GaDOE
Georgia Department of Education

DATA REVIEW

What is Data Review?

Data review is a crucial step in the item development process, as Georgia educators consider field test item statistics in conjunction with item content and the characteristics of the tested students to determine if an item should be accepted or rejected for future use.

Field test item statistics help answer questions such as:

Is the item too difficult?
Is the item functioning properly?
Is there any evidence of potential bias?

Student performance data from field testing are analyzed for various statistical properties including item difficulty, item discrimination, and differential item functioning, or DIF. Items with extreme statistics or odd response patterns are flagged for review by committees of Georgia educators.

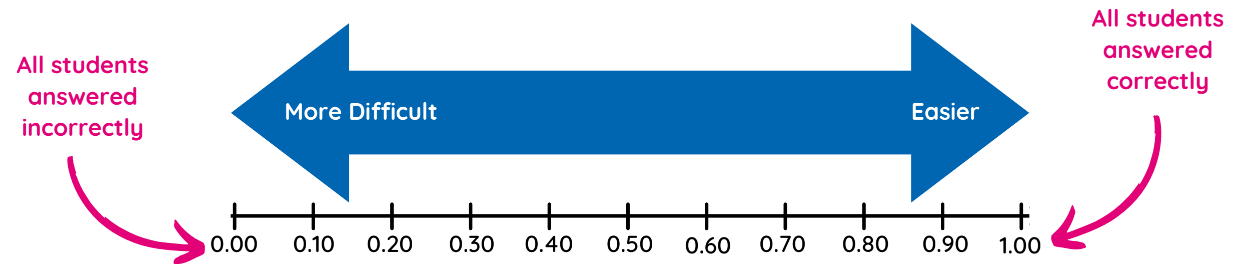
Evaluation by the data review committee focuses on the quality of the item in measuring the intended standard, not simply meeting given statistical criteria.

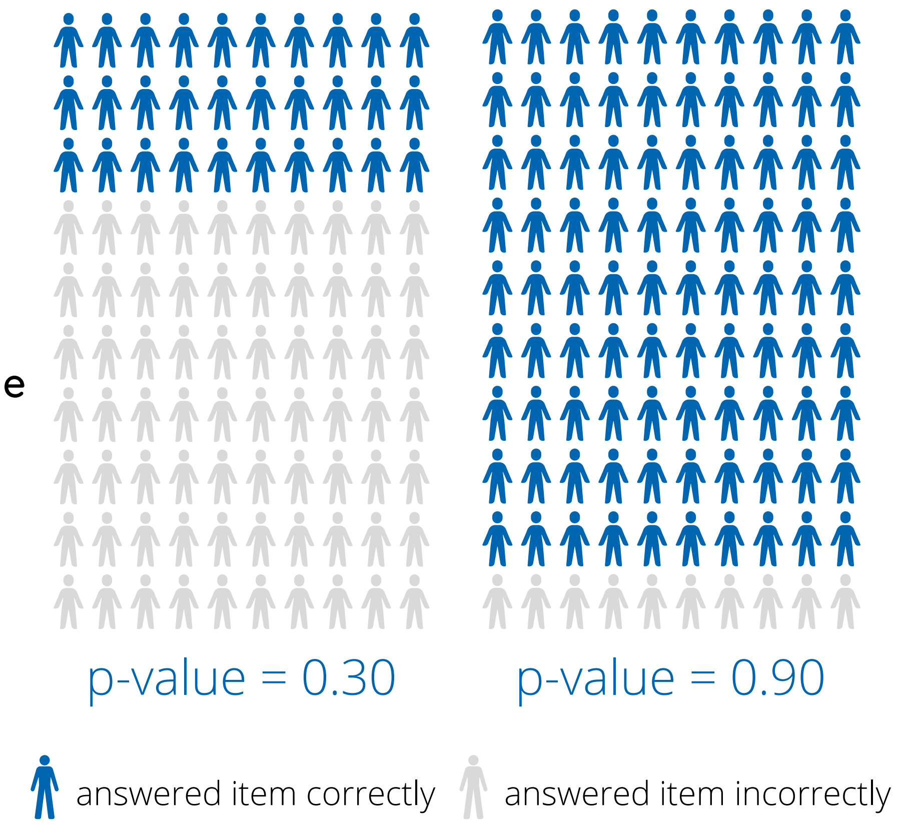
Item Difficulty (p-value)

