010 | 1.056 | 1.058 | .038 |
| 14 | -.004 | 1.049 | 1.028 | .036 |
| 15 | -.006 | 1.025 | 1.027 | .009 |
| 16 | -.034 | 1.225 | 1.199 | .050 |
| 17 | -.027 | 1.168 | 1.157 | .031 |
| 18 | -.036 | 1.165 | 1.174 | .039 |
| 19 | .008 | .997 | .951 | .037 |
| 20 | .006 | .957 | .963 | .006 |
| 21 | -.001 | 1.009 | 1.006 | .012 |
| 22 | .011 | .940 | .950 | .018 |
| 23 | .113 | .529 | .567 | .142 |
| 24 | -.049 | 1.206 | 1.241 | .060 |
| 25 | .019 | .897 | .907 | .027 |
| 26 | .006 | .951 | .968 | .025 |
| 27 | -.052 | 1.239 | 1.246 | .054 |
| 28 | .059 | .730 | .754 | .072 |
| 29 | .012 | .936 | .934 | .013 |
| 30 | -.078 | 1.344 | 1.382 | .086 |
| 31 | .061 | .716 | .739 | .074 |
| 32 | -.025 | 1.101 | 1.110 | .028 |
| 33 | -.016 | 1.030 | 1.062 | .039 |
| 34 | -.129 | 1.853 | 1.861 | .135 |
| 35 | -.012 | 1.058 | 1.055 | .019 |
| 36 | -.008 | 1.020 | 1.039 | .032 |
| 37 | -.058 | 1.313 | 1.340 | .059 |
| 38 | -.019 | 1.056 | 1.080 | .033 |
| 39 | -.011 | 1.016 | 1.048 | .035 |
| 40 | -.026 | 1.082 | 1.144 | .057 |
| 41 | -.019 | 1.094 | 1.107 | .021 |
| 42 | -.033 | 1.215 | 1.246 | .042 |

(table continues)

| Item # | $STD_{\to\infty}(\theta)$ | $MH_{\to\infty}(\theta)$ $\alpha = .50$ | $MH_{\to\infty}(\theta)$ $\alpha = .09$ | $RMWSD_{\to\infty}(\theta)$ |
|---|---|---|---|---|
| 43 | .003 | .935 | .975 | .048 |
| 44 | -.095 | 1.481 | 1.530 | .107 |
| 45 | -.086 | 1.436 | 1.476 | .094 |
| 46 | -.088 | 1.485 | 1.539 | .093 |
| 47 | -.093 | 1.444 | 1.504 | .102 |
| 48 | -.068 | 1.336 | 1.412 | .082 |
| 49 | -.081 | 1.495 | 1.548 | .088 |
| 50 | -.121 | 1.678 | 1.728 | .130 |
| 51 | -.097 | 1.634 | 1.686 | .099 |
| 52 | -.010 | 1.051 | 1.082 | .020 |
| True Total (F/B) | 1/2 | 0/0 | 0/0 | 5 |

**TABLE 6**

**Asymptotic DIF for Test C**

| Item # | $\underset{\to\infty}{\text{STD}}(\theta)$ | $\underset{\to\infty}{\text{MH}}(\theta)$ | $\underset{\to\infty}{\text{RMWSD}}(\theta)$ |
|---|---|---|---|
| 1 | .030 | .812 | .032 |
| 2 | .022 | .903 | .026 |
| 3 | .051 | .706 | .056 |
| **4** | **-.110** | **2.025** | **.119** |
| 5 | .053 | .743 | .056 |
| 6 | -.033 | 1.190 | .035 |
| 7 | -.081 | 1.577 | .087 |
| 8 | .017 | .914 | .023 |
| 9 | -.068 | 1.360 | .084 |
| 10 | -.037 | 1.217 | .059 |
| 11 | .012 | .935 | .015 |
| 12 | .047 | .774 | .059 |
| 13 | -.063 | 1.366 | .067 |
| 14 | .040 | .766 | .054 |
| 15 | -.074 | 1.545 | .082 |
| 16 | .035 | .836 | .038 |
| 17 | .043 | .795 | .048 |
| 18 | .087 | .607 | .100 |
| 19 | -.002 | 1.011 | .006 |
| 20 | .055 | .721 | .065 |
| 21 | -.016 | 1.079 | .047 |
| 22 | .064 | .723 | .083 |
| 23 | .003 | .987 | .037 |
| 24 | .010 | .946 | .019 |
| 25 | .055 | .717 | .062 |
| 26 | .017 | .914 | .032 |
| 27 | .005 | .977 | .010 |
| 28 | -.097 | 1.629 | .108 |
| 29 | -.040 | 1.238 | .051 |
| 30 | .038 | .812 | .045 |
| 31 | -.054 | 1.318 | .059 |
| 32 | -.081 | 1.537 | .099 |
| 33 | .020 | .894 | .034 |
| 34 | .015 | .917 | .024 |
| 35 | .018 | .907 | .028 |
| 36 | .014 | .921 | .046 |
| 37 | -.028 | 1.161 | .043 |
| 38 | .014 | .928 | .024 |
| 39 | -.037 | 1.292 | .042 |
| 40 | -.058 | 1.379 | .077 |
| True Total (F/B) | 0/1 | 0/1 | 3 |

*Note: The value of $\underset{\to\infty}{\text{MH}}(\theta)$ did not change for different $\alpha$ values because $g_F(\theta)$ and $g_B(\theta)$ were identical.

TABLE 7

Average Rank Correlation Coefficients for Test A

| Sample Size F/B | MH | STD | RMWSD |
|---|---|---|---|
| 2000:2000 | .95 | .95 | .57 |
| 1000:1000 | .91 | .90 | .40 |
| 500:500 | .84 | .83 | .22 |
| 250:250 | .74 | .72 | .08 |
| 100:100 | .55 | .51 | .01 |
| 50:50 | .41 | .37 | .01 |
| 200:2000 | .81 | .80 | .17 |
| 100:1000 | .70 | .69 | .07 |
| 50:500 | .56 | .55 | .03 |

TABLE 8

Average Rank Correlation Coefficients for Test B

| Sample Size F/B | MH | STD | RMWSD |
|---|---|---|---|
| 2000:2000 | .93 | .93 | .52 |
| 1000:1000 | .89 | .89 | .41 |
| 500:500 | .82 | .81 | .28 |
| 250:250 | .70 | .68 | .16 |
| 100:100 | .51 | .52 | .05 |
| 50:50 | .37 | .44 | -.05 |
| 200:2000 | .78 | .78 | .30 |
| 100:1000 | .67 | .67 | .18 |
| 50:500 | .54 | .54 | .10 |

TABLE 9

Average Rank Correlation Coefficients for Test C

| Sample Size F/B | MH | STD | RMWSD |
|---|---|---|---|
| 2000:2000 | .95 | .94 | .66 |
| 1000:1000 | .91 | .90 | .45 |
| 500:500 | .85 | .84 | .30 |
| 250:250 | .76 | .75 | .12 |
| 100:100 | .59 | .59 | .04 |
| 50:50 | .42 | .46 | .05 |
| 200:2000 | .82 | .81 | .19 |
| 100:1000 | .71 | .70 | .12 |
| 50:500 | .60 | .59 | .06 |

TABLE 10

Average (Across Items) DIF by Test

| Sample Size Ratio (F:B) | Test A MH | Test A STD | Test A RMWSD | Test B MH | Test B STD | Test B RMWSD | Test C MH | Test C STD | Test C RMWSD |
|---|---|---|---|---|---|---|---|---|---|
| → ∞(α=.50) | **1.043** | **-.006** | **.059** | **1.110** | **-.020** | **.053** | **1.052** | **-.003** | **.052** |
| 2000:2000 | 1.021 | .000 | .095 | 1.026 | .001 | .109 | 1.016 | .000 | .094 |
| 1000:1000 | 1.025 | .000 | .122 | 1.029 | .001 | .144 | 1.045 | .000 | .122 |
| 500:500 | 1.032 | .001 | .165 | 1.037 | .004 | .195 | 1.051 | .001 | .165 |
| 250:250 | 1.045 | .004 | .230 | 1.050 | .011 | .265 | 1.066 | .004 | .236 |
| 100:100 | 1.101 | .026 | .365 | 1.103 | .043 | .398 | 1.118 | .039 | .389 |
| 50:50 | 1.280 | .089 | .468 | 1.270 | .110 | .497 | 1.291 | .125 | .512 |
| → ∞(α=.09) | **1.049** | **-.006** | **.059** | **1.127** | **-.020** | **.053** | **1.052** | **-.003** | **.052** |
| 200:2000 | 1.037 | .000 | .182 | 1.045 | .001 | .201 | 1.056 | .000 | .184 |
| 100:1000 | 1.061 | .000 | .242 | 1.066 | .002 | .268 | 1.077 | .000 | .246 |
| 50:500 | 1.111 | .001 | .308 | 1.111 | .004 | .340 | 1.120 | .001 | .319 |

TABLE 11

**Average (Across Items) DIF Standard Deviations by Test**

| Sample Size Ratio (F:B) | Test A | | | Test B | | | Test C | | |
|---|---|---|---|---|---|---|---|---|---|
| | MH | STD | RMWSD | MH | STD | RMWSD | MH | STD | RMWSD |
| 2000:2000 | .079 | .014 | .012 | .077 | .016 | .014 | .079 | .014 | .011 |
| 1000:1000 | .116 | .020 | .016 | .106 | .022 | .019 | .113 | .019 | .016 |
| 500:500 | .166 | .028 | .022 | .159 | .032 | .027 | .161 | .028 | .021 |
| 250:250 | .246 | .041 | .033 | .237 | .047 | .035 | .236 | .040 | .032 |
| 100:100 | .451 | .072 | .051 | .432 | .077 | .051 | .444 | .073 | .052 |
| 50:50 | 1.031 | .098 | .069 | .958 | .103 | .066 | 1.044 | .106 | .069 |
| 200:2000 | .190 | .031 | .023 | .181 | .033 | .024 | .194 | .032 | .024 |
| 100:1000 | .291 | .044 | .032 | .273 | .047 | .033 | .289 | .045 | .032 |
| 50:500 | .463 | .063 | .043 | .449 | .070 | .046 | .431 | .063 | .043 |

TABLE 12

Average Relative Efficiency of DIF Estimates Between Adjacent Sample Sizes

| Efficiency Ratio | Test A | | | Test B | | | Test C | | |
|---|---|---|---|---|---|---|---|---|---|
| | MH | STD | RMWSD | MH | STD | RMWSD | MH | STD | RMWSD |
| 1000/2000 | 1.09 | 1.10 | .93 | .97 | .97 | 1.00 | 1.02 | 1.03 | .96 |
| 500/1000 | 1.04 | 1.00 | .99 | 1.12 | 1.09 | 1.01 | 1.05 | 1.03 | .94 |
| 250/500 | 1.16 | 1.10 | 1.13 | 1.14 | 1.10 | .89 | 1.09 | 1.10 | 1.14 |
| 100/250 | 1.42 | 1.24 | 1.03 | 1.37 | 1.07 | .85 | 1.44 | 1.32 | 1.08 |
| 50/100 | 2.88 | .95 | .93 | 2.64 | .90 | .86 | 3.22 | 1.07 | .90 |

