



EduStat Software

Metric qualities of a test

It is important to check whether a test used to carry out the learning assessment meets the standards by guaranteeing quality. EduStat allows this analysis using the following two techniques: i) item analysis and ii) differential item functioning (DIF).

The classical item analysis provides statistics allowing to decide on the metric quality of the test. Three statistical indices may be produced: i) the difficulty of the items; li) the discrimination of the items and iii) the internal consistency of the test.

Moreover, the differential item functioning seeks analysis to identify items tending to be separated from the others for the same dominant theoretical ability and this, for all subgroups of students. The analysis thus makes it possible to keep in an evaluation only the items having a sufficiently uniform functioning and thus be able to make a comparison between various subgroups of pupils on the basis of a comparable scale and without distortion. EduStat provides this form of analysis using two techniques: (i) omnibus analysis and (ii) analysis according to two groups of students.

This document contains useful information for compilations to determine the metric quality of a test. The database "**TestEN.edu**" will be used to carry out the proposed exercises. This database contains 70 variables and 450 records; It should be noted that this database includes data on a mathematics test. We will produce statistical compilations related to the following EduStat options:

- Item analysis
- DIF analysis - "Omnibus"
- DIF Analysis – Two groups

Several documents are available when using the EduStat software. If you would like to have more information on how to proceed to the quality analysis of items, the following documents can be found in the documentation. To access it, choose the "Documentation" option from the "Help" menu. A window with five options appears. It is possible to open one of the three options allowing to have information on the production of quality of items and tests by choosing these options: Slideshows, Videos (not available in English), Procedural documents. It is also possible to consult the "User's Guide" (not available in English). We advise you to read some of these documents before carrying out the activities presented in this text. Here are some references.

User's Guide - Not available in English

Viewing Slide Shows

- Item analysis : Classical item analysis
- DIF analysis – Two group - Differential item analysis
- DIF analysis– DIF - Omnibus

Viewing Videos – Not available in English

Viewing Procedural documents Items analysis:

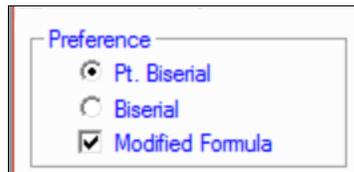
- Item analysis - Classical item analysis
- DIF analysis – Differential item analysis
- DIF analysis– DIF - Omnibus

❖ *Classical item analysis*

The execution of an item analysis must be carried out in relation to a single discipline (here, mathematics). If compilations have to be made in relation to several disciplines part of a same database, treatment must therefore be carried out for each discipline. Here is the approach to use.

Choice of different calculation parameters

- Preference for the calculation technique



Preference

- Pt. Biserial
- Biserial
- Modified Fomula

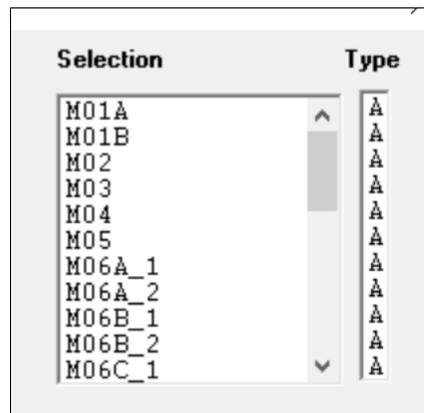
- Processing missing values



Missing Values

- Ignore
- For each variable
- Global
- Included
- Excluded

- Selection of variables (the items)



Selection	Type
M01A	A
M01B	A
M02	A
M03	A
M04	A
M05	A
M06A_1	A
M06A_2	A
M06B_1	A
M06B_2	A
M06C_1	A

Report

- Statistics for each item

Items	Number	Proportion	Pt.Bis.
M01A	450	45.778	0.378
M01B	450	13.333	0.296
M02	450	67.556	0.379
M03	450	49.556	0.491
M04	450	17.556	0.273
M05	450	35.111	0.072
M06A_1	450	74.667	0.418

