



Research Memorandum
ETS RM-16-17

**LRDIF_ES: A SAS Macro for
Logistic Regression Tests for
Differential Item Functioning of
Dichotomous and Polytomous Items**

Jianbin Fu

Lora Monfils

December 2016

ETS Research Memorandum Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Senior Research Scientist

Heather Buzick
Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Research Director

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Distinguished Presidential Appointee

Anastassia Loukina
Research Scientist

John Mazzeo
Distinguished Presidential Appointee

Donald Powers
Managing Principal Research Scientist

Gautam Puhan
Principal Psychometrician

John Sabatini
Managing Principal Research Scientist

Elizabeth Stone
Research Scientist

Matthias von Davier
Senior Research Director

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ayleen Gontz
Senior Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Memorandum series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Memorandum series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

**LRDIF_ES: A SAS Macro for Logistic Regression Tests for Differential Item Functioning of
Dichotomous and Polytomous Items**

Jianbin Fu and Lora Monfils
Educational Testing Service, Princeton, New Jersey

December 2016

Corresponding author: J. Fu, E-mail: jfu@ets.org

Suggested citation: Fu, J., & Monfils, L. (2016). *LRDIF_ES: A SAS macro for logistic regression tests for differential item functioning of dichotomous and polytomous items* (Research Memorandum No. RM-16-17). Princeton, NJ: Educational Testing Service.

Find other ETS-published reports by searching the ETS ReSEARCHER
database at <http://search.ets.org/researcher/>

To obtain a copy of an ETS research report, please visit
<http://www.ets.org/research/contact.html>

Action Editor: James Carlson

Reviewers: Guangming Ling and Lixiong Gu

Copyright © 2016 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, MEASURING THE POWER OF LEARNING are registered trademarks of Educational Testing Service (ETS). All other trademarks are the property of their respective owners.



Abstract

A SAS macro, LRDIF_ES, has been developed to apply logistic regression analyses to test differential item functioning (DIF) for dichotomous and polytomous items (Swaminathan & Rogers, 1990; Zumbo, 1999). The macro can be used to detect overall DIF, uniform DIF, and nonuniform DIF. Effect sizes are calculated and can be used to classify DIF into A, B, or C categories. Additional explanatory variables can be included in the logistic regression analyses as needed. The macro has the functionality to produce a line plot for each item and grouping variable, where the mean item score and the lower and upper boundaries of the 95% confidence interval of the mean score for each group are plotted against the criterion (matching) variable. The underlying statistical procedure and the syntax of this macro are described, and sample output is provided.

Key words: differential item functioning, logistic regression, SAS macro

Acknowledgments

Thanks are due to Carolyn Wentzel, James Carlson, Guangming Ling, and Lixiong Gu for their helpful suggestions and edits on early versions of this paper. Thanks also to Ayleen Gontz for copyediting this paper.

In this paper we describe a SAS macro, LRDIF_ES,¹ which applies logistic regression analyses to test differential item functioning (DIF) for dichotomous and polytomous items (Swaminathan & Rogers, 1990; Zumbo, 1999). The macro can be used to detect overall DIF, uniform DIF, and nonuniform DIF. Effect sizes are calculated and can be used to classify DIF into A, B, or C categories. Additional explanatory variables can be included in logistic regression analyses if needed. The macro has an option to produce a line plot for each item and grouping variable, where the mean item score and the lower and upper boundaries of the 95% confidence interval of the mean score for each group are plotted against the criterion (matching) variable.

Note that this SAS macro is a modification and extension of Christensen's (2004) SAS macro, known as "Irdif." The original macro is designed to conduct only uniform DIF for both dichotomous and polytomous items and does not compute effect sizes.