TABLE 13

DIF Identification (F/B) for the MH Statistic: Test A

| | Sample Size Ratio (F:B) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
| 1 | 0/0 | 0/0 | 0/0 | 0/8 | 1/16 | 7/26 | 0/0 | 1/2 | 1/7 |
| 2 | 0/0 | 1/0 | 3/0 | 11/0 | 21/2 | 39/6 | 2/0 | 6/0 | 14/0 |
| 3 | 0/0 | 0/0 | 0/0 | 0/0 | 2/3 | 11/8 | 0/0 | 0/1 | 2/3 |
| 4 | 0/0 | 0/0 | 0/0 | 0/4 | 1/18 | 3/21 | 0/1 | 0/9 | 1/7 |
| 5 | 0/35 | 0/39 | 0/43 | 0/46 | 0/57 | 0/58 | 0/48 | 0/49 | 0/58 |
| 6 | 0/0 | 0/0 | 0/0 | 0/0 | 10/0 | 11/10 | 0/0 | 1/0 | 7/1 |
| 7 | 19/0 | 23/0 | 37/0 | 34/0 | 46/0 | 41/1 | 50/0 | 47/0 | 54/0 |
| 8 | 0/0 | 0/0 | 0/0 | 2/0 | 7/0 | 21/8 | 0/0 | 3/0 | 8/0 |
| 9 | 0/0 | 0/0 | 0/0 | 0/1 | 0/3 | 2/19 | 0/0 | 0/1 | 1/7 |
| 10 | 0/0 | 0/0 | 0/0 | 0/0 | 3/4 | 9/14 | 0/0 | 0/2 | 0/3 |
| 11 | 0/0 | 0/0 | 0/1 | 0/1 | 0/15 | 7/20 | 0/1 | 0/7 | 2/10 |
| 12 | 0/0 | 0/0 | 0/0 | 0/0 | 8/4 | 17/8 | 0/0 | 0/0 | 2/3 |
| 13 | 0/0 | 0/0 | 0/0 | 1/0 | 2/0 | 17/6 | 0/0 | 2/0 | 9/0 |
| 14 | 0/0 | 0/0 | 0/0 | 0/0 | 3/0 | 8/16 | 0/0 | 2/0 | 2/8 |
| 15 | 0/0 | 0/0 | 0/0 | 1/0 | 5/3 | 14/7 | 0/0 | 2/1 | 6/2 |
| 16 | 0/0 | 0/0 | 0/0 | 0/2 | 1/8 | 13/20 | 0/0 | 0/4 | 1/9 |
| 17 | 0/0 | 6/0 | 5/0 | 12/0 | 30/1 | 32/2 | 4/0 | 17/0 | 21/1 |
| 18 | 0/0 | 1/0 | 1/0 | 5/0 | 22/2 | 27/8 | 2/0 | 2/0 | 10/1 |
| 19 | 100/0 | 97/0 | 95/0 | 92/0 | 76/0 | 71/0 | 91/0 | 83/0 | 73/0 |
| 20 | 0/0 | 0/0 | 0/0 | 0/0 | 9/4 | 17/7 | 0/0 | 1/0 | 10/2 |
| 21 | 0/0 | 0/0 | 1/0 | 5/0 | 16/1 | 32/3 | 3/0 | 10/0 | 15/0 |
| 22 | 0/0 | 0/0 | 0/0 | 0/0 | 7/1 | 19/10 | 0/0 | 3/0 | 5/1 |
| 23 | 0/0 | 0/0 | 0/1 | 0/4 | 1/14 | 3/26 | 0/1 | 0/5 | 2/13 |
| 24 | 0/0 | 0/0 | 0/0 | 0/4 | 0/13 | 5/17 | 0/0 | 0/6 | 2/14 |
| 25 | 0/0 | 1/0 | 12/0 | 26/0 | 29/0 | 33/2 | 13/0 | 18/0 | 31/0 |
| 26 | 0/0 | 0/0 | 0/0 | 0/0 | 1/4 | 9/17 | 0/0 | 0/0 | 1/1 |
| 27 | 0/0 | 0/0 | 0/0 | 0/0 | 1/5 | 11/19 | 0/0 | 0/0 | 1/4 |
| 28 | 0/0 | 0/0 | 0/0 | 4/0 | 14/0 | 30/9 | 1/0 | 5/0 | 10/0 |
| 29 | 0/0 | 0/0 | 0/0 | 0/1 | 0/11 | 6/29 | 0/1 | 0/4 | 2/20 |
| 30 | 0/0 | 0/0 | 0/0 | 0/0 | 4/4 | 14/15 | 0/0 | 0/0 | 2/3 |
| 31 | 0/0 | 0/0 | 0/0 | 2/0 | 3/0 | 25/9 | 1/0 | 1/0 | 10/2 |
| 32 | 0/0 | 0/0 | 0/2 | 0/5 | 1/15 | 4/22 | 0/0 | 0/10 | 0/16 |
| 33 | 0/0 | 0/0 | 0/0 | 1/0 | 6/5 | 20/9 | 0/0 | 0/0 | 6/6 |
| 34 | 0/0 | 0/0 | 0/0 | 5/0 | 23/1 | 20/6 | 0/0 | 9/0 | 12/2 |
| 35 | 0/0 | 0/0 | 0/0 | 0/0 | 4/7 | 10/18 | 0/0 | 0/2 | 3/10 |
| 36 | 0/0 | 0/0 | 0/1 | 0/1 | 2/7 | 12/18 | 0/1 | 0/10 | 2/10 |
| 37 | 0/0 | 0/0 | 0/0 | 0/1 | 5/3 | 14/20 | 0/1 | 1/4 | 4/11 |
| 38 | 0/0 | 0/0 | 0/0 | 0/2 | 0/14 | 5/26 | 0/0 | 0/5 | 2/11 |
| 39 | 0/0 | 0/0 | 0/0 | 0/3 | 3/5 | 11/19 | 0/0 | 0/2 | 2/9 |
| 40 | 0/0 | 0/0 | 0/2 | 0/4 | 1/13 | 6/28 | 0/3 | 0/10 | 1/27 |
| AVE. per Test | 1.2/ .4 | 1.3/ .4 | 1.4/ .5 | 2.0/ .9 | 3.7/ 2.6 | 6.6/ 5.9 | 1.7/ .6 | 2.2/ 1.0 | 2.6/ 2.6 |

TABLE 14

DIF Identification (F/B) for the STD Statistic: Test A

| | Sample Size Ratio (F:B) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
| 1 | 0/0 | 0/2 | 0/3 | 0/15 | 8/8 | 54/0 | 0/5 | 0/10 | 1/15 |
| 2 | 0/0 | 0/0 | 1/0 | 9/0 | 40/0 | 79/0 | 1/0 | 4/0 | 11/0 |
| 3 | 0/0 | 0/0 | 0/0 | 1/1 | 19/3 | 65/1 | 0/0 | 0/3 | 3/9 |
| 4 | 0/0 | 0/5 | 0/3 | 0/17 | 5/25 | 30/6 | 0/8 | 0/17 | 3/18 |
| 5 | 0/100 | 0/97 | 0/91 | 0/74 | 2/56 | 16/31 | 0/91 | 0/83 | 0/75 |
| 6 | 0/0 | 0/0 | 0/0 | 0/0 | 28/1 | 62/2 | 1/0 | 1/0 | 11/2 |
| 7 | 78/0 | 59/0 | 62/0 | 56/0 | 73/0 | 92/0 | 63/0 | 66/0 | 60/0 |
| 8 | 0/0 | 1/0 | 5/0 | 15/0 | 33/0 | 73/1 | 5/0 | 12/0 | 22/1 |
| 9 | 0/0 | 0/0 | 0/0 | 1/4 | 8/4 | 44/3 | 0/0 | 1/3 | 1/13 |
| 10 | 0/0 | 0/0 | 0/0 | 0/1 | 16/7 | 55/3 | 0/0 | 3/6 | 6/14 |
| 11 | 0/1 | 0/5 | 0/9 | 0/13 | 7/19 | 31/8 | 0/9 | 0/23 | 5/21 |
| 12 | 0/0 | 0/0 | 0/0 | 0/0 | 26/3 | 54/3 | 0/0 | 0/1 | 4/4 |
| 13 | 0/0 | 1/0 | 3/0 | 15/1 | 41/3 | 62/1 | 8/0 | 18/0 | 26/2 |
| 14 | 0/0 | 0/0 | 0/0 | 1/0 | 12/2 | 28/3 | 0/0 | 2/1 | 3/10 |
| 15 | 0/0 | 0/0 | 0/0 | 1/0 | 19/6 | 60/3 | 0/0 | 2/2 | 10/1 |
| 16 | 0/0 | 0/0 | 0/0 | 0/4 | 8/8 | 40/6 | 0/1 | 1/4 | 2/12 |
| 17 | 0/0 | 7/0 | 5/0 | 19/0 | 48/0 | 68/0 | 9/0 | 17/0 | 23/0 |
| 18 | 0/0 | 0/0 | 0/0 | 2/0 | 32/1 | 58/1 | 0/0 | 3/0 | 13/0 |
| 19 | 14/0 | 27/0 | 34/0 | 52/0 | 56/0 | 65/0 | 30/0 | 46/0 | 42/0 |
| 20 | 0/0 | 0/0 | 0/0 | 2/0 | 18/3 | 57/3 | 0/0 | 1/0 | 12/1 |
| 21 | 0/0 | 2/0 | 7/0 | 12/0 | 39/0 | 61/0 | 8/0 | 17/0 | 21/0 |
| 22 | 0/0 | 0/0 | 1/0 | 6/0 | 43/2 | 66/2 | 1/0 | 7/0 | 14/2 |
| 23 | 0/0 | 0/0 | 0/2 | 0/8 | 3/8 | 26/12 | 0/1 | 0/5 | 3/14 |
| 24 | 0/0 | 0/0 | 0/3 | 0/11 | 4/13 | 31/9 | 0/1 | 0/9 | 2/19 |
| 25 | 0/0 | 0/0 | 4/0 | 18/0 | 31/0 | 62/0 | 7/0 | 13/0 | 22/0 |
| 26 | 0/0 | 0/0 | 0/2 | 0/3 | 8/10 | 34/3 | 0/0 | 1/1 | 6/5 |
| 27 | 0/0 | 0/0 | 0/0 | 1/5 | 4/9 | 29/7 | 0/1 | 0/2 | 4/15 |
| 28 | 0/0 | 0/0 | 2/0 | 6/0 | 34/0 | 55/2 | 4/0 | 9/0 | 13/0 |
| 29 | 0/0 | 0/1 | 0/3 | 0/10 | 6/15 | 24/8 | 0/7 | 0/18 | 2/28 |
| 30 | 0/0 | 0/0 | 0/0 | 2/1 | 16/5 | 52/6 | 0/0 | 6/0 | 7/3 |
| 31 | 0/0 | 0/0 | 0/0 | 7/0 | 20/0 | 48/1 | 4/0 | 3/0 | 14/0 |
| 32 | 0/2 | 0/3 | 0/13 | 0/16 | 5/24 | 21/17 | 0/15 | 0/26 | 0/29 |
| 33 | 0/0 | 0/0 | 0/0 | 3/0 | 11/9 | 44/1 | 0/0 | 0/0 | 9/3 |
| 34 | 0/0 | 0/0 | 0/0 | 5/0 | 29/3 | 44/2 | 3/0 | 13/0 | 13/2 |
| 35 | 0/0 | 0/0 | 0/1 | 0/0 | 6/5 | 30/4 | 0/0 | 0/1 | 3/10 |
| 36 | 0/0 | 0/0 | 0/1 | 0/0 | 6/0 | 24/5 | 0/0 | 0/5 | 3/4 |
| 37 | 0/0 | 0/0 | 0/0 | 0/0 | 6/0 | 23/9 | 0/0 | 1/0 | 1/4 |
| 38 | 0/0 | 0/0 | 0/0 | 0/6 | 4/17 | 20/12 | 0/2 | 1/9 | 2/14 |
| 39 | 0/0 | 0/0 | 0/0 | 0/3 | 10/4 | 25/3 | 0/0 | 2/1 | 5/7 |
| 40 | 0/0 | 0/3 | 0/6 | 0/11 | 1/9 | 23/8 | 0/5 | 0/20 | 1/28 |
| AVE. per Test | .9/ 1.0 | 1.0/ 1.2 | 1.2/ 1.4 | 2.4/ 2.0 | 7.8/ 2.8 | 18.4/ 1.9 | 1.4/ 1.5 | 2.5/ 2.5 | 4.0/ 3.8 |

TABLE 15

DIF Identification for the RMWSD Statistic: Test A

| | Sample Size Ratio (F:B) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
| 1 | 40 | 94 | 100 | 100 | 100 | 100 | 99 | 100 | 100 |
| 2 | 5 | 51 | 95 | 100 | 100 | 100 | 99 | 100 | 100 |
| 3 | 11 | 86 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 4 | 61 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 5 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 6 | 3 | 60 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 7 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | 45 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 9 | 4 | 77 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 10 | 52 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 11 | 47 | 96 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 12 | 4 | 75 | 98 | 100 | 100 | 100 | 100 | 100 | 100 |
| 13 | 57 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 14 | 0 | 67 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 15 | 11 | 75 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 16 | 9 | 76 | 99 | 100 | 100 | 100 | 100 | 100 | 100 |
| 17 | 33 | 94 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 18 | 43 | 82 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 19 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 20 | 5 | 80 | 99 | 100 | 100 | 100 | 100 | 100 | 100 |
| 21 | 17 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 22 | 17 | 93 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 23 | 13 | 86 | 99 | 100 | 100 | 100 | 100 | 100 | 100 |
| 24 | 21 | 94 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 25 | 53 | 92 | 98 | 100 | 100 | 100 | 99 | 100 | 100 |
| 26 | 6 | 88 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 27 | 13 | 89 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 28 | 13 | 84 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 29 | 89 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 30 | 22 | 96 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 31 | 6 | 78 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 32 | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 33 | 2 | 69 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 34 | 26 | 83 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 35 | 1 | 66 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 36 | 11 | 82 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 37 | 0 | 57 | 97 | 100 | 100 | 100 | 100 | 100 | 100 |
| 38 | 29 | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 39 | 1 | 80 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 40 | 46 | 89 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| AVE. per Test | 12.1 | 34.2 | 39.8 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 |

