What Is a Standards-Based, Computer-Adaptive Test?

Webinar
Agenda

• Which ILEARN tests will be adaptive and when?
• What is a computer-adaptive test (CAT)?
• What does it mean for a CAT to be standards-based?
• What is the role of the adaptive algorithm in selecting items for each student?
• How does a CAT handle item difficulty and cognitive complexity?
• How are CATs scored?
## Which ILEARN tests will be adaptive and when?

<table>
<thead>
<tr>
<th>Subject</th>
<th>Test Type</th>
<th>Grade(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILEARN ELA/Literacy</td>
<td>Computer-Adaptive</td>
<td>Grades 3–8</td>
</tr>
<tr>
<td>ILEARN Mathematics</td>
<td>Computer-Adaptive</td>
<td>Grades 3–8</td>
</tr>
<tr>
<td>ILEARN Science</td>
<td>Fixed-Form (Year 1)</td>
<td>Grades 4 and 6</td>
</tr>
<tr>
<td>ILEARN Social Studies</td>
<td>Fixed-Form</td>
<td>Grade 5</td>
</tr>
<tr>
<td>ILEARN Biology</td>
<td>Fixed-Form (Year 1)</td>
<td>High School</td>
</tr>
<tr>
<td>ILEARN US Government (optional)</td>
<td>Fixed-Form</td>
<td>High School</td>
</tr>
</tbody>
</table>

Note that tests for ILEARN Science in grades 4 and 6 and Biology will also eventually transition to being computer-adaptive once the item bank has grown large enough to accommodate that transition.
What is a computer-adaptive test?

• In a CAT, each test question is selected by a computer algorithm that is based on the student’s performance on previously administered items.

• A CAT is tailored to the student’s ability based on whether the student responded successfully to the preceding item or sets of items.
  • Targeting test information to the student’s ability increases the precision of resulting test scores, especially for very low- and high-ability students.

• Each time a student answers a question, his or her response helps to determine the next question or set of questions that will be presented to the student.
What is a **standards-based, computer-adaptive test**?

- The goal of a standards-based, computer-adaptive test is to enact a complex blueprint that ensures breadth of coverage of the state’s content standards, as well as the depth of knowledge of those standards as also defined in the standards.
- Within the constraint of matching the blueprint, items are selected to maximize test information at the student’s estimated ability level.
- The difficulty of the test will adjust to each student’s skills, providing a better measure of what the student knows and can do.
- Adaptive tests measure the **same content** for all students on the basis of the test blueprint.
What advantages do CATs have over fixed-form tests?

- **More precise**: CATs more precisely measure the ability of students who struggle and those who excel.
- **More secure**: Maintaining test security on a CAT is easier than on paper-based tests, because the items administered vary from student to student.
- **More engaging**: Because item selection is targeted to the individual’s level of mastery, the items are maximally engaging to students.
How does a CAT work?

The delivery of an adaptive test depends on three things:

- The **test blueprint** defines the characteristics and content of the items that will be delivered.

- The **item pool** represents the sample of knowledge and skills available to measure student achievement of the state standards.

- The **adaptive algorithm** refers to the sequence of procedures that selects the items according to the blueprint.
The Importance of the Blueprint

ILEARN

Measures student performance against a fixed set of criteria or learning standards

Blueprint

Ensures that each test administration is appropriately constructed to make valid inferences about student achievement

Indiana’s Academic Standards

Define what students need to know, understand, and be able to do
The Importance of the Blueprint (cont.)

Test blueprints ensure that:

• Students are administered tests that represent the breadth and depth of the standards;

• overall test score and subscale performance-level reporting is reliable; and

• student test scores and performance levels would be similar across repeated administrations of a similar test.
What is the adaptive algorithm?

- Indiana’s computer-adaptive tests will be administered using a standards-based adaptive algorithm.
- The algorithm is designed to meet a complex set of content constraints and, within these constraints, vary item difficulty (not depth of knowledge) to adapt to the student’s current performance.
How does the adaptive algorithm work?

The adaptive algorithm seeks to optimize all of the following criteria:

• Match to the ILEARN blueprint
• Precision of overall scores
• Precision of content strand or other reporting category scores
How does the algorithm select the first test item?

Test Begins – Students are administered the first item on the test.

• The first time the system encounters a student, it assigns the state mean as the student’s initial ability estimate.
How does the algorithm select subsequent test items?

