

Computer-Adaptive Testing (CAT)

FAQs on aimswebPlus

What is the value of computer-adaptive testing (CAT), and what are its limitations?

CAT approaches improve tests that will be given to many different students with varying abilities, such as in the context of universal screening. Instead of administering the exact same set of questions to all students, the questions on a CAT form are uniquely selected for each individual student. This means that with each question a student answers the next questions delivered will be more and more appropriate for evaluating the student's individual ability level.

The advantages of these approaches do come with limitations. CAT works best for measures that are continuous and computer-administered, as opposed to a performance-based measure which uses a digital record form (DRF) to actively record responses by an examiner. Furthermore, when each student's test is different, teachers and test administrators will not be able to memorize exactly what questions are on a test. This is an advantage for test security, but it makes it harder to draw specific comparisons across students based on single items.

Why are we developing computer adaptive testing versions of Reading Comprehension and Math Concepts and Applications?

Transitioning our Reading Comprehension (RC) and Math Concepts & Applications (CA) measures to CAT is a powerful way to improve these measures because of their purpose and measurement approach. Both measures are computer-administered without a time limit, and they share

the purpose of evaluating a student's ability level in a complex skill. Therefore, CAT versions of these measures can more precisely assess a student's content knowledge and their level of ability within various content areas.

How can I fairly compare student performance if each student's test is unique?

All CAT versions of RC and CA share a common test blueprint based on the design of the existing fixed forms. The blueprints define how many questions on each form assess a distinct skill. Therefore, all students who take a CAT measure will see a similar portion of items assessing each skill of the test's blueprint. While assessing the same skills according to the blueprint, questions will vary for each student because the difficulty of the questions will change according to the student's performance during the test. Since item difficulties will vary between students, it will not be fair to compare students based only on percent correct.

The *raw* score from a CAT is a value called *estimated student ability*. Estimated student ability integrates two parts of a student's performance: item accuracy and item difficulty. Therefore, CAT scores should be interpreted as the level of content knowledge the student has shown at various levels of difficulty. A higher score means a student has likely mastered basic content and is starting to show proficiency with more advanced content. A lower score means a student is only starting to show proficiency with more basic content.

CAT versus Linear Fixed Forms: What's the difference on the aimswebPlus platform?

Once an account is set up to use CAT versions of RC and CA, test administration will be the same as the linear fixed forms. Students log on to TestNav and take the test on the computer just as they would take the fixed form versions of these measures.

Because results from both the CAT and fixed forms are based on the same vertical scale, CAT estimated student ability scores can be converted to the same scaled scores and national percentiles as the fixed form scores. Evaluating growth between tests is also the same, using ROI scores and student growth percentiles (SGPs).

Will there be specific differences in the questions/content of the CAT forms compared to Fixed Forms?

The development of the computer-adaptive forms allows us to add a few additional features to the questions and content we use:

1. New content: Developing our CAT measures involves building a large bank of questions across the test blueprint's skills, knowledge, and abilities. For each RC and CA skill we measure, we added new questions to the list of items present on our fixed forms. This added content gives our CAT measures the flexibility of assessing each part of the test blueprint at various levels of difficulty, complexity, and depth of knowledge.

2. New response styles: The fixed forms are entirely composed of multiple-choice questions. Beyond the benefits of multiple-choice questions, there are skills, abilities, and content that are challenging or impossible to assess with a multiple-choice question. The CAT forms include *technology-enhanced* questions which involve interactive response styles. For example, a CA question may ask a student to identify the location of coordinates by clicking on the specific locations on the graph. Or an RC question may ask a student to recall the events of a story by dragging and dropping the events into the

correct order. (Note for the 2022–2023 school year, 2nd grade CAT forms will not include technology-enhanced questions given the importance of simplifying the testing experience for our younger students.)

3. Passage structure (RC only): On the fixed RC forms, students read 6 passages and answer 4 questions for each passage. RC CAT will be organized around 4 passages. The first two passages will be of the same length and structure as the current fixed test and delivered with 4 questions. The latter two passages will be slightly longer and will be presented with 8 questions.

- The maximum length of these longer passages changes by grade. For 2nd grade, the passages are all very similar in length. By 8th grade, the difference in passage length between short and long texts will be more noticeable.
- Even with these changes, we still maintain the same design of the fixed forms where half of the passages are literary (fictional) stories or poems, and half are informational (nonfictional) texts.

Can we select CAT for certain schools and continue to use fixed form for others?

The choice to use CAT is a setting at the account level. Whether an account represents multiple schools or just one school, everyone under the same account needs to have the same setting.

Can we select CAT for Reading Comprehension but fixed form for Math Concepts & Applications (or vice versa)?

Yes.

How long should we estimate the testing to take for each CAT measure?