Differential Item Functioning Using Logistic Regression

Methods of DIF analysis based on logistic regression provide a model-based approach to identify uniform and nonuniform DIF. For dichotomously scored items, the probability of a correct response is modeled as a function of ability and group membership as well as other explanatory variables (e.g., family income, parents' years of education) if desired. Under this formulation, the DIF procedure using logistic regression compares the following three models (Swaminathan & Rogers, 1990):

$$\text{Model 1: } \text{logit}(P_1) = \ln\left(\frac{P_1}{1-P_1}\right) = \beta_{1(0)} + \beta_{11}X + \boldsymbol{\beta}'_{14}\mathbf{E}; \quad (1)$$

$$\text{Model 2: } \text{logit}(P_2) = \ln\left(\frac{P_2}{1-P_2}\right) = \beta_{2(0)} + \beta_{21}X + \beta_{22}G + \boldsymbol{\beta}'_{24}\mathbf{E}; \quad (2)$$

$$\text{Model 3: } \text{logit}(P_3) = \ln\left(\frac{P_3}{1-P_3}\right) = \beta_{3(0)} + \beta_{31}X + \beta_{32}G + \beta_{33}XG + \boldsymbol{\beta}'_{34}\mathbf{E}; \quad (3)$$

Where P denotes the probability of a test taker answering an item incorrectly, X denotes the criterion variable, G denotes group membership, \mathbf{E} is a vector including additional explanatory variables, and $\beta_{k(0)}$, β_{k1} , β_{k2} , β_{k3} , and $\boldsymbol{\beta}_{k4}$ are the associated regression parameters for model k . The model subscript on these terms of the model emphasizes that, for example, the P_k and the

parameters for X differ from model to model. Note that, for simplicity, the subscripts for item and test-taker notation are excluded from these formulas.

For polytomously scored items, the ordinal logistic regression analysis (Zumbo, 1999) with the cumulative probability function is used to test for DIF. Specifically, the following three ordinal logistic models are compared, similar to those used for dichotomous items:

$$\text{Model 1: } \text{logit}[P_1(Y \leq j)] = \ln \left[\frac{P_1(Y \leq j)}{P_1(Y > j)} \right] = \beta_{1(j)} + \beta_{11}X + \beta'_{14}\mathbf{E}; \quad (4)$$

$$\text{Model 2: } \text{logit}[P_2(Y \leq j)] = \ln \left[\frac{P_2(Y \leq j)}{P_2(Y > j)} \right] = \beta_{2(j)} + \beta_{21}X + \beta_{22}G + \beta'_{24}\mathbf{E}; \quad (5)$$

$$\text{Model 3: } \text{logit}[P_3(Y \leq j)] = \ln \left[\frac{P_3(Y \leq j)}{P_3(Y > j)} \right] = \beta_{3(j)} + \beta_{31}X + \beta_{32}G + \beta_{33}XG + \beta'_{34}\mathbf{E}; \quad (6)$$

where an item has possible integer scores of $0, 1, 2, \dots, M-1$, $P(Y \leq j)$ denotes the probability of getting a score of j or lower on an item by a test taker, and $j = 0, 1, 2, \dots, M-2$. A higher score means better performance. In general, j can be any real number; the logistic regression model treats each unique j as an ordered class (or level). Note that each item score category j is associated with a unique intercept $\beta_{k(j)}$, while the other regression parameters are the same across item score categories. For both dichotomous and polytomous items, Models 1, 2, and 3 are also referred to as a no DIF model, a uniform DIF model, and a nonuniform DIF model, respectively. The logistic regression models are estimated in the macro by *Proc Logistic* for both dichotomously scored items and polytomously scored items.

Note that for a dichotomously scored item the target probability that the logistic regression estimates is the probability of answering an item incorrectly, which is different from the logistic regression with the target probability as answering an item correctly that readers may be accustomed to. This impacts the interpretation of the signs of model parameters; readers should keep this in mind when interpreting the direction of DIF. Similarly, the target probability in the regression model for a polytomously scored item is the probability of obtaining an item score or below, to be consistent with that for a dichotomously scored item.