**Test Continues** – Students are administered additional items.

- For subsequent questions the algorithm first identifies a subset of items that best satisfy the blueprint requirements.
- From that subset the algorithm then identifies a further subset with item difficulties that maximize test information to student ability.
- Finally, it randomly selects from the best items, providing a measure of exposure control.

**Test Ends** – The maximum number of items has been administered.
Cognitive Complexity vs. Difficulty

- Every student has the chance to demonstrate the same range of cognitive skills, even if they receive items of varying difficulty.
- **Difficulty** refers to the likelihood that a student will respond correctly.
- **Cognitive complexity** (also called Depth of Knowledge, or DOK) refers to the mental processes required to meet the demands of a task.
Cognitive Complexity vs. Difficulty

DOK 1
Recall & Reproduction
Who? What? Where? When?

DOK 2
Basic Application of Skills and Concepts
How did it happen? Why did it happen? How does it work? Why does it work that way?

DOK 3
Strategic Thinking
How can you use it? Why can you use it? What is the cause? What is the effect? What is the reason? What is the result?

DOK 4
Extended Thinking
What is the impact? What is the influence? What is the relationship? What if? What would happen? What could happen? What do you think, feel, believe?
Examples of Low-Difficulty Tasks

• Low cognitive complexity:
  An advertisement notes that four out of five dentists recommend sugarless gum. Define the word recommend.

• High cognitive complexity:
  An advertisement notes that four out of five dentists recommend sugarless gum. What is the ad author’s intent?
Examples of High-Difficulty Tasks

• Low cognitive complexity: In what year was the epic poem “The Song of Hiawatha” written?

• High cognitive complexity: Compare how American-Indian characters are portrayed in Longfellow’s “The Song of Hiawatha” and Sherman Alexie’s *The Lone Ranger and Tonto Fistfight in Heaven*. 
How are CATs scored?

• Student performance is reported as a scale score, which is based on
  • the difficulty of the administered test items, and
  • the student’s pattern of correct and incorrect responses to those items.

• CAT item selection is designed to maximize test information near each student’s ability.

• Very high- and very low-performing students may both respond correctly to about 50% of test items.
  • The high achiever is being administered much more difficult items, so they will receive a higher scale score.
How are CATs scored? (cont.)

• Students can and will correctly answer items above their ability, and may incorrectly answer items below their ability.
  • However, the probability of such response patterns decreases as the item difficulty moves away from the student’s ability.

• Performance level is determined from the scale score.

• For ILEARN, educators will set the cut scores associated with proficiency levels when standard-setting takes place.
CAT Myth #1: Students may encounter test items that go beyond the standards or are not grade-appropriate.

Why it’s a myth:

• A standards-based CAT follows the test blueprint
• The blueprint determines whether administration of off-grade items is allowed
• For ILEARN, students will receive only items for their tested grade.
CAT Myth #2: Adaptive tests are longer.

Why it’s a myth:

• A CAT follows the test blueprint, which specifies the number of items to administer.
• ILEARN blueprints will satisfy constraints on testing time delineated by the recent ISTEP+ Panel.
• Note that ILEARN online tests will be untimed.
CAT Myth #3: Students may deliberately answer items incorrectly, ensuring a test administration of easy items, and then go back and provide a correct response to all of the items in order to elevate their score.

Why it’s a myth:

• The student will actually get a lower score.
• Overall test score is determined by difficulty of the items to which the student responded correctly and incorrectly.
• If the student changes responses to be correct, the items are still too easy to support a high test score.
Upcoming Trainings: Spring 2018

• Spring 2018:
  • Understanding Indiana’s New Assessment Webinar
    • May 23, 4:00pm ET
    • June 4, 11:00am ET
    • June 11, 1:00pm ET
  • Why Is It Important to Assess? Webinar
    • May 22, 4:00pm ET
    • June 8, 11:00pm ET
    • June 14, 1:00pm ET
  • Student Interface Training Module
    • Available on the Indiana Portal
Upcoming Trainings: Fall 2018

• Fall 2018:
  • Online Testing System Webinars
  • Assessment Literacy Webinars
  • Online Test Administrator Certification Course

• All training resources will be available on the Indiana Portal
More Information:

Indiana Portal: [http://indiana.portal.airast.org](http://indiana.portal.airast.org)

Indiana Assessment Help Desk
- Toll-Free Phone: 1.866.298.4256
- Email: airindianahelpdesk@air.org

Indiana Department of Education:
- INassessments@doe.in.gov
- 317-232-9050
- Follow us #IDOEILEARN
Thank you! Questions?