

Computer Adaptive Testing Design

Grade 9 Assessment of Mathematics

Technical Summary

What is the new design for the assessment?

EQAO is introducing a new modernized platform for the Grade 9 Assessment of Mathematics, where students will complete a multi-stage computer adaptive test (msCAT). This form of computer testing adapts to a student's level of proficiency according to the student's achievement on a set of items (called a *module*). Due to their adaptive nature, msCATs meet students where they are, by presenting modules that are challenging but also aligned with the students' individual proficiency in mathematics.

The following table provides a snapshot of the key aspects of the digitalized Grade 9 Assessment of Mathematics.

Snapshot of Digitalized Grade 9 Assessment of Mathematics	
Delivery	Multi-stage computer adaptive test (msCAT)
Sessions	Two sessions
Time	60 minutes to complete each session ¹
Items	50 machine-scored items (operational)
Window	Flexible (depending on the academic term used: e.g., semester, quadmester, octomester)
Framework	Two-dimensional framework based on strands, overall expectations and skill categories
Report	Automated

How will this new assessment design be delivered?

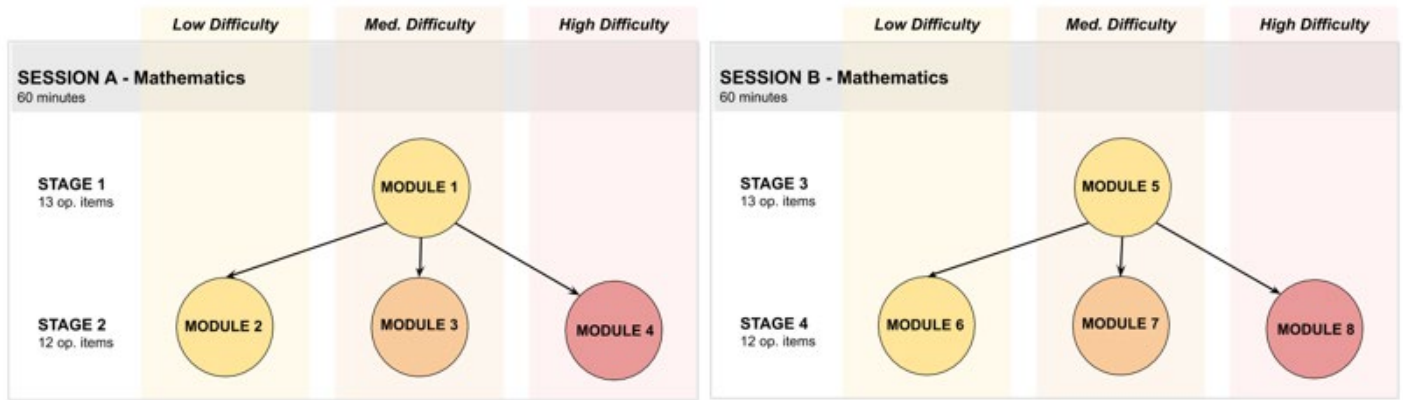
The new Grade 9 mathematics msCAT is composed of two 60-minute sessions (Sessions A and B) that students can attempt one after the other or at different times.

Students begin Session A by attempting a module that consists of items of medium average difficulty (Module 1 of Stage 1 below).² Depending on their achievement on this first module, students are presented with a new module in Stage 2 of low (Module 2), medium (Module 3) or high (Module 4) average difficulty.

In Stage 3, students start fresh, again attempting a module of medium average difficulty. Depending on their responses, students are presented with a new module of differing average difficulty in Stage 4. Once finished, students receive their results (i.e., Level 1, 2, 3 or 4) based on their achievement through the test.

¹ Additional time can be granted to any student requiring it.

² Stages 1 and 3 contain medium-difficulty modules. These modules consist of low-, medium- and high-difficulty items but are on average of medium difficulty.



What results might students taking this new assessment receive?

When Modules 1 and 5 are completed, the student's ability will be computed. Preset cut-points determine the average difficulty of the next module presented to the student: low, medium or high.

Depending on the pathway the student takes through the stages of the test, different results are possible. To illustrate this, the table below shows the average likelihood of achieving different levels on each possible pathway. For example, if a student is presented with a medium-difficulty module in Stage 2 and a high-difficulty module in Stage 4, it is very likely that this student will receive a Level 3. The table below outlines possible outcomes based on session results. These outcomes have been demonstrated using simulation studies.

Session Outcomes		Average Likelihood of Achieving Level				
Session A	Session B	Below Level 1	Level 1	Level 2	Level 3	Level 4
Low	Low	20%	45%	30%	5%	0%
Low	Medium	0%	15%	80%	5%	0%
Low	High	0%	15%	70%	10%	5%
Medium	Low	0%	15%	80%	5%	0%
Medium	Medium	0%	0%	35%	65%	0%
Medium	High	0%	0%	10%	75%	15%
High	Low	0%	15%	70%	10%	5%
High	Medium	0%	0%	10%	75%	15%
High	High	0%	0%	0%	40%	60%

How has EQAO validated this new assessment design?

In preparation for transforming Ontario's approach to large-scale assessments, EQAO worked with national and international assessment, measurement and technology experts. This engagement has helped validate key aspects of the transition to online adaptive assessments.

EQAO also conducted a simulation study to evaluate the effectiveness and reliability of the msCAT delivery model for the modernized Grade 9 Assessment of Mathematics. Using previous EQAO data, the study simulated student responses and achievement on the new msCAT design. The study validated the adoption of a four-stage msCAT delivery model with a total of 50 items for each test. It also allowed for the use of a simplified psychometric model (1-parameter Item Response Theory) to estimate student achievement with a high level of accuracy, leaving aside statistical calculations that account for guessing and other complexities that have proven difficult for schools to work with in the past.