**TABLE 16**

**Bias Identification (F/B) for the MH Statistic: Test B**

| | Sample Size Ratio (F:B) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
| 1 | 0/0 | 0/0 | 0/0 | 4/0 | 8/1 | 23/5 | 0/0 | 7/0 | 9/1 |
| 2 | 0/0 | 1/0 | 5/0 | 9/0 | 25/0 | 30/6 | 10/0 | 15/0 | 32/0 |
| 3 | 0/0 | 0/0 | 1/0 | 1/0 | 10/0 | 22/7 | 1/0 | 1/0 | 12/1 |
| 4 | 0/0 | 0/0 | 0/0 | 0/0 | 2/9 | 4/13 | 0/0 | 0/2 | 0/6 |
| 5 | 0/0 | 0/0 | 0/0 | 2/0 | 8/1 | 22/5 | 0/0 | 2/0 | 10/3 |
| 6 | 0/0 | 0/0 | 0/0 | 0/0 | 2/3 | 14/5 | 0/0 | 0/0 | 1/2 |
| 7 | 0/0 | 0/0 | 0/0 | 0/1 | 0/4 | 5/15 | 0/0 | 0/1 | 0/6 |
| 8 | 0/0 | 0/0 | 0/0 | 0/0 | 4/1 | 15/6 | 0/0 | 1/0 | 7/1 |
| 9 | 0/0 | 0/0 | 0/0 | 2/0 | 12/0 | 13/6 | 1/0 | 2/0 | 5/0 |
| 10 | 0/0 | 0/0 | 0/0 | 0/2 | 2/3 | 9/21 | 0/0 | 1/0 | 5/4 |
| 11 | 0/0 | 0/0 | 0/0 | 1/0 | 8/1 | 18/5 | 0/0 | 2/0 | 14/1 |
| 12 | 0/0 | 0/0 | 2/0 | 6/0 | 14/0 | 30/2 | 2/0 | 6/0 | 14/0 |
| 13 | 0/0 | 0/0 | 0/0 | 0/0 | 7/1 | 14/8 | 1/0 | 3/0 | 9/4 |
| 14 | 0/0 | 0/0 | 0/0 | 0/0 | 5/1 | 10/14 | 0/0 | 0/0 | 7/1 |
| 15 | 0/0 | 0/0 | 0/0 | 0/0 | 5/4 | 14/11 | 0/0 | 0/0 | 2/1 |
| 16 | 0/0 | 0/0 | 0/0 | 0/0 | 2/7 | 7/12 | 0/0 | 0/0 | 0/10 |
| 17 | 0/0 | 0/0 | 0/0 | 0/0 | 4/1 | 13/14 | 0/0 | 0/0 | 4/1 |
| 18 | 0/0 | 0/0 | 0/0 | 0/0 | 2/2 | 6/15 | 0/0 | 0/1 | 3/5 |
| 19 | 0/0 | 0/0 | 0/0 | 1/0 | 17/7 | 35/19 | 1/0 | 6/0 | 11/3 |
| 20 | 0/0 | 0/0 | 1/0 | 2/0 | 13/3 | 28/9 | 0/0 | 6/1 | 16/1 |
| 21 | 0/0 | 0/0 | 0/0 | 0/0 | 9/1 | 15/4 | 0/0 | 0/0 | 4/2 |
| 22 | 0/0 | 0/0 | 0/0 | 2/0 | 6/0 | 14/6 | 0/0 | 0/0 | 9/1 |
| 23 | 76/0 | 67/0 | 63/0 | 70/0 | 53/0 | 57/1 | 50/0 | 54/0 | 55/0 |
| 24 | 0/0 | 0/0 | 0/0 | 0/1 | 0/3 | 6/8 | 0/0 | 0/1 | 0/8 |
| 25 | 0/0 | 0/0 | 0/0 | 4/0 | 4/3 | 18/4 | 0/0 | 6/0 | 7/1 |
| 26 | 0/0 | 0/0 | 0/0 | 1/0 | 6/1 | 17/5 | 0/0 | 2/0 | 3/0 |
| 27 | 0/0 | 0/0 | 0/0 | 0/0 | 1/7 | 6/16 | 0/0 | 0/1 | 0/2 |
| 28 | 0/0 | 1/0 | 3/0 | 7/0 | 25/0 | 33/1 | 7/0 | 10/0 | 25/0 |
| 29 | 0/0 | 0/0 | 0/0 | 2/0 | 17/3 | 24/10 | 1/0 | 3/0 | 11/1 |
| 30 | 0/0 | 0/0 | 0/0 | 0/0 | 1/8 | 3/17 | 0/0 | 0/2 | 1/6 |
| 31 | 0/0 | 0/0 | 6/0 | 12/0 | 19/0 | 39/2 | 6/0 | 16/0 | 23/1 |
| 32 | 0/0 | 0/0 | 0/0 | 0/0 | 2/9 | 14/15 | 0/0 | 0/1 | 1/3 |
| 33 | 0/0 | 0/0 | 0/0 | 1/0 | 5/2 | 12/9 | 0/0 | 1/0 | 1/1 |
| 34 | 0/2 | 0/5 | 0/17 | 0/19 | 0/37 | 1/44 | 0/17 | 0/36 | 0/29 |
| 35 | 0/0 | 0/0 | 0/0 | 0/0 | 1/3 | 11/20 | 0/0 | 0/0 | 5/4 |
| 36 | 0/0 | 0/0 | 0/0 | 1/0 | 3/2 | 16/8 | 0/0 | 0/0 | 5/4 |
| 37 | 0/0 | 0/0 | 0/0 | 0/1 | 0/5 | 11/21 | 0/1 | 0/2 | 1/10 |
| 38 | 0/0 | 0/0 | 0/0 | 0/0 | 2/2 | 15/9 | 0/0 | 0/1 | 0/1 |
| 39 | 0/0 | 0/0 | 0/0 | 1/0 | 6/0 | 16/11 | 0/0 | 0/0 | 6/5 |
| 40 | 0/0 | 0/0 | 0/0 | 1/0 | 5/5 | 11/21 | 0/0 | 0/1 | 2/8 |
| 41 | 0/0 | 0/0 | 0/0 | 0/0 | 2/10 | 17/15 | 0/0 | 1/1 | 0/8 |
| 42 | 0/0 | 0/0 | 0/0 | 0/1 | 3/13 | 9/25 | 0/0 | 0/4 | 0/12 |
| 43 | 0/0 | 0/0 | 0/0 | 0/0 | 7/1 | 17/7 | 0/0 | 0/0 | 6/4 |

(table continues)

| Item # | Sample Size Ratio (F:B) 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
|---|---|---|---|---|---|---|---|---|---|
| 44 | 0/0 | 0/0 | 0/1 | 0/6 | 0/7 | 5/26 | 0/1 | 0/9 | 0/14 |
| 45 | 0/0 | 0/0 | 0/1 | 0/2 | 0/8 | 2/25 | 0/1 | 0/6 | 1/11 |
| 46 | 0/0 | 0/0 | 0/2 | 0/3 | 0/26 | 0/29 | 0/0 | 0/11 | 0/17 |
| 47 | 0/0 | 0/0 | 0/0 | 0/6 | 1/15 | 2/23 | 0/0 | 0/10 | 0/16 |
| 48 | 0/0 | 0/0 | 0/1 | 0/1 | 2/14 | 9/19 | 0/1 | 0/9 | 1/13 |
| 49 | 0/0 | 0/0 | 0/0 | 0/4 | 0/18 | 4/27 | 0/2 | 0/9 | 0/27 |
| 50 | 0/0 | 0/0 | 0/8 | 0/9 | 0/21 | 2/41 | 0/6 | 0/11 | 0/25 |
| 51 | 0/0 | 0/0 | 0/3 | 0/12 | 0/24 | 4/38 | 0/11 | 0/27 | 0/25 |
| 52 | 0/0 | 0/0 | 0/0 | 1/0 | 5/12 | 21/19 | 0/0 | 1/8 | 3/11 |
| AVE. per Test | .8/ .0 | .7/ .1 | .8/ .3 | 1.3/ .7 | 3.3/ 3.1 | 7.6/ 7.2 | .8/ .4 | 1.2/ 1.6 | 3.3/ 3.2 |

TABLE 17

DIF Identification (F/B) for the STD Statistic: Test B

| | Sample Size Ratio (F:B) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
| 1 | 0/0 | 0/0 | 7/0 | 18/0 | 44/3 | 70/0 | 4/0 | 19/1 | 14/2 |
| 2 | 8/0 | 20/0 | 29/0 | 37/0 | 72/0 | 86/0 | 25/0 | 26/0 | 38/1 |
| 3 | 0/0 | 0/0 | 5/0 | 14/0 | 53/1 | 79/2 | 1/0 | 5/1 | 22/1 |
| 4 | 0/0 | 0/1 | 0/5 | 3/3 | 14/9 | 44/4 | 0/3 | 0/6 | 4/27 |
| 5 | 0/0 | 0/0 | 3/0 | 12/0 | 28/3 | 51/3 | 3/0 | 8/1 | 19/5 |
| 6 | 0/0 | 0/0 | 0/0 | 2/2 | 17/2 | 56/2 | 0/0 | 5/2 | 8/5 |
| 7 | 0/0 | 0/0 | 0/2 | 2/7 | 15/8 | 48/1 | 0/3 | 1/7 | 5/18 |
| 8 | 0/0 | 0/0 | 0/0 | 15/0 | 36/1 | 70/0 | 0/1 | 7/1 | 13/5 |
| 9 | 0/0 | 2/0 | 15/0 | 24/0 | 57/0 | 66/1 | 5/0 | 23/0 | 33/1 |
| 10 | 0/0 | 0/0 | 0/0 | 4/2 | 26/1 | 61/2 | 0/0 | 2/0 | 7/8 |
| 11 | 0/0 | 0/0 | 1/0 | 14/0 | 38/0 | 76/0 | 7/0 | 10/0 | 27/4 |
| 12 | 23/0 | 29/0 | 36/0 | 48/0 | 62/1 | 86/1 | 30/0 | 36/0 | 35/0 |
| 13 | 0/0 | 0/0 | 0/0 | 4/0 | 29/1 | 60/1 | 1/0 | 5/0 | 15/8 |
| 14 | 0/0 | 0/0 | 0/0 | 9/0 | 32/1 | 63/2 | 2/0 | 4/0 | 13/5 |
| 15 | 0/0 | 0/0 | 2/0 | 11/0 | 33/5 | 69/1 | 2/0 | 5/0 | 14/2 |
| 16 | 0/0 | 0/0 | 0/0 | 0/1 | 10/6 | 37/3 | 0/0 | 0/1 | 1/14 |
| 17 | 0/0 | 0/0 | 0/0 | 3/1 | 30/1 | 71/2 | 0/1 | 3/0 | 8/5 |
| 18 | 0/0 | 0/0 | 0/1 | 1/0 | 30/4 | 53/2 | 0/0 | 3/5 | 6/13 |
| 19 | 0/0 | 0/0 | 1/0 | 6/0 | 48/0 | 86/0 | 0/0 | 5/0 | 15/2 |
| 20 | 0/0 | 0/0 | 2/0 | 11/0 | 47/0 | 86/0 | 1/0 | 9/0 | 17/4 |
| 21 | 0/0 | 0/0 | 2/0 | 6/0 | 45/2 | 76/1 | 1/0 | 4/0 | 13/3 |
| 22 | 0/0 | 0/0 | 5/0 | 17/0 | 38/2 | 66/1 | 4/0 | 12/3 | 20/2 |
| 23 | 95/0 | 92/0 | 84/0 | 86/0 | 85/0 | 91/0 | 81/0 | 72/0 | 75/0 |
| 24 | 0/0 | 0/0 | 0/2 | 0/3 | 10/7 | 40/2 | 0/2 | 0/4 | 2/15 |
| 25 | 0/0 | 0/0 | 0/0 | 12/0 | 27/4 | 55/0 | 2/0 | 14/0 | 18/1 |
| 26 | 0/0 | 0/0 | 0/0 | 8/1 | 27/0 | 59/2 | 1/0 | 6/0 | 8/1 |
| 27 | 0/0 | 0/0 | 0/5 | 0/11 | 10/9 | 47/6 | 0/1 | 1/13 | 4/16 |
| 28 | 11/0 | 20/0 | 36/0 | 35/0 | 58/1 | 86/0 | 32/0 | 34/0 | 41/1 |
| 29 | 0/0 | 0/0 | 5/0 | 17/0 | 56/1 | 89/0 | 2/0 | 9/0 | 16/2 |
| 30 | 0/0 | 0/3 | 0/5 | 0/12 | 14/9 | 51/8 | 0/7 | 0/17 | 4/20 |
| 31 | 10/0 | 21/0 | 36/0 | 43/0 | 63/1 | 74/0 | 27/0 | 34/0 | 44/2 |
| 32 | 0/0 | 0/0 | 0/0 | 3/2 | 22/6 | 47/6 | 0/1 | 2/2 | 6/11 |
| 33 | 0/0 | 0/0 | 0/0 | 4/1 | 26/4 | 55/1 | 1/0 | 6/2 | 8/4 |
| 34 | 0/74 | 0/54 | 0/58 | 0/41 | 0/28 | 20/14 | 0/53 | 0/61 | 0/48 |
| 35 | 0/0 | 0/0 | 0/0 | 2/1 | 25/4 | 53/3 | 1/0 | 4/1 | 10/6 |
| 36 | 0/0 | 0/0 | 0/0 | 8/0 | 29/3 | 39/5 | 0/0 | 3/1 | 10/9 |
| 37 | 0/0 | 0/1 | 0/2 | 0/7 | 3/9 | 29/6 | 0/4 | 0/9 | 4/16 |
| 38 | 0/0 | 0/0 | 1/0 | 4/0 | 20/3 | 60/4 | 1/0 | 6/4 | 5/8 |
| 39 | 0/0 | 0/0 | 1/0 | 1/0 | 26/1 | 38/3 | 0/0 | 2/0 | 11/7 |
| 40 | 0/0 | 0/0 | 0/0 | 1/2 | 7/9 | 29/7 | 0/0 | 1/2 | 5/9 |
| 41 | 0/0 | 0/0 | 0/0 | 1/0 | 9/5 | 38/6 | 0/0 | 3/2 | 4/11 |
| 42 | 0/0 | 0/0 | 0/0 | 0/1 | 6/6 | 21/8 | 0/0 | 0/2 | 1/8 |
| 43 | 0/0 | 0/0 | 1/0 | 4/0 | 25/3 | 52/4 | 0/0 | 1/1 | 20/8 |

