Why Use a Computer Adaptive Test to Assess Students' Reading and Math Abilities?

A computer adaptive test (CAT) functions like an experienced content-area teacher. A teacher may ask the student a few questions to learn about their background knowledge and their instructional needs. They use their content area expertise to select a question of appropriate difficulty. The teacher then evaluates the student's answer and asks another question to refine what they know about the student. This process continues until the teacher concludes they've isolated the student's knowledge and needs. CATs use an algorithm to individualize, or adapt, the assessment for each student as illustrated in Figure 1. A CAT generates the same information as a conventional assessment in just one-third of the time and, thus, is very efficient.

How Does a Computer Adaptive Test Work?

The CAT software uses research-based estimates of item difficulties to select the most optimal items based on the student's response profile during the assessment. In Figure 1, there were 33 assessment items available (blue dots) that span the ability range, but the CAT needed only 10 items to estimate the student's ability.

Item difficulty and student ability are depicted on the bottom axis. In the case of Adaptive Reading (aReading), items and abilities are on a scale of 300 (basic) to 700 (advanced skills and abilities).

The order that items are selected is defined on the left axis. The assessment begins with an item of middle difficulty, 500. The estimate of student ability increases with correct responses and, so, the estimated ability and item difficulty increases.

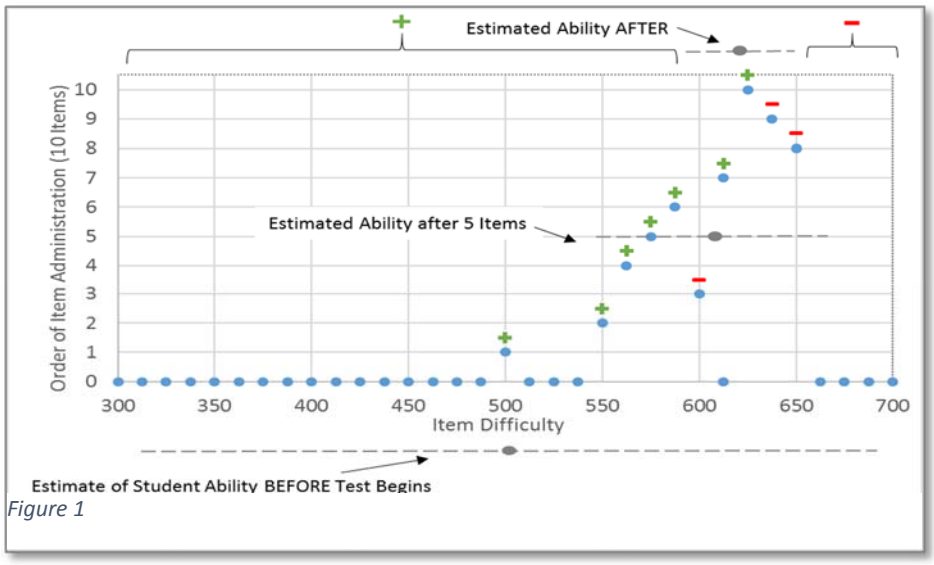

Figure 1

The probability of a correct response is estimated for all of the other items. As shown in the upper right of Figure 1, the estimate of student ability is used to predict likely responses to all items and to generate instructional recommendations. For example, aReading provides specific estimates of each student's reading level, profile (skills that are mastered, developing, or pending), and instructional targets.

Why use Adaptive Reading (aReading) or Adaptive Math (aMath)?

Adaptive Reading (aReading) and Math (aMath) take an average of 15 to 30 minutes to administer, which is **half the time of other CATs with equivalent or improved reliability and validity**.

Assessment items are sampled from more than 3,000 active standards-aligned items, which are continuously refreshed. Research and development of aReading and aMath was funded by the US Department of Education (2006–16).

