

SAT Exam Question: Quality Evaluation

Sample Analysis: Using Item Response Theory (IRT) to Identify Strong and Weak Questions

Prepared by: Joel Artifex, PhD

Prepared for: Smart City Department of Education

Date: January 12, 2022

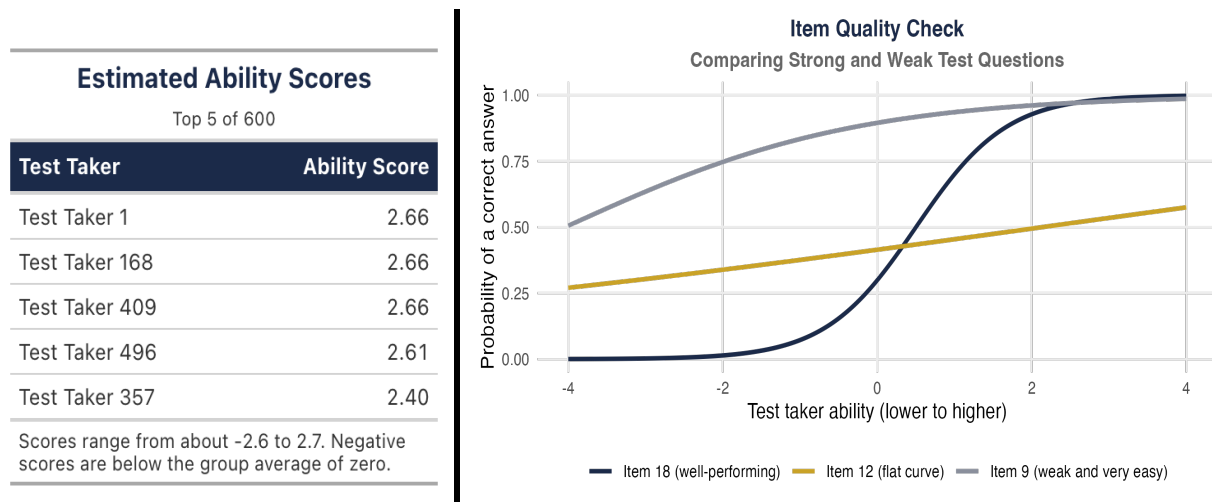
I- Executive Summary

This analysis uses the SAT12 dataset, consisting of 32 multiple-choice questions answered by 600 students, to examine how well each question separates stronger students from weaker ones. Most questions were of good quality, but seven did not clearly distinguish between stronger and weaker students and are candidates for review.

II- Methodology

1PL (Rasch), 2PL, and 3PL item response theory models were fit to the data and compared using model fit statistics (AIC and BIC). The 2PL was selected, given that it offered the best balance of fit and stability.

III- Result and Interpretation



Seven questions (Items 4, 9, 12, 21, 23, 30, and 32) did not clearly distinguish between stronger and weaker students and are candidates for review. Items 9 and 12 are shown above as examples, alongside Item 18 as an example of a well-performing question.

IV- Recommendation

Review the seven flagged questions, starting with Items 12 and 32, which show almost no ability to distinguish stronger from weaker students. The ability score estimates can also be used to group students for intervention implementations or other purposes.