The presence of nonuniform DIF is tested by comparing Model 2 to Model 3 using the likelihood ratio test:

$$\chi^2 = [-2 \ln L(\text{Model2})] - [-2 \ln L(\text{Model3})],$$

where $-2 \ln L(\)$ refers to the negative two times the log likelihood value of a model, and χ^2 follows the chi-square distribution with one degree of freedom (df). Similarly, the presence of uniform DIF is tested using the likelihood ratio test with 1 df between Model 1 and Model 2:

$$\chi^2 = [-2 \ln L(\text{Model1})] - [-2 \ln L(\text{Model2})].$$

A simultaneous test of overall DIF (uniform DIF or nonuniform DIF) can be conducted using the likelihood ratio test with 2 df between Model 1 and Model 3:

$$\chi^2 = [-2 \ln L(\text{Model1})] - [-2 \ln L(\text{Model3})].$$

In addition to the likelihood ratio test between two models, effect size is used for DIF classification so as to avoid practically trivial but statistically significant results (French & Maller, 2007). Effect size is indicated by the difference of the Nagelkerke R^2 between two models (Gómez-Benito, Hidalgo, & Padilla, 2009). Jodoin and Gierl (2001) proposed a DIF classification rule for dichotomous items (Table 1) using the logistic regression procedure as described previously and indicated that this criterion was consistent to the DIF classification rule in the Mantel-Haenszel procedure used by Educational Testing Service (ETS). We have not seen a corresponding DIF classification rule for polytomous items using logistic regression.

Therefore, as a starting point, we may just use available empirical data to determine a rule for polytomous items. (See Kim, Cohen, Alagoz, & Kim, 2007, for an example of applying the DIF procedure to polytomous items using logistic regression.)

Table 1. Differential Item Functioning (DIF) Classification Rule Using Logistic Regression for Dichotomous Items

DIF category	Criterion
A (Negligible)	χ^2 test is not significant at .05 level or $\Delta R^2 < .035$
B (Slight to Moderate)	χ^2 test is significant at .05 level and $.035 \leq \Delta R^2 < .070$
C (Moderate to Large)	χ^2 test is significant at .05 level and $\Delta R^2 \geq .070$

Note. ΔR^2 is the Nagelkerke R^2 difference between two logistic regression models.

Usually, we are only concerned with items classified as having C-DIF. These items should be further reviewed by content specialists to determine if sources of sensitivity or bias can be identified; this may result in the items being revised accordingly and field tested again or the items being discarded. We first assess if an item has nonuniform C-DIF. If yes, we say that this item has nonuniform C-DIF regardless of evidence of uniform C-DIF. If nonuniform C-DIF is not found, we check whether the item has uniform C-DIF. The distinction between nonuniform DIF and uniform DIF may be helpful for identifying the cause of item DIF and thus appropriate revisions of the item.

Assume all grouping variables are coded as 0 for the focal group and 1 for the reference group. If an item appears to have uniform DIF, a positive β_{22} value indicates the item favors the focal group and a negative value indicates the item favors the reference group. If the item exhibits nonuniform DIF, positive values in both β_{32} and β_{33} indicate the item favors the focal group and negative values in both variables indicate the item favors the reference group. If the signs of the two parameter estimates are in the opposite direction, then the item favors one group or the other within different ranges of the criterion variable.

The advantages of using logistic regression for DIF identification include not only the ability to model uniform and/or nonuniform DIF, but also the ability to include additional explanatory variables beyond ability and group membership in the model.

Syntax

```
%macro lrdif_es(data, items, group, criterion, exo_con=, exo_cat=,
sig_level=0.05, RS_D_crit_B=0.035, RS_D_crit_C=0.070, RS_P_crit_B=0.035,
RS_P_crit_C=0.070, foc_size=100, ref_size=100, combined_size=400,
outfile=LRDIF, plot=NO, size_ratio=0, out_parameter=0);
```

Arguments

data: the input data.

items: the variable names in the data containing item scores (could be any real number; however, typically are integers starting from 0; i.e., 0, 1, 2, . . .) on which DIF is to be conducted; multiple item score variables can be included and each variable name must be no longer than 20 characters (this can be changed in the code).

group: the variable names in the data indicating groups on which DIF is to be conducted; each variable is coded as 1 for reference group and 0 for focal group; multiple grouping variables can be included, and the combined length of each grouping variable name and the criterion name must be no longer than 19 characters.

criterion: the matching variable for DIF; usually the total test score.

exo_con: additional continuous explanatory variables, included only if needed.

exo_cat: additional categorical explanatory variables; should be numeric variables, included only if needed.

sig_level: the level of significance for model comparisons; the default value is 0.05.