(table continues)

| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
|---|---|---|---|---|---|---|---|---|---|
| | Sample Size Ratio (F:B) | | | | | | | | |
| 44 | 0/3 | 0/9 | 0/14 | 0/22 | 14/13 | 44/9 | 0/17 | 0/30 | 0/31 |
| 45 | 0/1 | 0/8 | 0/10 | 0/14 | 9/13 | 35/8 | 0/17 | 0/23 | 2/27 |
| 46 | 0/2 | 0/13 | 0/18 | 0/20 | 0/26 | 19/12 | 0/26 | 0/23 | 1/25 |
| 47 | 0/4 | 0/13 | 0/21 | 0/20 | 10/20 | 23/7 | 0/18 | 0/43 | 1/34 |
| 48 | 0/0 | 0/1 | 0/6 | 0/11 | 9/12 | 34/7 | 0/4 | 0/15 | 3/20 |
| 49 | 0/0 | 0/3 | 0/8 | 0/17 | 0/16 | 19/12 | 0/7 | 0/15 | 0/31 |
| 50 | 0/47 | 0/45 | 0/44 | 0/35 | 1/24 | 21/17 | 0/44 | 0/43 | 1/53 |
| 51 | 0/6 | 0/13 | 0/21 | 0/29 | 4/20 | 13/15 | 0/25 | 0/38 | 0/27 |
| 52 | 0/0 | 0/0 | 0/0 | 0/0 | 1/21 | 15/8 | 0/0 | 1/0 | 2/2 |
| AVE. per Test | 1.5/ 1.4 | 1.8/ 1.7 | 2.7/ 2.2 | 4.9/ 2.7 | 14.0/ 3.1 | 27.5/ 2.0 | 2.3/ 2.3 | 3.9/ 3.8 | 6.6/ 5.8 |

**TABLE 18**

**DIF Identification for the RMWSD Statistic: Test C**

| | Sample Size Ratio (F:B) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
| 1 | 69 | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 2 | 98 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 | 64 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 4 | 95 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 5 | 47 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 6 | 53 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 7 | 83 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | 49 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 9 | 100 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 10 | 33 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 11 | 61 | 96 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 12 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 13 | 18 | 84 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 14 | 50 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 15 | 66 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 16 | 36 | 96 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 17 | 44 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 18 | 55 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 19 | 63 | 93 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 20 | 41 | 99 | 100 | 100 | 100 | 100 | 99 | 100 | 100 |
| 21 | 55 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 22 | 68 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 23 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 24 | 53 | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 25 | 47 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 26 | 13 | 94 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 27 | 87 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 28 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 29 | 61 | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 30 | 88 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 31 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 32 | 58 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 33 | 57 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 34 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 35 | 41 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 36 | 34 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 37 | 68 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 38 | 54 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 39 | 29 | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 40 | 39 | 95 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 41 | 31 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 42 | 14 | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 43 | 53 | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

(table continues)

| Item # | Sample Size Ratio (F/B) 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
|---|---|---|---|---|---|---|---|---|---|
| 44 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 45 | 87 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 46 | 93 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 47 | 95 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 48 | 71 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 49 | 79 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 50 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 51 | 94 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 52 | 4 | 74 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| AVE. per Test | 32.9 | 50.9 | 52.0 | 52.0 | 52.0 | 52.0 | 52.0 | 52.0 | 52.0 |

**TABLE 19**

**DIF Identification (F/B) for the MH Statistic: Test C**

| | Sample Size Ratio (F:B) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
| 1 | 0/0 | 0/0 | 0/0 | 4/0 | 15/3 | 19/4 | 0/0 | 4/0 | 6/0 |
| 2 | 0/0 | 0/0 | 0/0 | 0/0 | 4/1 | 21/6 | 0/0 | 1/0 | 5/1 |
| 3 | 0/0 | 0/0 | 0/0 | 10/0 | 21/2 | 30/4 | 8/0 | 8/0 | 24/0 |
| 4 | 0/60 | 0/62 | 0/63 | 0/63 | 0/54 | 3/54 | 0/57 | 0/51 | 0/60 |
| 5 | 0/0 | 0/0 | 0/0 | 3/0 | 18/1 | 31/4 | 3/0 | 6/0 | 19/0 |
| 6 | 0/0 | 0/0 | 0/0 | 0/0 | 2/10 | 6/23 | 0/0 | 0/2 | 0/6 |
| 7 | 0/0 | 0/1 | 0/11 | 0/21 | 0/30 | 2/32 | 0/19 | 0/22 | 0/31 |
| 8 | 0/0 | 0/0 | 0/0 | 0/0 | 6/1 | 16/7 | 0/0 | 1/0 | 13/1 |
| 9 | 0/0 | 0/0 | 0/0 | 0/1 | 1/11 | 2/20 | 0/0 | 0/8 | 0/10 |
| 10 | 0/0 | 0/0 | 0/0 | 0/1 | 2/7 | 11/16 | 0/0 | 0/3 | 4/10 |
| 11 | 0/0 | 0/0 | 0/0 | 1/0 | 4/3 | 9/12 | 0/0 | 1/0 | 7/2 |
| 12 | 0/0 | 0/0 | 0/0 | 3/0 | 13/1 | 20/9 | 1/0 | 6/0 | 12/0 |
| 13 | 0/0 | 0/0 | 0/0 | 0/4 | 0/15 | 4/21 | 0/1 | 0/0 | 0/15 |
| 14 | 0/0 | 0/0 | 0/0 | 1/0 | 14/1 | 19/7 | 1/0 | 8/0 | 17/0 |
| 15 | 0/0 | 0/1 | 0/4 | 0/16 | 0/34 | 1/37 | 0/7 | 0/19 | 0/32 |
| 16 | 0/0 | 0/0 | 1/0 | 0/0 | 7/1 | 12/11 | 0/0 | 0/0 | 9/0 |
| 17 | 0/0 | 0/0 | 0/0 | 1/0 | 8/0 | 35/3 | 1/0 | 3/0 | 10/1 |
| 18 | 0/0 | 0/0 | 11/0 | 21/0 | 46/0 | 37/1 | 18/0 | 28/0 | 37/0 |
| 19 | 0/0 | 0/0 | 0/0 | 1/0 | 3/3 | 15/16 | 0/0 | 1/0 | 3/0 |
| 20 | 0/0 | 0/0 | 1/0 | 9/0 | 15/1 | 22/9 | 4/0 | 7/0 | 19/0 |
| 21 | 0/0 | 0/0 | 0/0 | 0/0 | 0/7 | 8/15 | 0/0 | 1/0 | 2/6 |
| 22 | 0/0 | 0/0 | 2/0 | 1/0 | 21/0 | 23/2 | 1/0 | 11/0 | 19/1 |
| 23 | 0/0 | 0/0 | 0/0 | 0/0 | 4/3 | 5/13 | 0/0 | 0/0 | 4/3 |
| 24 | 0/0 | 0/0 | 0/0 | 1/0 | 10/3 | 9/13 | 0/0 | 0/0 | 4/2 |
| 25 | 0/0 | 0/0 | 3/0 | 5/0 | 19/0 | 32/4 | 1/0 | 10/0 | 14/1 |
| 26 | 0/0 | 0/0 | 0/0 | 1/0 | 4/0 | 17/8 | 0/0 | 11/0 | 2/4 |
| 27 | 0/0 | 0/0 | 0/0 | 0/0 | 3/4 | 15/11 | 0/0 | 0/0 | 2/1 |
| 28 | 0/0 | 0/3 | 0/6 | 0/13 | 0/20 | 0/39 | 0/13 | 0/20 | 0/28 |
| 29 | 0/0 | 0/0 | 0/0 | 0/0 | 3/13 | 7/28 | 0/0 | 0/4 | 0/8 |
| 30 | 0/0 | 0/0 | 0/0 | 3/0 | 6/1 | 26/6 | 0/0 | 3/0 | 14/0 |
| 31 | 0/0 | 0/0 | 0/0 | 0/3 | 0/7 | 4/26 | 0/1 | 0/7 | 1/9 |
| 32 | 0/0 | 0/0 | 0/3 | 0/7 | 0/19 | 4/28 | 0/7 | 0/17 | 0/20 |
| 33 | 0/0 | 0/0 | 0/0 | 0/0 | 10/2 | 18/3 | 0/0 | 2/0 | 2/1 |
| 34 | 0/0 | 0/0 | 0/0 | 0/0 | 8/1 | 21/5 | 0/0 | 1/0 | 6/5 |
| 35 | 0/0 | 0/0 | 0/0 | 1/0 | 7/0 | 21/9 | 0/0 | 0/0 | 3/3 |
| 36 | 0/0 | 0/0 | 0/0 | 1/0 | 8/1 | 19/8 | 0/0 | 0/0 | 6/3 |
| 37 | 0/0 | 0/0 | 0/0 | 0/0 | 1/5 | 13/18 | 0/0 | 0/3 | 0/6 |
| 38 | 0/0 | 0/0 | 0/0 | 0/0 | 7/3 | 13/14 | 0/0 | 0/0 | 5/2 |
| 39 | 0/0 | 0/0 | 0/0 | 0/1 | 0/12 | 8/26 | 0/2 | 0/4 | 1/12 |
| 40 | 0/0 | 0/0 | 0/1 | 0/4 | 1/19 | 7/36 | 0/4 | 0/7 | 0/16 |
| AVE. per Test | .0/ .6 | .0/ .7 | .2/ .9 | .7/ 1.4 | 2.9/ 3.0 | 5.8/ 6.1 | .3/ 1.1 | 1.0/ 1.7 | 2.7/ 2.8 |