- Summary of results according to difficulty and discrimination

Discrimination	Difficulty Index										Frequencies
	00>10	10>20	20>30	30>40	40>50	50>60	60>70	70>80	80>90	90>100	
0.00> 0.10	0	0	0	1	1	0	0	0	0	0	2
0.10> 0.20	1	1	0	0	2	0	1	0	0	0	5
0.20> 0.30	2	2	2	0	0	0	0	0	0	0	6
0.30> 0.40	1	2	1	0	2	0	1	0	0	0	7
0.40> 0.50	0	0	0	2	4	1	0	2	1	0	10
0.50> 0.60	0	0	0	0	0	0	0	0	0	0	0
0.60> 0.70	0	0	0	0	0	0	0	0	0	0	0
0.70> 0.80	0	0	0	0	0	0	0	0	0	0	0
0.80> 0.90	0	0	0	0	0	0	0	0	0	0	0
0.90> 1.00	0	0	0	0	0	0	0	0	0	0	0
Frequencies	4	5	3	3	9	1	2	2	1	0	30

Results for the entire test

	Mean	Std-Dev	Alpha	Std-Error
Global	36.459	16.541	0.805	7.304

- Results obtained according to groups of items

The abilities	Mean	Std-Dev	Alpha	Std-Error
Knowledge	40.370	18.808	0.041	18.420
Understanding	32.500	31.581	0.646	18.789
Application	36.078	19.441	0.815	8.357

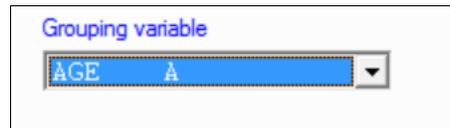
It should be noted that it is preferable to retain three decimal places in order to better appreciate the results of the different groupings.

❖ *Differential Item Analysis – Omnibus*

When we process the classical item analysis, it is necessary to choose a single discipline. The technique assumes the presence of several values in relation to the selected variable. In this exercise, we will use the variable "AGE", which includes three subgroups of students: i) students aged 8 and under, ii) 9-year-old students, and iii) students aged 10 years and over. Here's how to get the desired compilations.

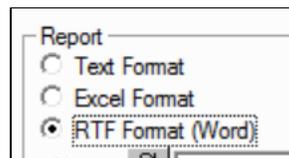
- Sub group variable

The variable to be used as subgroups of students must be selected.

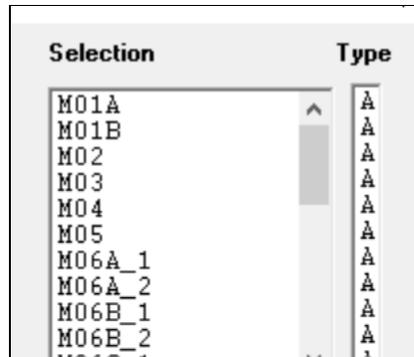


- Select report format

If we wish to have a report including an illustration of the location of the indices linked to each item according to the regression line, it is necessary to select RTF format.



- Selection of variables associated with a discipline; we retained the items of mathematics.



Report

The report includes a table providing different statistical coefficients as well as a graphical illustration.

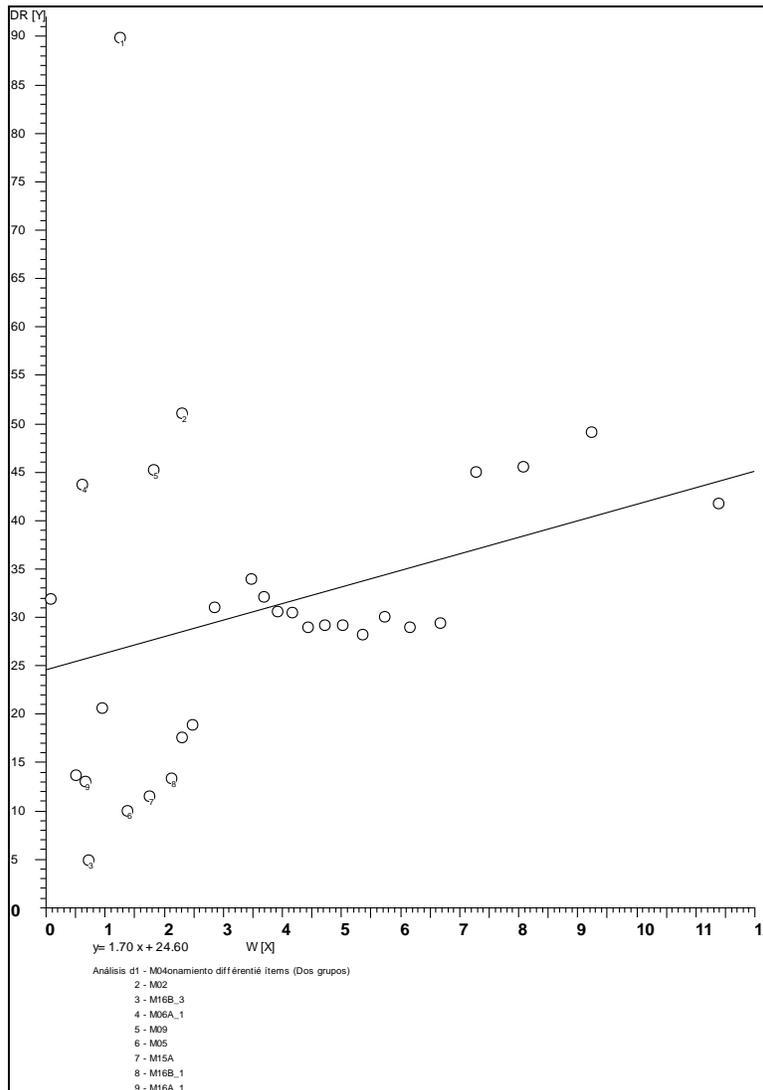
A graph illustrates the adequacy of the theoretical distribution to the empirical distribution which is evaluated by the coordinates of the values Rank (the "X" axis) and the DR values (the "Y" axis). A good match between the values "Rank" and "DR" is obtained by a graph in which the points (x, y) draw a linear line. Any point with a significant deviation from this linear line is interpreted as potentially marginalized. The chart includes the list of the nine most "biased" items. It is possible to locate them in relation to the graphic illustration; these are the items most likely to be biased that deviate most from the regression line.

