AUTHOR Pomplun, Mark; And Others
TITLE An Initial Evaluation of the Use of Bivariate Matching in DIF Analyses for Formula Scored Tests.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-92-63
PUB DATE Nov 92
NOTE 82p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Blacks; Criteria; Females; *Item Bias; Males; *Racial Differences; Sample Size; Scoring; *Scoring Formulas; *Sex Differences; *Test Items; Whites
IDENTIFIERS *Bivariate Matching; Mantel Haenszel Procedure; *Rights and Formula Scoring; Scholastic Aptitude Test

ABSTRACT

This study evaluated the use of bivariate matching as a solution to the problem of studying differential item functioning (DIF) with formula scored tests. Using Scholastic Aptitude Test verbal data with large samples, both male/female and black/white group comparisons were investigated. Mantel-Haenszel (MH) delta-(D) DIF values and DIF category classifications based on bivariate matching were compared with MH D-DIF values and categories based on rights scored and formula scored matching criteria. When large samples were used, values based on the bivariate matching criterion were ordered very similarly to MH D-DIF values based on the other criteria. With small samples, the values based on the bivariate matching criterion displayed only moderate correlations with values from the other criteria. DIF category classifications based on the bivariate matching criterion showed fewer high DIF items than those based on the rights or formula scored matching criteria. A secondary study of the differences between formula and rights scored criteria in DIF analyses of formula scored tests showed differences related to item difficulty that were greater for comparisons with large ability differences. An appendix contains 30 scatterplots. Nine tables and four figures present details of the analyses. (Contains seven references.) (Author/SLD)

RESEARCH REPORT

AN INITIAL EVALUATION OF THE USE OF BIVARIATE MATCHING IN DIF ANALYSES FOR FORMULA SCORED TESTS

Mark Pomplun
Patricia Baron
Fred McHale

Educational Testing Service
Princeton, New Jersey
November 1992

An Initial Evaluation of the Use of Bivariate Matching in DIF Analyses for Formula Scored Tests[1] [2]

by

Mark Pomplun

Patricia Baron

and

Fred McHale

Draft Final Report

Educational Testing Service

September 1992

[1] This research was supported by the Educational Testing Service through Program Research and Planning Council funding.

[2]The authors thank Paul Holland for his help in the design of this study. Thė authors also wish to thank Neil Dorans, Dan Eignor, Gary Marco, and Rebecca Zwick for their helpful reviews.

Abstract

This study evaluated the use of bivariate matching as a solution to the problem of studying DIF with formula scored tests. This problem specifically involves including in or excluding from a formula scored matching criterion formula scored items to-be-studied for a DIF analysis. Using SAT Verbal data with large and small samples both Male-Female and Black-White group comparisons were investigated. MH D-DIF values and DIF category classifications based on bivariate matching were compared with MH D-DIF values and categories based on rights scored and formula scored matching criteria. When large samples were used, MH D-DIF values based on the bivariate matching criterion were ordered very similarly to MH D-DIF values based on the other criteria. The DIF category classifications were almost identical. However, with small samples the MH D-DIF values based on the bivariate matching criterion displayed only moderate correlations with MH D-DIF values from the other criteria. In addition, the DIF category classifications based on the bivariate matching criterion showed fewer high DIF items than those based on the rights or formula scored matching criteria. As a secondary result, this study documented the differences between formula and rights scored criteria in DIF analyses of formula scored tests. These results showed that the substitution of a rights scored criterion in a DIF analysis of a formula scored test resulted in MH D-DIF values that were ordered very similarly as those based on the proper formula scored criterion. The MH D-DIF values based on the rights score criterion were, however, different in magnitude from those based on the formula scored criterion. The differences were related to item difficulty and were greater for comparisons with larger ability differences.

An Initial Evaluation of the Use of Bivariate Matching in DIF Analyses for Formula Scored Tests

INTRODUCTION

Differential item functioning (DIF) analysis provides a measure of the difference in item (test question) performance between two comparable groups of examinees. The groups are matched with respect to the construct measured by the test. A well known and widely used measure of DIF is provided by the Mantel-Haenszel (MH) statistic (Holland and Thayer, 1988). It provides a chi-square based significance test to help evaluate the presence or abrence of DIF in an item for two examinee groups of interest. The MH procedure is a statistically powerful technique developed for binary variables.

Because the MH procedure is appropriate for binary variables, two problems arise when one attempts to apply it to formula-scored items, one at the individual item level and the other at the score matching criterion level. Formula scoring is when instead of scoring items as simply as right or wrong, items are scored to correct for guessing. The formula score used in this study is formula score = number right - (1/(k-1) number wrong) where k is the number of item response options. A problem arises at the individual item level when performing MH DIF analyses on formula scored items because the items are not scored in a binary fashion, i. e., as right or wrong, and as such, cannot be analyzed through the MH procedure. For a formula-scored item, there are three possible scores [right = 1, no response = 0, wrong = (-1/(k-1))], rather than two. To allow analysis of formula-scored items through the MH procedure, it is common practice to score such items as simply right or wrong, with no response counted as wrong.

Another problem in DIF analyses of formula scored items occurs at the formula score matching criterion level[3]. According to Holland and Thayer (1988), a statistical bias affects the MH DIF statistic when the item being analyzed is not included in the matching criterion. Hence, the item being studied should be included in the matching criterion when using the MH procedure. For a rights scored test, the inclusion of the item being studied is efficiently accomplished through the use of the correct response frequencies at each level of the criterion variable. The correct response frequencies for the studied item can be easily added to the criterion when the item is scored right or wrong. However, when items are formula scored, besides the adaition of response frequencies for correct responses, it is also necessary to subtract (-1/(k-1)) from the total formula score for incorrect responses. As a result, the inclusion of another item to the total formula score cannot be accomplished by simply adding correct response frequencies to those that already exist.

The present solutions to including and excluding formula scored items in a formula scored matching criterion involve practical and logical difficulties. The practical difficulty arises because there is currently no straightforward method to include the to-be-studied formula scored item in a formula score criterion. As a result, some testing programs attempting to include the item in the criterion use a labor and cost intensive process which involves creating a new formula score criterion. Other testing programs simply substitute a rights scored criterion for the proper formula score criterion. This leads to the logical inconsistency of using number right scoring at the criterion level when

[3] When formula score is referred :o at the criterion level, it is always a formula score rounded to an integer.

performing a DIF analysis on a test that was administered under formula scored directions.

The purpose of this study is to evaluate a possible solution to the problems brought about by the need to include a formula scored item to-be-studied in a formula score criterion for a DIF analysis. A solution has been proposed by Paul Holland (1990) that is based on matching examinees on a bivariate criterion. Test takers are matched jointly on their number of correct answers and their combined number of omitted and not-reached responses. The bivariate matching criterion, in contrast to the formula score criterion, allows a straightforward adjustment for inclusion of the formula scored item-to-be-analyzed in the formula score criterion.

BACKGROUND INFORMATION

DIF Procedure

The Mantel-Haenszel statistic was adapted by Holland and Thayer (1988) as an approach to detecting DIF with rights only scored tests. When the total test score is the matching criterion, the basic data used for computation of the statistic are contained in a series of 2 x 2 contingency tables (Kulick & Hu, 1989). For each item at each score level s, data from two groups of examinees can be arranged as a 2 x 2 table:

	Right	Wrong	Total
Focal group	R_{fs}	W_{fs}	N_{fs}
Reference group	R_{rs}	W_{rs}	N_{rs}
Total group	R_{ts}	W_{ts}	N_{ts}

R_{fs} is the number of persons in the focal group at score level s who answered the item correctly; W_{fs} is the number in the focal group at s who answered the item incorrectly; and N_{fs} , the sum of R_{fs} and W_{fs} , is the total number in the focal group at s. R_{rs} , W_{rs} , and N_{rs} are the corresponding numbers of persons in the reference group at s. $R_{ts} = R_{rs} + R_{fs}$, $W_{ts} = W_{rs} + W_{fs}$, and $N_{ts} = N_{rs} + N_{fs}$ are the corresponding numbers of people in the total group at s.

At each score level, the Mantel-Haenszel approach uses an odds-ratio

$$\alpha_s = (R_{rs}/W_{rs}) \;/\; (R_{fs}/W_{fs})$$

to compare the reference group with the focal group. At a given score level, s, the odds-ratio is a measure of the advantage or disadvantage that reference group members have on an item relative to the matched focal group members. If $\alpha > 1$, the reference group has an advantage on the item at score level s; if $\alpha < 1$, the advantage lies with the focal group at score level s. The Mantel-Haenszel odds-ratio estimate is a weighted average of the odds-ratios across all score levels:

$$\alpha_{MH} = \frac{\sum_{s=1}^{S} M_s \, \alpha_s}{\sum_{s=1}^{S} M_s} ,$$

where $$M_s = W_{rs} R_{fs} / N_{ts}$$

Thus,
$$\alpha_{MH} = \frac{\sum_{s=1}^{S} R_{rs}\ W_{fs}\ /\ N_{ts}}{\sum_{s=1}^{S} R_{fs}\ W_{rs}\ /\ N_{ts}}$$

The weight M_s is based on the frequencies for a given 2 x 2 table.

For the test used in this study, the Scholastic Aptitude Test (SAT), the delta item statistic is used as an index of item difficulty. The delta scale has a mean of 13 and a standard deviation of 4. The MH odds-ratio can be converted to the delta scale to judge an item's difficulty for a focal group relative to a reference group on this scale. Holland and Thayer (1988) showed that the α_{MH} can be converted to an approximate difference in deltas between the reference and focal groups (focal - reference) as follows:

$$\text{MH D-DIF} = -2.35\ \ln(\alpha_{MH})$$

where ln is the natural logarithm.

At present, DIF analysis of a set of items commonly occurs in two steps: (1) analysis with an initial criterion; and then (2) analysis with a purified criterion, from which high DIF items identified in the first step are removed. The initial criterion usually includes all operational items from the test. In order to create an unbiased or purified criterion, all high DIF items are removed from the initial criterion. The purified criterion is then used in the second DIF analysis. For number-right scored tests, the DIF statistics from the purified analysis are used to determine the differential item functioning for all the items. (For the removed items, the MH statistic for each of these items is calculated by adding the respective item back into the criterion.) However, because of the problem of adding the formula-scored item to-be-studied to the formula score criterion, for formula scored tests it is current practice to use DIF statistics from the initial analysis to evaluate the high DIF items identified in the initial analysis, and to use the DIF statistics from the purified analysis to determine the differential item functioning for the remaining items. If there are no high DIF items in the initial analysis, the DIF statistics from the first analysis are used to evaluate all the items.

The Problem of Including and Excluding Items from the Criterion

Holland and Thayer (1988) showed why the studied item needs to be included in the criterion in a Mantel-Haenszel DIF analysis. They also showed how easy it is to include the item in the criterion in a DIF analysis in which the number right score is the criterion:

> If the K 2x2 tables have been assembled for a number right score S as the matching criterion that does not include the studied item and we wish to include it in the score, then the 2x2[4] tables need to be altered to these.

		Score on Studied Item		
		1	0	Total
Group	R	A_{j-1}	B_j	n'_{Rj}
	F	C_{j-1}	D_j	n'_{Fj}
	Total	m_{1j-1}	m_{0j}	T'_j

[4] In these tables, T'_j is the total number of reference and focal group members in the jth matched set; n'_{Rj} is the number of these who are in R_j and of these A_{j-1} answered the studied item correctly. The other entries have similar definitions.

> The values of MH-CHISQ and $\hat{\alpha}_{MH}$ are then computed from these tables. Similarly, if S contains the score of the studied item and we wish to eliminate it this is done by using these 2x2 tables.

		Score on Studied Item		
		1	0	Total
Group	R	A_{j+1}	B_j	n''_{Rj}
	F	C_{j+1}	D_j	n''_{Fj}
	Total	m_{1j+1}	m_{0j}	T''_{j}

> Thus it is a simple matter to compute either $\hat{\alpha}_{MH}$ or MH-CHISQ including or excluding the studied item from a number-right-score matching criterion. If the matching criterion is a formula-score or a grouped, number-right-score then it is not easy to adjust for the inclusion of the studied item into the criterion, without recalculating the entire set of 2x2 tables. (Holland & Thayer, 1988, pp 141-142)

A problem arises, then, when item inclusion in the criterion is needed for DIF analysis of a formula-scored test. For rights-scored tests, the adjustment for the studied item can be easily accomplished since the criterion score and item score are on the same metric, 1 for each correct response, and 0 for each incorrect response. This allows for a straightforward adjustment of the 2 x 2 table at each score level to add in the frequencies for the studied item. For formula scored tests, however, the item score is 1 for each correct response, -1/(k-1) (k is the number of response options) for each incorrect response, and 0 for each omit/not reached response. With the item scored three different ways, it is not straightforward to adjust the 2 x 2 tables to include the data on the to-be-studied formula scored item, especially when the rounded formula score is the matching criterion.

The problem of excluding formula scored items from a criterion occurs when operational items undergo DIF analysis. In this case, high DIF items frequently need to be dropped from the initial formula scored criterion. The item to-be-studied is usually already part of the criterion when DIF analyses are initially conducted. As a result, in the initial analysis, no adjustments need to be made because the item to-be-studied is already part of the criterion. For the purified analysis, however, items flagged for high DIF need to be removed from the criterion and the analysis done again. The correct criterion in a purified DIF analysis is a criterion with all high DIF items removed from the criterion. However, at the present time, no straightforward method exists to exclude high DIF items from a formula score criterion. As a result, for the purified DIF analysis of formula scored tests, a new formula score matching criterion is created by excluding the high DIF items, thereby necessitating the recalculation of the entire set of 2x2 tables.

The problem of including formula scored items in the criterion occurs in particular when pretest items undergo DIF analysis. For this DIF analysis, the criterion is usually the score on the operational test items that are administered with the pretest items. The purified criterion resulting from the DIF analysis of the operational items then needs to be adjusted each time an item is studied because of the need to include the studied item in the criterion. As just discussed, this adjustment is difficult for a formula scored criterion. As a result, it is common practice to use a number-right criterion for pretest DIF analyses, even though the items were actually administered under formula scored directions.

The Problem of Using a Rights Scored Criterion for Formula-Scored Tests

The use of a rights scored DIF criterion when a test is administered under formula scored conditions results in several inconsistencies. One clear inconsistency is that the actual items will be scored right or wrong (right =1, wrong = 0) when included in the criterion rather than formula scored (right = 1, no response = 0, wrong = -1/(k-1)) as was done for creating scores. Another inconsistency is that the number right criterion is inconsistent with the instructions given to examinees. If an examinee had been told a test was to be scored rights-only instead of by formula, it is likely that many fewer items would have been omitted than were actually omitted under the formula scoring criterion.

The effect on MH D-DIF statistics of using a rights scored criterion for formula scored tests has not been formally studied. Informal analysis of operational DIF results suggests that there is a relationship between item difficulty and the MH D-DIF statistics produced with rights versus formula score criteria. When the criterion is changed from formula-scored to rights-scored, MH D-DIF statistics appear to become slightly more negative for easy items and slightly more positive for harder items.

A Multivariate Criterion

Holland (1990) suggested a possible solution to the problem of excluding or including formula scored items in the DIF criterion:

The idea is to match examinees jointly on their number of correct answers and their number of omitted and not-reached responses (herein simply called "omits"). For a set of J items, the number of right (r), wrong (w) and omitted (m) responses satisfies the equation:

$$r + w + m = J. \tag{1}$$

The pair (r, m) satisfies these inequalities.

$$0 \leq r \leq J \ , \ 0 \leq m \leq J$$

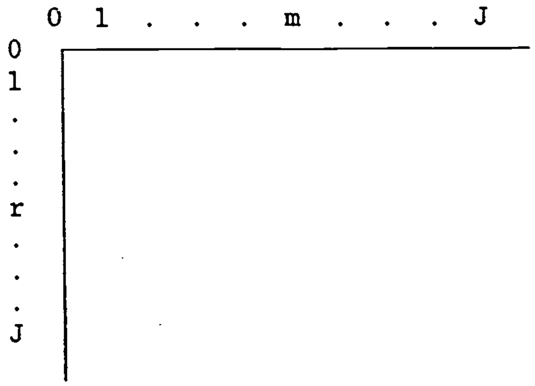
$$0 \leq r + m \leq J. \tag{2}$$

For a formula-scored test in which all items have the same number of options (i.e. SAT-V but not SAT-M), examinees with the same pair (r, m) will have the same formula score, due to the equation

$$r - w/k = (k + 1)r/k + m/k - J/k, \tag{3}$$

where k + 1 is the number of answer options on the items in the test.

The pair (r, m) satisfies (2) and may be arranged into a triangular array, i.e.



It may be useful to note that the columns of this triangular array can be stored in a long vector via the relationship

$$location\ (r,\ m) = 1 + r + m(J + 1) - m(m -1)/2 \tag{4}$$

when (r, m) satisfies (2).

[Note: A restriction on the discussion given here is the assumption that k is the same for all items. This can be accomplished in practice by only using items with the *same number of answer options* in the DIF analysis of a given studied item.]

Note that from (4) we have

$$\mathit{location}\ (r - 1, m) = \mathit{location}\ (r,\ m) - 1, \tag{5}$$

for $r = 1, \ldots, J$, and

$$\mathit{location}\ (r,\ m - 1) = \mathit{location}\ (r,\ m) + m - J - 2, \tag{6}$$

for m = 1,...,J. These relationships are useful for adding a not-yet-included study item into the matching criterion (see below).

. . .The basic data for a DIF analysis is this 2 by 3 table indexed by the pair (r, m):

	Right	Wrong	Omits
Reference	A_{rm}	$B^{(1)}_{rm}$	$B^{(2)}_{rm}$
Focal	C_{rm}	$D^{(1)}_{rm}$	$D^{(2)}_{rm}$

(7)

where $(r,\ m)$ satisfies (2).
This table may be collapsed to a 2 x 2 table of the form

	Right	Wrong
Reference	A_{rm}	B_{rm}
Focal	C_{rm}	D_{rm}

(8)

where $B_{rm} = B^{(1)}_{rm} + B^{(2)}_{rm}$ or $B_{rm} = B^{(1)}_{rm}$ depending on the treatment of omits in the DIF analysis. Similarly, $D_{rm} = D^{(1)}_{rm} + D^{(2)}_{rm}$ or $D_{rm} = D^{(1)}_{rm}$ depending on the treatment of omits.

Once a table of the form (8) is in hand, we would simply use it in a DIF analysis like the ones we do now. MH will be ok, but standardization may begin to break down because of the finer matching.

Now suppose that we have computed table (7) *excluding* the studied item from (r, m) and that we wish to include the score on the studied item in the computation of (r, m), i.e. the pretest case. Hence, we want a new 2 by 3 table of the form:

	Right	Wrong	Omits
Reference	A^{*}_{rm}	$B^{(1)*}_{rm}$	$B^{(2)*}_{rm}$
Focal	C^{*}_{rm}	$D^{(1)*}_{rm}$	$D^{(2)*}_{rm}$

(9)

where (r, m) refers to the criterion that includes the studied item and satisfies (2) with J replaced by $J + 1$.

Remembering that (r, m) in (8) is computed excluding the score on the studied item while (r, m) in (9) includes the score on the studied item, the relation between the entries in (7) and (9) is given by these equations:

$$\begin{gathered} A^{*}_{rm} = A_{r-1m} \ , \ C^{*}_{rm} = C_{r-1m} \\ B^{(1)*}_{rm} = B^{(1)}_{rm} \ , \ D^{(1)*}_{rm} = D^{(1)}_{rm} \\ B^{(2)*}_{rm} = B^{(2)}_{rm-1} \ , \ D^{(2)*}_{rm} = D^{(2)}_{rm-1} \end{gathered} \tag{10}$$

for $r = 1,\ldots,J + 1 \quad m = 1,\ldots,J + 1$.

Note that in (10) we can set $A_{-1,m} = C_{-1,m} = B^{(2)}_{r,-1} = D^{(2)}_{r,-1} = 0$ in order to allow r and m to range over $r = 0,\ldots,J + 1$ and $m = 0,\ldots,J + 1$, with $0 \leq r + m \leq J + 1$. (Holland, 1990, pp 1-4)

This kind of matching permits easy adjustment whenever a studied item needs to be included in or excluded from the criterion. The adjustments are straightforward because the 2 x 2 tables at each matching level have been expanded to include number of omits as well as number of rights and wrongs, and are now 2 x 3 tables. As a result, when items are included or excluded, the adjustments can be made in the number of rights and number of omits at each bivariate score level. Thus, the 2 x 3 tables can be easily corrected to include the analyzed item. For calculation of the MH statistic, these 2 x 3 tables are collapsed into 2 x 2 tables with omits and not reached items either counted as wrong or excluded. As a result, Holland's method allows the straightforward inclusion or exclusion of the formula scored item-to-be-studied.

Two possible problems with Holland's approach have been identified. First, because the multivariate matching requires matching in many more cells than univariate matching, Petersen (1990) has questioned the stability of the DIF indices generated from the Holland method when there are small sample sizes (and, as a result, many cells with zero examinees). Second, Holland (1990) viewed the blurred distinction between omits and not reached items as a possible problem. Recall that for the matching on omits, the number of omits and not reached items are summed. Theoretically, examinees could be matched on omits separately and not reached separately. Holland felt, however, that adding another variable, so that omitted and not reached items could be separated, would be "gilding one too many lily". A problem with not treating not reached items separately is that they cannot be excluded from analyses while omits are included. When tests are speeded, Dorans, Schmitt and Curley (1988) found that including not reached items as wrong in a DIF analysis resulted in larger negative DIF for minority groups affected by the speededness.

This study conducted an initial evaluation of bivariate matching as a solution to the problem of studying DIF on formula scored tests. The bivariate approach matches examinees jointly on their number of correct responses and their number of omitted and not reached responses. This bivariate matching allows a straightforward adjustment whenever a studied item needs to be included in or excluded from a criterion. This matching provides a solution to the problem of including or excluding the item to-be-studied in the formula score criterion in a DIF analysis of a formula scored item. However, this approach may have some problems. This study was an initial evaluation of bivariate matching. It compared the MH D-DIF values and resulting DIF categories from the use of a bivariate matching criterion with those from both rights and formula scored criteria.

METHOD

Because this was an initial evaluation of the bivariate method, only selected analyses were conducted. These analyses were of MH D-DIF values from two comparisons employing different matching criteria, with two sample sizes and administrations, and involving final form and pretest data. The MH D-DIF values from the different criteria were examined for their consistency and the consistency of their resulting DIF classifications.

Matching Criteria

Four different criteria for the DIF analyses were contrasted. The first three criteria were variations of current ETS DIF procedures, and the fourth was the proposed bivariate method. The first criterion was a number right score with omits and not reached counted as wrong in the analysis of individual items. Although counting not reached as wrong does not correspond to what is done when MH is commonly used, this resulted in a more direct comparison with the bivariate criterion and was done for most analyses. In the text and tables that follow, this criterion is referred to as the rights criterion.