RS_D_crit_B: the cut-off value of R^2 difference between DIF A category and DIF B category for dichotomous items; the default value is 0.035 (see Table 1).

RS_D_crit_C: the cut-off value of R^2 difference between DIF B category and DIF C category for dichotomous items; the default value is 0.070 (see Table 1).

RS_P_crit_B: the cut-off value of R^2 difference between DIF A category and DIF B category for polytomous items; the default value is 0.035.

RS_P_crit_C: the cut-off value of R^2 difference between DIF B category and DIF C category for polytomous items; the default value is 0.070.

foc_size, ref_size, combined_size: minimum sample size requirements for focal group, reference group, and two groups combined; if the sample size of the focal group is smaller than **foc_size**, the sample size of the reference group is smaller than **ref_size**, or the sample size of the groups combined is smaller than **combined_size**, then the DIF analysis is not conducted; the default values for **foc_size**, **ref_size**, and **combined_size** are 100, 100, and 400, respectively.

size_ratio: maximum sample size ratio of the large group over the small group. If the ratio of the large group over the small group is larger than **size_ratio**, then the DIF analysis is not

conducted. If `size_ratio < 1`, then `size_ratio` is not treated as a prerequisite for conducting DIF. The default is 0.

`out_parameter`: the name of the output SAS file containing parameter estimates for all logistic models. See Table 2 for the variables included in this file. The name of a library reference (`libref`) can be added to the front of the file name so that the output file is a permanent SAS data file. If `out_parameter= 0`, then this file is not output. The default is 0.

`outfile`: the name of the SAS output file containing DIF results for each item in “items” and each grouping variable in “group.” The variables in the output file are defined in Table 3. The name of a library reference (`libref`) can be added to the front of the file name so that the output file is a permanent SAS data file. Note that if the sample size criteria are not met (see above), or all test takers get the same score on the target item in the overall valid sample in a DIF run, then all the variables are missing except for Item, Group, Score_Category, N_ref, and N_foc.

`plot`: “Yes” or “No.” “Yes” produces a line plot for each item in “items” and each grouping variable in “group,” where the mean item score and the lower and upper boundaries of the 95% confidence interval of the mean score for each group in the grouping variable are plotted against the criterion variable. If the sample size for a group at a criterion value is smaller than 20, or if all item scores for a group at a criterion value are the same, then the lower and upper boundary of the mean score for this group at the criterion value are set to missing. Some labels in the plot are defined in Table 4; the label names can be modified in the code if needed.

Table 2. Variables in the Parameter Estimate File (&out_parameter)

Variable	Definition
Item	The name of the item variable on which DIF is conducted
Group	The name of the grouping variable on which DIF is conducted
Model	“Model 1,” “Model 2,” and “Model 3” corresponding to Equations 1–6
Variable	The parameter name (term) in the logistic model
ClassVal0	Class level corresponding to target item score level; only relevant to the intercepts of polytomous items; representing the subscript j of $\beta_{1(j)}$, $\beta_{2(j)}$, and $\beta_{3(j)}$ in Equations 4–6
DF	Degree of freedom
Estimate	Parameter estimate
StdErr	Standard error of parameter estimate
WaldChiSq	Wald chi-square statistic
ProbChiSq	Probability of Wald chi-square statistic
<u>_ESTTYPE_</u>	Estimation method used for model parameters

Note. DIF = differential item functioning.

Table 3. Variables in the Output File (&outfile)