Efficiency, Reliability and Validity

A CAT adapts to sample student performance in a broad domain. Scores are reliable (stable) and valid (measures what is intended), and the assessment is highly efficient (takes less time) for screening. There is no more efficient way to automate the assessment of performance with similar levels of reliability and validity.

Supplemental Examples

In contrast to Figure 1, Figure 2 illustrates a response profile for a student with lower ability. It can be inferred that students with lower abilities receive less difficult items.

In both Figures, approximately 10 items are required to generate a reliable and valid score. If the assessment did not adapt, then more items would be required. Considering the range of student abilities span -4 to +4, which 10 items would you select if the test were not adaptive? You would need 30 items to span the ability range to get an equivalent prediction of a student's ability.

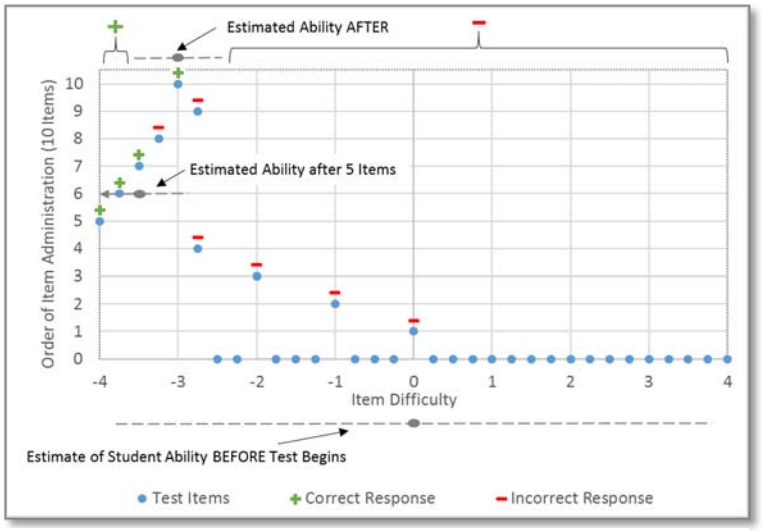

Figure 2

Figure 3 illustrates a non-adaptive assessment, which is familiar to most educators. It has a fixed set of items for all students. This figure illustrates the same response profile of the high-ability student (e.g., Figure 1). An experienced teacher with content-area expertise could infer that the higher ability student would probably get the easy items correct. The CAT functions in a similar way.

Notice, this non-adaptive illustration requires 30 items rather than 10 to cover the ability range. **This non-adaptive assessment has 300% more items and will take 300% longer to administer.**

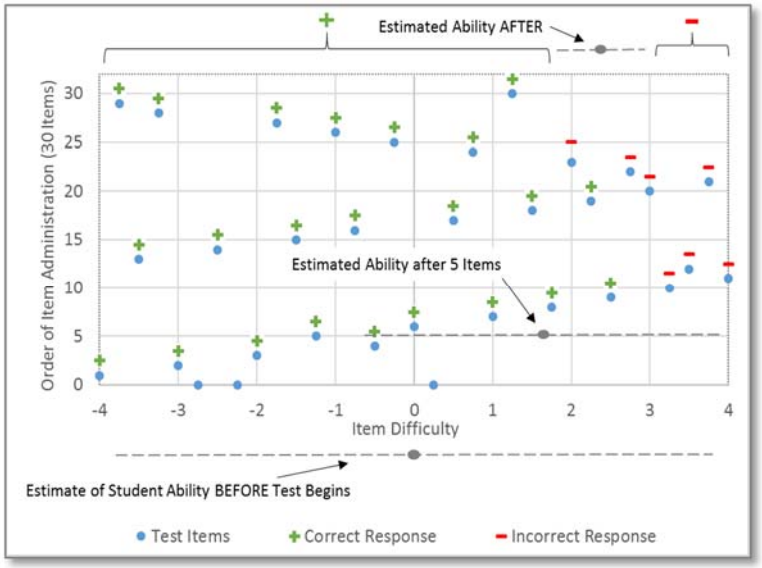

Figure 3