The second criterion was a formula score criterion with omits and not reached both included as wrong for analysis of individual items. In the text and tables that follow, this is referred to as the formula score criterion with not reached counted as wrong (NR=W). The third DIF criterion was a formula score criterion, with omitted items counted as wrong, but not reached items excluded from the analysis. This criterion is referred to as the formula score criterion with not reached not included (NR=NI). (This corresponds to the approach currently used when MH is applied to SAT data.)

The fourth criterion was Holland's method which used the bivariate distribution of rights and omits plus not reached for matching, with omits and not reached counted as wrong for analysis of individual items. In the text and tables that follow, this is referred to as the bivariate criterion. As noted earlier, the bivariate method does not allow for separate treatment of omitted and not reached items. Including both omitted and not reached item responses as wrong was thought to be more plausible than excluding both from the analysis, and as a result, only this alternative was evaluated.

In this initial evaluation of the bivariate matching method, comparisons were stressed between MH D-DIF values from the different criteria rather than regarding one method as the baseline. This was decided for two reasons. One, both rights and formula scores are used as criteria in DIF analyses of formula scored tests. As a result, it is important to compare the results from the bivariate method with both of these criteria. Two, it is not clear which method should be the criterion. The formula score is consistent with the test instructions and is the rights score adjusted for guessing. But the formula score criterion has additional error due to rounding to the nearest integer after the adjustment. The bivariate criterion is equivalent to matching on unrounded formula scores. As a result, the bivariate criterion could be regarded as a better measure than the other criteria of the ability that the test is measuring. However, as Zwick[5] pointed out, from a Mantel-Haenszel perspective, it is not

[5] Zwick pointed out that Holland and Thayer's results on the inclusion of the studied items are based on the fact that, in the Rasch model, number right score is a sufficient statistic for the latent ability, theta. It has not been shown that the formula score or bivariate scores are sufficient statistics for the item response model that is assumed for formula-scored tests.

clear that the formula score or the bivariate scores are the correct matching variable.

Comparisons and Data

DIF analyses were conducted on two comparisons, Male and Female examinees and White and Black examinees. These two particular comparisons were chosen for the study because when the SAT data was analyzed operationally using all focal groups, the Male/Female comparison had the most high DIF items and the White/Black comparison showed the largest ability difference. SAT-Verbal final form and pretest data with five response alternatives for all items were used for these analyses.

DIF analyses for the Male/Female comparison were done separately on March and on May administrations of the SAT with two different sample sizes. One sample was the full group of test takers from the administration and the other sample was very close in size to the minimum recommended for a DIF analysis. DIF analyses for the White/Black comparison were done on the March and May administrations but with only the full group of test takers.

The pretest DIF analysis was done for only the Male/Female comparison and used minimum samples from the May administration. To model a pretest situation, analyses were conducted on a set of five SAT pretest items with an external criterion made up of purified operational SAT items. Only three of the four criteria were used for these DIF analyses; the rights scored criterion with omits and not reached counted as wrong for analysis of individual items, the formula score criterion with omits and not reached as wrong, and the bivariate matching method with omits and not reached counted as wrong.

Evaluation Indices

Agreement between DIF indices was evaluated through correlation coefficients and the consistency of DIF classifications. The Pearson correlation coefficient was used to correlate the MH D-DIF statistics based on the different criteria. The consistency of DIF classifications was calculated for the Male/Female comparison and the White/Black comparison. To assess classification consistency, the number of items were identified that fell in the same DIF category using the different criteria. The classification scheme used in establishing the categories is the scheme that has been adopted by and is currently in use at the Educational Testing Service. This classification scheme (Petersen, 1987), which uses the value of the MH D-DIF statistic and its corresponding standard error, is as follows:

	Absolute Value and Significance of Mantel-Haenszel Delta Difference
Category	(MH D-DIF) Statistic
A	MH D-DIF not significantly different from 0 (.05 level) OR Absolute value less than 1
B	MH D-DIF significantly different from 0 (.05 level) AND EITHER (1) Absolute value at least 1 but less than 1.5 OR (2) Absolute value at least 1 but not significantly greater than 1 (.05 level)
C	Absolute value of MH D-DIF at least 1.5 and significantly greater than 1 (.05 level)

To facilitate the evaluation of the results of the DIF analyses for the same items using the different matching criteria, MH D-DIF values for a particular DIF criterion were plotted against the D-DIF values from each of the other DIF criteria. In the figures contained in Appendix A, a reference line was determined by setting the standard deviates from each of the two sets of MH D-DIF values equal. The slope of the line is S_y/S_x, where S_y is the standard deviation of MH D-DIF values from criterion y, and S_x is the standard deviation of criterion x. The intercept of the line is M_y - AM_x, where M_y is the mean of MH D-DIF values from criterion y, M_x is the mean of the criterion x, and A is the slope of the line. When means and standard deviations for the two sets of MH D-DIF values are very similar, this line will closely follow the 45 degree line across the plot. Each figure also contains the classification consistency results based on the two different matching criteria in a table in the right hand corner. This table indicates the number of items consistently and inconsistently classified by the two matching criteria. The symbols from the table are displayed in the figures.

Analyses

These correlation coefficients and DIF classifications were the primary indices used in the following analyses. The MH D-DIF values and DIF classifications were first compared from the different criteria for the large samples. Then, the same comparisons were made for the small samples. To assess stability across sample size, the MH D-DIF values and DIF classifications from each criterion for the small samples were compared with those for the same items from the large samples.

In addition, this study afforded an opportunity to document differences between MH D-DIF values from a rights and a formula score criterion in DIF analyses of formula scored tests. As a result, because informal analysis of DIF results using rights versus formula scored matching criteria suggested an effect by item difficulty, differences in MH D-DIF values from these two criteria were plotted against item difficulty.

RESULTS

Table 1 displays the correlations between the MH D-DIF values from the analyses using the different matching criteria. For each criterion comparison, bivariate vs. formula, bivariate vs. rights, and formula vs. rights, correlations are given for different samples. The samples sizes for the large sample comparisons varied from 257,414 for the White May sample to 18,484 for the Black March sample. Sample sizes for the small sample comparisons ranged from 419 for the Male May sample to 200 for the Female March sample. Sample sizes for the pretest comparison were 416 for the Male sample and 120 for the Female sample.

As Table 1 shows, correlations between MH D-DIF values from the different criteria with large sample sizes are very high, from .952 to .999. With large sample sizes, it appears that the MH D-DIF statistics based on the different matching criteria are ordering items with respect to DIF in a very similar fashion. For the smaller sample sizes, the correlations between the MH D-DIF values based on the different criteria are lower, particularly the correlations involving the bivariate criterion. However, for the small sample formula score-rights criteria comparison, the correlations are still quite high, .960 and .971. Correlations for the small sample pretest comparisons are also vary high, .985

and .990. It should be noted that the correlations for the pretest analysis are based on only five items.

Tables 2 through 8 summarize the MH D-DIF values and the DIF classifications for the same items from use of the different matching criteria. These tables display the number of 'B' and 'C' items and MH D-DIF summary statistics from the use of the different matching criteria with the different samples. A positive MH D-DIF value indicates that the item is easier for the focal group than the reference group after matching on the criterion score. A negative MH D-DIF value indicates that the item is more difficult for the focal group than the reference group after matching on the criterion score. The mean and standard deviation of the MH D-DIF values and the MH D-DIF standard errors are displayed. In operational DIF analyses, mean DIF values of plus or minus .10 are considered reasonable. The standard error represents the stability of the MH D-DIF value and is closely related to sample sizes at the levels of the matching variable. In all analyses (except the pretest analysis in Table 8), 85 items (the total set) from the SAT-Verbal were analyzed. Scatterplots for the MH D-DIF values with tables of item classifications are in Appendix A.

Table 2 (see also Appendix Figures A-1 to A-4) shows that for the Male-Female comparison with the large March sample there is no difference in the number of items classified 'B' and 'C'. Mean MH D-DIF values are slightly different between the rights and formula score matching criteria. The means and standard deviations of the MH D-DIF standard errors are identical. Appendix Figures A-1 to A-4 also show the close correspondence between the MH D-DIF values derived from the different matching criteria. The differences in these figures appear most related to whether not reached items are treated as wrong or excluded from the analysis.

Table 3 (see also Appendix Figures A-5 to A-8) shows that for the Male-Female comparison in the large May sample there are slight differences in the number of items classified 'C'and 'B'. Mean 'H D-DIF values from the rights and formula score criteria are also slightly different. The mean and variability of the standard errors of the MH D-DIF values are identical. Appendix Figures A-5 to A-8 also show that the different score criteria yield very similar MH D-DIF values. The bivariate criterion MH D-DIF values are slightly different from the others. Figure A-8 indicates that the MH D-DIF values from the rights criterion are slightly more negative than those from the formula score criterion.

Table 4 (see also Appendix Figures A-9 to A-12) shows the classifications and statistics for the large March sample White-Black comparison. The rights criterion results in one more 'B-' item than the other criteria. Slight differences exist between the MH D-DIF means and standard deviations across the matching criteria. The MH D-DIF standard error means and standard deviations are identical. Figures A-9 to A-12 in Appendix A indicate that the bivariate MH D-DIF values are very similar to those from the formula score criterion with not reached items as wrong and show the most scatter with the MH D-DIF values from the rights criterion. As shown in Figure A-12, the largest differences are between the rights and formula score criteria with not reached items as wrong.

Table 5 (see also Appendix Figures A-13 to A-16) displays the same information for the large May sample White-Black comparison. Here, the rights criterion results in the same number of 'B' and 'C' items as the other criteria. Slight differences again exist between the MH D-DIF means and standard deviations. The MH D-DIF standard error means and standard deviations are again identical. Figures A-13 to A-16 in Appendix A indicate that the bivariate MH D-DIF values are very similar to the formula score criterion with not reached items as wrong and show the most divergence with those from the rights criterion. The largest differences are between the rights and formula score with not reached it3ms as wrong and are shown in A-16.

The results displayed in Tables 1 through 5 strongly suggest that MH D-DIF values based on a bivariate matching criterion are closely related to those from the other criteria. They appear to have the same order of DIF magnitude. The MH D-DIF values from the bivariate criterion identify items having extreme amounts or levels of DIF in the same way as MH D-DIF values based on formula and rights scored matching criteria. The Figures indicate that the MH D-DIF values from the bivariate criterion are most related to those from the formula score criterion with not reached items as wrong and least similar to those from the rights criterion. However, all MH D-DIF values from the different criterion based on the large samples are very similar.

Tables 6 through 8 show the same information as Tables 2 through 5, but for the small samples. For the small samples, there are clear differences in the number of 'B' and 'C' items identified using the different matching criteria and clear differences in the summary statistics. Figures A-17 to A-24 in Appendix A clearly show these differences in contrast to Figures A-1 through A-16 for the large samples.

Table 6 and Table 7 display the values from the Male-Female comparison for the March and May small sample analyses, respectively. Table 6 and Table 7 indicate that the bivariate criterion results in fewer 'B's and 'C's, only about half as many as the other matching criteria. Use of the rights criterion results in the most 'B's and 'C's for the March sample. This criterion also resulted in a high mean MH D-DIF value. The formula score with not reached included as wrong matching criterion results in the most 'B's and 'C's for the May sample. Both

Tables 6 and 7 show that the variability of the MH D-DIF values from the bivariate criterion is much larger than from the other criteria. The variability of the MH D-DIF standard errors from the bivariate cri:erion is more than three times that for the other criteria.

This variability is clearly shown in the contrast between Appendix Figures A-17, A-18, A-19, and Appendix Figure A-20. Figures A-17, A-18, and A-19 show that negative MH D-DIF values from the rights and formula score criteria have a tendency to become more negative when a bivariate criterion is used. The positive MH D-DIF values from the rights and formula score criteria become more positive when the bivariate criterion is used. These tendencies are also shown in Figures A-21, A-22, and A-23. Figure A-20 and Figure A-24 show that the MH D-DIF values from the rights and formula score criteria are much more similar than those from the bivariate criterion. Figure A-20 also shows that the MH D-DIF values from the rights criterion are slightly more positive than those from the formula score with not reached as wrong criterion. Figure A-24 does not confirm this tendency.

Table 8 (see also Appendix Figures A-25 and A-26) displays the DIF classifications and the MH D-DIF values for the set of five pretest items (means and standard deviations were not calculated because there were only five items). The bivariate matching criterion results in the fewest 'B's and 'C's. The MH D-DIF values from the bivariate criterion are different from those of the other two criteria and the standard errors much larger. Appendix Figures A-25 and A-26 display a tendency for the bivariate MH D-DIF values to be more negative for items with negative MH D-DIF values from the other criteria.

Table 9 disp'ays the correlations between the MH D-DIF values derived with each matching criterion from the large and small samples of the March administration for the Male-Female comparison. As expected because of the larger standard errors, MH D-DIF values from the large and small samples with the bivariate matching criterion have the lowest correlation of the four different matching criteria. The correlations for the formula and rights scored matching criteria are of the same magnitude as those reported by Wright (1987) for large and small sample stability for focal groups of similar sizes.

Appendix Figures A-27 to A-30 compare the MH D-DIF values from small samples with those from large samples using the same criterion for the same items. The figures indicate that items with negative MH D-DIF values in the large samples have more negative MH D-DIF values in the small samples. Those items with positive MH D-DIF values in the large samples have more positive MH D-DIF values in the small samples. Although all four criteria showed these tendencies, Appendix Figure A-30 shows that the tendencies were most pronounced for the bivariate criterion.

The results from the large samples lead one to conclude that MH D-DIF values using a bivariate criterion are ordered in the same way as those based on formula and rights scored matching criteria. However, with small sample sizes, the bivariate matching criterion produces MH D-DIF values that are different from and with larger standard errors than those based on the other matching criteria. These values are also different from those obtained in the larger samples with the bivariate criterion. There appears a clear tendency for MH D-DIF values from small samples to be more negative for items that had negative MH D-DIF values in the large samples, and to be more positive for items identified as positive MH D-DIF in the large sample analysis. The results also indicate that the formula and rights scored criteria produce more stable MH D-DIF values across different sample sizes than did the bivariate criterion.

In addition, the results of the above analyses show that MH D-DIF values from a rights scored criterion and a formula scored criteria are very similar but slightly different. Figures 1 through 4 illustrate how the differences between MH D-DIF values from rights and formula score with not reached as wrong matching criteria are influenced by item difficulty. These figures display item difficulty on the y-axis (in the delta metric) and the difference on the x-axis between the MH D-DIF values from use of rights and formula scored matching criteria (Difference = Formula MH D-DIF - Rights MH D-DIF). The same difference axis is used for all four graphs.

Figure 1 shows the differences between the MH D-DIF values from a rights score criterion and a formula score criterion for the Male-Female comparison for the March large sample. (When Rights MH D-DIF values are subtracted from Formula Scored MH D-DIF values, positive differences are produced from more negative Rights values.) This figure displays a trend for positive differences to increase in size with decreasing item difficulty (lower delta values). But the differences, which range from .00 to .13, are small. Figure 2 displays the same information from the Male-Female comparison for the May large sample analysis. Again, MH D-DIF values for easy items based on a rights scored matching criterion tend to be more negative (produce more positive differences) than those based on a formula scored matching criterion. However, the actual differences are small.

Figures 3 and 4 show the same differences for the White-Black comparison. Figure 3 displays the White-Black MH D-DIF values from the large sample March administration and Figure 4 from large sample May administration. Here the trend was more pronounced with larger differences in D-DIF values, up to .36. These figures indicate that when a rights scored criterion is used instead of the formula scored criterion, more difficult items display more positive DIF values and less difficult items show more negative MH D-DIF values.

Conclusions

In this study, a solution was evaluated for the problem of studying DIF for formula scored tests. A bivariate matching criterion had been proposed that provided a straightforward method of including and excluding studied items from the formula score matching criterion in DIF analyses of formula scored tests. The bivariate method was evaluated by comparing its MH D-DIF values and resulting DIF classifications with those from rights and formula scored matching criteria in DIF analyses of formula scored tests. The results suggest that the problem of including or excluding items from the criteria in a DIF analysis of a formula scored test was solved for DIF analyses only when very large samples are available.

The bivariate method was effective with large samples but performed less well with smaller samples. With the large samples used in this study the MH D-DIF values from the bivariate criterion were ordered in the same way as those for the rights and formula score criteria and produced nearly identical DIF classifications. With smaller sample sizes, however, the price of the bivariate matching was an increase in the variability of the MH D-DIF values. This produced less stability across different sample sizes and less consistency with the MH D-DIF values from other criteria. However, because this study used very large and very small samples, further research is needed to determine the exact sample size at which the bivariate matching produces MH D-DIF values consistent with those from the other criteria.

In addition, this study provided some clarification as to the effects of substituting a rights score for the formula score criterion in these analyses. The use of the formula score as the matching criterion for a formula scored test is logically satisfying in contrast to matching on the rights score. However, the price of using this matching is the difficulty involved in adding or dropping items from the criterion. These difficulties have led several testing programs to substitute a rights score criterion in place of the formula score criterion.

The primary reason for using a rights score as the matching criterion for a formula scored test is convenience. While the use of the rights score as the criterion is not logically consistent with a test given under formula scored directions, it is easy to include or exclude items from the criterion when using this method. The results of this study indicate that the price of its easy use is a difference in the MH D-DIF values in comparison to those calculated using the formula score as the criterion. The size of these differences appears related to the size of the ability difference between the groups in the comparison. The results indicate that the differences are largest when matching is most needed, i.e., when large ability differences exist. The direction of the differences between MH D-DIF values derived using a rights scored criterion in contrast to a formula score criterion are related to an item's difficulty.

The results of this study suggest the need for further research into two areas. The first area is the identification of the smallest sample size at which bivariate matching produces MH D-DIF values that are consistent with those from large sample sizes and with those from the formula score criterion. It is possible that additional research will show that the bivariate method produces reasonable results at some practical sample size. This would allow testing programs that met the identified minimum sample size to use bivariate matching as a solution to the problem of DIF analyses of formula scored tests. The second area that needs further research is the effect of a rights score criterion substituted for the formula score criterion on MH D-DIF values in the DIF analysis of a formula scored test. Even with the large samples used in this study, the MH D-DIF values for the White/Black comparison from the rights score criterion displayed moderate differences from those of the formula score criterion. These differences could be much larger in the sample sizes ordinarily used in operational DIF analyses. Further research is needed into this issue to ensure that substitution of a rights score for a formula score criterion does not produce a sizable bias in MH D-DIF estimates.

References

Dorans, N. J., Schmitt, A. P., & Curley, W. E. (1988, April). *Differential speededness: Some items have DIF because of where they are, not what they are.* Paper presented at the annual meeting of the NCME, New Orleans, LA.

Holland, P. W. (1990, June). *DIF and formula scores.* Unpublished memorandum. Princeton, NJ: Educational Testing Service.

Holland, P. W., & Thayer, D. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.

Kulick, E., & Hu, P. G. (1989). *Examining the relationship between differential item functioning and item difficulty* (RR-89-18). Princeton, NJ: Educational Testing Service.

Petersen, N. S. (1990, August). Personal communication.

Petersen, N. S. (1987, September). *DIF procedures for use in statistical analysis.* Unpublished memorandum. Princeton, NJ: Educational Testing Service.

Wright, D. (1987). An empirical comparison of the Mantel-Haenszel and Standardization methods of detecting differential item performance. In A. P. Schmitt & N. J. Dorans (Eds.), *Differential item functioning on the Scholastic Aptitude Test* (RM-87-1). Princeton, NJ: Educational Testing Service.

Table 1. Correlations Between MH D-DIF Values from Different Criteria

		Criterion Comparison			
		Bivariate vs Formula			Rights vs Formula
Sample	Sample Sizes	Formula NR=W	Formula NR=NI	Bivariate vs Rights	Formula NR=W
Large					
March					
M-F	83,621/91,714	.999	.996	.998	.999
W-B	139,559/18,484	.997	.981	.971	.952
May					
M-F	145,237/164,039	.999	.998	.998	.999
W-B	257,414/20,875	.999	.985	.983	.975
Small					
March					
M-F	400/200	.660	.642	.662	.960
May					
M-F	419/220	.722	.695	.709	.971
Pretest					
M-F	416/120	.990	-	.985	-

Table 2. DIF Classifications and Summary Statistics for the March Large Sample Male-Female Comparison

	Criterion			
	Rights	Formula (NR=NI)	Formula (NR=W)	Bivariate
<u>Classification</u>				
C-	4	4	4	4
B-	3	3	3	3
B+	1	1	1	1
C+	2	2	2	2
<u>MH D-DIF</u>				
Mean	-0.09	-0.07	-0.05	-0.09
S.D.	0.77	0.75	0.76	0.77
<u>MH D-DIF SE</u>				
Mean	0.03	0.03	0.03	0.03
S.D.	0.01	0.01	0.01	0.01

Table 3. DIF Classifications and Summary Statistics for the May Large Sample Male-Female Comparison

	Criterion			
	Rights	Formula (NR=NI)	Formula (NR=W)	Bivariate
<u>Classification</u>				
C-	2	2	2	2
B-	2	1	1	2
B+	6	6	5	5
C+	1	1	2	1
<u>MH D-DIF</u>				
Mean	-0.07	-0.01	-0.02	-0.06
S.D.	0.73	0.73	0.73	0.74
<u>MH D-DIF SE</u>				
Mean	0.02	0.02	0.02	0.02
S.D.	0.01	0.01	0.01	0.01

Table 4. DIF Classifications and Summary Statistics for the March Large Sam;le White-Black Comparison

	Criterion			
	Rights	Formula (NR=NI)	Formula (NR=W)	Bivariate
Classification				
C-	0	0	0	0
B-	3	2	2	2
B+	0	0	0	0
C+	0	0	0	0
MH D-DIF				
Mean	0.00	0.01	-0.04	-0.01
S.D.	0.47	0.39	0.41	0.42
MH D-DIF SE				
Mean	0.05	0.05	0.05	0.05
S.D.	0.01	0.01	0.01	0.01

Table 5. DIF Classifications and Summary Statistics for the May Large Sample White-Black Comparison

	Criterion			
	Rights	Formula (NR=NI)	Formula (NR=W)	Bivariate
Classification				
C-	1	1	1	1
B-	0	0	0	0
B+	0	0	0	0
C+	0	0	0	0
MH D-DIF				
Mean	-0.02	-0.06	-0.03	-0.03
S.D.	0.48	0.45	0.44	0.44
MH D-DIF SE				
Mean	0.04	0.04	0.04	0.04
S.D.	0.01	0.01	0.01	0.01

Table 6. DIF Classifications and Summary Statistics for the March Small Sample Male-Female Comparison

	Criterion			
	Rights	Formula (NR-NI)	Formula (NR-W)	Bivariate
Classification				
C-	4	2	3	2
B-	5	5	4	2
B+	7	6	6	2
C+	1	1	2	1
MH D-DIF				
Mean	-0.13	-0.06	0.03	-0.05
S.D.	1.04	1.00	0.98	1.50
MH D-DIF SE				
Mean	0.56	0.56	0.56	1.21
S.D.	0.13	0.12	0.12	0.42

Table 7. DIF Classifications and Summary Statistics for the May Small Sample Male-Female Comparison

	Criterion			
	Rights	Formula (NR=NI)	Formula (NR=W)	Bivariate
<u>Classification</u>				
C-	2	1	1	1
B-	6	5	7	2
B+	6	5	7	3
C+	3	4	4	1
<u>MH D-DIF</u>				
Mean	-0.01	-0.06	0.02	-0.10
S.D.	0.93	0.99	1.01	1.26
<u>MH D-DIF SE</u>				
Mean	0.55	0.57	0.56	1.09
S.D.	0.13	0.14	0.14	0.42

Table 8. DIF Classifications and Summary Statistics for the May Pretest Sample Male-Female Comparison

	Criterion		
	Rights	Formula (NR-W)	Bivariate
Classification			
C-	1	1	1
B-	0	0	0
B+	0	1	0
C+	1	0	0
MH D-DIF Values*			
Item 206	-0.23 (.56)	-0.27 (.59)	-1.03 (1.01)
217	-2.36 (.71)	-2.72 (.79)	-3.77 (1.35)
231	1.46 (.78)	1.27 (.78)	1.22 (1.40)
235	2.44 (.87)	1.91 (.88)	2.79 (1.62)
239	-0.05 (.75)	0.34 (.79)	0.22 (1.31)

* Values in parentheses are standard errors.