Definition

For 1-point items:
the proportion of students selecting the correct response
If 79% answer correctly: p-value = 0.79

For items with more than one point possible:
the average item score divided by the maximum item score possible
If an item is worth two points and the average score is 1.5 points:
p-value = 1.5/2 = 0.75


All students answered incorrectly
More Difficult
Easier
All students answered correctly
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00



Interpretation

Relatively lower p-values (e.g., 0.30) correspond to more difficult items
Relatively higher p-values (e.g., 0.70) correspond to easier items

Items that are either very difficult (e.g., <0.10) or very easy (e.g., >0.90) provide little information about student differences in achievement.

Georgia builds tests with a wide range of p-values (generally 0.30-0.90) in order to effectively measure the achievement of students across all achievement levels.

Item Discrimination (Item-Total Correlation)

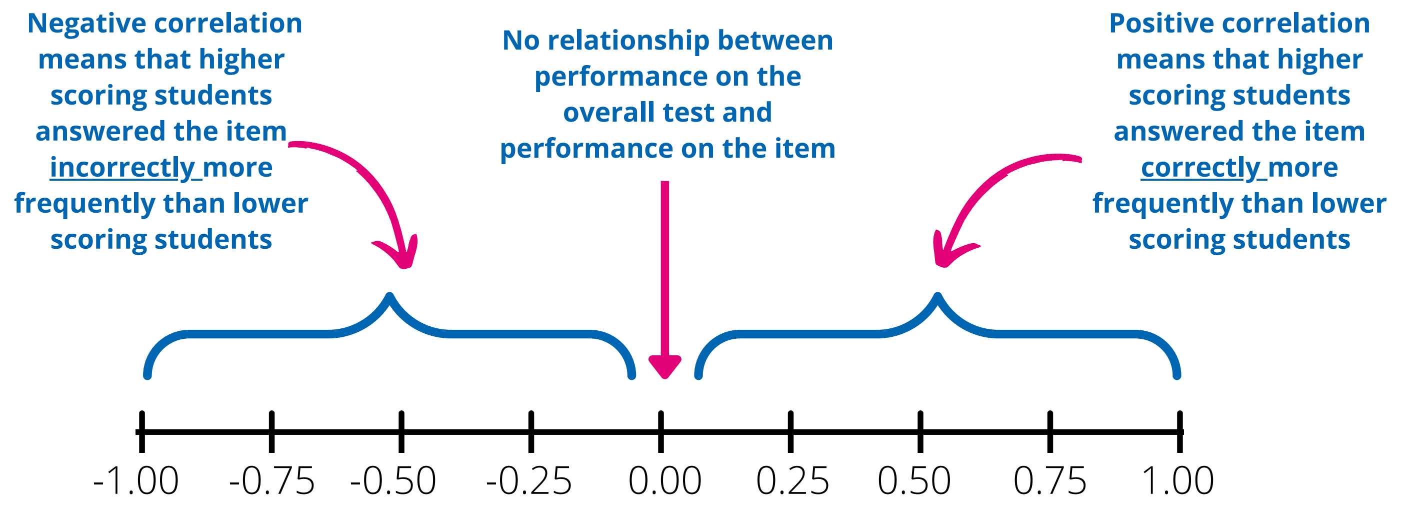
Definition

The relationship between performance on a specific item and performance on the overall test

The correlation coefficient can range from -1.00 to +1.00

Assumption

Students who score well on the overall assessment should also score well on an individual item


Negative correlation means that higher scoring students answered the item incorrectly more frequently than lower scoring students
No relationship between performance on the overall test and performance on the item
Positive correlation means that higher scoring students answered the item correctly more frequently than lower scoring students
-1.00
-0.75
-0.50
-0.25
0.00
0.25
0.50
0.75
1.00

Interpretation

Large positive values (e.g., >0.40) mean the item is a good discriminator between high- and low-achieving students.