TABLE 20

DIF Identification (F/B) for the STD Statistic: Test C

| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
|---|---|---|---|---|---|---|---|---|---|
| | Sample Size Ratio (F:B) | | | | | | | | |
| 1 | 0/0 | 0/0 | 0/0 | 6/0 | 40/0 | 88/0 | 0/0 | 2/0 | 4/0 |
| 2 | 0/0 | 0/0 | 0/0 | 4/0 | 40/3 | 75/0 | 1/0 | 4/0 | 15/3 |
| 3 | 0/0 | 0/0 | 1/0 | 12/0 | 52/1 | 84/0 | 8/0 | 8/0 | 22/0 |
| 4 | 0/75 | 0/74 | 0/71 | 0/66 | 1/29 | 38/8 | 0/66 | 0/59 | 0/59 |
| 5 | 0/0 | 1/0 | 6/0 | 16/0 | 50/1 | 84/0 | 9/0 | 12/0 | 30/1 |
| 6 | 0/0 | 0/0 | 0/1 | 0/3 | 14/7 | 59/2 | 0/1 | 0/5 | 0/16 |
| 7 | 0/13 | 0/22 | 0/31 | 0/33 | 4/17 | 45/5 | 0/33 | 0/39 | 0/44 |
| 8 | 0/0 | 0/0 | 0/0 | 3/1 | 28/0 | 72/1 | 0/0 | 4/1 | 18/2 |
| 9 | 0/0 | 0/5 | 0/16 | 0/19 | 6/16 | 35/0 | 0/17 | 0/27 | 0/34 |
| 10 | 0/0 | 0/0 | 0/2 | 0/4 | 9/4 | 53/2 | 0/3 | 0/10 | 5/18 |
| 11 | 0/0 | 0/0 | 1/0 | 3/0 | 28/2 | 66/1 | 1/0 | 4/0 | 14/3 |
| 12 | 0/0 | 1/0 | 3/0 | 13/0 | 41/2 | 84/0 | 7/0 | 15/0 | 20/0 |
| 13 | 0/0 | 0/1 | 0/11 | 0/19 | 7/15 | 52/2 | 0/14 | 0/23 | 2/29 |
| 14 | 0/0 | 0/0 | 1/0 | 5/0 | 40/0 | 82/0 | 1/0 | 8/0 | 17/0 |
| 15 | 0/0 | 0/13 | 0/13 | 0/23 | 1/25 | 38/5 | 0/17 | 0/31 | 0/38 |
| 16 | 0/0 | 0/0 | 3/0 | 3/0 | 33/2 | 73/2 | 2/0 | 7/0 | 19/1 |
| 17 | 0/0 | 0/0 | 2/0 | 8/0 | 39/1 | 79/0 | 4/0 | 12/1 | 20/1 |
| 18 | 19/0 | 22/0 | 31/0 | 47/0 | 75/0 | 87/0 | 39/0 | 47/0 | 50/0 |
| 19 | 0/0 | 0/0 | 0/0 | 2/0 | 21/3 | 60/3 | 0/0 | 2/0 | 6/4 |
| 20 | 0/0 | 1/0 | 5/0 | 15/0 | 50/0 | 79/1 | 7/0 | 11/0 | 24/0 |
| 21 | 0/0 | 0/0 | 0/0 | 1/2 | 15/7 | 51/2 | 0/0 | 2/2 | 6/14 |
| 22 | 0/0 | 4/0 | 15/0 | 26/0 | 59/0 | 77/0 | 14/0 | 31/0 | 34/1 |
| 23 | 0/0 | 0/0 | 0/0 | 3/1 | 24/3 | 63/2 | 0/0 | 1/2 | 10/6 |
| 24 | 0/0 | 0/0 | 0/0 | 5/0 | 29/0 | 58/2 | 0/0 | 1/0 | 8/2 |
| 25 | 0/0 | 1/0 | 5/0 | 11/0 | 58/0 | 76/1 | 10/0 | 16/0 | 22/1 |
| 26 | 0/0 | 0/0 | 1/0 | 4/1 | 32/1 | 58/1 | 0/0 | 4/0 | 9/5 |
| 27 | 0/0 | 0/0 | 0/0 | 0/2 | 19/7 | 50/4 | 0/0 | 3/3 | 7/5 |
| 28 | 0/32 | 0/37 | 0/49 | 0/40 | 1/21 | 23/14 | 0/37 | 0/44 | 0/39 |
| 29 | 0/0 | 0/0 | 0/2 | 0/6 | 14/6 | 37/3 | 0/2 | 0/10 | 0/14 |
| 30 | 0/0 | 0/0 | 2/0 | 8/0 | 43/0 | 65/0 | 3/0 | 10/0 | 20/0 |
| 31 | 0/0 | 0/0 | 0/2 | 0/7 | 7/9 | 33/8 | 0/9 | 0/17 | 2/19 |
| 32 | 0/2 | 0/8 | 0/16 | 0/18 | 1/19 | 27/13 | 0/18 | 0/32 | 0/26 |
| 33 | 0/0 | 0/0 | 0/0 | 2/0 | 31/1 | 61/1 | 1/0 | 6/0 | 12/0 |
| 34 | 0/0 | 0/0 | 0/0 | 4/1 | 21/2 | 42/4 | 1/0 | 3/1 | 14/5 |
| 35 | 0/0 | 0/0 | 0/0 | 4/0 | 22/2 | 58/1 | 1/0 | 5/0 | 8/2 |
| 36 | 0/0 | 0/0 | 1/0 | 6/0 | 19/3 | 56/1 | 0/0 | 3/0 | 9/2 |
| 37 | 0/0 | 0/0 | 0/0 | 0/1 | 8/6 | 47/10 | 0/0 | 0/5 | 1/7 |
| 38 | 0/0 | 0/0 | 1/0 | 1/0 | 23/2 | 59/2 | 1/0 | 3/1 | 11/3 |
| 39 | 0/0 | 0/0 | 0/0 | 0/2 | 3/5 | 24/6 | 0/0 | 0/2 | 1/8 |
| 40 | 0/0 | 0/1 | 0/4 | 0/15 | 5/16 | 27/7 | 0/8 | 0/13 | 2/20 |
| AVE. per Test | .2/ 1.2 | .3/ 1.6 | .8/ 2.2 | 2.1/ 2.6 | 10.1/ 2.4 | 23.2/ 1.2 | 1.1/ 2.2 | 2.2/ 3.3 | 4.4/ 4.3 |

TABLE 21

DIF Identification for the RMWSD Statistic: Test C

| | Sample Size Ratio (F:B) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Item # | 2000: 2000 | 1000: 1000 | 500: 500 | 250: 250 | 100: 100 | 50: 50 | 200: 2000 | 100: 1000 | 50: 500 |
| 1 | 3 | 50 | 99 | 100 | 100 | 100 | 99 | 100 | 100 |
| 2 | 12 | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 3 | 11 | 59 | 99 | 100 | 100 | 100 | 98 | 100 | 100 |
| 4 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 5 | 23 | 89 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 6 | 6 | 87 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 7 | 90 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | 0 | 73 | 99 | 100 | 100 | 100 | 100 | 100 | 100 |
| 9 | 85 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 10 | 23 | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 11 | 1 | 70 | 99 | 100 | 100 | 100 | 100 | 100 | 100 |
| 12 | 18 | 81 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 13 | 61 | 95 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 14 | 3 | 56 | 99 | 100 | 100 | 100 | 100 | 100 | 100 |
| 15 | 82 | 96 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 16 | 12 | 87 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 17 | 24 | 83 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 18 | 94 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 19 | 2 | 67 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 20 | 35 | 80 | 100 | 100 | 100 | 99 | 100 | 100 | 100 |
| 21 | 11 | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 22 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 23 | 9 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 24 | 0 | 69 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 25 | 27 | 88 | 99 | 100 | 100 | 100 | 100 | 100 | 100 |
| 26 | 9 | 86 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 27 | 10 | 95 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 28 | 98 | 98 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 29 | 24 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 30 | 8 | 91 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 31 | 34 | 97 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 32 | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 33 | 6 | 88 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 34 | 12 | 88 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 35 | 11 | 92 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 36 | 26 | 91 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 37 | 11 | 91 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 38 | 12 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 39 | 11 | 79 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| 40 | 66 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| AVE. per Test | 12.5 | 34.5 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 | 40.0 |