Table 9. Correlations Between Large and Small Sample March MH D-DIF Values for the Male Female Comparison

	Criterion			
	Bivariate	Formula Score		Rights
		NR-W	NR-NI	
Correlation	.506	.758	.779	.715

Figure 1

Difference in MH D-DIF values for 85 Verbal items

Male-Female Comparison with Large March Sample

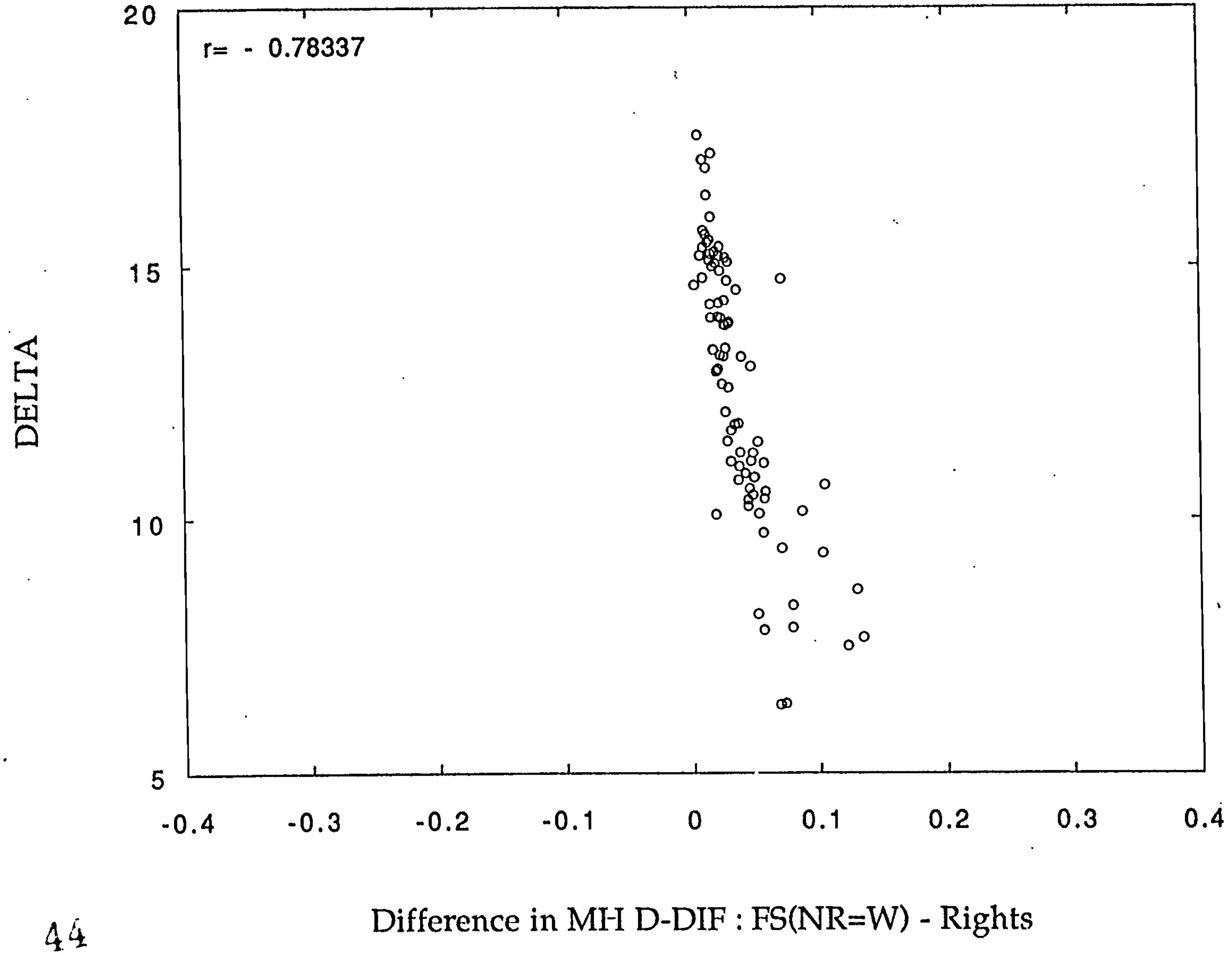


Figure 2

Difference in MH D-DIF values for 85 Verbal items
Male-Female Comparison with Large May Sample

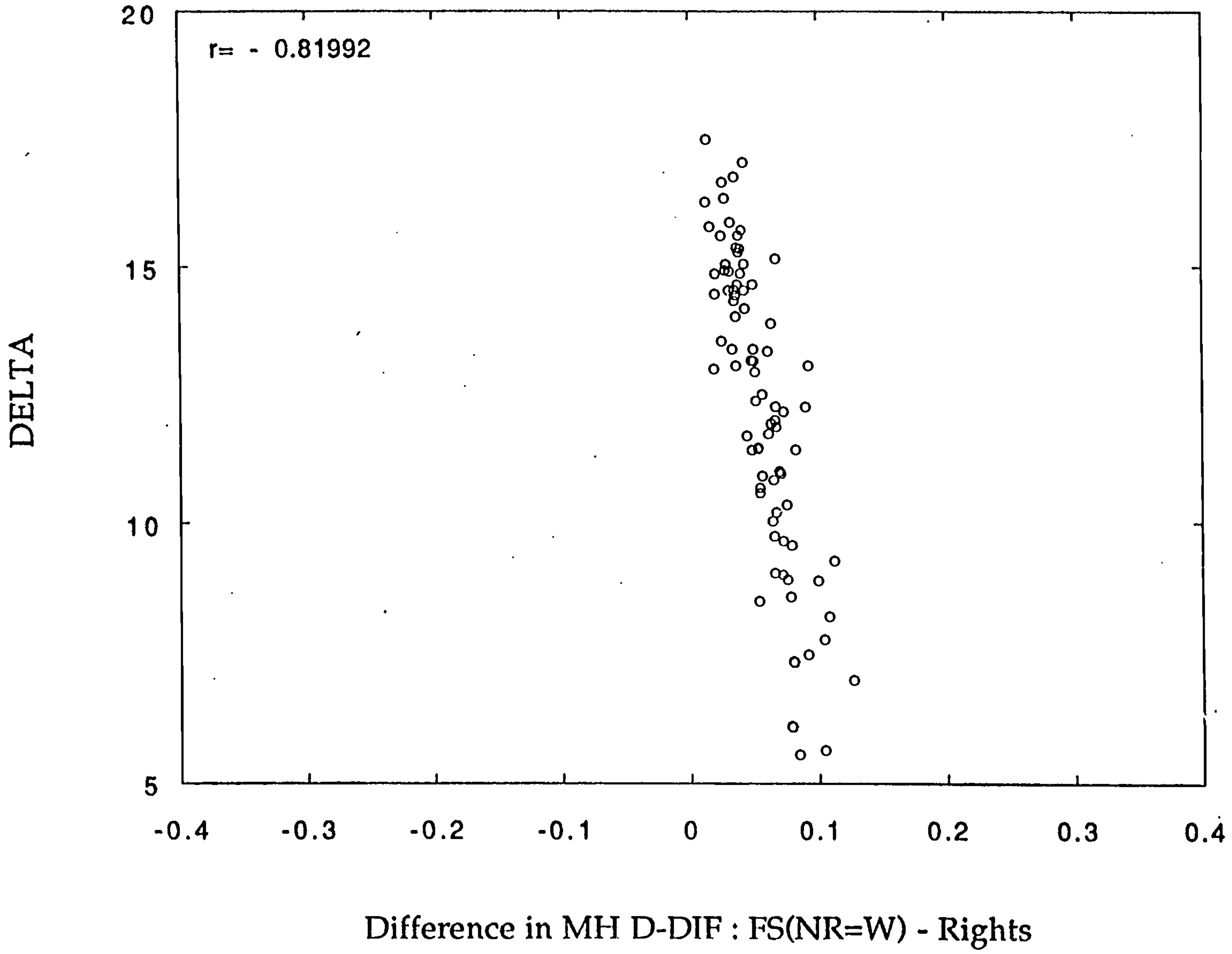


Figure 3
Difference in MH D-DIF values for 85 Verbal items
White-Black Comparison with Large March Sample

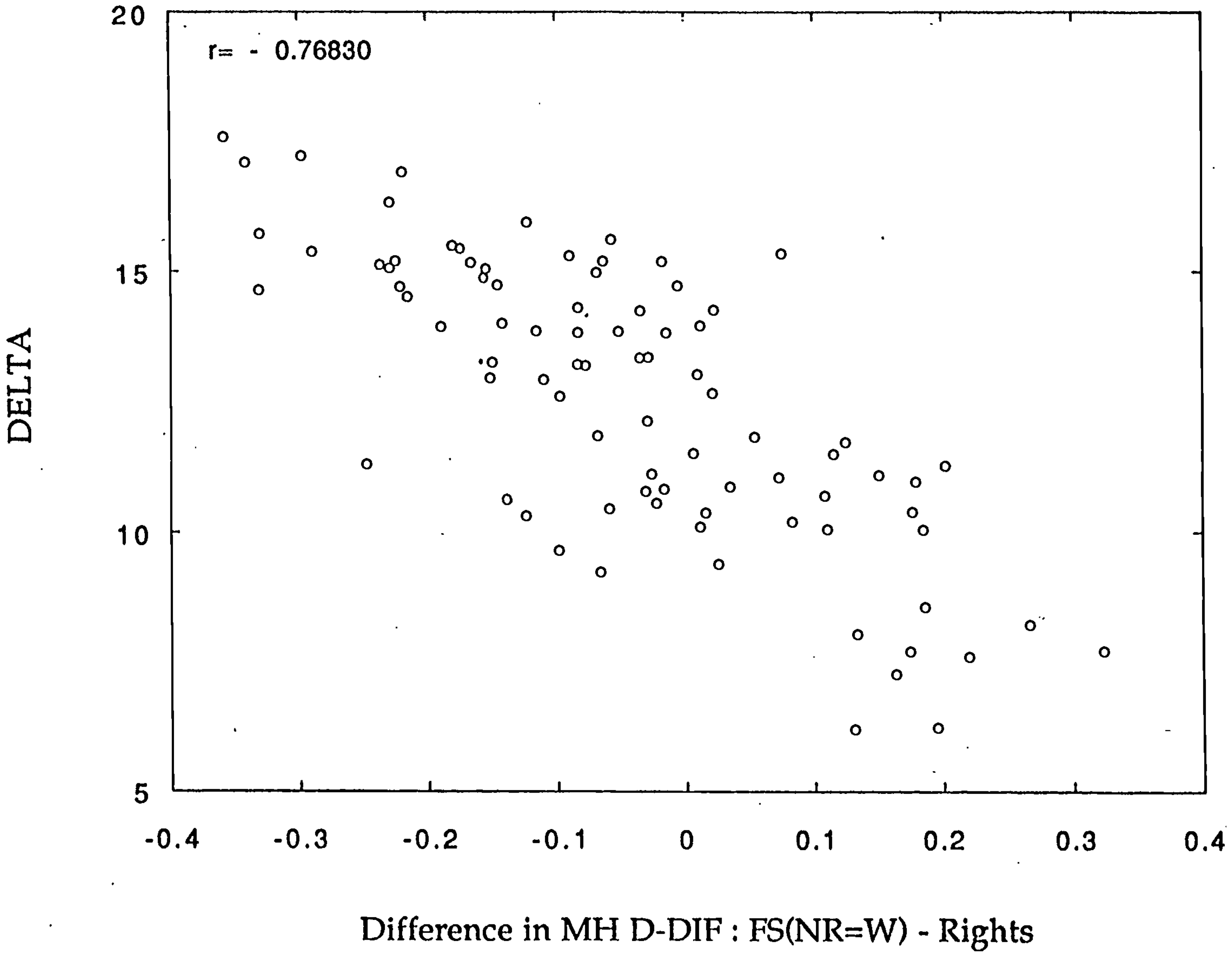
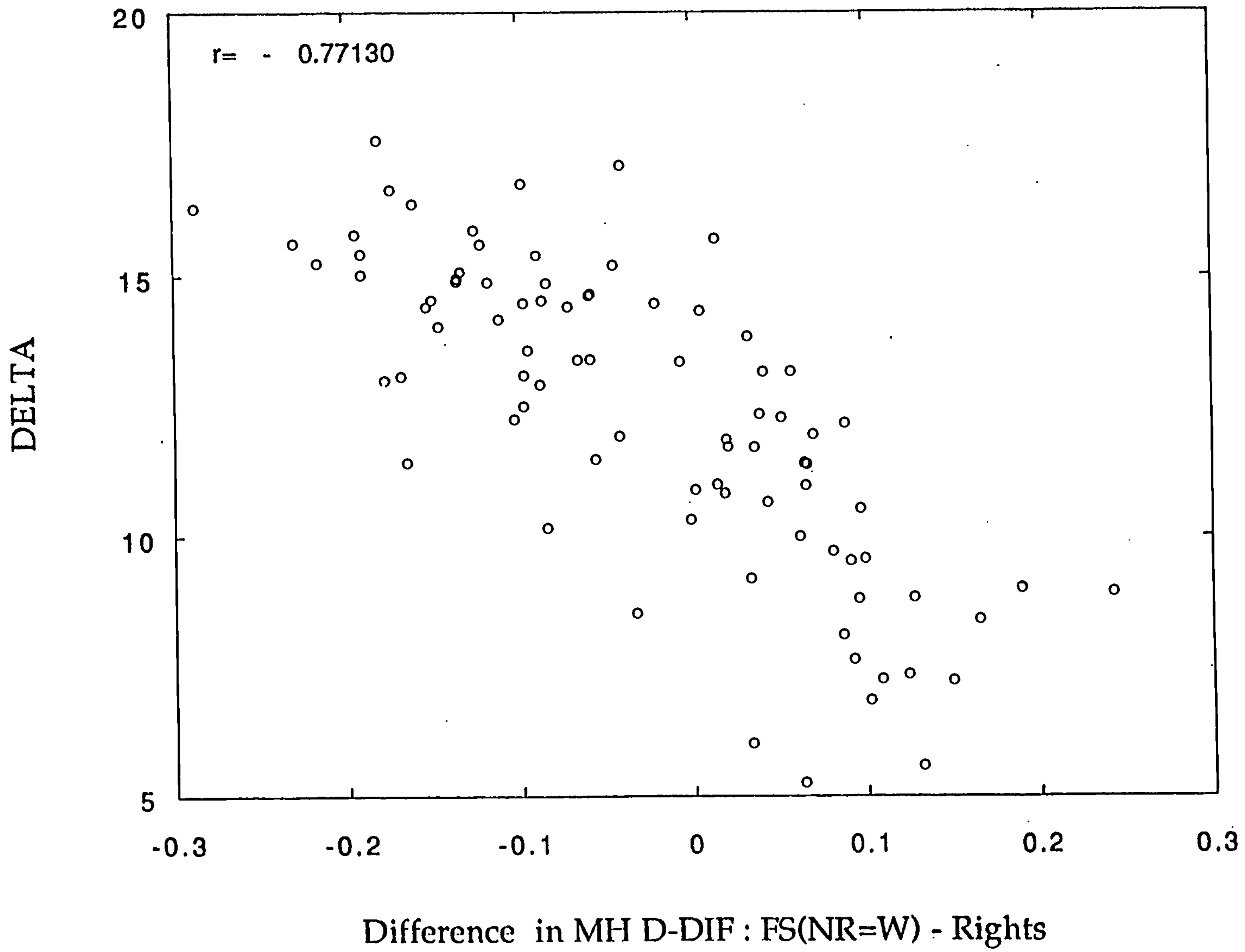
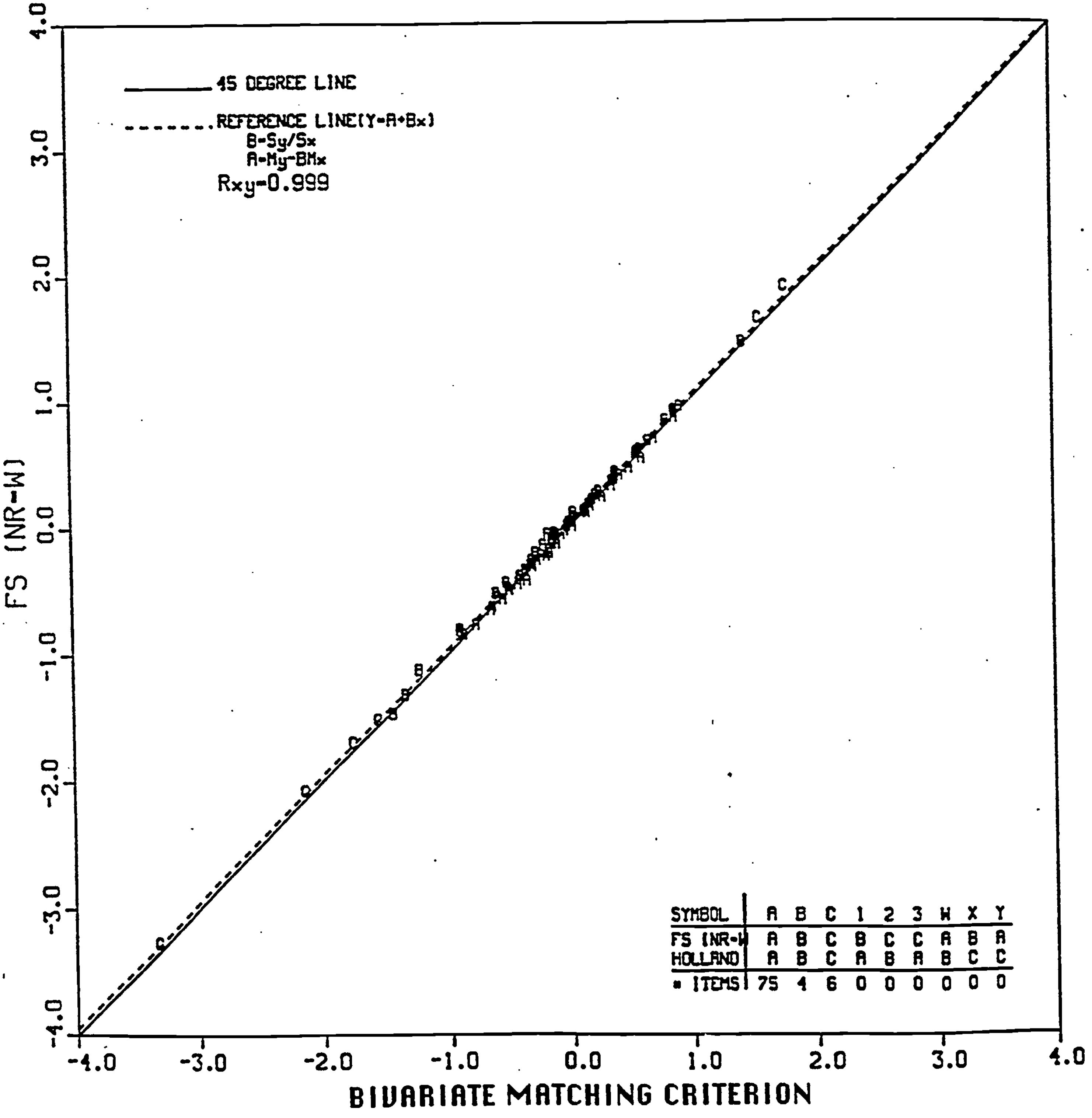


Figure 4
Difference in MH D-DIF values for 85 Verbal items
White-Black Comparison with Large May Sample

