

EDUCATIONAL TESTING SERVICE

By Ada Woo, PhD, and Marijana Dragan, MS

As the field of educational assessment continues to evolve, the need for more sophisticated methods of analyzing test data has become increasingly apparent. One such method is Differential Item Functioning (DIF) analysis, which allows researchers to identify and address bias in test results based on demographic characteristics such as gender, race, and ethnicity. This article explores the concept of DIF, its various types, and the methods used to detect and adjust for bias in test scores.

Detecting Bias

One of the primary goals of educational assessment is to ensure that test results are fair and unbiased. However, various factors can introduce bias into the testing process, leading to inaccurate scores and unfair comparisons. Detecting bias is a critical step in the development and validation of educational tests. This section discusses the methods used to identify and address bias in test results.

One common method for detecting bias is Differential Item Functioning (DIF) analysis. DIF analysis examines the performance of test items across different groups of test-takers, such as those defined by gender, race, or ethnicity. If an item performs differently for one group compared to another, it is considered biased. DIF analysis can be used to identify biased items and adjust the test scores to account for the bias.

Another method for detecting bias is the use of test equating. Test equating involves comparing the scores of two different tests to ensure that they are measuring the same construct and that the scores are comparable. This process can help to identify and address bias in test results.

Finally, the use of test re-revision can also help to detect and address bias. Test re-revision involves reviewing and modifying test items based on feedback from test-takers and other stakeholders. This process can help to identify and address bias in test results.

Differential Item Functioning (DIF) analysis is a statistical method used to identify and address bias in test results. It examines the performance of test items across different groups of test-takers, such as those defined by gender, race, or ethnicity. If an item performs differently for one group compared to another, it is considered biased. DIF analysis can be used to identify biased items and adjust the test scores to account for the bias.

There are three main types of DIF: uniform DIF, non-uniform DIF, and interaction DIF. Uniform DIF occurs when an item performs consistently better or worse for one group compared to another. Non-uniform DIF occurs when the performance of an item varies across different levels of ability within a group. Interaction DIF occurs when the performance of an item varies across different groups and levels of ability.

There are several methods for detecting DIF, including the Mantel-Haenszel method, the logistic regression method, and the generalized Mantel-Haenszel method. Each method has its own strengths and weaknesses, and researchers often use multiple methods to identify DIF.

Once DIF has been identified, there are several ways to address the bias. One approach is to remove the biased items from the test. Another approach is to adjust the test scores to account for the bias. Finally, test re-revision can be used to modify the biased items.

Differential Item Functioning Analyses

Differential Item Functioning (DIF) analysis is a statistical method used to identify and address bias in test results. It examines the performance of test items across different groups of test-takers, such as those defined by gender, race, or ethnicity. If an item performs differently for one group compared to another, it is considered biased. DIF analysis can be used to identify biased items and adjust the test scores to account for the bias.

There are three main types of DIF: uniform DIF, non-uniform DIF, and interaction DIF. Uniform DIF occurs when an item performs consistently better or worse for one group compared to another. Non-uniform DIF occurs when the performance of an item varies across different levels of ability within a group. Interaction DIF occurs when the performance of an item varies across different groups and levels of ability.

There are several methods for detecting DIF, including the Mantel-Haenszel method, the logistic regression method, and the generalized Mantel-Haenszel method. Each method has its own strengths and weaknesses, and researchers often use multiple methods to identify DIF.

Once DIF has been identified, there are several ways to address the bias. One approach is to remove the biased items from the test. Another approach is to adjust the test scores to account for the bias. Finally, test re-revision can be used to modify the biased items.