# FIGURE CAPTIONS

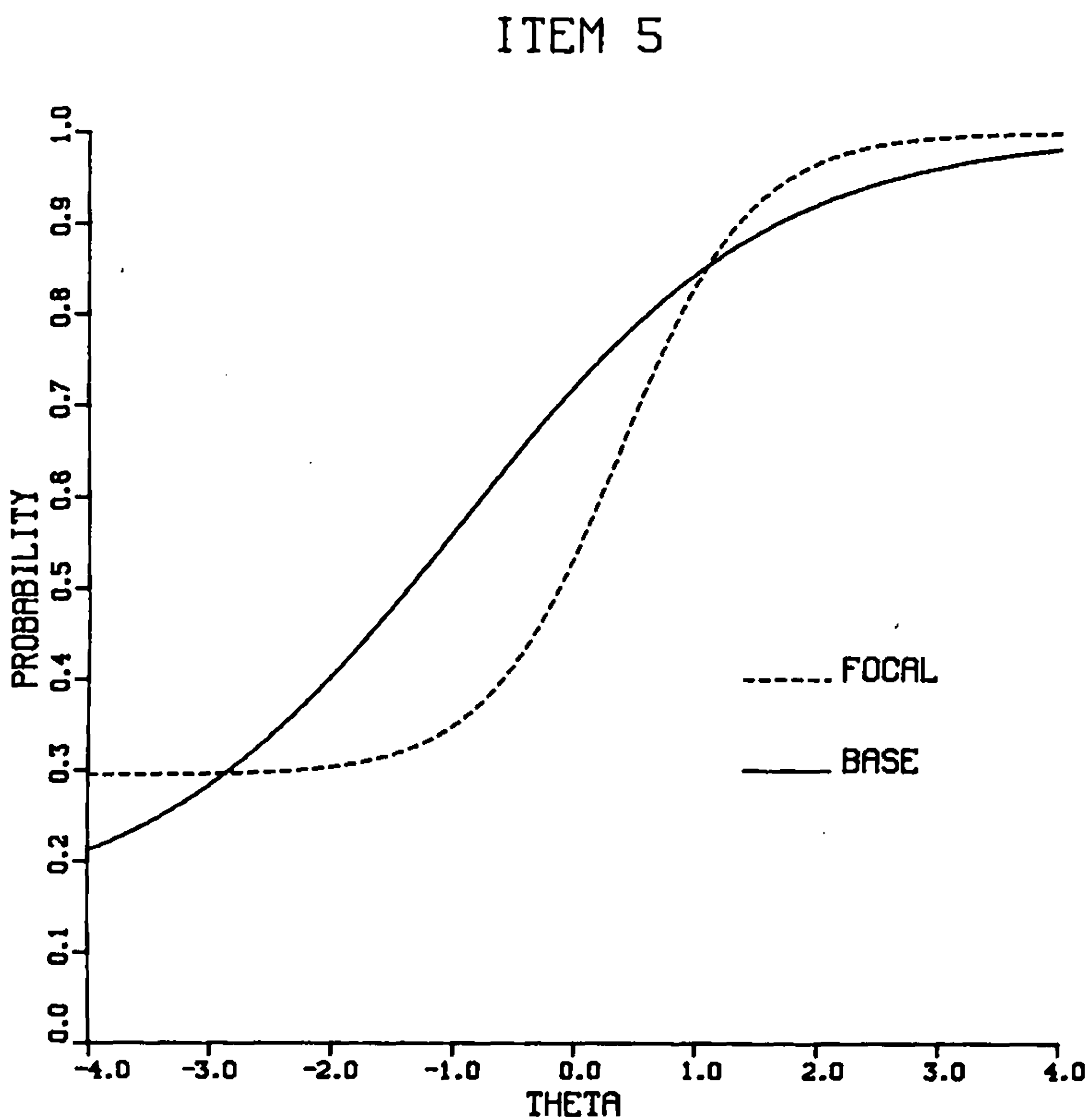


Figure 1. ICCs of F(blacks) and B(whites) for Item #5 of Test A

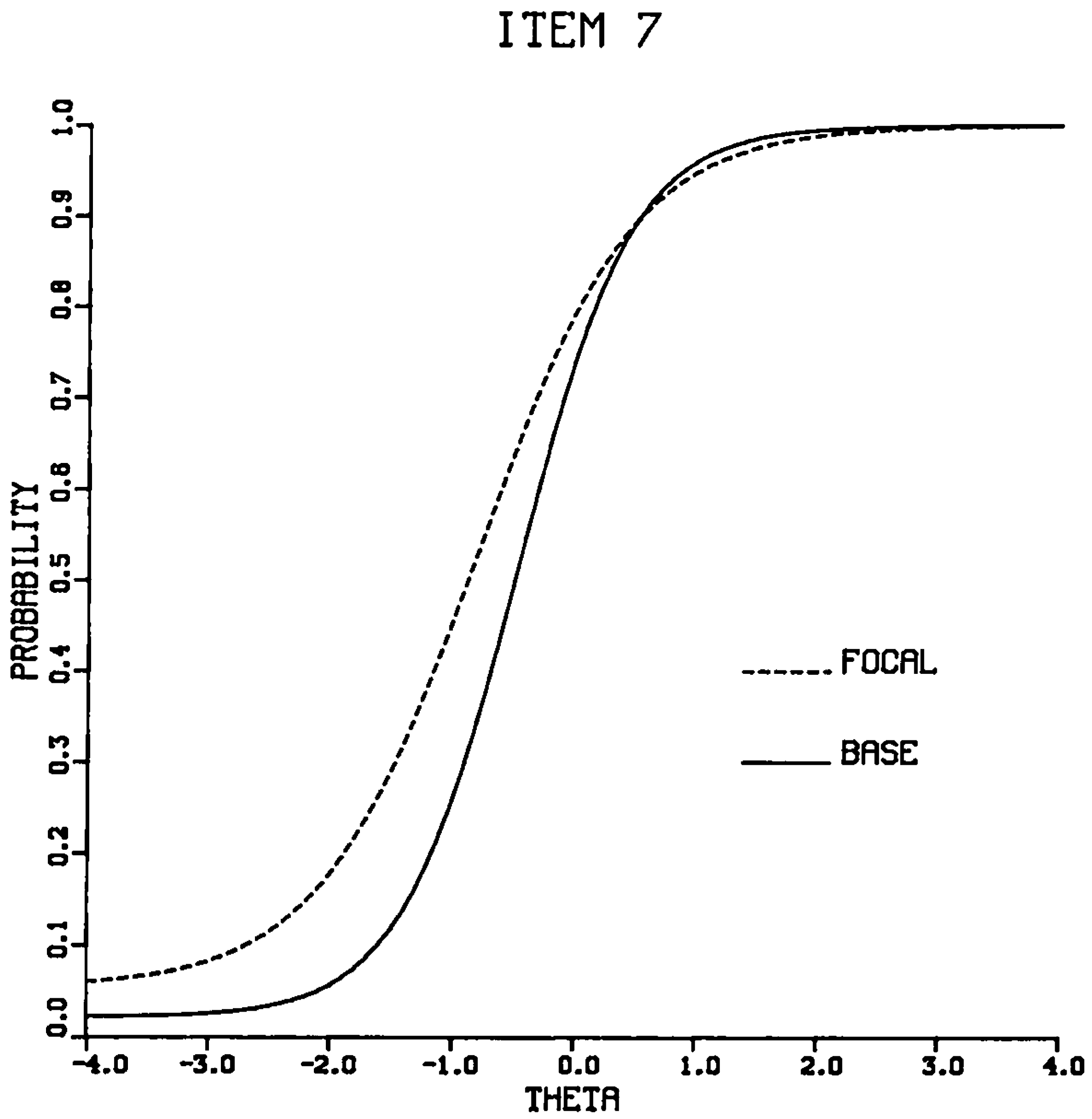


Figure 2. ICCs of F(blacks) and B(whites) for Item #7 of Test A

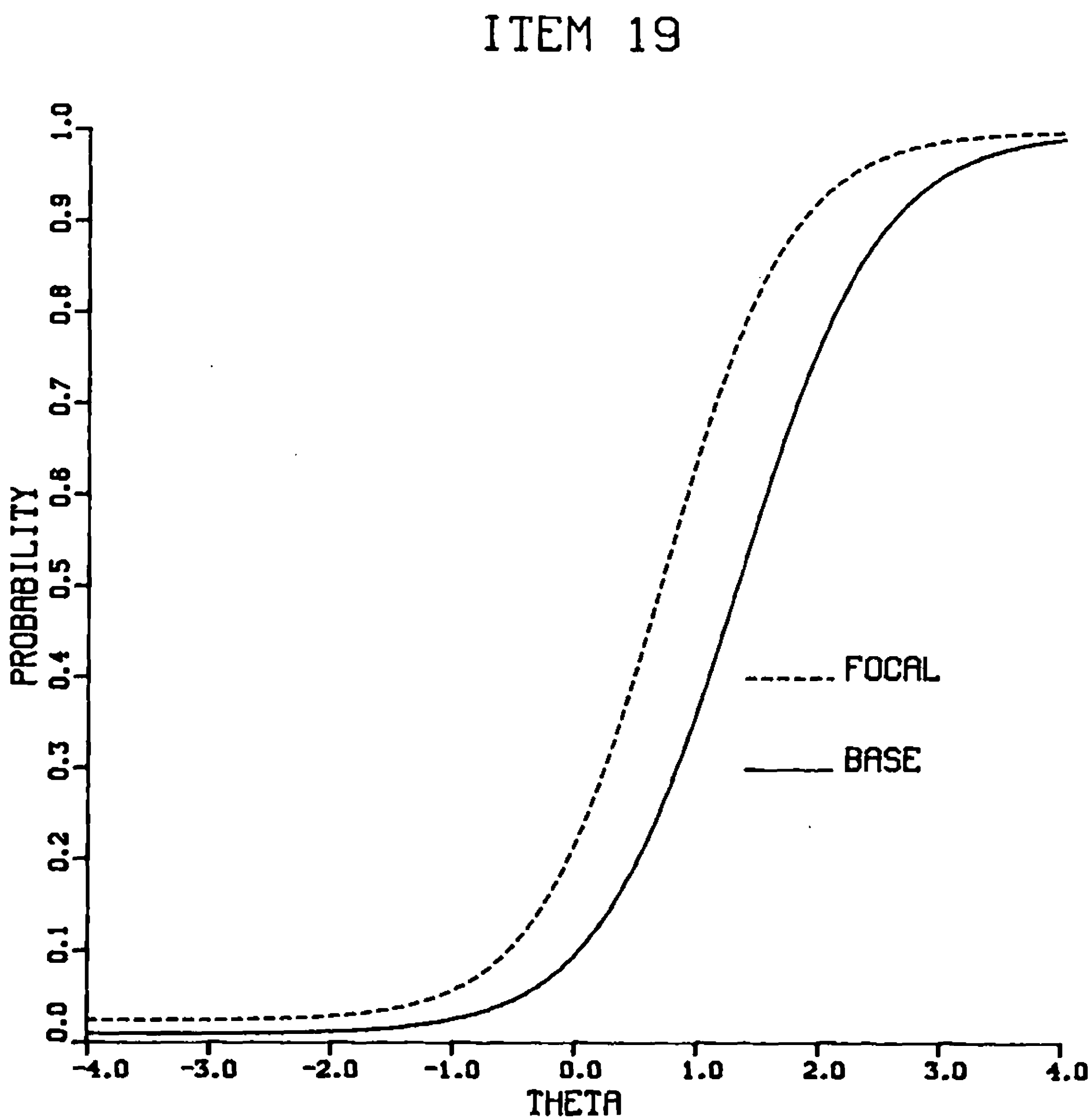


Figure 3. ICCs of F(blacks) and B(whites) for Item #19 of Test A

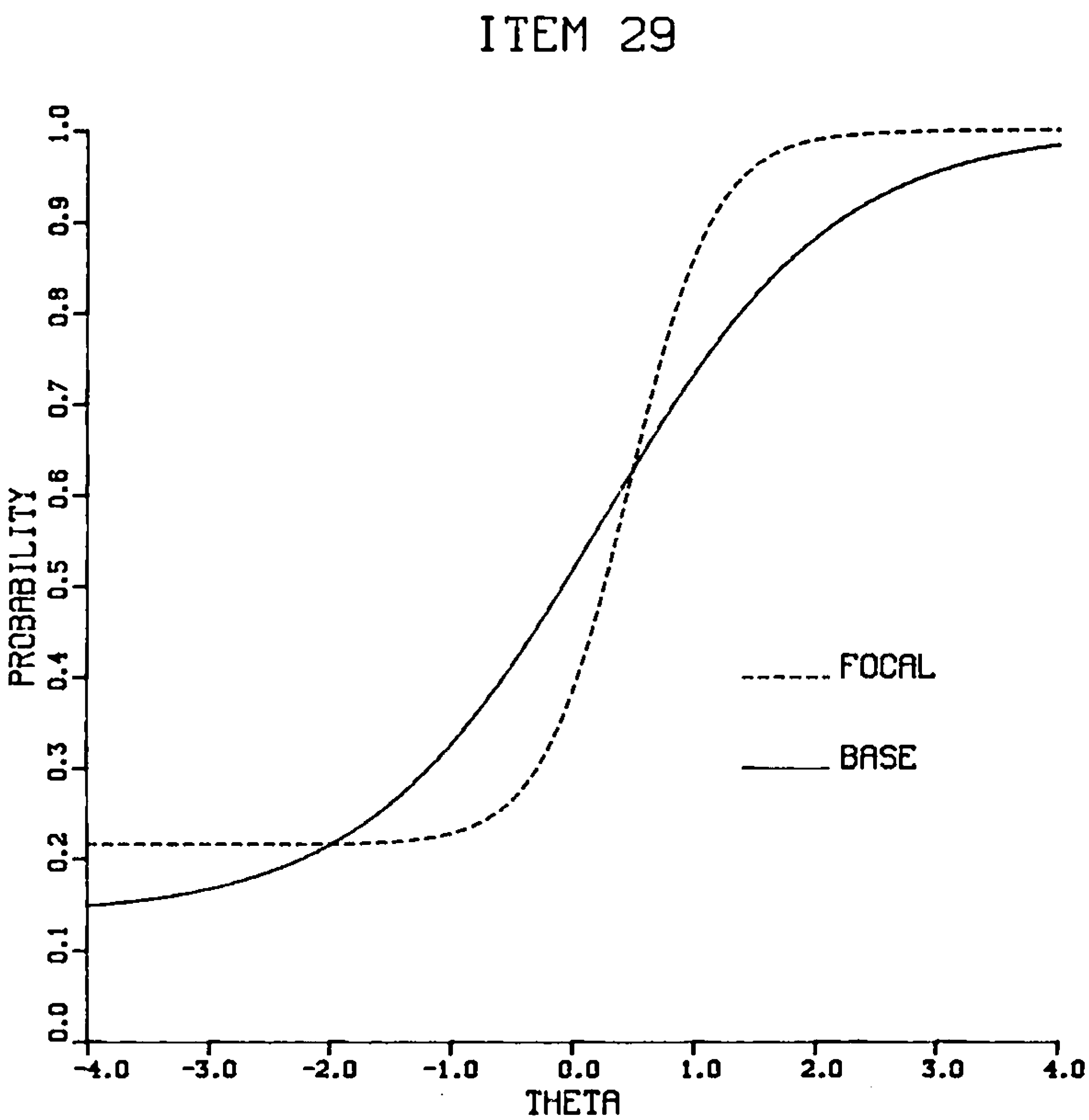


Figure 4. ICCs of F(blacks) and B(whites) for Item #29 of Test A