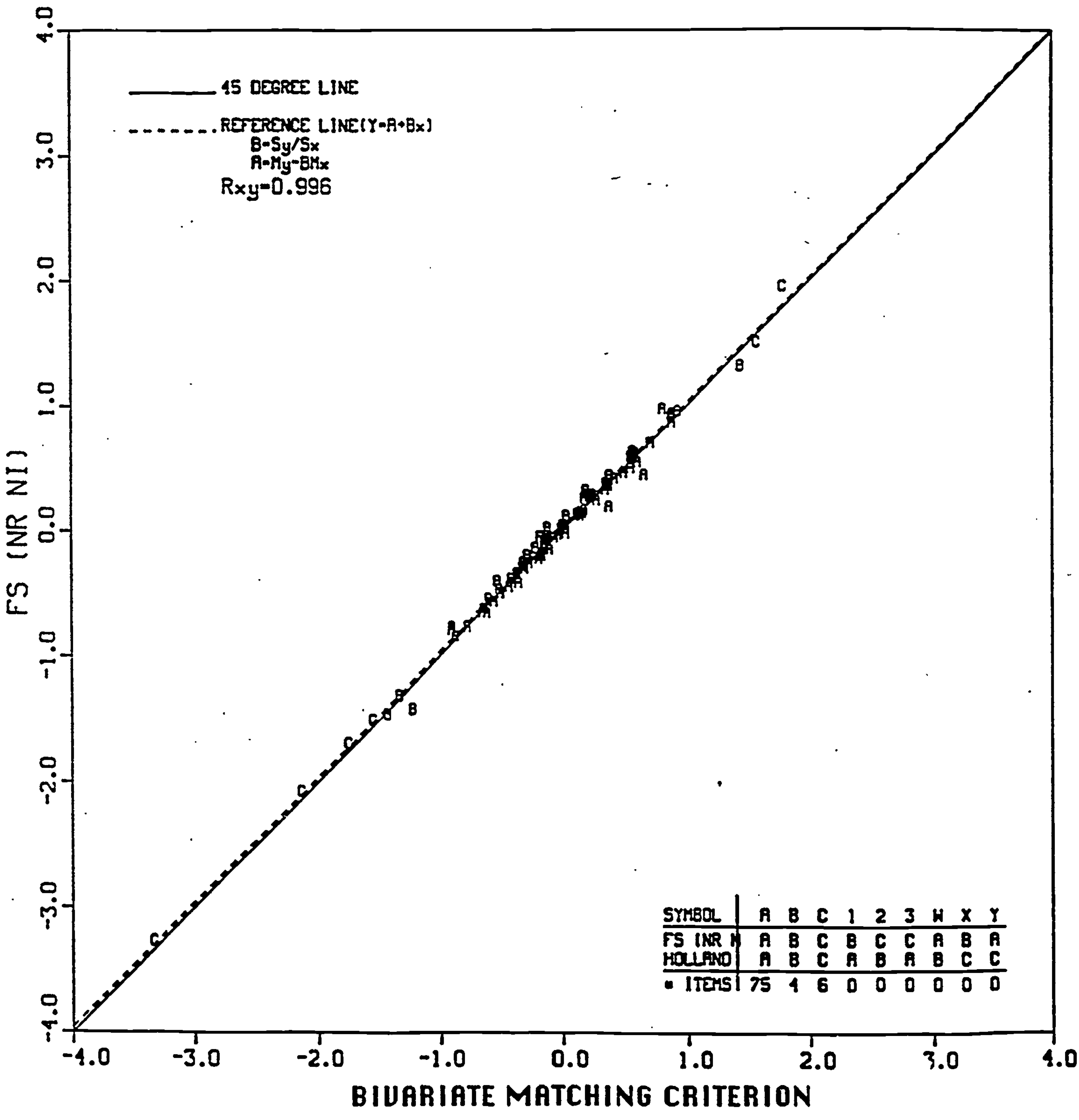


APPENDIX A

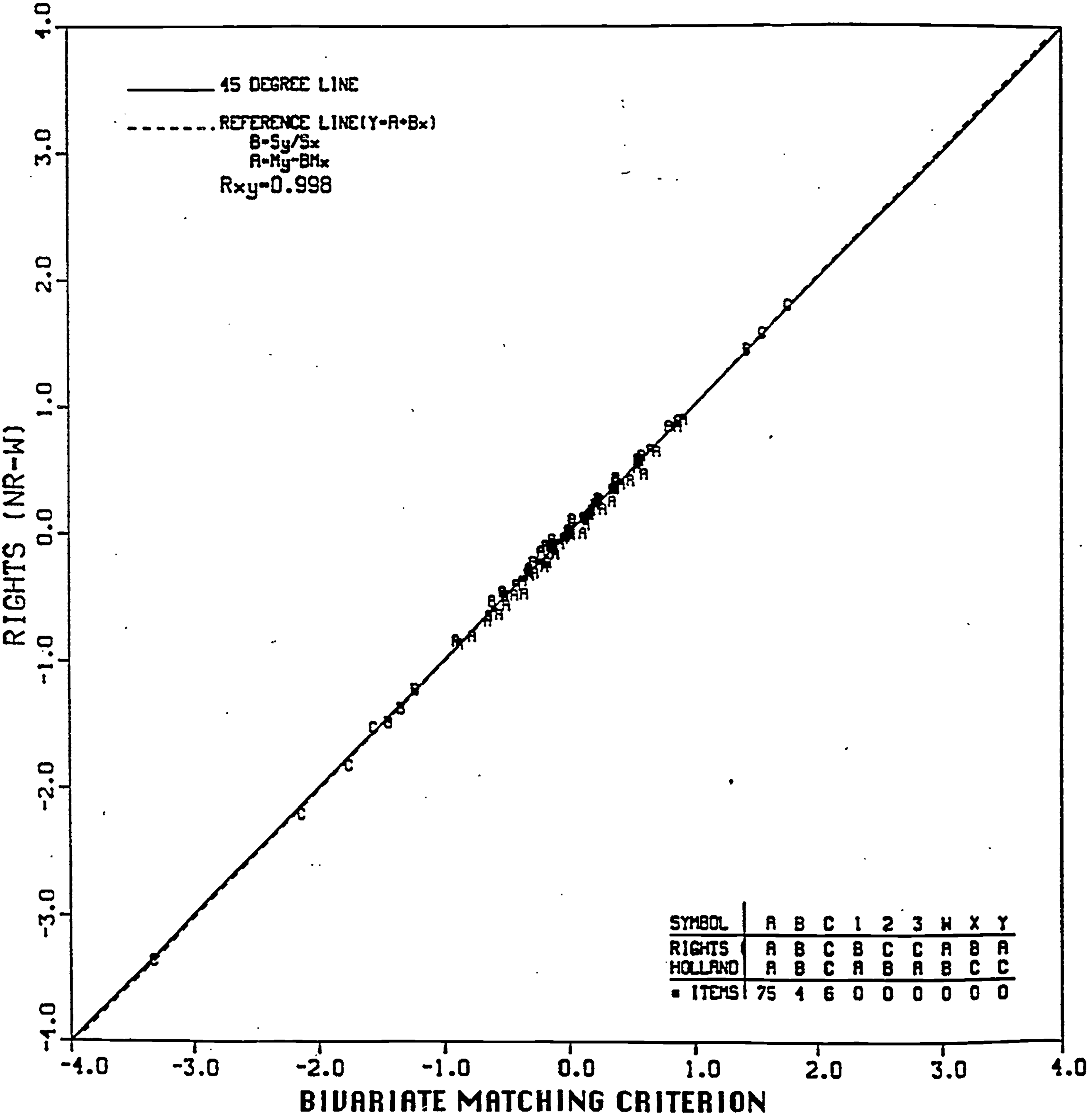
MH D-DIF SCATTERPLOT OF MALE/FEMALE ON LARGE GROUP FOR FS(NRW)-BV (03/88)



MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON LARGE GROUP FOR FS(NRNI)-BV (03/88)

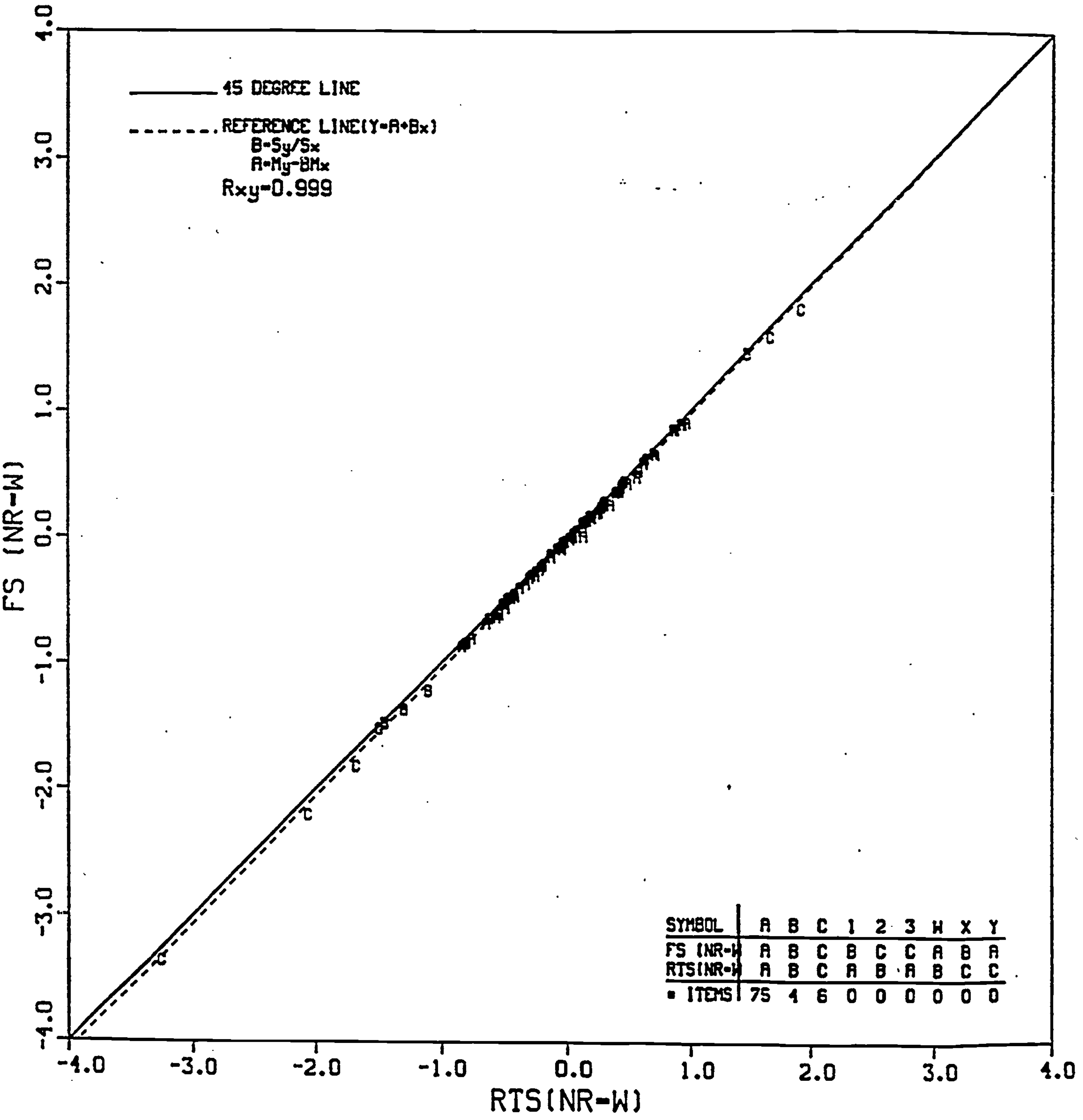


MH D-DIF SCATTERPLOT OF MALE/FEMALE ON LARGE GROUP FOR RTS-BV (03/88)



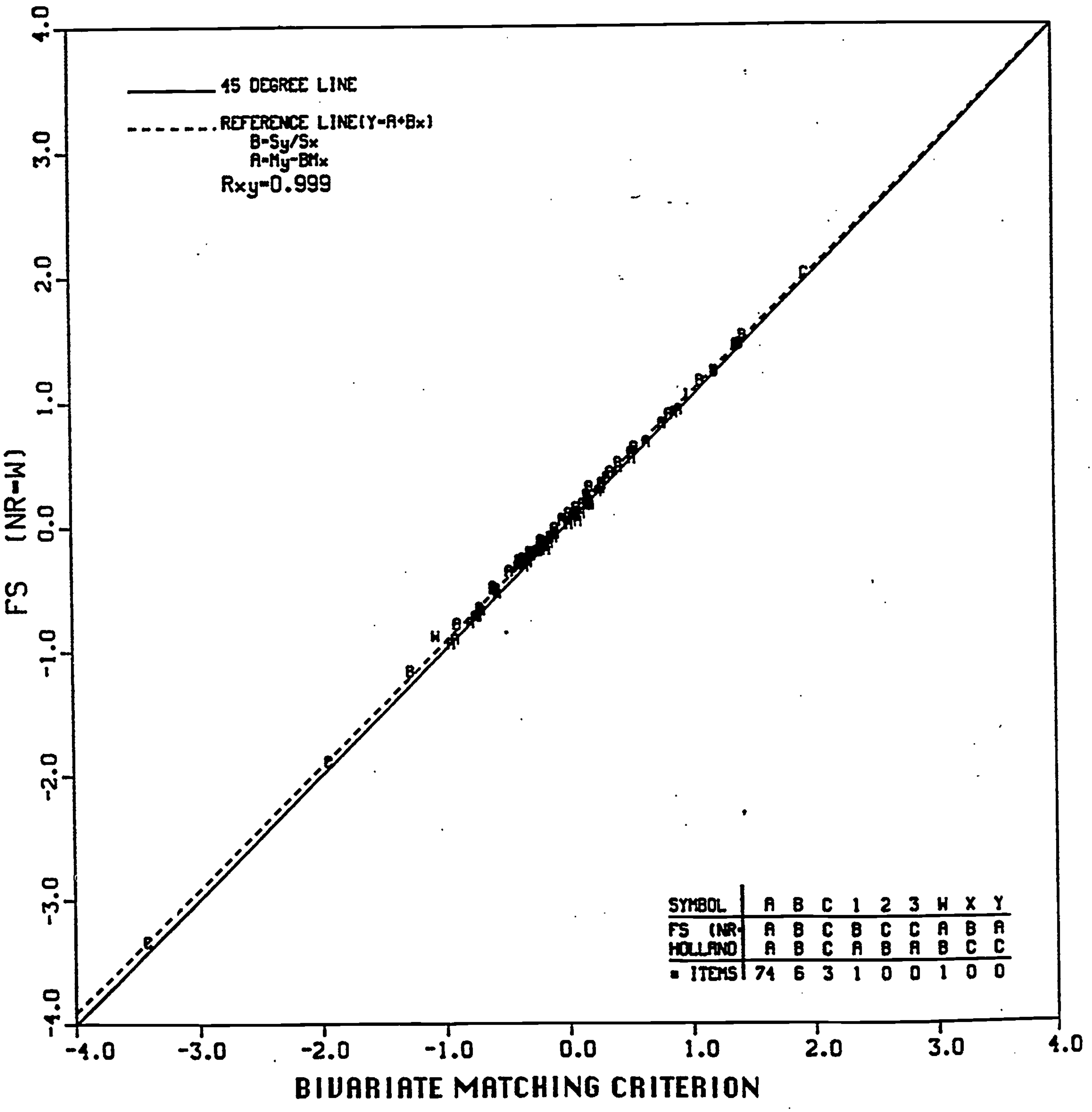
SYMBOL	A	B	C	1	2	3	W	X	Y
RIGHTS	A	B	C	B	C	C	A	B	A
HOLLAND	A	B	C	A	B	A	B	C	C
■ ITEMS	75	4	6	0	0	0	0	0	0

MH D-DIF SCATTERPLOT OF MALE/FEMALE ON LARGE GROUP FOR FS(NRW)-RTS (03/88)

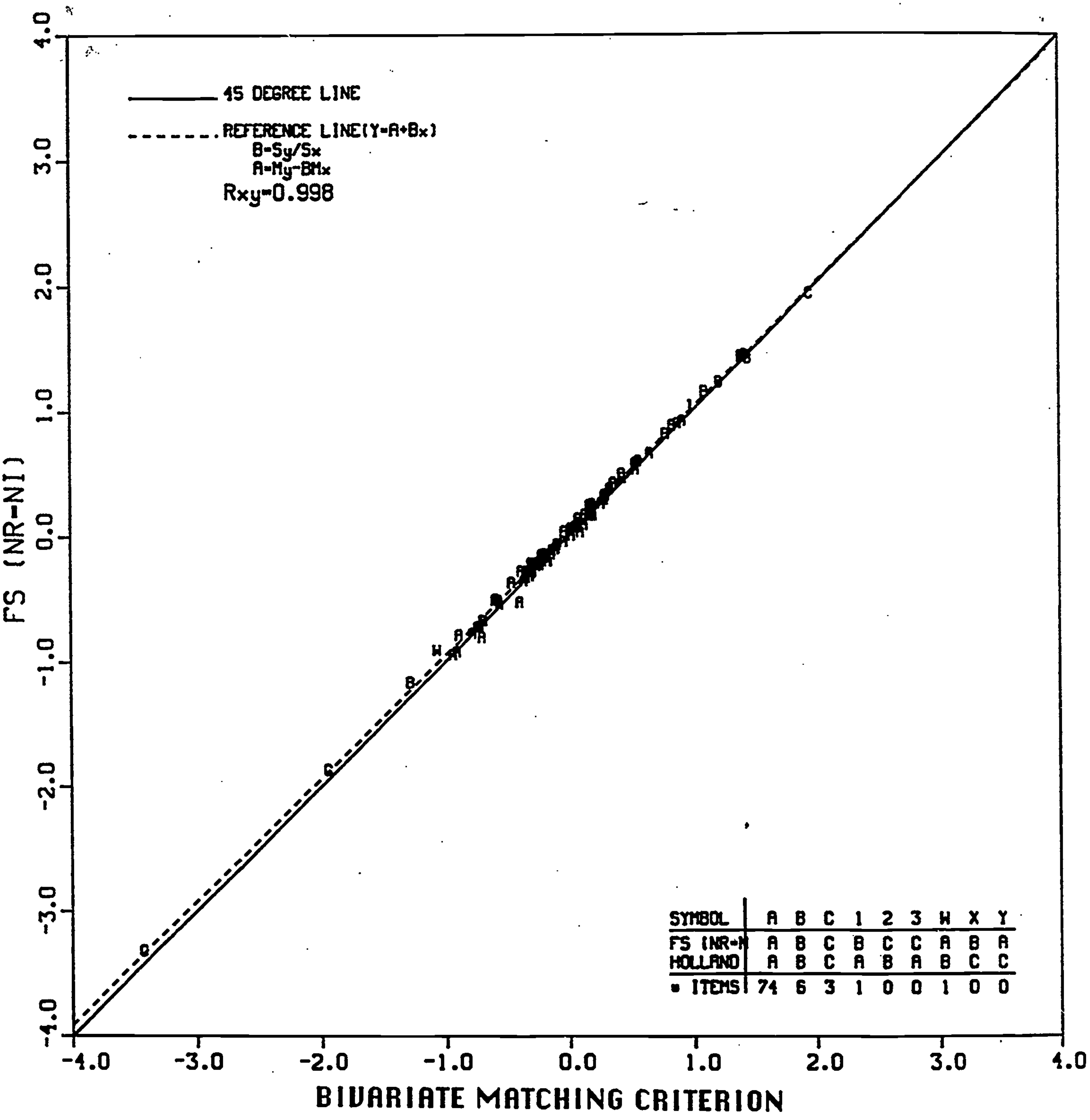


SYMBOL	A	B	C	1	2	3	H	X	Y
FS (NR-W	A	B	C	B	C	C	A	B	A
RTS(NR-W	A	B	C	A	B	A	B	C	C
■ ITEMS	75	4	6	0	0	0	0	0	0

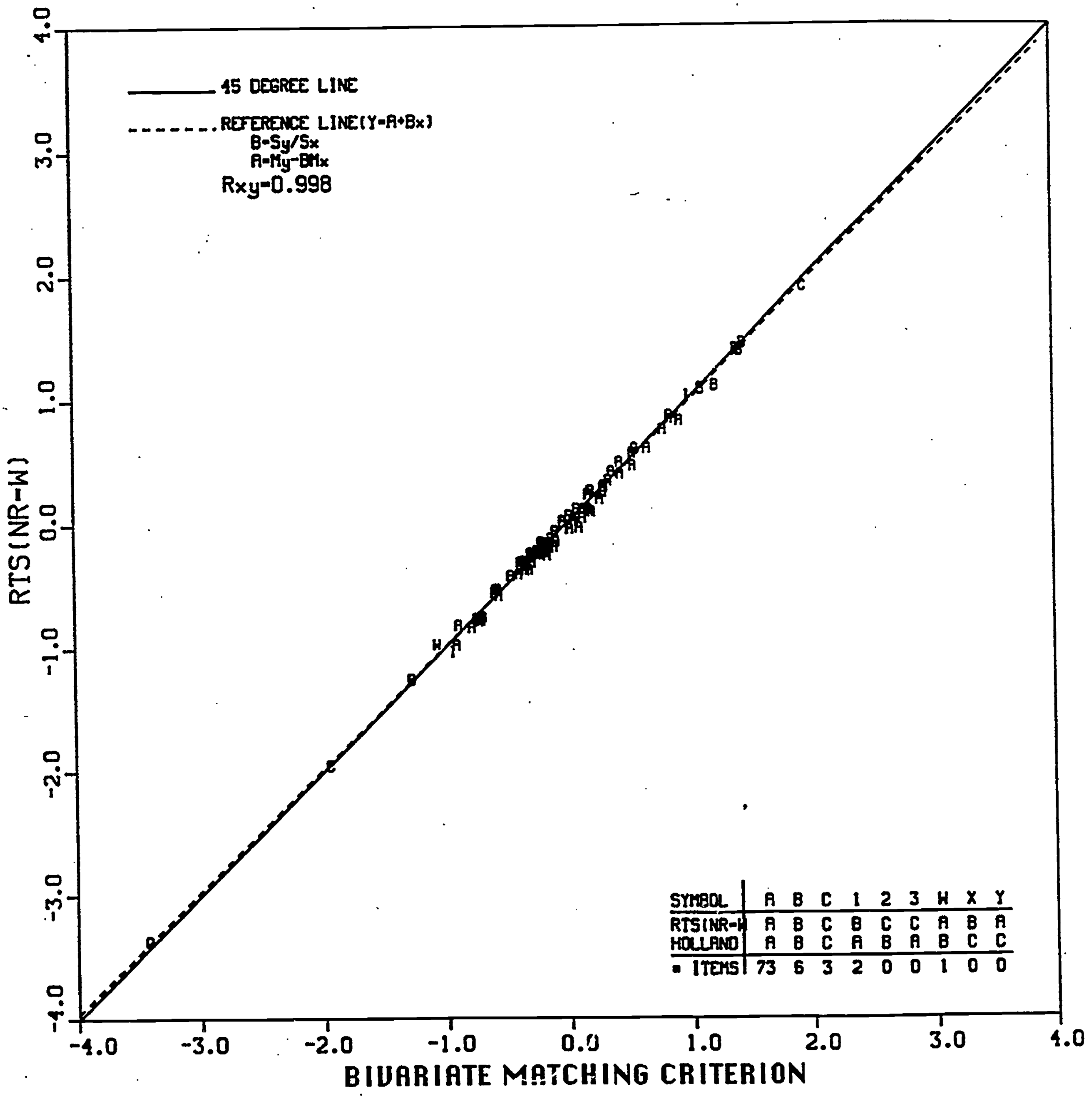
MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON LARGE GROUP FOR FS(NRW) - BV (05/88)