Close to zero or negative values (e.g., 0.01, -0.20) can indicate problems with the item content or students' opportunity to learn.

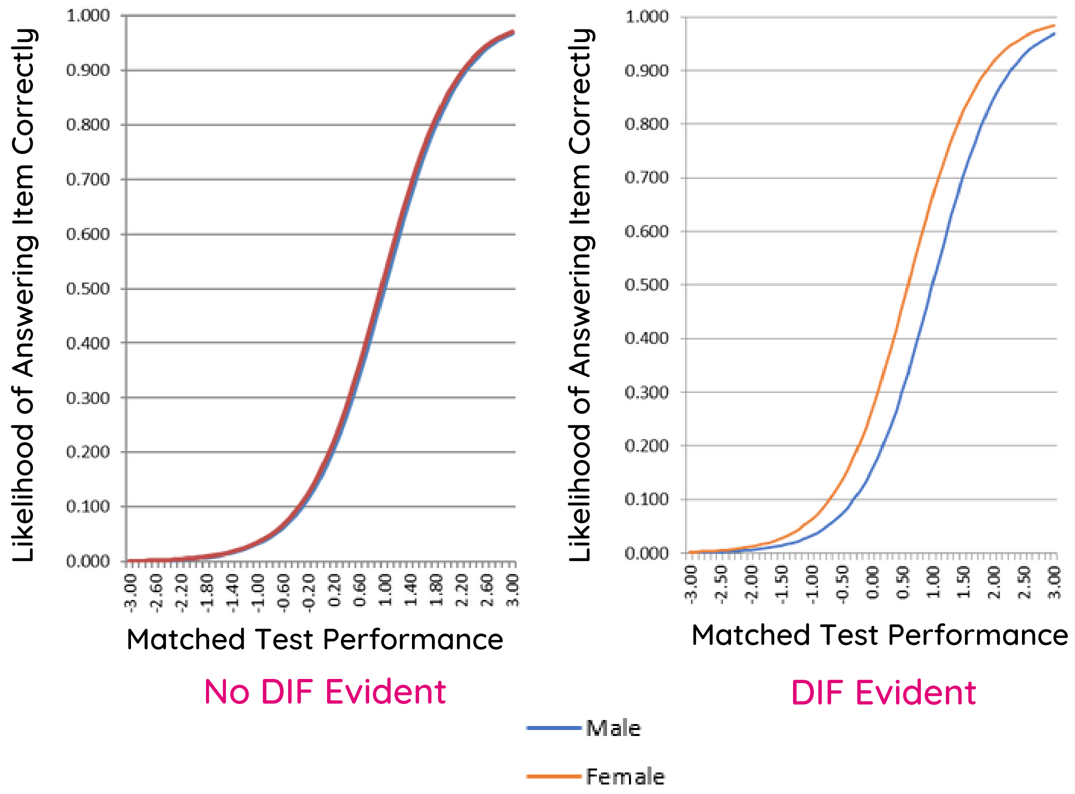
Differential Item Functioning (DIF)

Definition

DIF occurs when samples of students from different groups (e.g., gender or race/ethnicity) with the same ability level have different probabilities of responding correctly to an item.

DIF statistics quantify the difference in item performance between two groups - after matching group performance on the overall test.

DIF suggests a potential threat to validity but not all items with DIF are biased.



No DIF Evident

DIF Evident

Levels of DIF

Level A
Item with little or no difference in performance between matched groups of students

Level B
Items with small to moderate differences in performance between matched groups of students

Level C
Items with larger differences between matched groups of students

For general questions about the Georgia student assessment program, contact ga.assessment@doe.k12.ga.us.