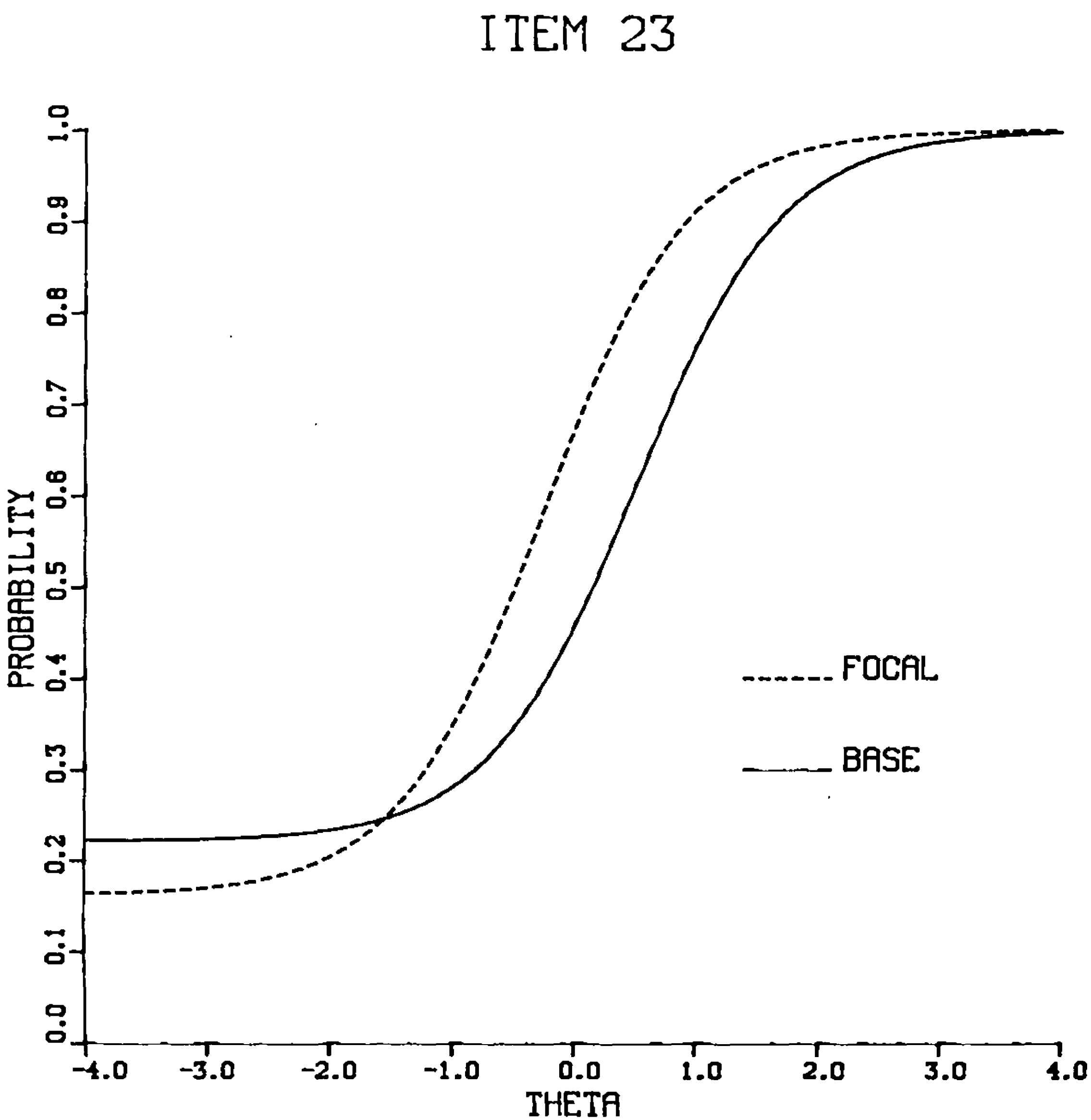


Figure 5. ICCs of F(blacks) and B(whites) for Item #23 of Test B

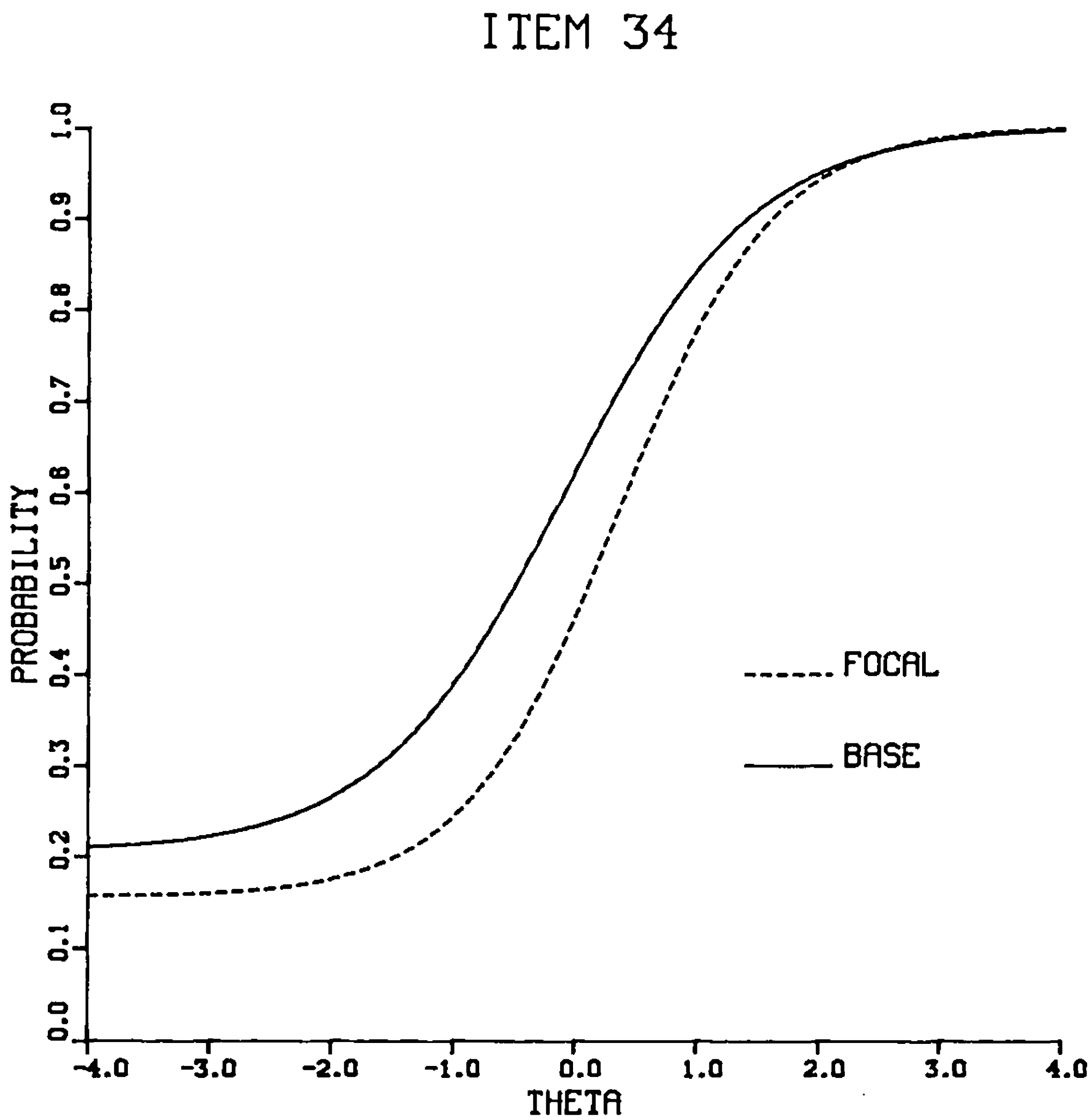


Figure 6. ICCs of F(blacks) and B(whites) for Item #34 of Test B

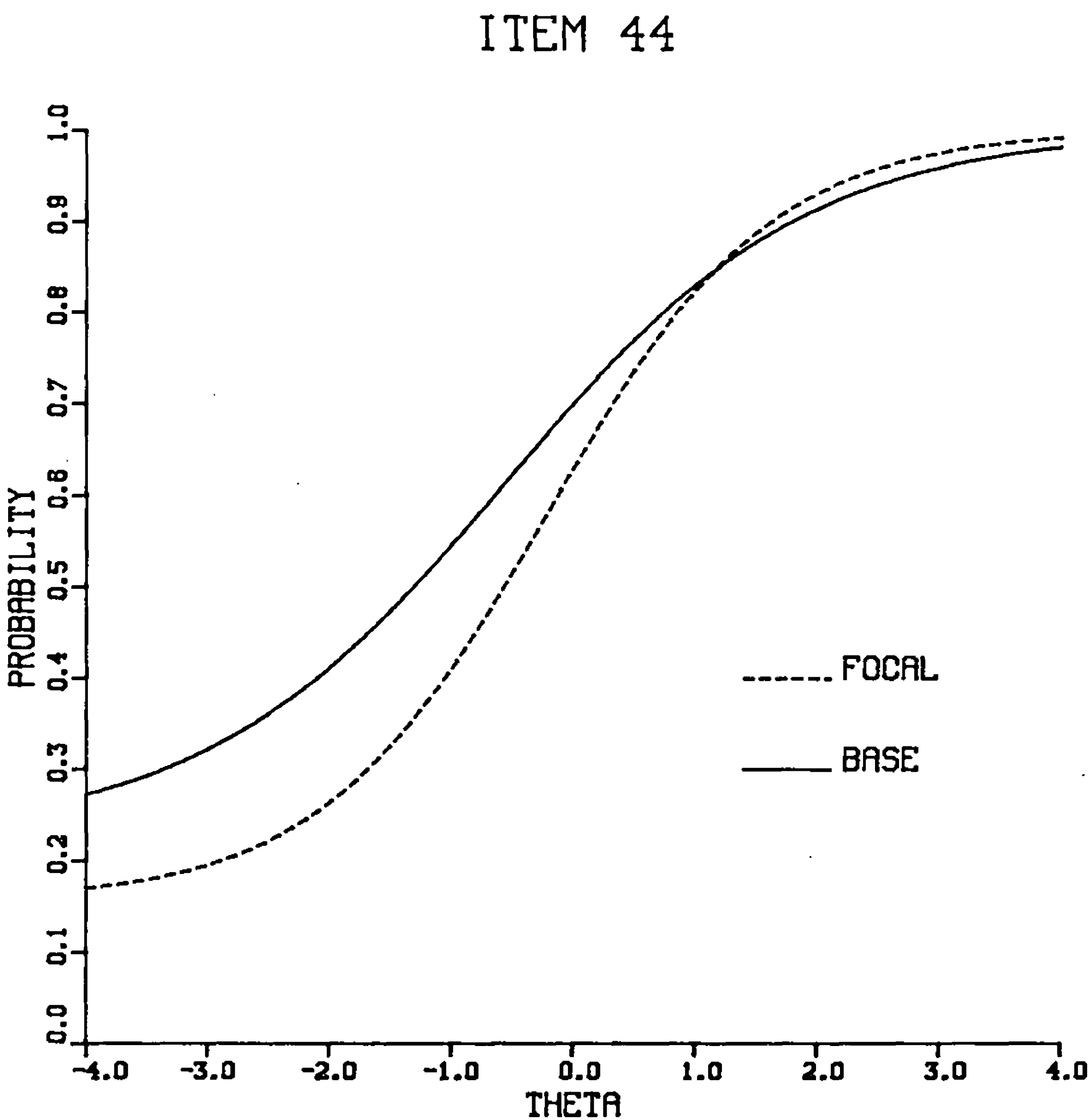


Figure 7. ICCs of F(blacks) and B(whites) for Item #44 of Test B

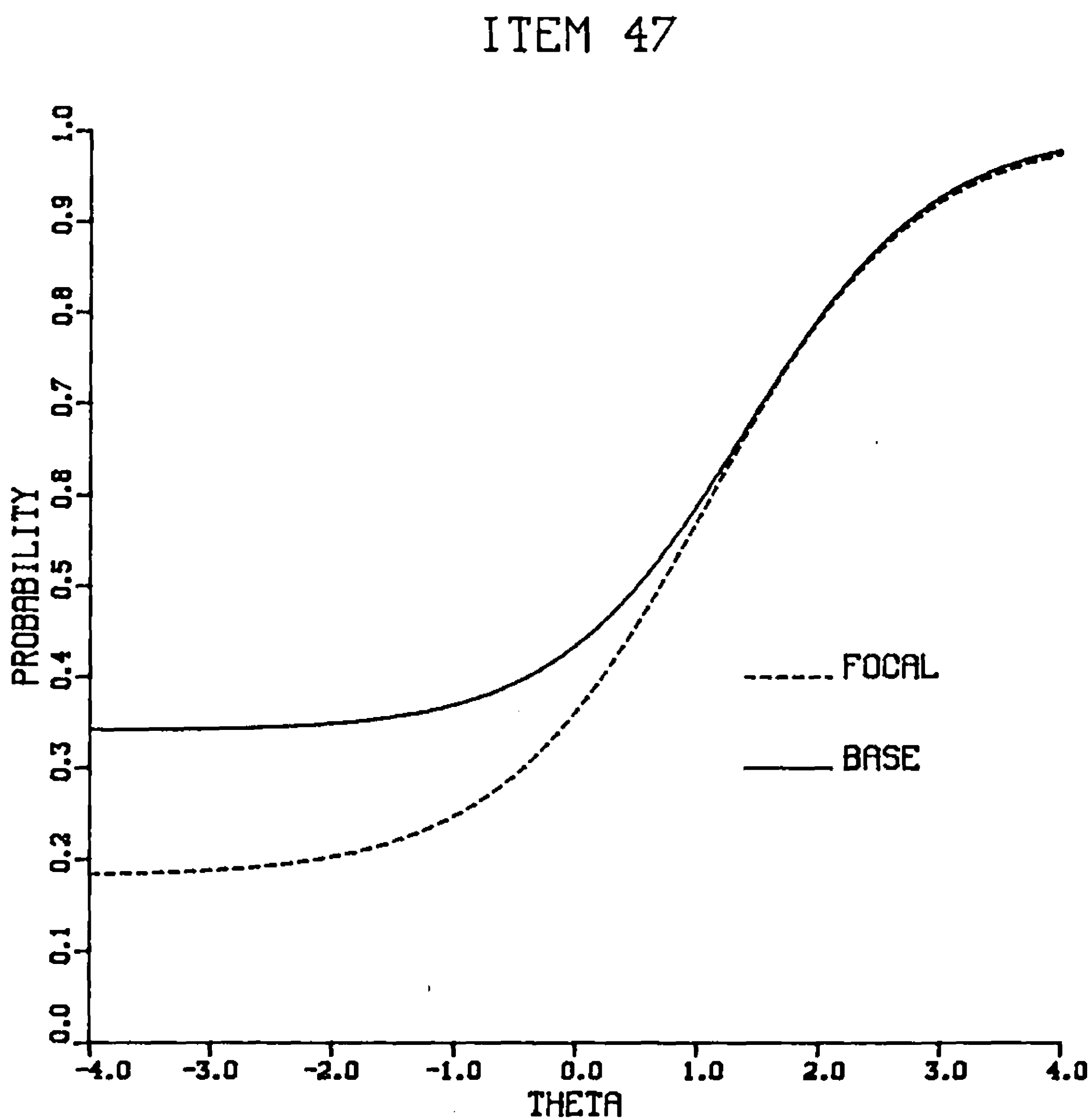


Figure 8. ICCs of F(blacks) and B(whites) for Item #47 of Test B