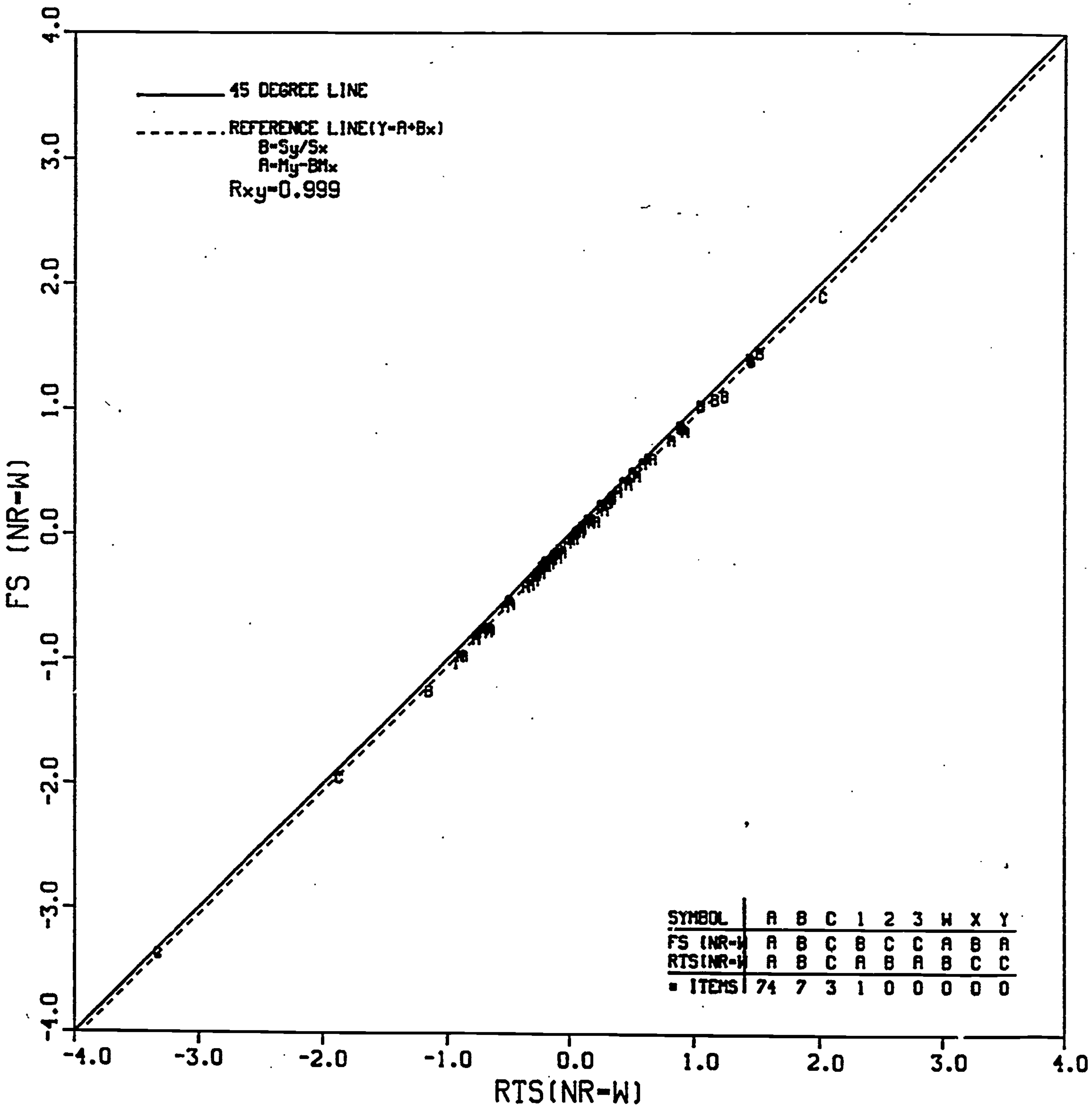
MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON LARGE GROUP FOR FS(NRNI) - BV (05/88)



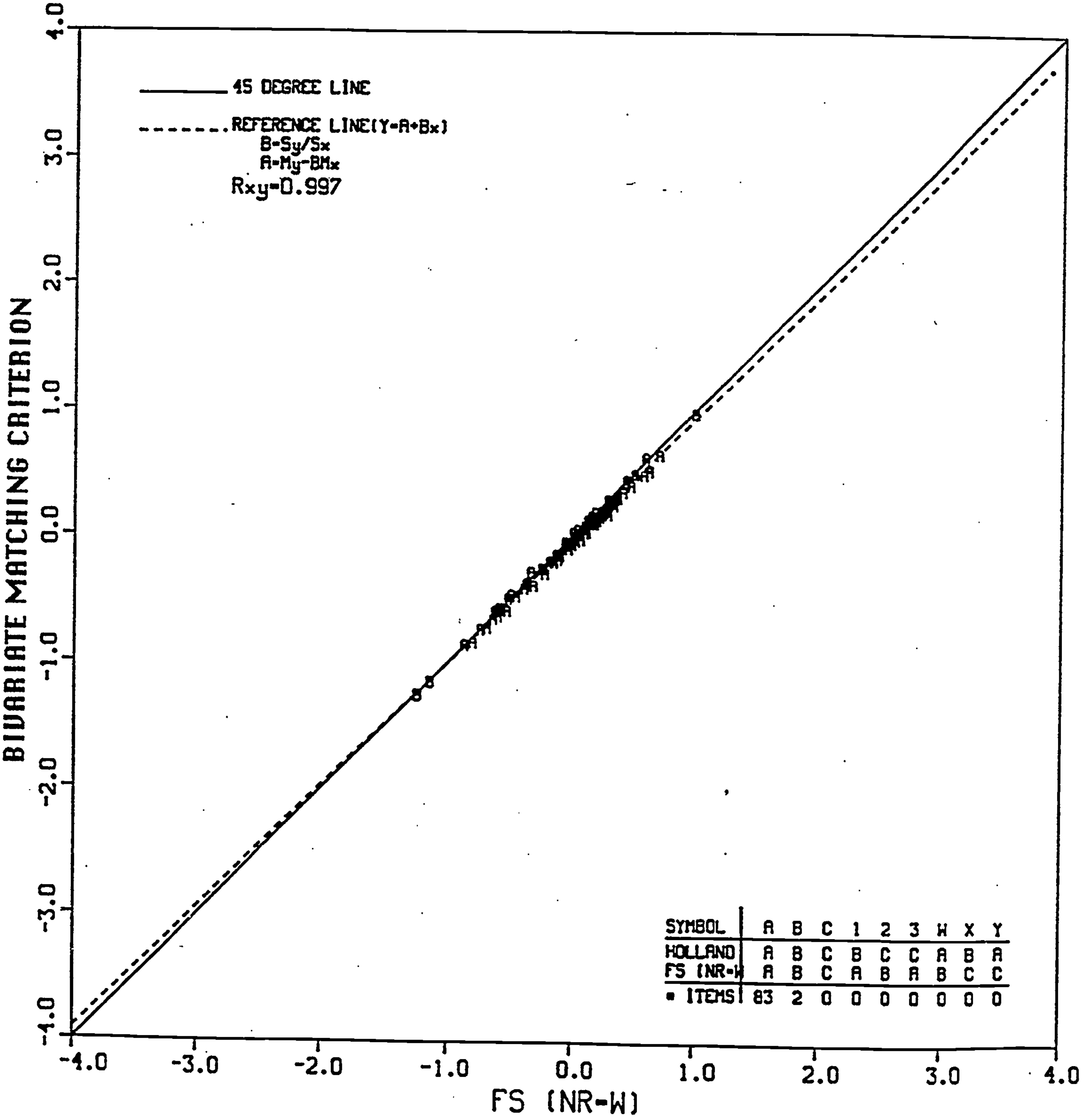
MH D-DIF SCATTERPLOT OF MALE/FEMALE ON LARGE GROUP FOR RTS - BU (05/88)



MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON LARGE GROUP FOR FS(NRW)-RTS (05/88)

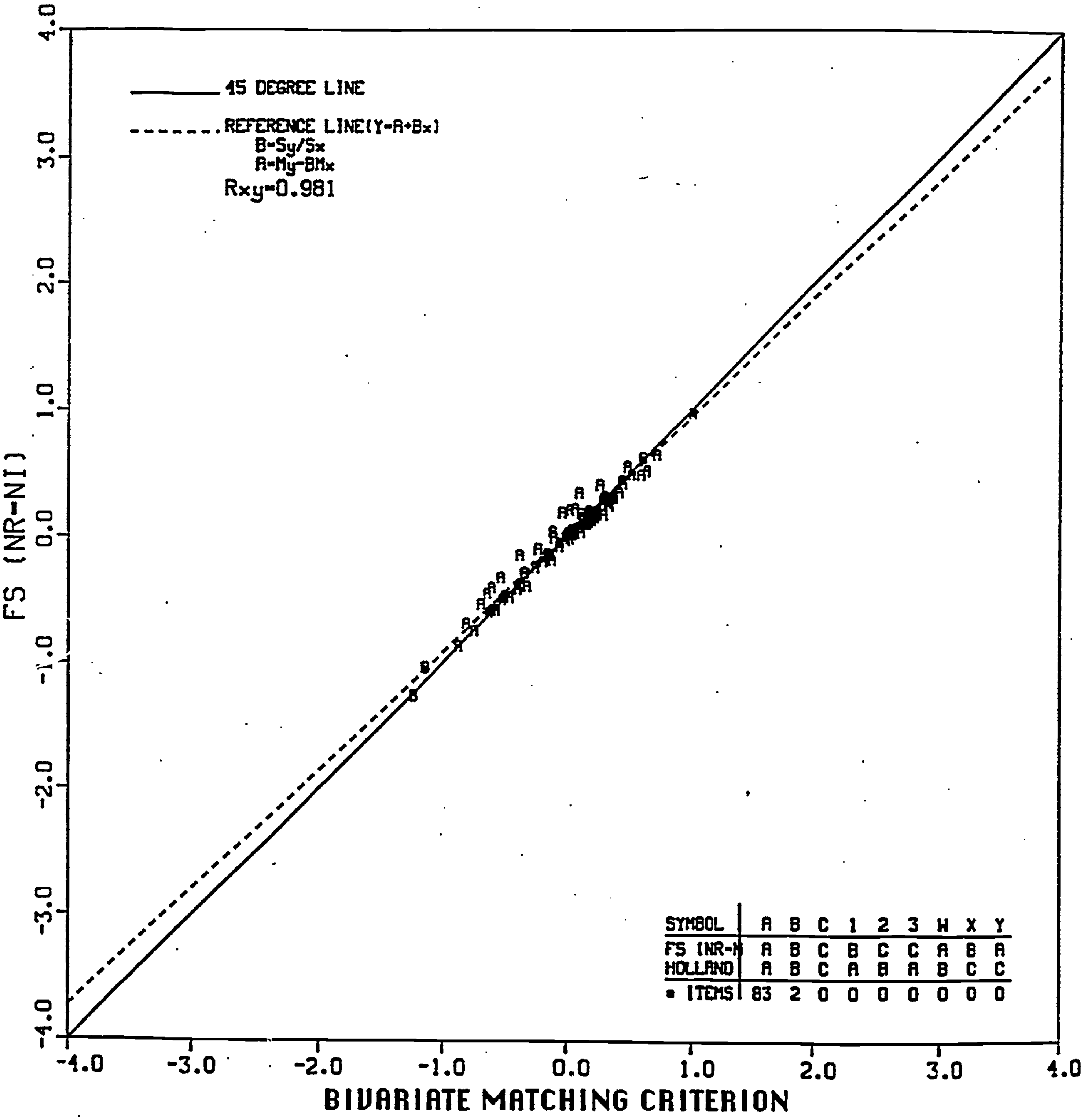


MH D-DIF SCATTERPLOT OF WHITE/BLACK ON LARGE GROUP FOR BV - FS(NRW) (03/88)



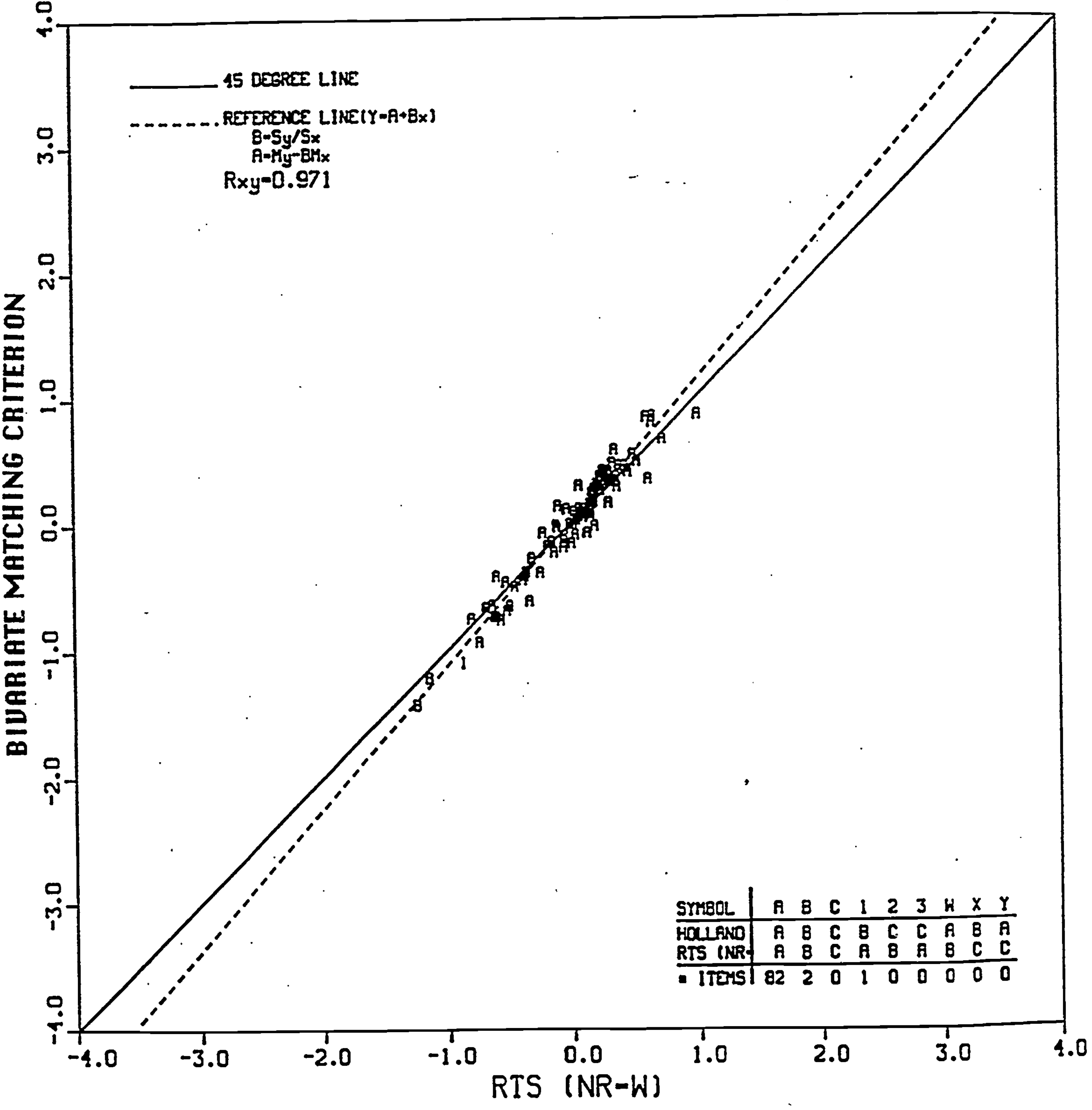
SYMBOL	A	B	C	1	2	3	W	X	Y
HOLLAND	A	B	C	B	C	C	A	B	A
FS (NR-W	A	B	C	A	B	A	B	C	C
• ITEMS	83	2	0	0	0	0	0	0	0

MH D-DIF SCATTERPLOT OF WHITE/BLACK ON LARGE GROUP FOR FS(NRNI) BV (03/88)

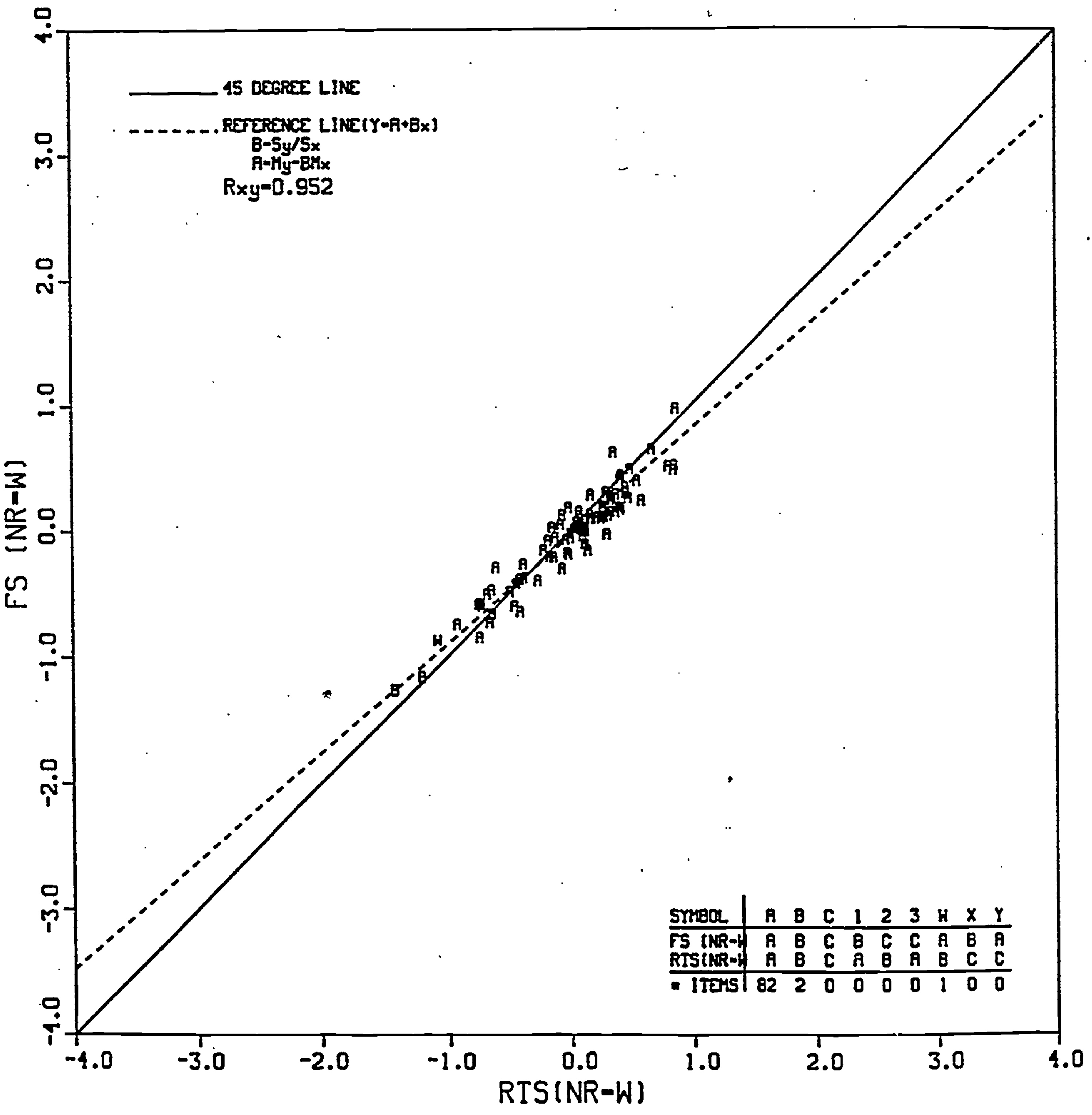


SYMBOL	A	B	C	1	2	3	H	X	Y
FS (NR-N	A	B	C	B	C	C	A	B	A
HOLLAND	A	B	C	A	B	A	B	C	C
• ITEMS	83	2	0	0	0	0	0	0	0

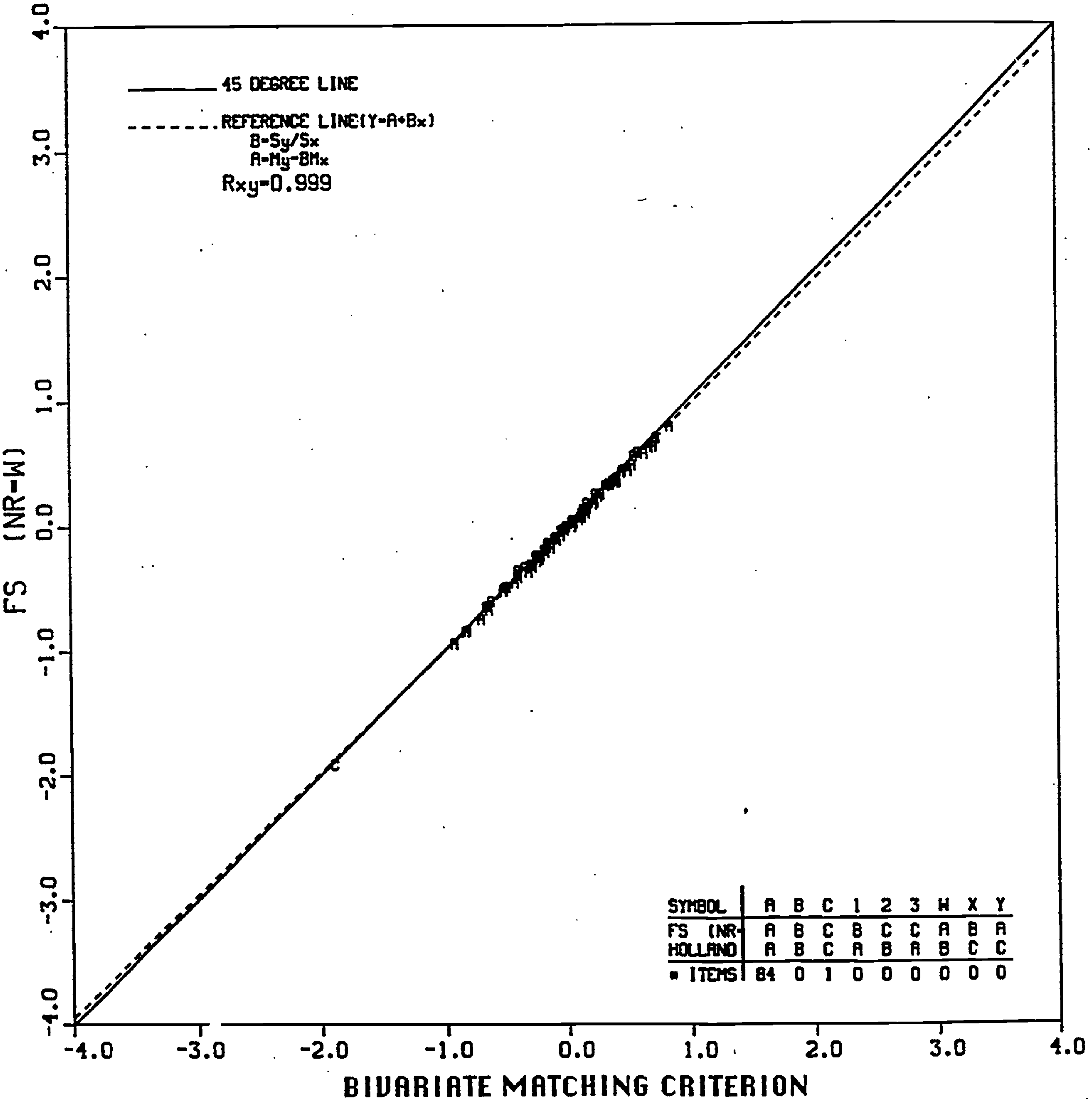
MH D-DIF SCATTERPLOT OF WHITE/BLACK
ON LARGE GROUP FOR BU - RTS (03/88)