The time to complete the CAT versions of the RC and CA measures should be consistent with the amount of time it takes to complete the fixed versions of these measures.

Is there a set number of questions on each CAT measure that students take?

Yes. The numbers of questions will be the same as the fixed forms for the 2022–2023 school year. The only unique case where test length is different between CAT and fixed forms is for the 2nd Grade RC. The CAT form of the test will be two questions shorter (22 questions) than the fixed forms (24 questions). By conducting additional research on the test's performance, we will explore whether we can reduce the length of the tests and maintain a robust level of validity and reliability for all grades.

Can we pilot the CAT with select students next year?

The choice to use CAT or fixed forms is decided at the account level and cannot be specified differently for specific students. Once an account starts testing students in a benchmark season, the system will be set to use the chosen version of RC or CA for the rest of the benchmark period.

Can we access a demo account to see what the changes look like before deciding on whether we will make the switch to CAT?

Demo accounts will not be available until CAT Forms are available August 1st. In July, we will release a series of screenshots with descriptions detailing the similarities and differences between choosing CAT or linear. In addition, our training team will show a demo of all BTS features--including CAT--on our monthly user Forum on July 13th. Final decisions about which form to use can be made up until Fall testing starts.

We understand that the application will default to the CAT approach for these two measures for the 2022–2023 school year. If we do not want to use the CAT, how do we change it back to the fixed form approach?

We will be sending out notifications and instructions prior to back-to-school 2022 (August 1) to ensure that all customers who wish to keep using the fixed forms know exactly how to set their account accordingly.

How will students be monitored in Reading Comprehension if the CAT is administered for the benchmark?

CAT RC is designed to be an improved/alternate version of the currently fixed form RC benchmarking measure. Therefore, it is intended to be used for benchmark screening three times a year. Existing progress monitoring measures are not impacted by this new option of CAT.

Do we still need to use the survey level assessments (SLAs) for students falling well below grade level if we transition to CAT Reading Comprehension?

Perhaps. The CAT forms of RC and CA are designed to improve the testing experience of students who have fallen below grade level because easier questions will be selected more often to evaluate their developing skills. This can improve the testing experience, but the student's final score will still be evaluated according to the norms (e.g., national percentile) for the assessed grade level.

Sometimes it is not valuable to know how poorly a student is doing relative to their peers, and it is more valuable to know where the student stands relative to students one grade-level below. In these cases, it will still be necessary to use SLA and give the student the test one grade level lower. (Depending on 2022-2023 research results, we are planning to further develop our CAT forms

to support off-grade level norms using on-grade level forms.)

How will the aimswebPlus reports be different? Will student performance be reported in percentiles as they are now or in grade level bands?

On the Benchmark Comparison screen or the Student Profile screen, test scores will be reported in the same way for CAT forms as fixed forms. CAT forms were designed to be “anchored” to the fixed forms. This means that both tests report the same scaled score and use the same norms to report the student’s national percentile score.

On the Math Skills Plan, Reading Skills Plan, and Math Skills Analysis reports, there will be differences. When a student takes a computer-adaptive test, the total number of questions answered correctly (or the percentage correct) does *not* tell you the most important information about the student’s ability. Instead, CAT reports use the *estimated student ability* score which integrates the student’s accuracy and the difficulty of the items. On the Skills Plan and Skills Analysis reports, bar charts break down the student’s overall score into separate *student ability scale scores* ranging from 0 to 100. Just like percentages seen on reports for Fixed forms, student ability scores indicate how well students are performing in different categories of the test. Very high scores tell you that the student is mastering content that is very advanced and too difficult for most peers. Very low scores tell you that the student is still struggling with basic content

that is commonly very easy for peers. A score of 50 tells you that the student is performing right in the middle relative to their peers, proficient with easier content but not yet mastering advanced content.

Can you give me a more specific example of how scores on these individual reports will be different?

On the Reading Skills plan report, the bar chart at the top of the report indicates how well a student is doing with reading comprehension when passages are either fictional literary stories or nonfiction informational texts. For fixed form reports, performance is reported with a bar chart showing the percentage of questions answered correctly when the passage was literary or informational. For CAT forms, these reports use the same bar charts to give the same information but with a different score. The bar charts present the CAT estimated student ability score. This value combines item accuracy and item difficulty. A very high student ability score (max 100) indicates that a student answered easy, medium, and difficult questions with good accuracy. The interpretation is that the student has shown proficiency with very advanced content and their mastery of grade-level content is far above average. Very low student ability scores indicate a student may have answered some easy (basic) questions correctly but did not show reliable proficiency with average difficulty content. The interpretation is that student mastery of grade-level content is far below average.

To learn more and order, visit
[PearsonAssessments.com/aimswebPlus](https://www.pearsonassessments.com/aimswebPlus)

866-313-6194 | [PearsonAssessments.com](https://www.pearsonassessments.com)