Variable	Definition
Item	The name of the item variable on which DIF is conducted
Group	The name of the grouping variable on which DIF is conducted
Score_Category	The number of item score categories in the data
N_ref	The number of students in the reference group (with group=1) used in the DIF calculation
N_foc	The number of students in the focus group (with group=0) used in the DIF calculation
Model 1	The -2 log likelihood of the no DIF model
Model 2	The -2 log likelihood of the uniform DIF model
Model 3	The -2 log likelihood of the nonuniform DIF model
RSQ1	The Nagelkerke R^2 of the no DIF model
RSQ2	The Nagelkerke R^2 of the uniform DIF model
RSQ3	The Nagelkerke R^2 of the nonuniform DIF model
M2Est	The parameter estimate of the grouping variable (“Group”) in the uniform DIF model. If the item appears to have uniform DIF, a positive value indicates the item favors the focal group with code 0 and a negative value indicates the item favors the reference group with code 1.
M3Est M3EstInt	The parameter estimates of the grouping variable (“Group”) and the interaction of the grouping variable and the criterion variable, respectively, in the nonuniform DIF model. If the item appears to have nonuniform DIF, positive values in both variables indicate the item favors the focal group with code 0 and negative values in both variables indicate the item favors the reference group with code 1. If the signs of the two parameter estimates are opposite, then the item favors one group or the other within different ranges of the criterion variable.
LRDIF	The -2 log likelihood difference between Model 1 and Model 3 (Model 1 - Model 3)
LRUIDIF	The -2 log likelihood difference between Model 1 and Model 2 (Model 1 - Model 2)
LRNUIDIF	The -2 log likelihood difference between Model 2 and Model 3 (Model 2 - Model 3)
PDIF	The probability of overall DIF. This is equal to 1 minus the cumulative probability of LRDIF that follows a chi-square distribution with two df.
PUIDIF	The probability of uniform DIF. This is equal to 1 minus the cumulative probability of LRUIDIF that follows a chi-square distribution with one df.
PNUIDIF	The probability of nonuniform DIF. This is equal to 1 minus the cumulative probability of LRNUIDIF that follows a chi-square distribution with one df.
RSDIF	The Nagelkerke R^2 difference between the nonuniform DIF model and the no DIF model: RSQ3-RSQ1
RSUIDIF	The Nagelkerke R^2 difference between the uniform DIF model and the no DIF model: RSQ2-RSQ1
RSNUIDIF	The Nagelkerke R^2 difference between the nonuniform DIF model and the uniform DIF model: RSQ3-RSQ2
DIF	Overall DIF classification based on Table 1 or similar rules. Blank = nonsignificant overall DIF chi-square test (i.e., PDIF > sig_level), * = significant overall DIF chi-square test (i.e., PDIF ≤ sig_level) but not classified as B or C DIF, B = B DIF, C = C DIF (note that blank and * indicate A DIF); + sign after *, B, and C indicates the item favors the focal group; – sign after *, B, and C indicates the item favors the reference group; and no sign after *, B, and C indicates the item favors one group or the other within different ranges of the criterion variable.

Variable	Definition
UIDIF	Uniform DIF classification based on Table 1 or similar rules. Blank = nonsignificant uniform DIF chi-square test (i.e., PUIDIF > sig_level), * = significant uniform DIF chi-square test (i.e., PUIDIF ≤ sig_level) but not classified as B or C DIF, B = B DIF, C = C DIF (note that blank and * indicate A DIF); + sign after *, B, and C indicates the item favors the focal group; and – sign after *, B, and C indicates the item favors the reference group.
NUIDIF	Nonuniform DIF classification based on Table 1 or similar rules. Blank = nonsignificant nonuniform DIF chi-square test (i.e., PNUIDIF > sig_level), * = significant nonuniform DIF chi-square test (i.e., PNUIDIF ≤ sig_level) but not classified as B or C DIF, B = B DIF, C = C DIF (note that blank and * indicate A DIF); + sign after *, B, and C indicates the item favors the focal group; – sign after *, B, and C indicates the item favors the reference group; and no sign after *, B, and C indicates the item favors one group or the other within different ranges of the criterion variable.

Table 4. Labels in the Output Plot

Label	Definition
Criterion Scores	The criterion variable
Foc--LCLM	Lower boundary of the 95% confidence limit of the mean for the focal group with code 0
Foc--UCLM	Upper boundary of the 95% confidence limit of the mean for the focal group with code 0
Ref--LCLM	Lower boundary of the 95% confidence limit of the mean for the reference group with code 1
Ref--UCLM	Upper boundary of the 95% confidence limit of the mean for the reference group with code 1

Notes

1. A valid case in a logistic regression DIF run should not have missing values on any of the following variables: target item score, target grouping variable, criterion score, and, if used, explanatory variables. Categorical explanatory variables should have numeric values in the input data.
2. The target grouping variable is a numeric variable with valid values 0 and 1; 0 refers to the focal group and 1 refers to the reference group.