MH D-DIF SCATTERPLOT OF WHITE/BLACK
ON LARGE GROUP FOR FS(NRW)-RTS (03/88)

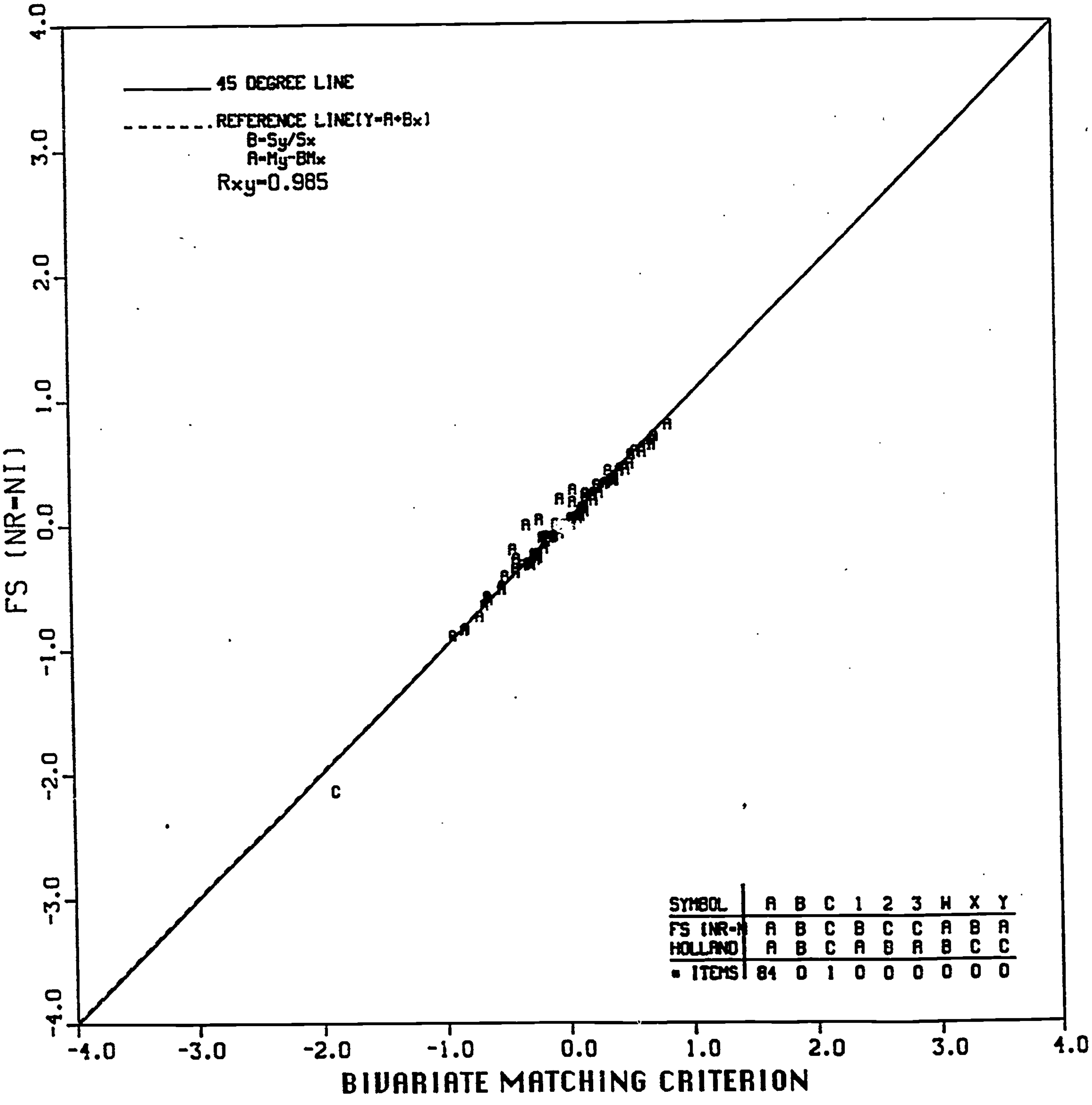


MH D-DIF SCATTERPLOT OF WHITE/BLACK
ON LARGE GROUP FOR FS(NRW) - BW (05/88)



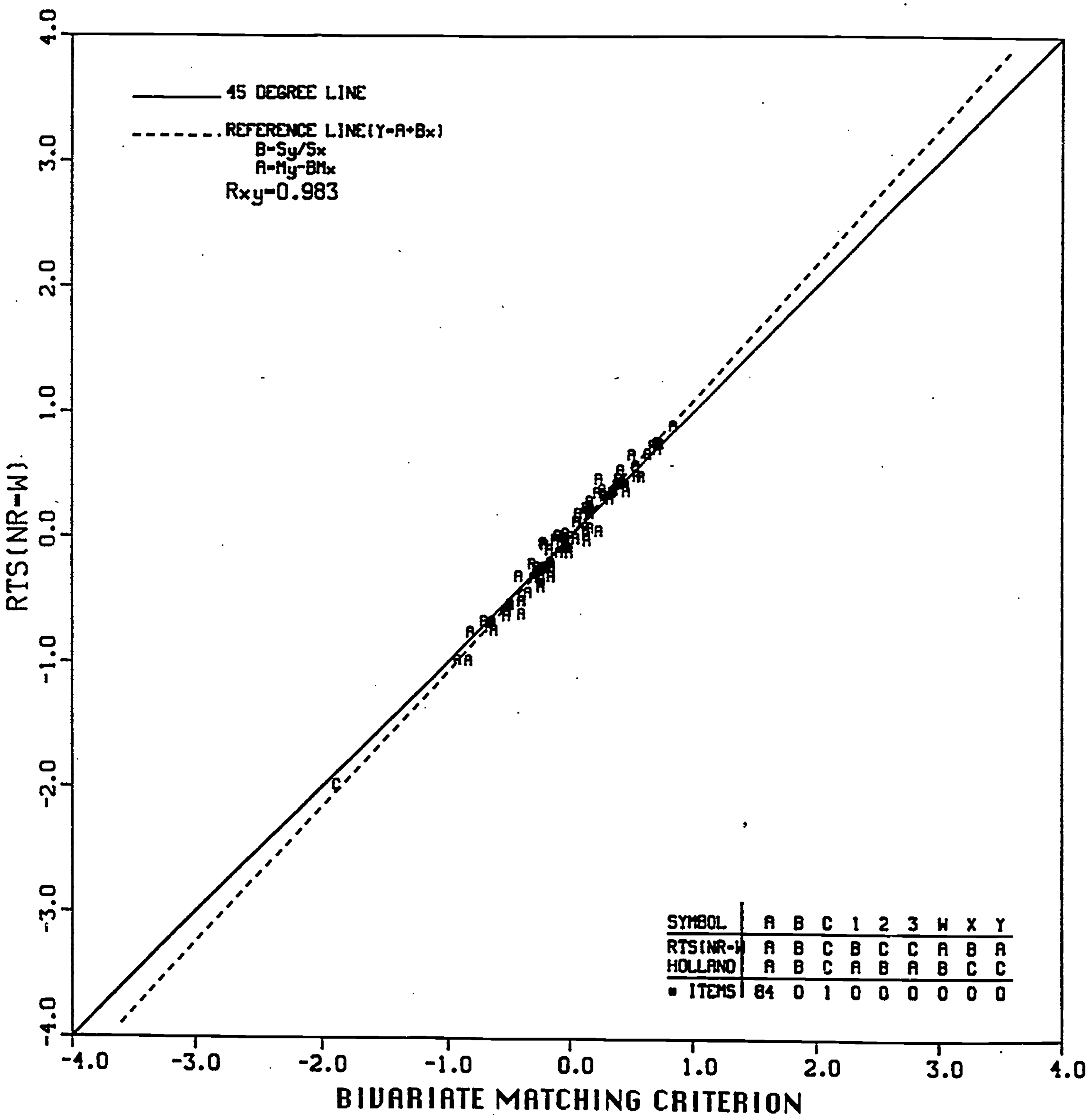
SYMBOL	A	B	C	1	2	3	W	X	Y
FS (NR-	A	B	C	B	C	C	A	B	A
HOLLAND	A	B	C	A	B	A	B	C	C
* ITEMS	84	0	1	0	0	0	0	0	0

MH D-DIF SCATTERPLOT OF WHITE/BLACK
ON LARGE GROUP FOR FS(NRNI) - BU (05/88)

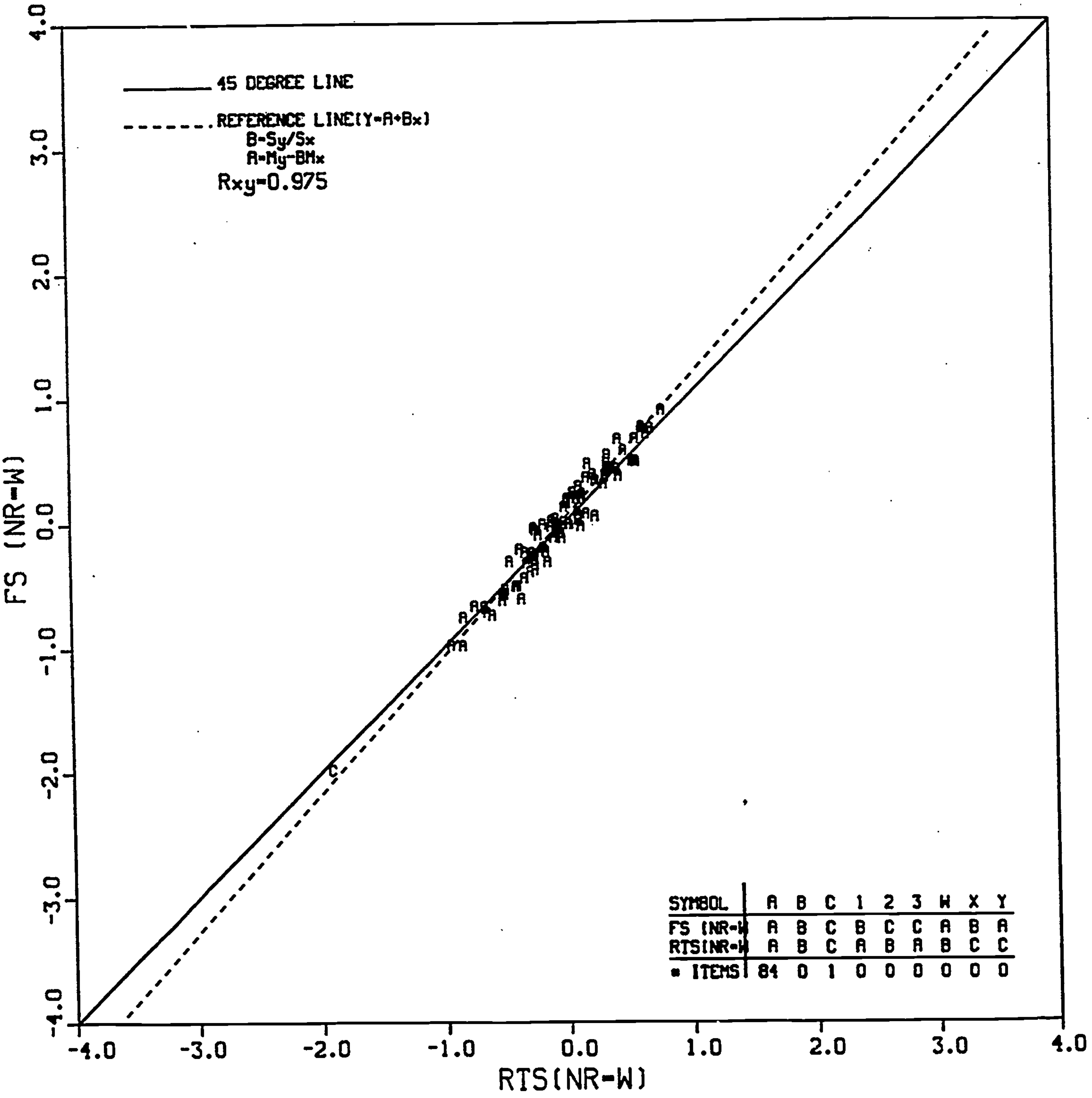


SYMBOL	A	B	C	1	2	3	H	X	Y
FS (NR-N	A	B	C	B	C	C	A	B	A
HOLLAND	A	B	C	A	B	A	B	C	C
# ITEMS	84	0	1	0	0	0	0	0	0

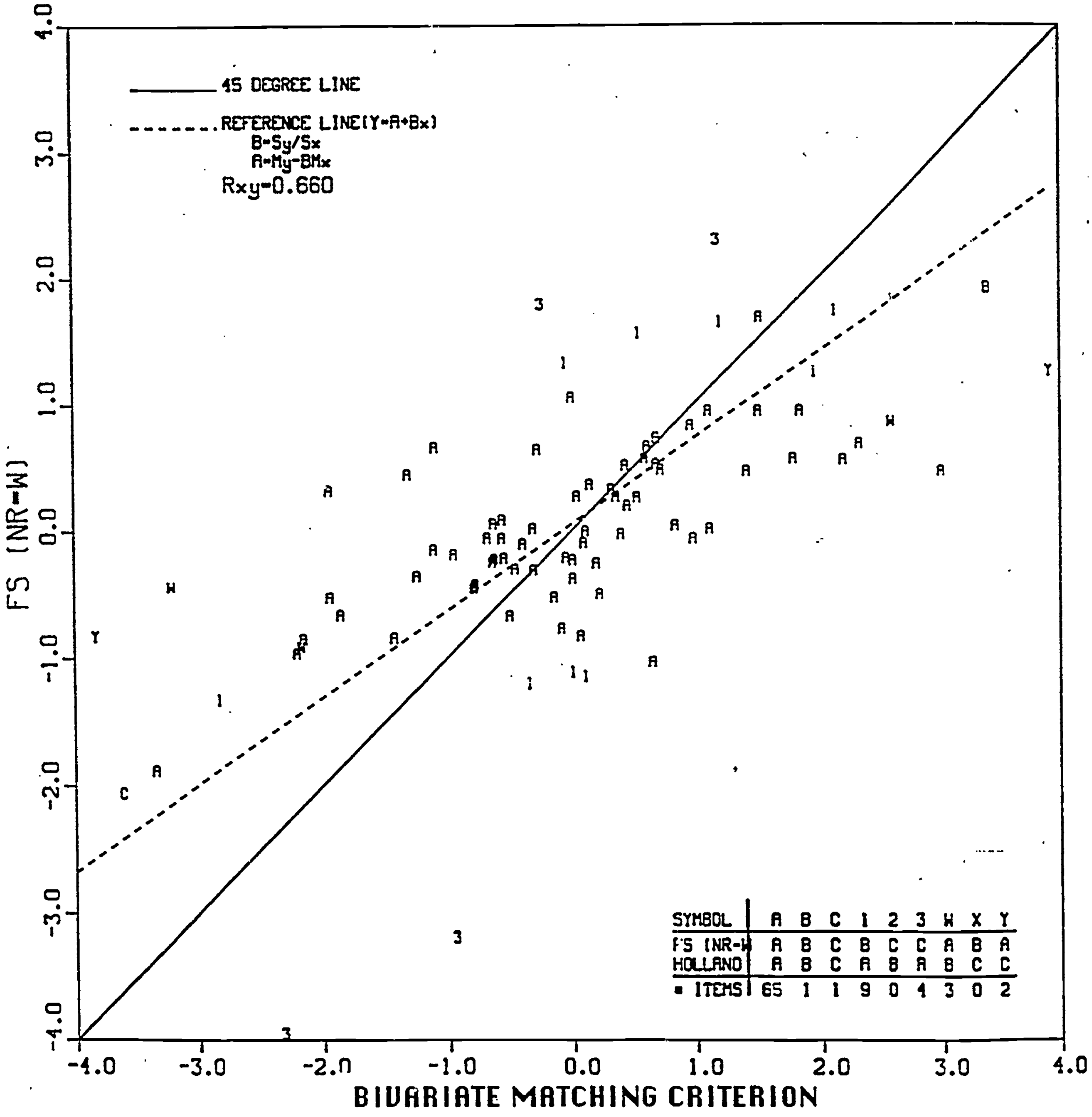
MH D-DIF SCATTERPLOT OF WHITE/BLACK
ON LARGE GROUP FOR RTS - BV (05/88)



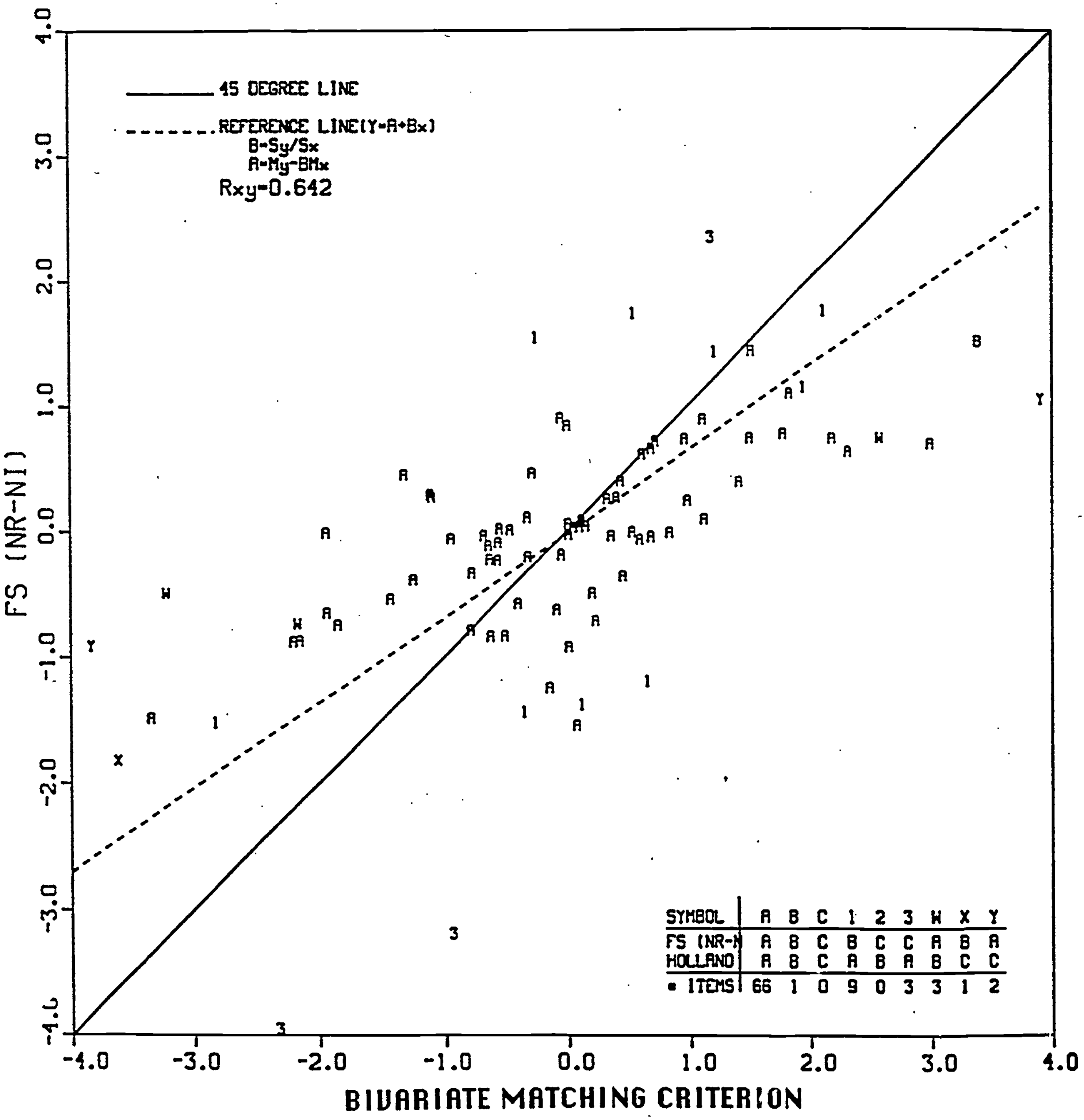
MH D-DIF SCATTERPLOT OF WHITE/BLACK
ON LARGE GROUP FOR FS(NRW)-RTS (05/88)



MH D-DIF SCATTERPLOT OF MALE/FEMALE ON SMLGRP FOR FS(NRW)-BV (03/88)

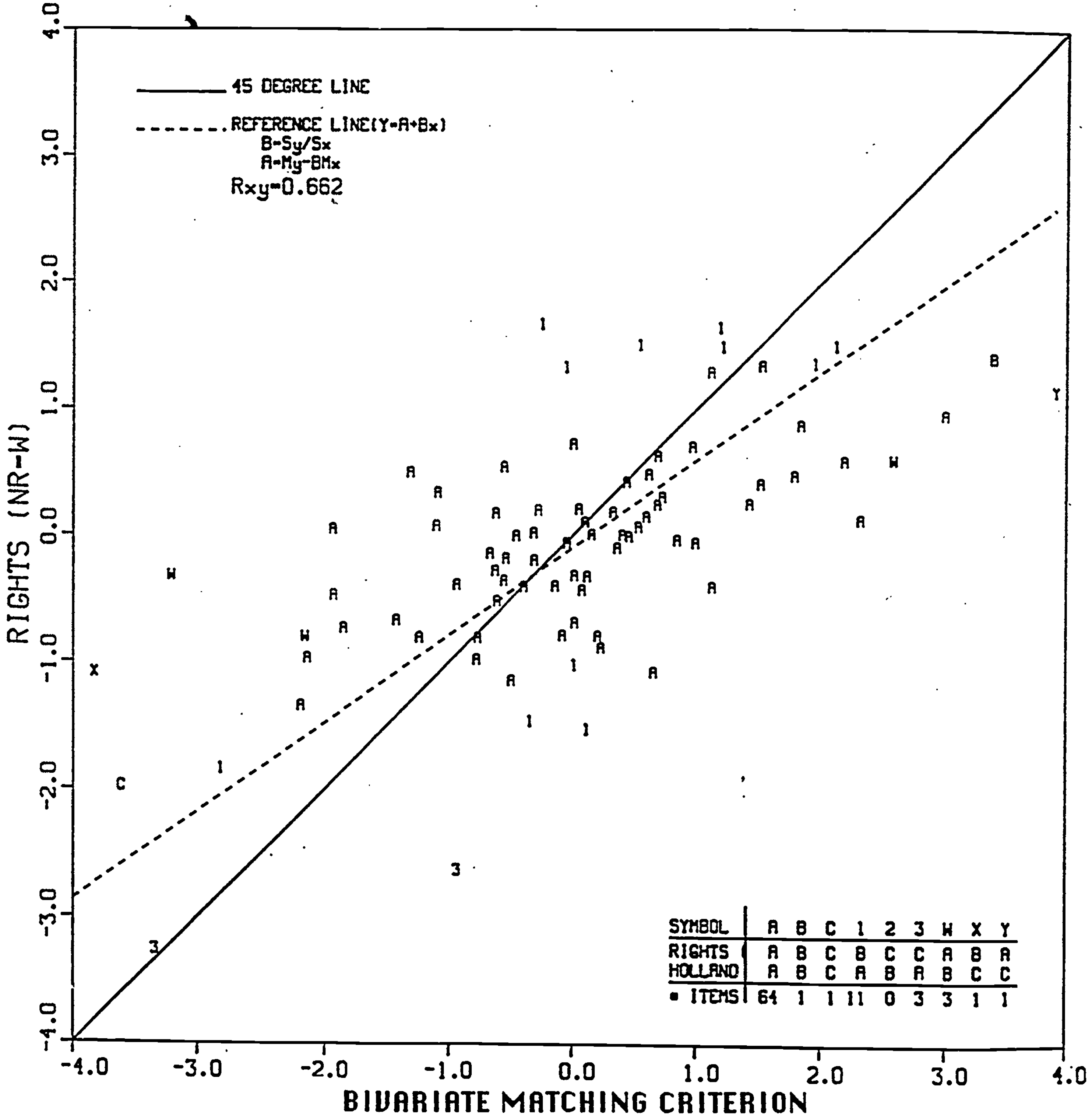


MH D-DIF SCATTERPLOT OF MALE/FEMALE ON SMLGRP FOR FS(NRNI)-BV (03/88)