- Statistical coefficients (only a few items are displayed)

Sort by Chi-Square

Item	Chi-SQ	P	W	DR	RS
M06A_2	475.235	0.978	11.390	41.723	1.406
M03	453.355	0.945	9.236	49.086	1.654
M01A	368.480	0.912	8.092	45.534	1.534
M14B	328.041	0.879	7.298	44.952	1.515
M14A	195.895	0.846	6.682	29.316	0.988
M06B_1	178.773	0.813	6.176	28.945	0.975
M16B_2	172.577	0.780	5.744	30.043	1.012

- Graphic illustration



This graph includes the following two parts: (i) the location of the items in relation to the regression line and (ii) the identification of the nine items that move away from this regression line. The more an item moves away from that line, the more "bias" there is. These are the items to be examined first.

❖ *Differential Item Analysis – Two groups*

Just as for the production of the classical items analysis, it is necessary to choose a single discipline when we select the variables. We must select a student subgroup variable that includes two subgroups; we will retain the variable "SEX". Here's what to do.

- Selection of sub groups

The codes representing each of the two subgroups of students must be chosen. In this database, code "1" refers to "Boys" and code "2" to "Girls". After making the choices, the software displays statistics on each subgroup. It will have been preferable to define each of these subgroups.

Reference Group	Description	Boys	253 Records (SEX=1)
Compared Group	Description	Girls	196 Records (SEX=2)

- Selection of variables (we selected mathematics items)

Selection	Type
M01A	A
M01B	A
M02	A
M03	A
M04	A
M05	A
M06A_1	A
M06A_2	A
M06B_1	A
M06B_2	A
M06C_1	A

Report

- First table: statistics produced

Reference Group: Boys

Compared Group: Girls

Item	Alpha	Delta	Std-Error	Confidence Interval		Cat
M13B	1.872	-1.473	0.627	0.243	2.703	B
M14B	1.785	-1.362	0.584	0.217	2.507	B
M15A	6.518	-4.405	1.996	0.493	8.318	B
M16B_2	0.604	1.185	0.538	0.130	2.239	B
M01A	0.889	0.275	0.519	-0.742	1.292	A
M01B	0.937	0.152	0.745	-1.308	1.612	A
M02	0.871	0.325	0.549	-0.751	1.401	A

Compared to each item, a series of statistics is available. The last column indicates the category occupied by each item according to the presence of possible bias:

C: priority item to be considered in relation to the possibility of bias;

B: second category of items likely to possess some differentiated functioning;

A: item that is not identified as likely to have differentiated operation.

- Summary table

Summary

Items	Delta value	Bias category	DIF Group
M13B	- 1.473	B	Girls
M14B	- 1.362	B	Girls
M15A	- 4.405	B	Girls
M16B_2	+ 1.185	B	Boys

As can be seen, the results obtained for four items may be slightly biased; three disadvantage girls and one boy.