Example

```
%lrdif_es(data=sample, items=item1-item8, group=gender black, exo_con=income age, exo_cat=parent_education state, criterion = score, outfile=results, plot=yes)
```

Sample Output

Table 5. SAS Output From DIF Analyses Using Logistic Regression

Item	Group	Score_ Category	N_ref	N_foc	Model1	Model2	Model3	RSQ1	RSQ2	RSQ3	M2Est	M3Es t	M3Es tInt	LRDI F	LRUIDI F	LRNUI DIF	PDIF	PUIDI F	PNUIDI F	RSDI F	RSUIDI F	RSNUI DIF	UIDI DIF	NUID F	NUID IF
item2	Male_Female	2	1276	1724	2684.808	2681.824	2675.731	.498	.499	.499	-.170	-.485	-.061	9.076	2.984	6.092	.011	.084	.014	.080	.001	.090	C-	C-	
item2	White_Black	2	938	2062	2684.808	2678.712	2678.048	.498	.499	.499	.192	.452	.052	6.760	6.096	.664	.034	.014	.415	.040	.045	.000	B+	B+	
item3	Male_Female	3	1276	1724	4286.991	4278.998	4278.992	.596	.597	.597	-.116	-.102	-.003	7.999	7.992	.007	.018	.005	.935	.000	.000	.000	*-	*-	
item3	White_Black	3	938	2062	4280.991	4280.867	4279.483	.596	.597	.597	-.031	.181	-.051	1.507	.123	1.384	.471	.726	.239	.000	.000	.000			

Note: See Table 3 for the definitions of table headings.

Item 2 is a dichotomous item. It demonstrates C-category overall DIF and C-category nonuniform DIF for the Male vs. Female comparison. This classification is based on the criteria in Table 1: the likelihood ratio tests for the respective overall DIF (PDIF = .011) and nonuniform DIF (PNUIDIF = .014) are significant at the .05 level, and the effect sizes for the overall DIF (RSDIF = .080) and nonuniform DIF (RSNUIDIF = .09) are larger than the cut-off value of C-category DIF for dichotomous items, .070 (i.e., RS_D_crit_C = .070). The regression parameters for the grouping variable (M3Est = -.485) and the interaction term (M3Est = -.061) in the nonuniform DIF model are both negative; consequently there is a – sign after C under NUIDIF, and the item favors male (i.e., reference group code = 1). Item 2 also has B-category overall DIF and nonuniform DIF on White versus Black. This classification is also based on the criteria in Table 1: the likelihood ratio tests for the respective overall DIF (PDIF = .034) and uniform DIF (PUIDIF = .014) are significant at the .05 level, and the effect sizes for the overall DIF (RSDIF = .040) and uniform DIF (RSUIDIF = .045) are larger than the cut-off value of B-category DIF, .035 (i.e., RS_D_crit_B = .035) but smaller than the cut-off value of C-category DIF, .070 (i.e., RS_D_crit_C = .070) for dichotomous items. The regression parameter for the grouping variable (M2Est = .192) in the uniform DIF model is positive, so that there is a + sign after B in the UIDIF column, and the item favors Black test takers (i.e., focus group code = 0).

Item 3 is a polytomous item and does not exhibit B- or C-category DIF for Male versus Female or White versus Black comparisons by using the criteria in Table 1. However, the chi-square tests for overall DIF and uniform DIF for Male versus Female

are significant at the .05 level, and thus there is a * in the DIF and UIDIF columns for this item. The regression parameters, M2Est, M3Est, and M3EstInt, are all negative, so that there is a – sign after * under DIF and UIDIF.

The following is a mean item score plot for Item 3 in the Black (focal) and White (reference) groups corresponding to the fourth row in Table 5.