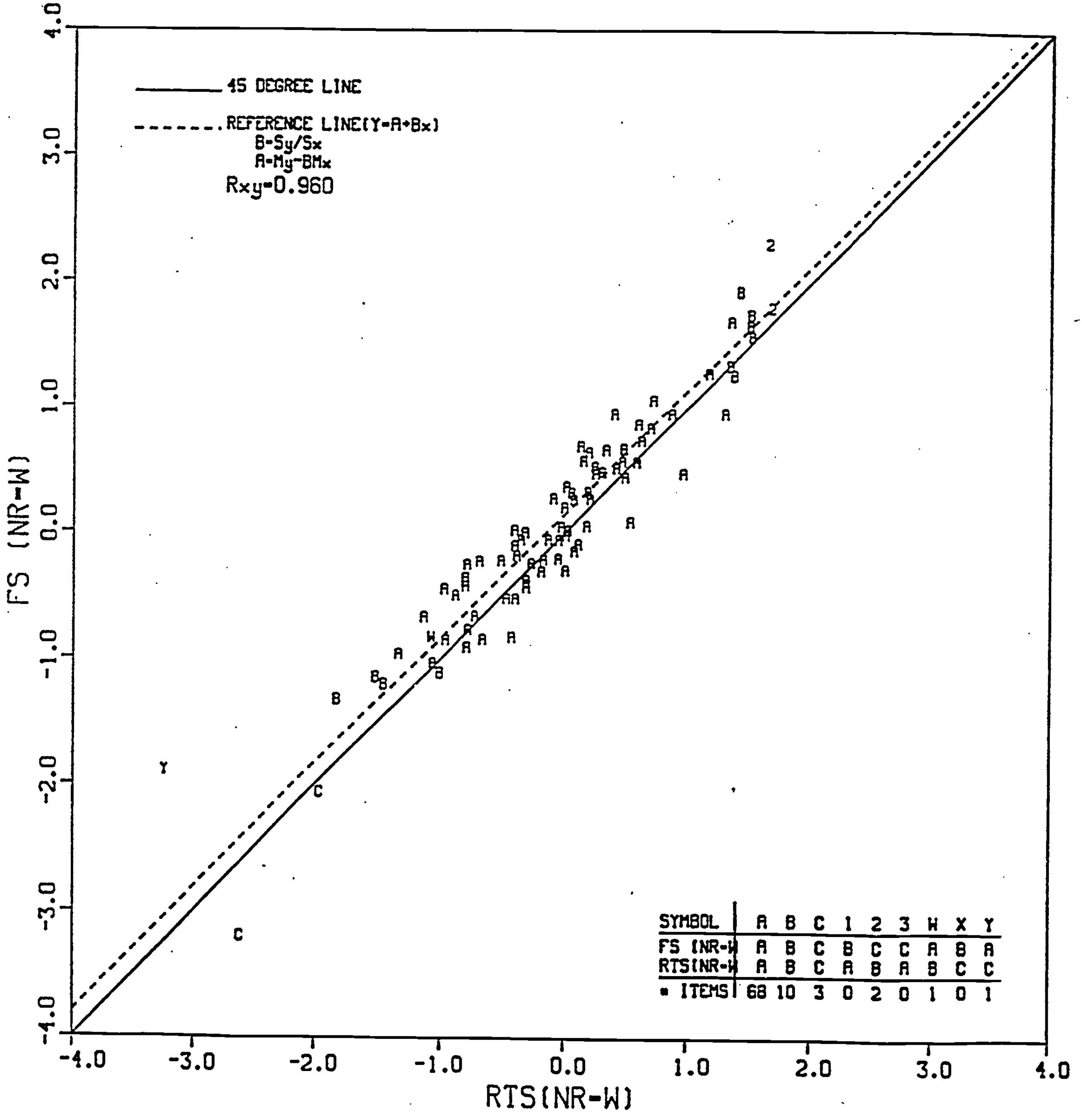
SYMBOL	A	B	C	1	2	3	W	X	Y
FS (NR-N	A	B	C	B	C	C	A	B	A
HOLLAND	A	B	C	A	B	A	B	C	C
• ITEMS	66	1	0	9	0	3	3	1	2

MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON SMLGRP FOR RTS-BV (03/88)



1 MH Values out of range

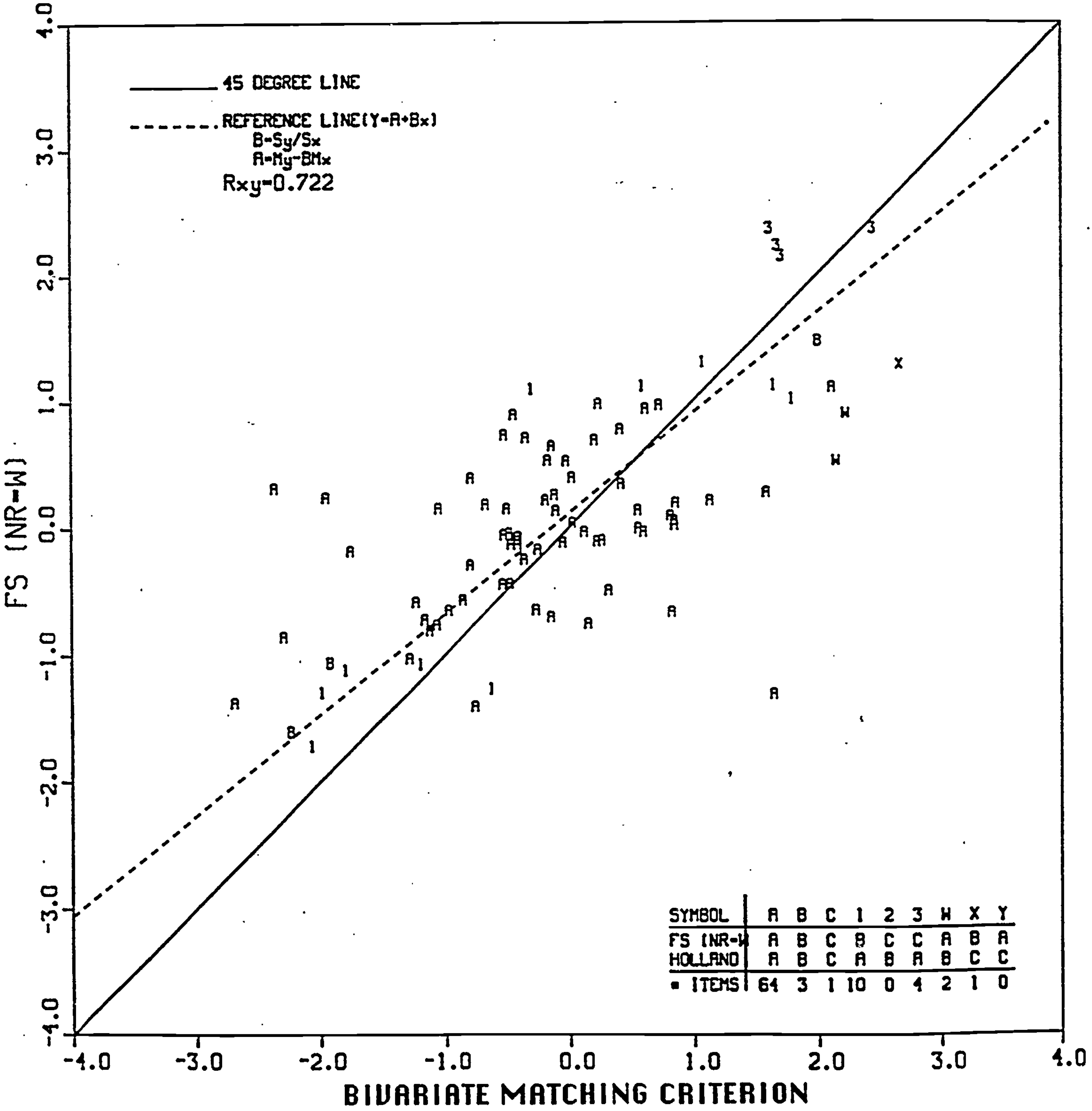
MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON SMLGRP FOR FS(NRW)-RTS (03/88)



SYMBOL	A	B	C	1	2	3	H	X	Y
FS (NR-W	A	B	C	B	C	C	A	B	A
RTS(NR-W	A	B	C	A	B	A	B	C	C
# ITEMS	68	10	3	0	2	0	1	0	1

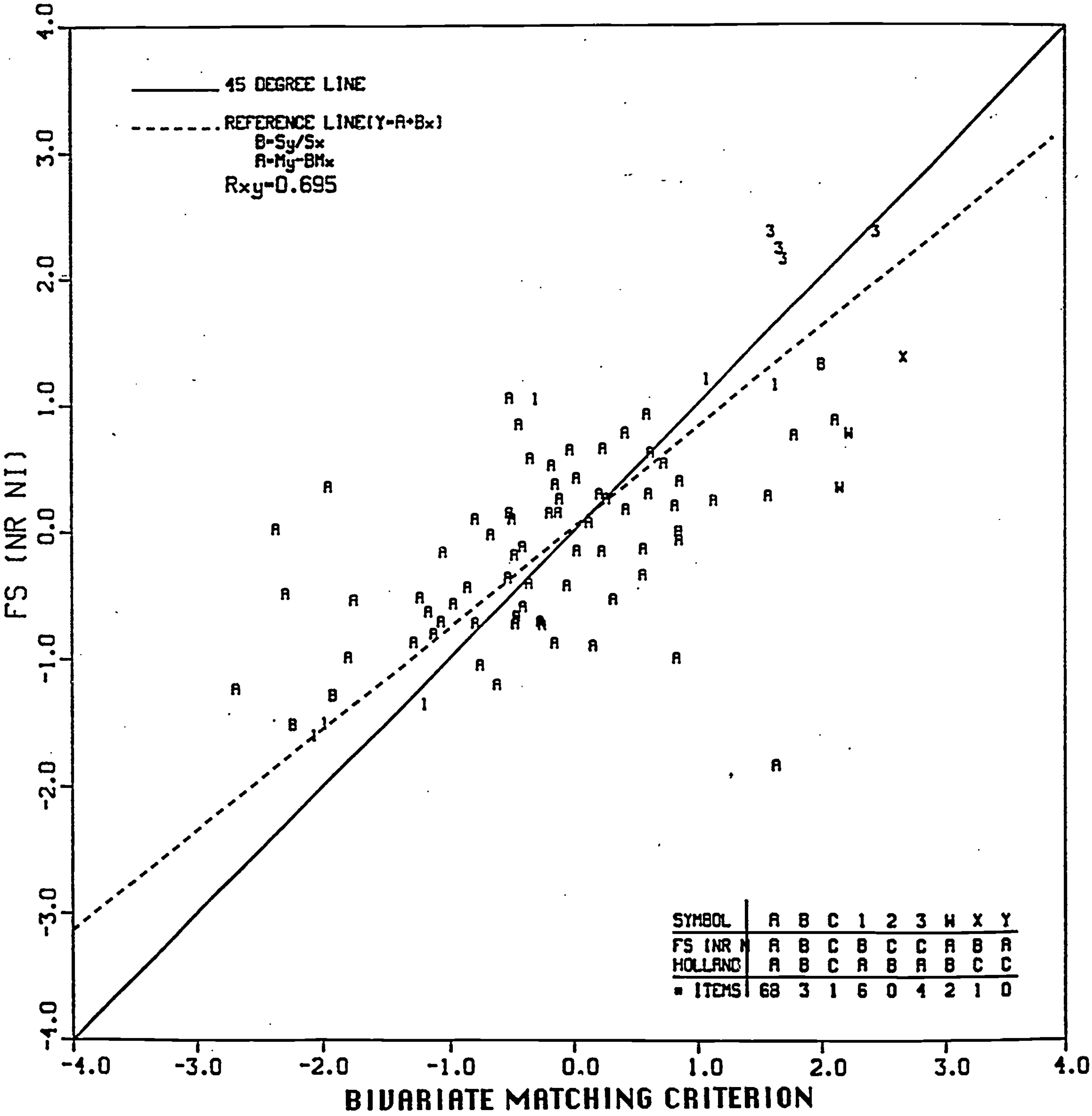
1 MH Values out of range

MH D-DIF SCATTERPLOT OF MALE/FEMALE ON SMLGRP FOR FS(NRW)-BV (05/88)



1 MH Values out of range

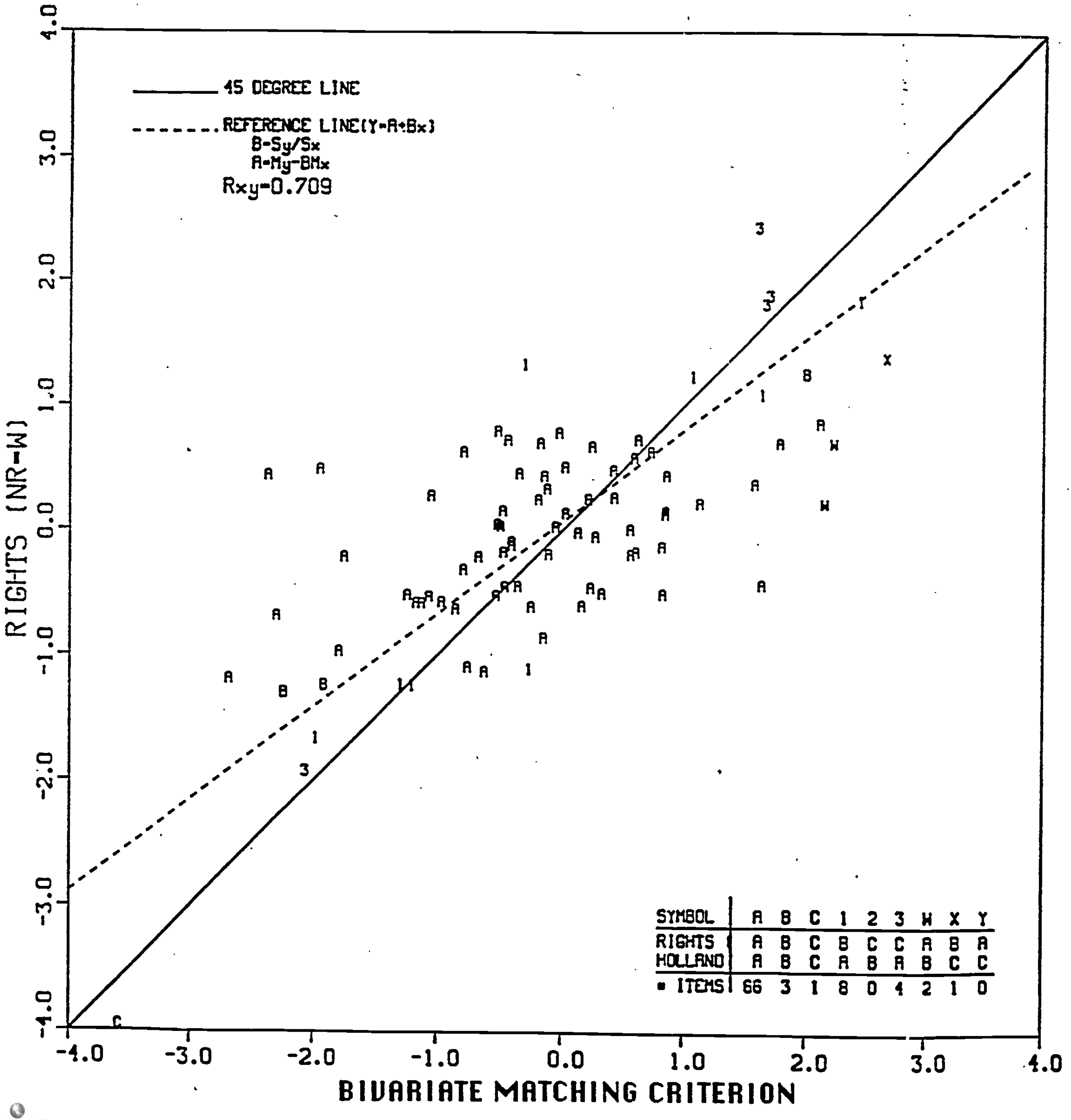
MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON SMLGRP FOR FS(NRNI)-BV (05/88)



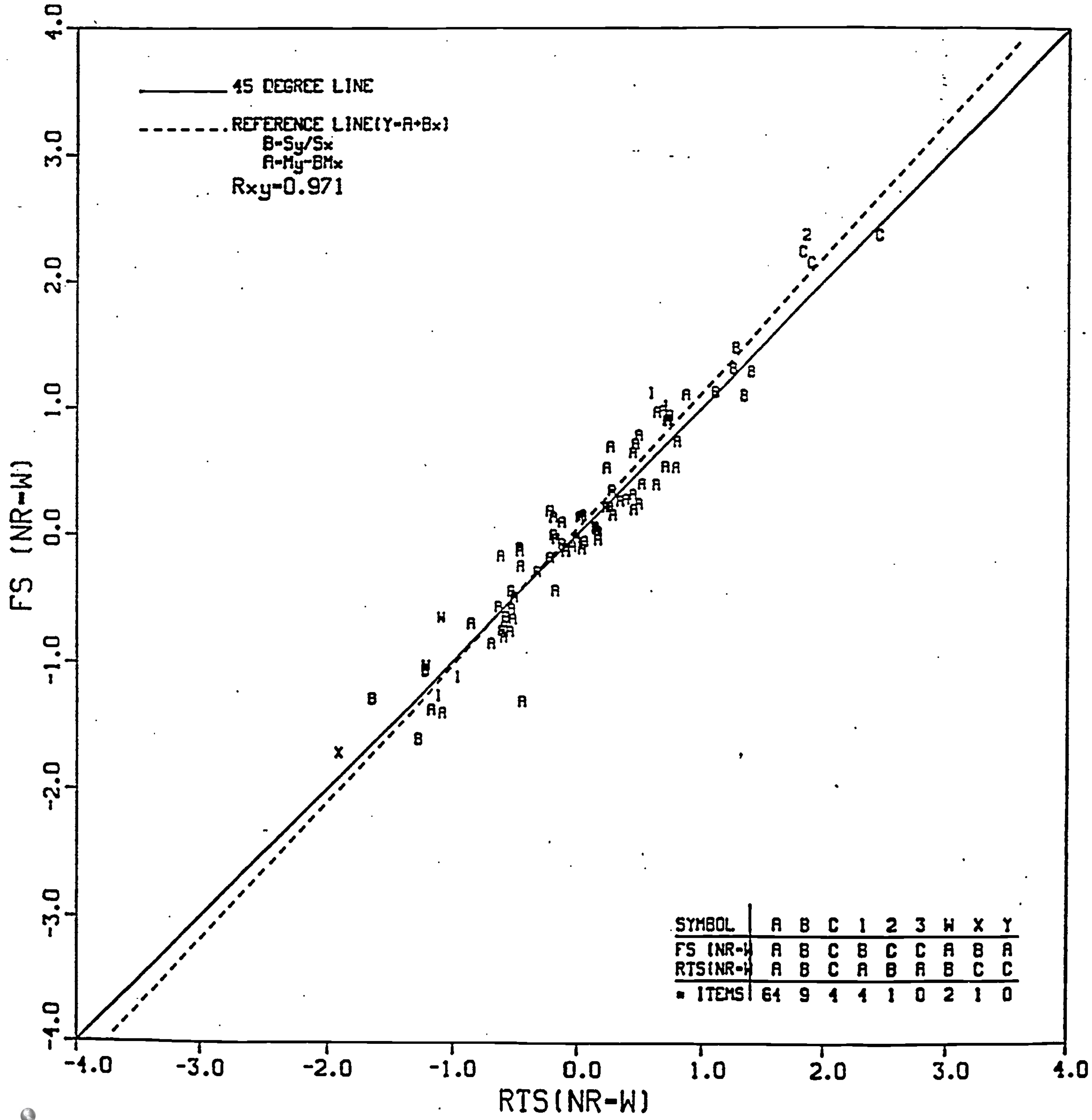
SYMBOL	A	B	C	1	2	3	W	X	Y
FS (NR N	A	B	C	B	C	C	A	B	A
HOLLAND	A	B	C	A	B	A	B	C	C
# ITEMS	68	3	1	6	0	4	2	1	0

1 MH Values out of range

MH D-DIF SCATTERPLOT OF MALE/FEMALE ON SMLGRP FOR RTS-BV (05/88)

