

RICAS IRT Model

STRUCTURE OF THE IRT MODEL

Item Response Theory (IRT) is a collection of psychometric models that are widely used to standardize and model the results of tests that measure student achievement and/or ability. This type of model has been in use in education for at least fifty years.

Usually an IRT model assumes that, for a given student, the probability that they will correctly answer a specific item is a function of an ability parameter (θ) associated with the student together with a set of parameters known as the *IRT parameters* associated with that test item.

The formulas for computing the probability of a correct response given θ and the IRT parameters are documented in the 2018 MCAS and MCAS-Alt Technical Report, Section 3.6.1. [1]

For each item on the test, the IRT parameters are estimated during the test calibration process. For the 2018 RICAS, the IRT parameter estimates are posted on the Massachusetts Department of Elementary and Secondary Education's website. [2].

The θ values are assumed to have a standard normal distribution in the population (i.e., a bell curve distribution with mean zero and standard deviation one).

Dichotomous Items. Dichotomous items receive a score of either one or zero depending on the answer the student gives.

A three-parameter logistic (3PL) model is used for dichotomous items. For each item and student, the probability that the student receives a score of 1 is determined by the values of the three IRT parameters:

- a discrimination
- b difficulty
- c pseudo guessing

in combination with the student ability parameter θ and a normalizing constant D whose value is 1.701.

The formula for computing the probability of receiving a score of 1 on item i given a_i, b_i, c_i and D for student j with ability parameter θ_j is:

$$P_i(1|\theta_j, a_i, b_i, c_i) = c_i + (1 - c_i) \left(\frac{\exp [Da_i(\theta_j - b_i)]}{1 + \exp [Da_i(\theta_j - b_i)]} \right)$$

Example: Item IA00323 from the 2018 ELA Grade 3 test (paper version) has the following estimated parameter values: ([2] p.22)

- a (discrimination) = 1.06404
- b (difficulty) = -0.76098
- c (pseudo guessing) = 0.13650

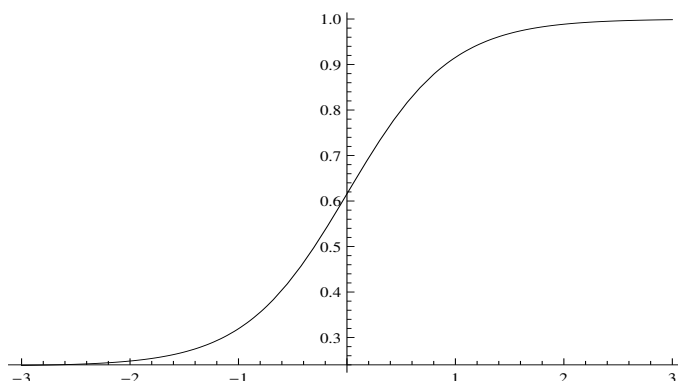
For a student at the median ability in the population ($\theta = 0$), the probability of a correct answer on this item is predicted to be:

$$0.13650 + (1 - 0.13650) \left(\frac{\exp [1.701 \cdot 1.06404(0 + 0.76098)]}{1 + \exp [1.701 \cdot 1.06404(0 + 0.76098)]} \right) = .826$$

For a student at the 25th percentile of the ability distribution ($\theta = -0.67$), the formulas assigns the probability 0.604 to a correct answer.

The difficulty parameter b is low for this item, indicating that it is relatively easy.

As the ability parameter θ varies from -3 to 3 , the probability a correct answer varies from approximately zero to approximately one, along an S-shaped curve similar to the one in the figure:



If you draw a horizontal line segment with length 1 on the graph from $y = 0$ to $y = 1$ at any value of θ on the horizontal axis, the portion

of that line that is above the curve is the probability of an incorrect answer, and the portion below the graph is the probability of a correct answer.

Most of the items on the RICAS are dichotomous, but there are a few that can receive scores other than zero or one. These are called *polytomous* items. These have more than three IRT parameters, and a more complicated probability formula, but the formula still produces a predicted probability for each possible score.

HOW WE MIGHT USE THIS

It is my understanding that RIDE does not currently have a psychometrician on staff, so I think we have an opportunity for East Greenwich to take the lead on using this information.

Using the individual item level scores for every student, it should be possible to produce an estimate of that student's θ value (with a confidence interval), which we could then use to compute the probability of a correct answer on each item.

If a student missed a question that they had a low probability of answering correctly, one could argue that they performed about as expected.

On the other hand, if they miss a question that the model says they should have a high probability of answering correctly, then that could be interpreted to be an indication that they were not adequately prepared on the part of the curriculum the item tested.

Educators could use that information to tailor individual instruction targeting the student's weaknesses.

REFERENCES

- [1] MCAS 2018 NextGen Technical Report: Section 3.6.1, p.54
- [2] Table M-3:Next Generation MCAS and MCAS Alt: IRT Parameters for Dichotomous Items - ELA Grade 3 Paper. MCAS 2018 NextGen Technical Report: Appendix M - Plots and IRT Parameters