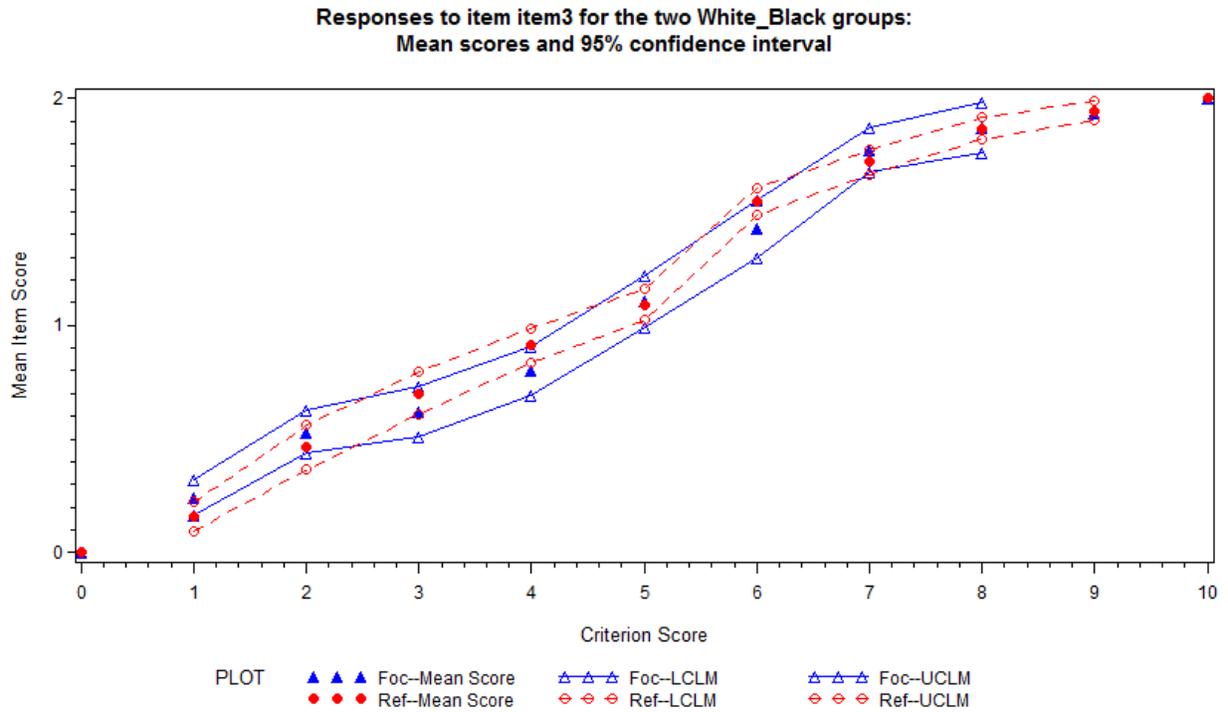


Figure 1. An example of mean item score plot. See Table 4 for the definitions of plot legends.

In Figure 1, the solid blue triangles and red circles represent mean item scores for the focus and reference groups, respectively, at each criterion score point. The empty blue triangles and red circles represent lower and upper boundaries of the 95% confidence limit of the mean item scores for the focus and reference groups, respectively, at each criterion score point. The boundaries of the 95% confidence limit across criterion score points are connected by the blue solid and red dash lines for the focus and reference groups, respectively, to show the trend of mean item score change across criterion scores. From Figure 1, one can observe that the Black and White groups perform similarly on Item 3 across criteria scores as the mean item scores of one group are within the 95% confidence limit of the other group across criterion scores.

References

- Christensen, K. B. (2004). *Irdif: A SAS macro for logistic regression tests for differential item functioning*. Retrieved from <http://192.38.117.59/~kach/>
- French, B., & Maller, S. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement, 67*, 373–393.
- Gómez-Benito, J., Hidalgo, M. D., & Padilla, J. L. (2009). Efficacy of measures of the effect size in logistic regression: An application in the detection of DIF. *Methodology--European Journal of Research Methods for the Behavioral and Social Sciences, 5*, 18–25.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329–349.
- Kim, S-H., Cohen, A. S., Alagoz, C., & Kim, S. (2007). DIF detection and effect size measures for polytomously scored items. *Journal of Educational Measurement, 44*, 93–116.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361–370.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression as a unitary framework for binary and likert-type (ordinal) item scores*. Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Notes

¹In “LRDIF_ES,” LR stands for logistic regression, DIF stands for differential item functioning, and ES stands for effect sizes.