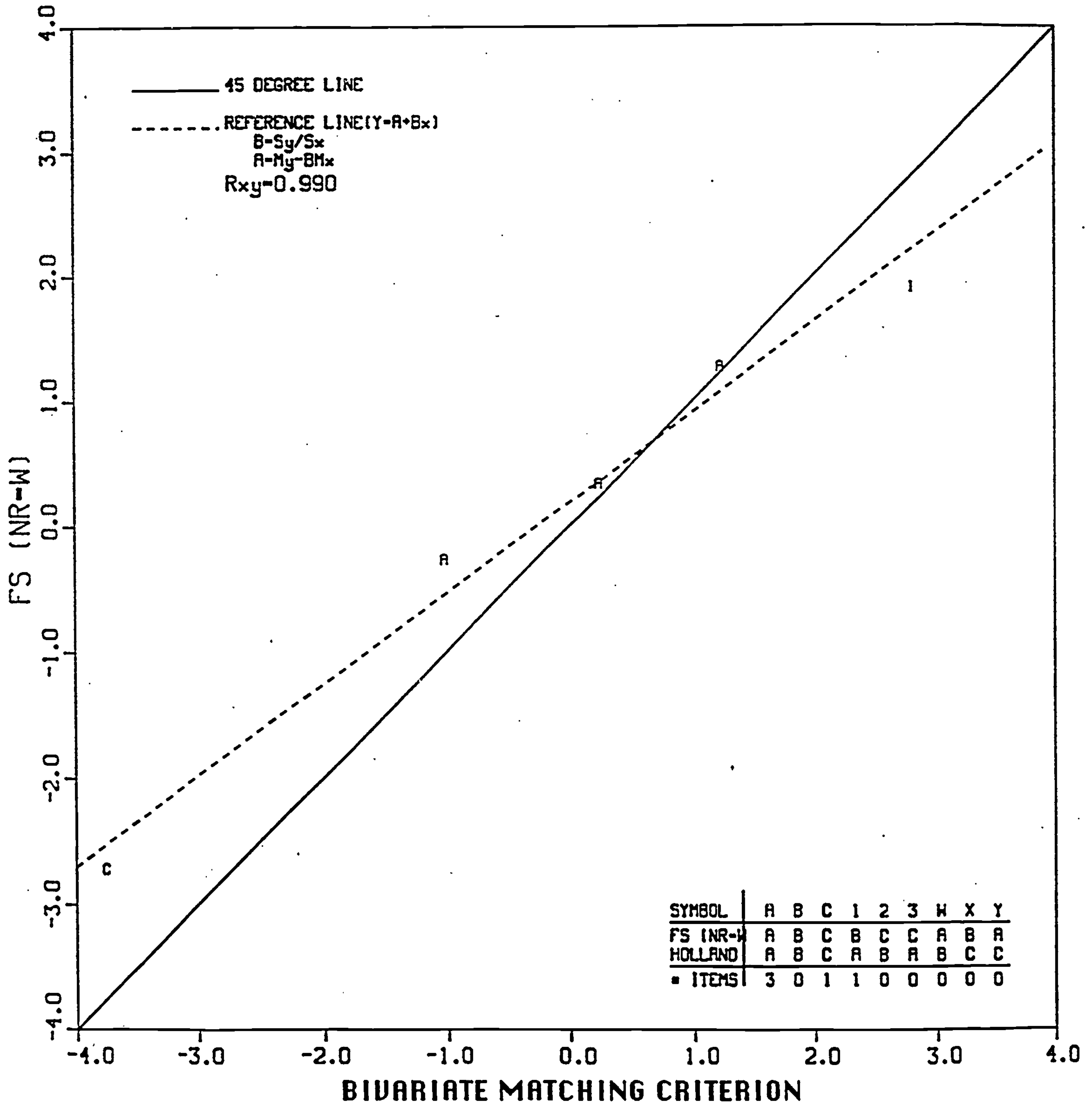


MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON SMLGRP FOR FS(NRW)-RTS **(05/88)**

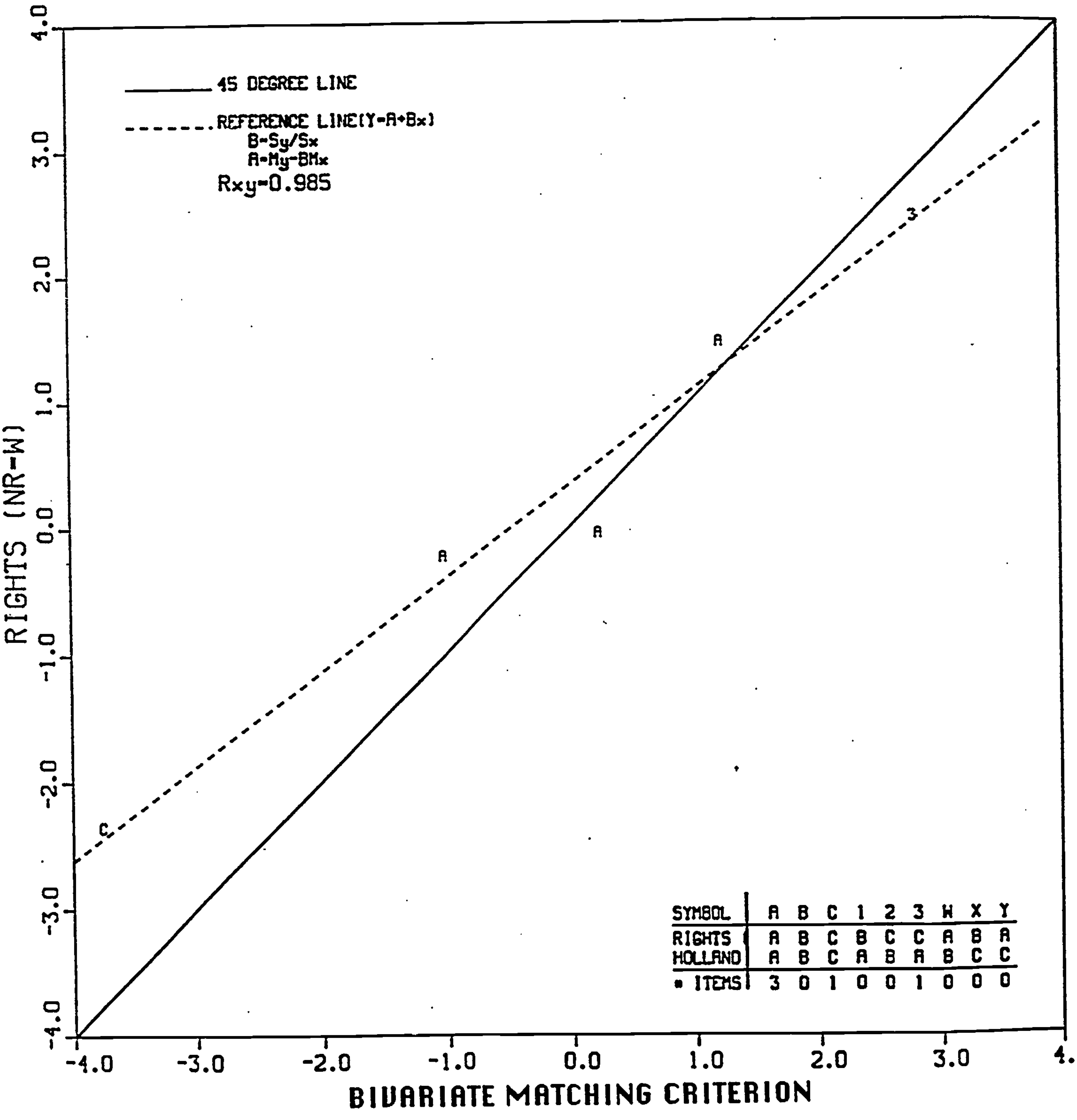


1 MH Values out of range

MH D-DIF SCATTERPLOT OF MALE/FEMALE ON PRETEST FOR FS(NR-W)-BV (05/88)

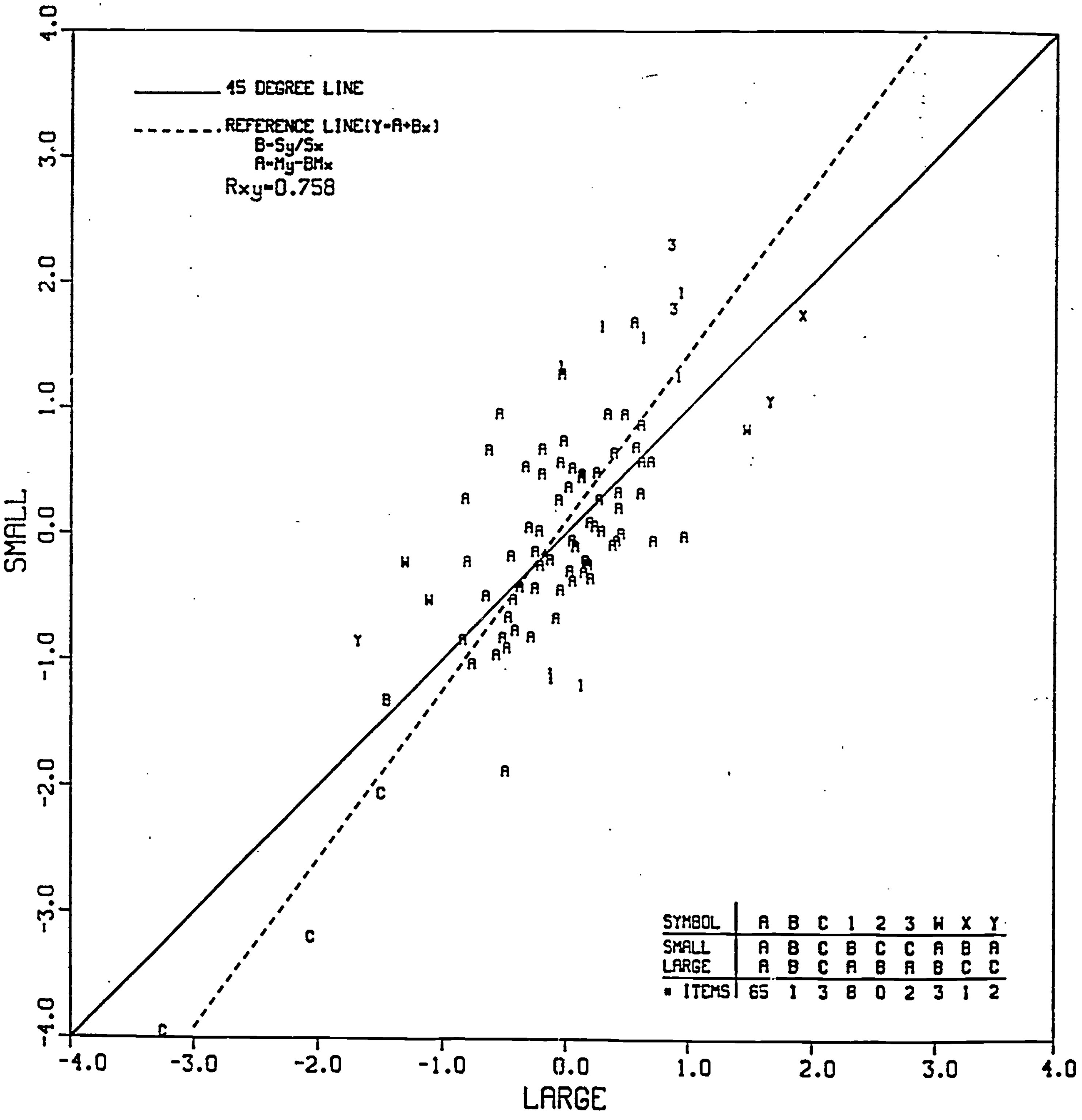


MH D-DIF SCATTERPLOT OF MALE/FEMALE ON PRETEST FOR RTS-BV (05/88)

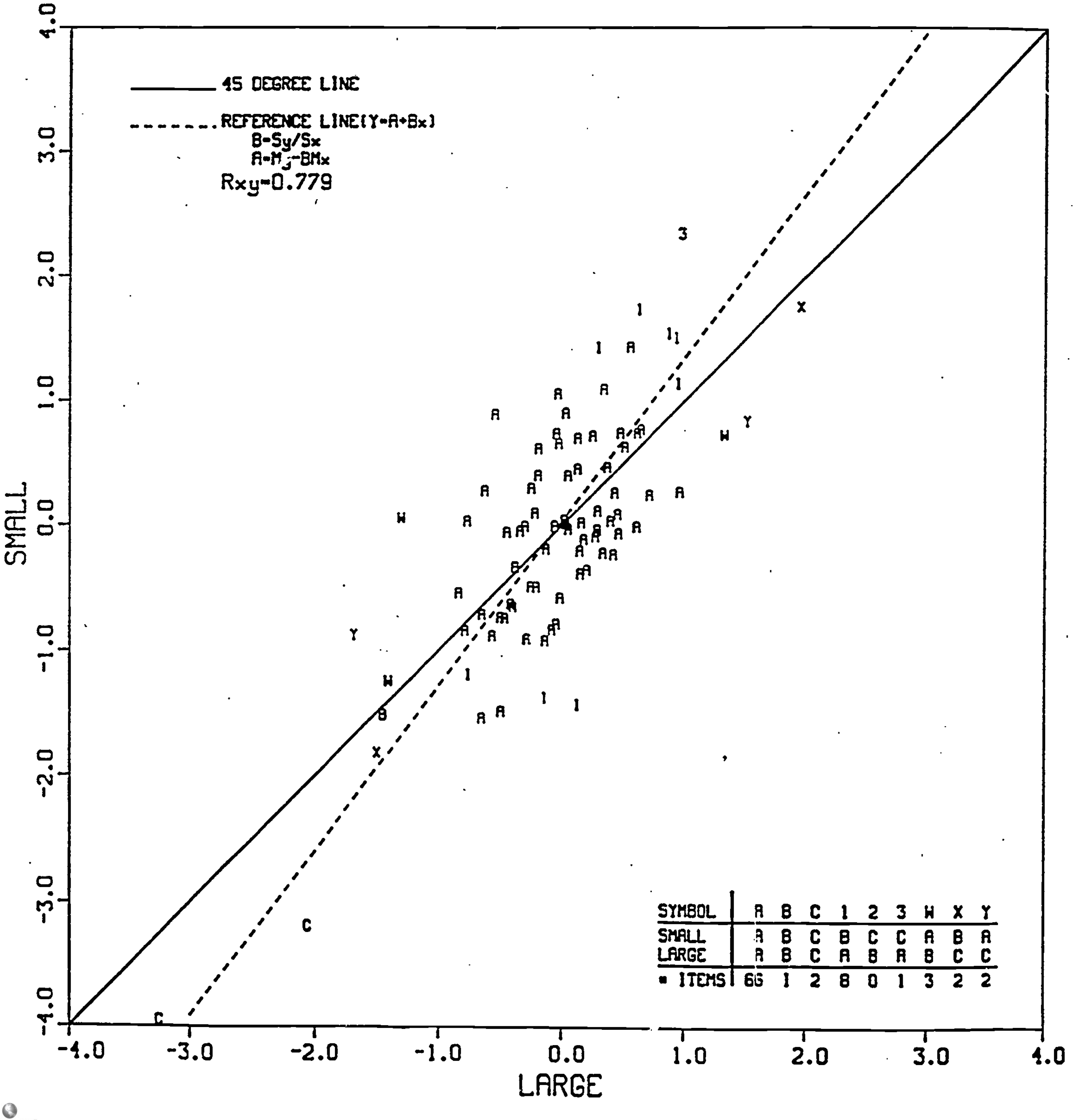


SYMBOL	A	B	C	1	2	3	W	X	Y
RIGHTS	A	B	C	B	C	C	A	B	A
HOLLAND	A	B	C	A	B	A	B	C	C
# ITEMS	3	0	1	0	0	1	0	0	0

MH D-DIF SCATTERPLOT OF MALE/FEMALE ON FS(NR-W) FOR SML-LRG (03/88)

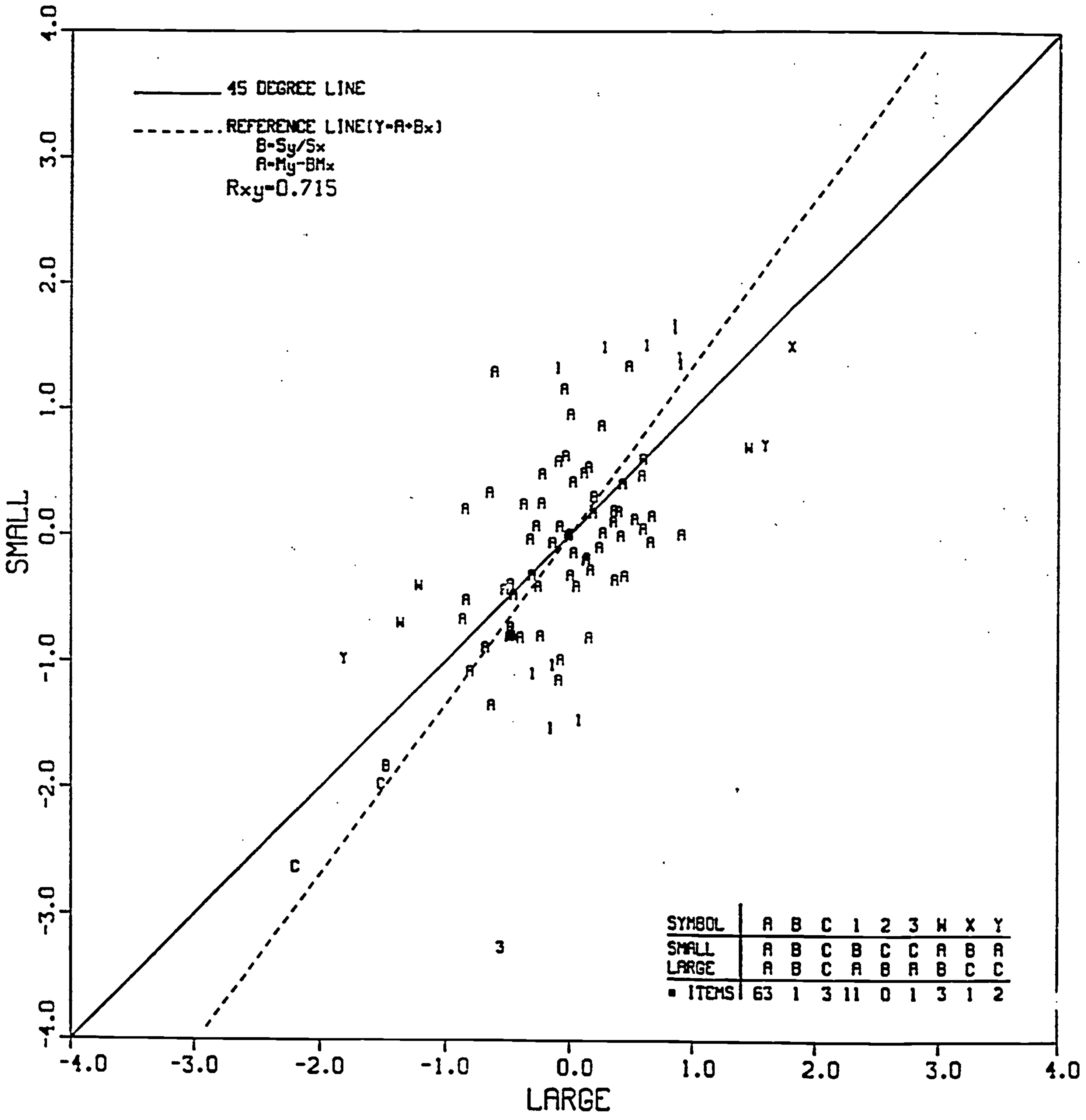


MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON FS(NR NI) FOR SML-LRG (03/88)



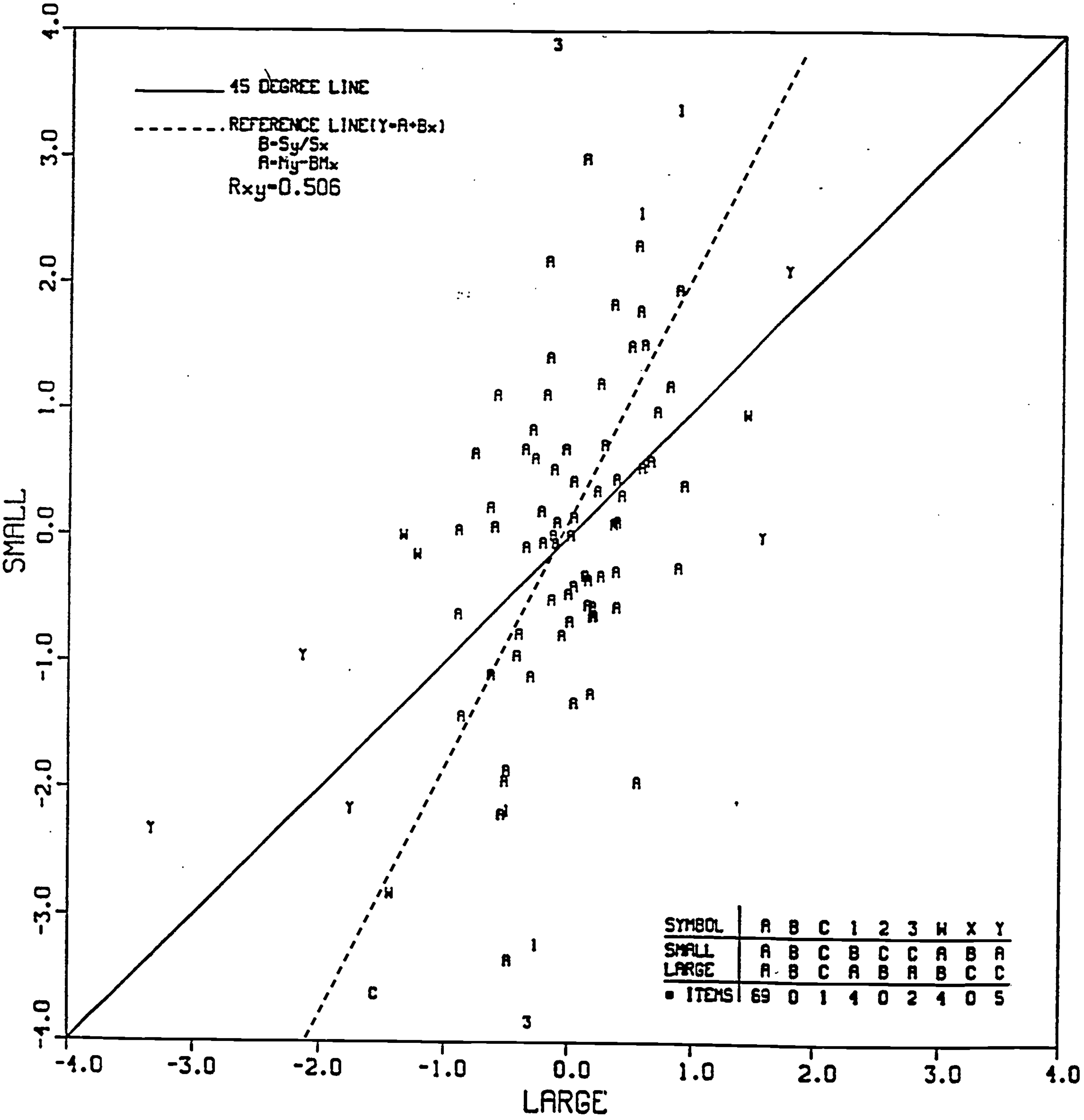
SYMBOL	A	B	C	1	2	3	W	X	Y
SMALL	A	B	C	B	C	C	A	B	A
LARGE	A	B	C	A	B	A	B	C	C
• ITEMS	66	1	2	8	0	1	3	2	2

MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON RIGHTS(NR-W) FOR SML-LRG (03/88)



1 MH Values out of range

MH D-DIF SCATTERPLOT OF MALE/FEMALE
ON BIVARIATE FOR SML-LRG (03/88)



SYMBOL	A	B	C	1	2	3	W	X	Y
SMALL	A	B	C	B	C	C	A	B	A
LARGE	A	B	C	A	B	A	B	C	C
# ITEMS	69	0	1	4	0	2	4	0	5