Item Efficiency for Conversational Items
Wallace Judd
1/29/2024

[A 'conversational' exam is one in which the device speaks a prompt to the candidate, and the candidate responds by speaking to the device.

We use 'conversational' because if we used 'audio tests', the title might indicate we were testing audio equipment.

Also, browser searches for 'audio tests' would confuse these tests for testing of decibel levels, noise, and frequency response.]

Statistic Incorporating Latency

With a correct answer, the response latency should be a good indicator of either:

1. The candidate's familiarity with the item content or
2. The candidate's fluency with the logic or calculations necessary to get the answer.

So how should latency be incorporated.? Obviously, one raw measure is the inverse of the time required to articulate a correct answer.

But to include the relevance of the response to the test total score, the statistic should include both the pass-fail relevance and the point-biserial correlation to total score. Since at a typical cutscore of 0.70, the range of the first term is +/- 0.3, and in performance tests item point-biserials can range to over 0.6+, the 2 constant just gives equal weight to both terms.

A statistic that combines all these measures is called Item Efficiency:

$$\text{Efficiency}_i = (2 * \text{ABS}[\text{Cut}_T - \text{pVal}_i] * (\text{pBis}_i)) / t_i$$

Where

Efficiency_i	= The item efficiency score
2	= A constant to equate the weight of both terms
Cut_T	= Test Cutscore
pVal_i	= pValue of item i
pBis_i	= Point Biserial correlation of item to total test score
t_i	= Response latency in seconds

Efficiency combines both measures of test relevance and an indicator of item response speed. We believe it will be used as a standard measure of items constructed for and administered in conversational tests.