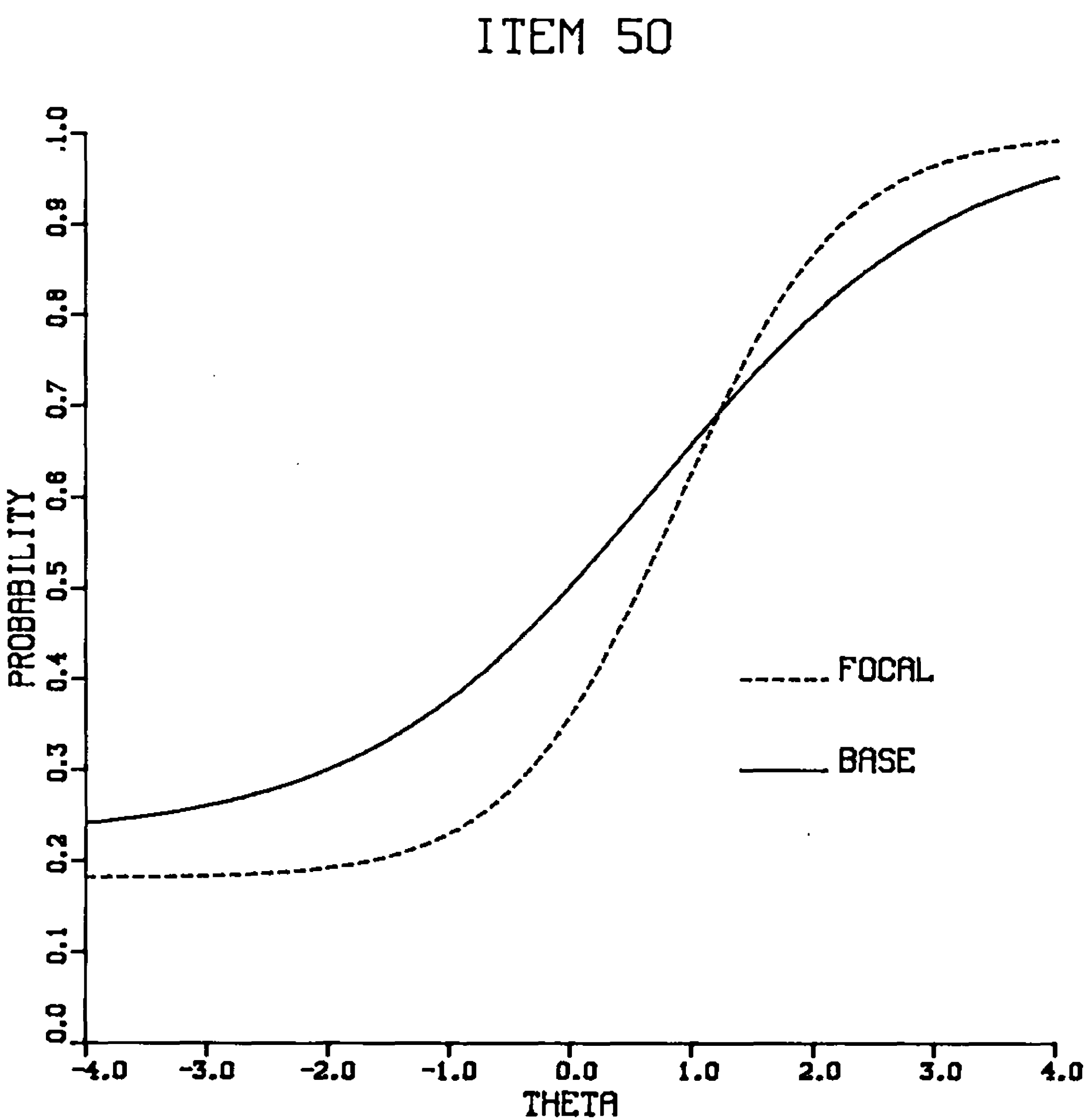


Figure 9. ICCs of F(blacks) and B(whites) for Item #50 of Test B

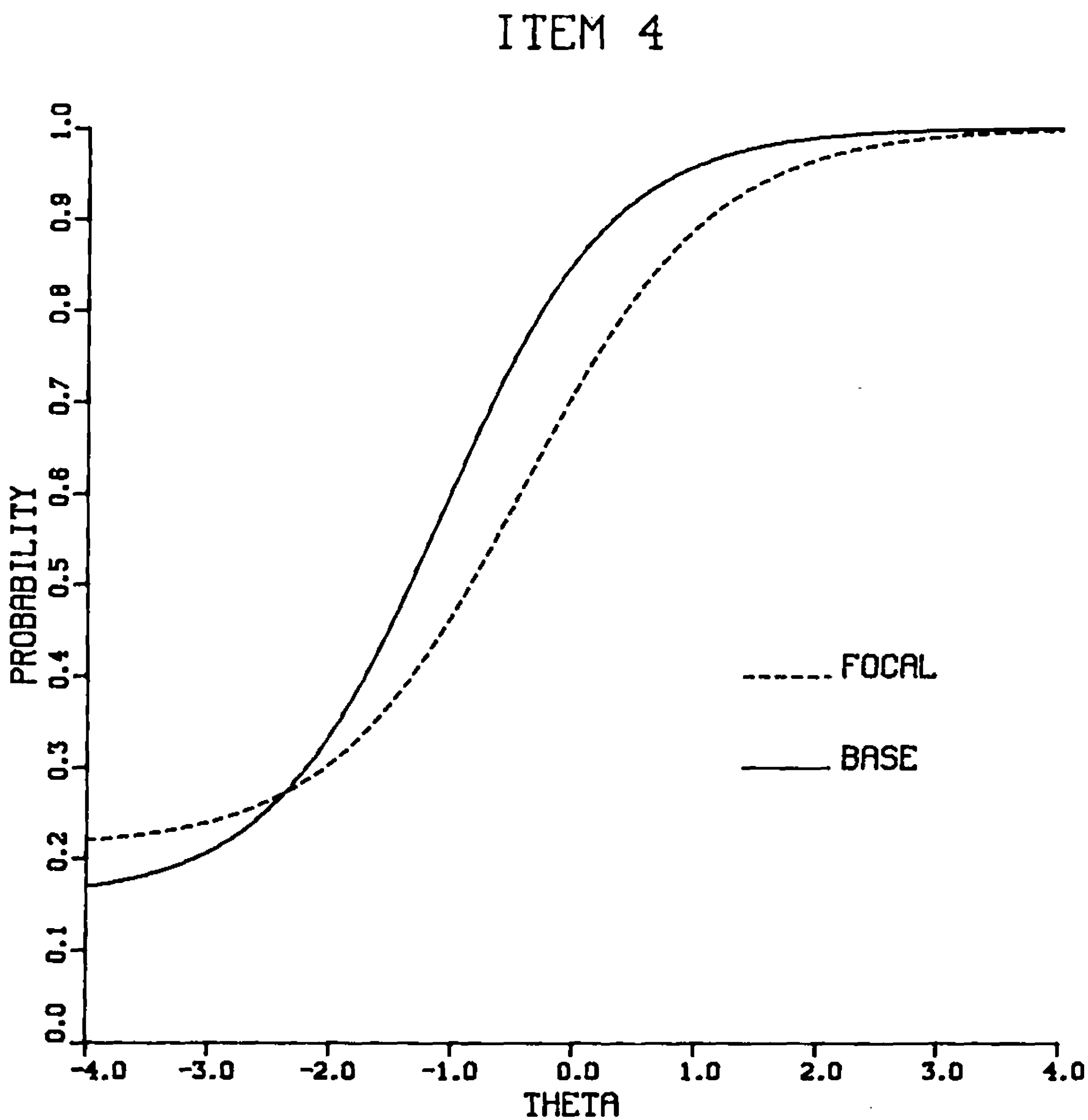


Figure 10. ICCs of F(females) and B(males) for Item #4 of Test C

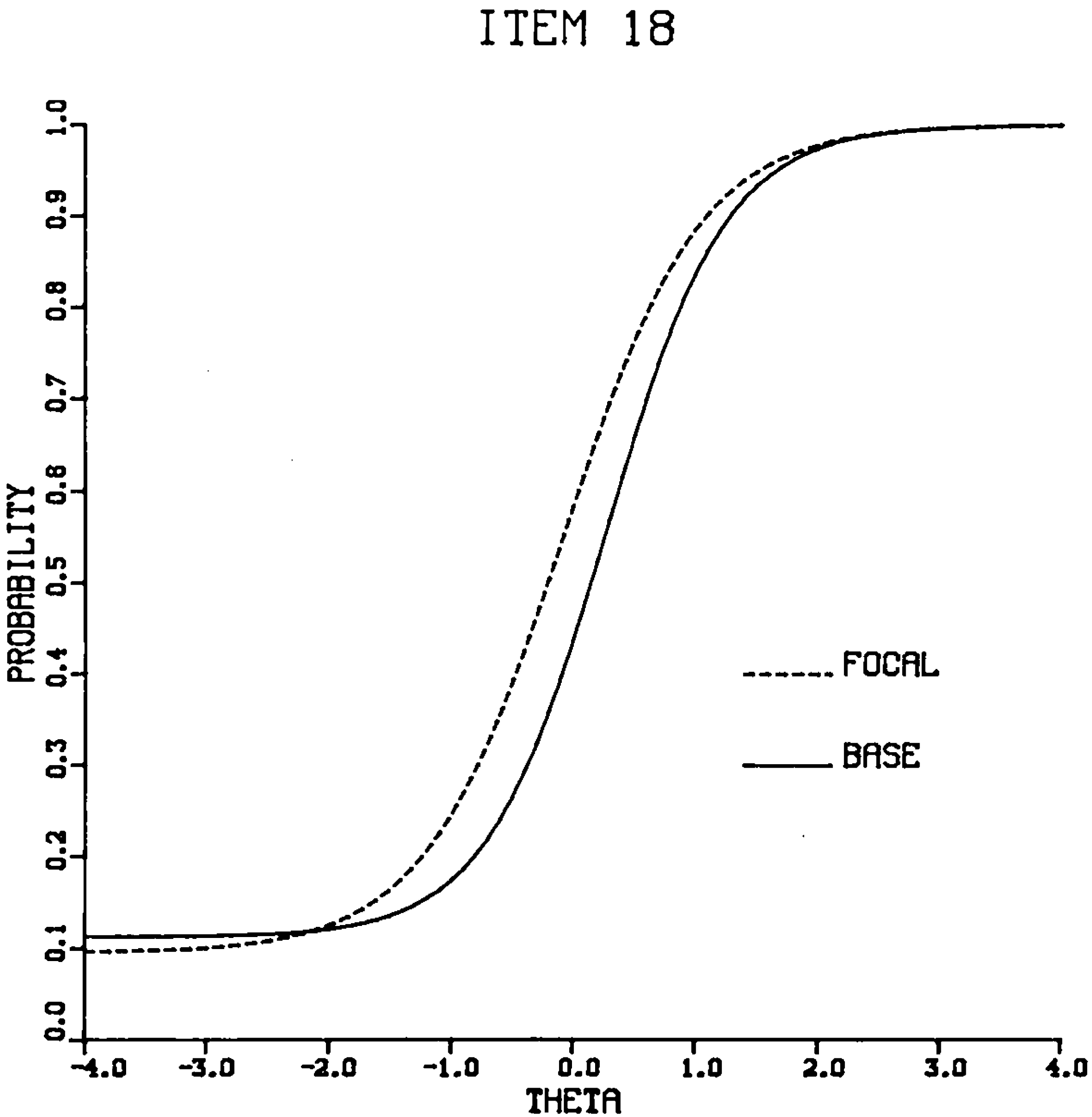


Figure 11. ICCs of F(females) and B(males) for Item #18 of Test C

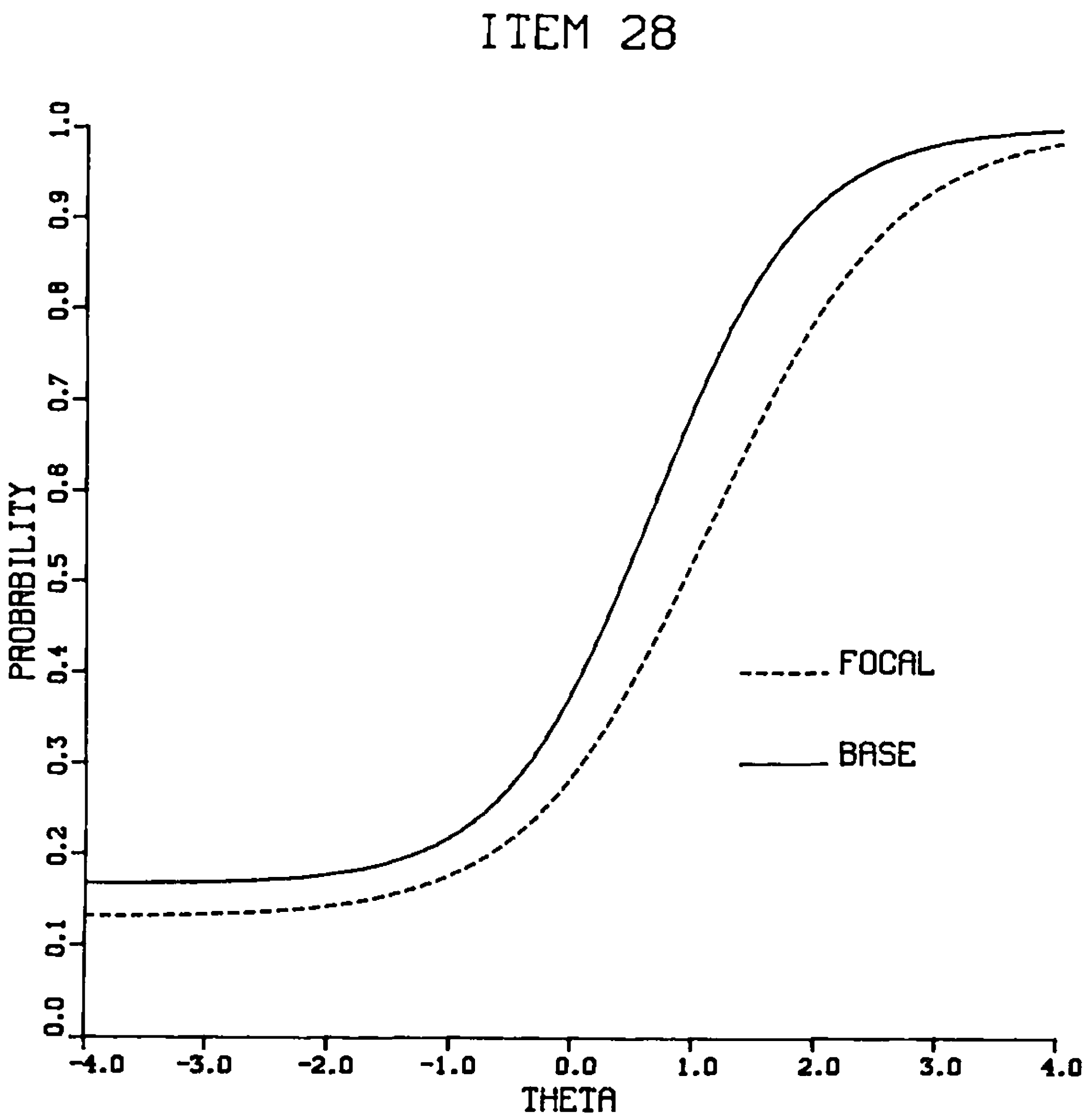


Figure 12. ICCs of F(females) and B(males) for Item #28 of Test C