

# Performance Test *p*-value Calculations

Wallace Judd, PhD



# Performance Test $p$ -value Calculations

Wallace Judd, PhD

# Performance Test $p$ -value Calculations

## Definition

The  $p$ -value is the percentage of correct responses to the item. It represents the difficulty of the item for the sample of test takers whose item scores are under consideration.

$$p\text{-value} = \text{sum of item scores} / \text{number of candidates}$$

The actual calculation of the  $p$ -value may vary in different situations: dichotomous scoring (i.e., binary 1 | 0), polytomous scoring, and performance scoring.

## Examples

Sample calculations for all the following explanations are available in the Excel spreadsheet:

[p-Val Calcs.xlsx](#)

## Dichotomous

This is the simplest data form for calculating a  $p$ -value because all items are scored either 0 or 1.

The  $p$ -value is simply the average of item scores.

$$p\text{-value} = \text{sum of item scores} / \text{number of candidates}$$

## Polytomous

In some tests, items are scored such that an item may be worth more than one point. For example, on an item worth 3 total points, a test taker could earn 0, 1, 2, or 3 points, depending on their performance.

In this case, the average score for the item could serve as an indicator of the item difficulty.

$$\text{difficulty} = \text{sum of item scores} / \text{number of candidates}$$

## Some Items Skipped

In many tests, not all items are administered to each candidate. In these adaptive testing scenarios, it is important to use the correct number of candidates: include only those who saw or *could have seen* the item.

Typically, candidates who saw or could have seen the item and failed to answer it (correctly or at all) get a zero, whereas candidates who weren't intended to see the item get a blank or other exclusion code.

This means that if many candidates don't get to certain items that appear toward the end of a test form, the  $p$ -value for the item will reflect only the item difficulty for that same at that position in the test.

## Performance Items

On performance tests with polytomous items, the average score may not be the best indicator of item difficulty. What is of interest is what percent of the points available candidates, on average, earned.

In this case, the  $p$ -value is the average of item scores:

$$p\text{-value} = \text{sum of item scores} / \text{maximum score} / \text{number of candidates}$$

With this approach, a time-consuming 10-point item and a quicker 1-point item can be compared on the same scale. If 90 percent of candidates earn full points on the first item and 90 percent earn the point on the second item, both items will be considered “easy,” with a  $p$ -value of 0.90. The different levels of effort required will be reflected in the points, not the  $p$ -value.

Elsewhere, we consider an item efficiency parameter, which considers how much candidate information is gleaned per unit of time.



3400 Monroe Avenue, Suite 336  
Rochester, N.Y. 14618-5513  
[www.AuthenticTesting.com](http://www.AuthenticTesting.com)  
+1-703-777-7321