EQAO’s Board of Directors

Dr. Brian L. Desbiens, Chair
June 1, 2005–March 16, 2014

Jerry Ponikvar, Vice-Chair
July 4, 1997–May 16, 2014

Roland Boudreau
June 2, 2011–June 1, 2014

Hélène Chayer
June 25, 2008–May 11, 2013

Dave Cooke
June 25, 2008–May 11, 2013

Dr. Dieudonné Detchou

Dr. Elizabeth (Lee) Ford-Jones
June 2, 2011–June 1, 2014

Abirami Jeyaratnam
October 3, 2012–October 2, 2014

Dr. Bette M. Stephenson, Director Emeritus

EQAO’s Executive Team

Marguerite Jackson, Chief Executive Officer

Marianne Mazzorato, Chief Assessment Officer

Richard Jones, Director, Assessment and Reporting

Michael Kozlow, Director, Data and Support Services

François Lavictoire, Executive Coordinator, French Language

Mary Low, Executive Coordinator, Operations

Tony Saini, Director, Corporate and Public Affairs (Acting)
What Is the Quality of EQAO Assessments?

W. Todd Rogers, Scholar-in-Residence, EQAO, University of Alberta

June 2013
The processes EQAO uses to develop and score its assessments conform to well-recognized standards of educational testing and provide valid and valuable information for Ontario’s education context. The agency’s steadfast commitment to ensuring the tests reflect the curriculum accurately, provide relevant results, and are comparable from year to year means Ontarians can be confident the results are a reliable, independent measure of student achievement over time.

Dany Laveault, University of Ottawa
In response to the Royal Commission on Learning’s call for province-wide testing in 1994, the Ontario Government created the Education Quality and Accountability Office (EQAO) as an independent Crown agency in 1996. The Commission proposed that system-wide testing be conducted in Ontario to check student learning at a few critical transition points and to assure the public that there was at least one common measure of achievement for all students at these points.

EQAO develops, administers, and scores province-wide assessments of student achievement each year in both of Canada’s official languages. The results, including the change in student performance from one year to the next, are reported to the Ontario government, school boards, schools, students, parents/guardians, and the general public. EQAO’s province-wide assessments measure the cumulative knowledge and skills students have acquired in three subject areas at four key points of their journey through school:

- Reading, writing, and mathematics up to the end of the primary division (end of Grade 3);
- Reading, writing, and mathematics up to the end of the junior division (end of Grade 6);
- Mathematics up to the end of the first year of secondary school (Grade 9); and
- Literacy (reading and writing) up to the end of the first year of secondary school (Grade 10).

Approximately 1,100,000 tests are administered to about 580,000 students in approximately 4,300 English-language and French-language schools in the province each year.
How Good Is EQAO at Doing What It Was Charged to Do? Very good!
EQAO develops items, scores student responses, analyzes the scores, and reports the results in ways that meet standards for educational testing; provides trustworthy and useful information for the Ontario educational system; and helps and encourages school officials to work with the results for their school board and schools to enhance student learning. EQAO procedures follow the Principles for Fair Student Assessment Practices for Education in Canada (1993). This document is widely endorsed by Canada’s education communities and is used to guide classroom and large-scale assessments. EQAO also follows the Standards for Educational and Psychological Testing (American Educational Research Association, National Council on Measurement in Education, & American Psychological Association, 1999), which is used to implement and evaluate testing programs world-wide.

EQAO has an Assessment Advisory Committee comprising school trustees, school board staff, principals, teachers, parents, and students. This committee meets twice a year and provides a forum for its members to share expertise and advise EQAO regarding English- and French-language assessment activities. Members pass on EQAO information to their constituent organizations and represent their organizations when providing EQAO with input or raising concerns. Members also provide advice on the best ways to communicate assessment information and results to parents, students, teachers, school boards, the media, and the general public in order to promote the use of the results to improve student learning.
EQAO also has a Psychometric Expert Panel. Its members are recognized national and international experts in large-scale testing. The panel meets twice a year to review the agency’s assessment procedures and confirm current procedures or suggest changes to ensure that EQAO’s procedures are current and correctly applied. EQAO also retains scholars-in-residence, who act as senior measurement specialists. They develop a deep understanding of the agency’s work; mentor the staff; review items, analyses, and reports; provide advice about EQAO research; and provide objective feedback about EQAO’s procedures throughout the year.
Seeking information from all of the above-mentioned groups on a regular basis makes EQAO unique in Canada, and helps ensure that EQAO assessment results are meaningful and can be trusted and used with confidence.

Before proceeding further, it is necessary to point out that there is potential for error whenever we measure. For example, how many of us have cut a piece of wood either too short or too long or hemmed a dress either too long or too short? What did we do just before we made the cut? We measured, but made an error. There are actually two kinds of measurement error. One type is random error, which we would get if we had many people measure the same piece of wood. Not everyone’s measurement would be exactly the same. The second type is constant or systematic error, which we would get if the tape measure began at 5 centimetres and not at 0 centimetres. All who used the tape measure would measure the length about 5 centimetres too long (for example, 96 centimetres instead of 91 centimetres). Random error affects the precision or reliability of the observed measurements. Constant error affects the accuracy of the measurements, which leads to an invalid or incorrect interpretation of the observed measurements.

The possibility of both types of measurement error exists every time we measure student achievement. Random error affects the reliability of the measurements, which leads to imprecise interpretations of students’ observed scores. Constant error affects the accuracy of the measurements, which leads to an invalid interpretation of students’ observed scores. It is important to note that EQAO employs best practices that minimize both types of errors. That is, the results derived from the tests administered each year are precise (reliable) and can be accurately (validly) interpreted in terms of what students in Ontario know and can do.
What Does EQAO Assess?
What Does EQAO Assess?

Ontario students are expected to acquire the learner expectations set by the Ontario Ministry of Education (www.edu.gov.on.ca). Learner expectations describe the knowledge and skills that students are expected to acquire, demonstrate, and apply in their class work.

The learner expectations that each EQAO assessment covers are listed below.

- The Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions measure how well Grades 3 and 6 students have met the reading, writing and mathematics learner expectations outlined in *The Ontario Curriculum, Grades 1–8: Language* (revised 2006) and *The Ontario Curriculum, Grades 1–8: Mathematics* (revised 2005).

- The academic and applied versions of the Grade 9 Assessment of Mathematics measure how well students have met the learner expectations for Grade 9 outlined in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005) for the academic and applied courses.

- The Ontario Secondary School Literacy Test (OSSLT) assesses Grade 10 students’ literacy skills based on the reading and writing learner expectations across all subjects in *The Ontario Curriculum*, up to the end of Grade 9.

EQAO uses the same *test specifications* for each assessment across years so that the assessments across years measure the same knowledge and skills. Maintaining this consistency allows EQAO to measure changes in student performance from one year to the next. Test specifications guide the development of items that measure the knowledge and skills students are to learn. They also specify the numbers of multiple-choice items (students select their answers) and open-response items (students produce their answers) so that clusters of learner expectations for each assessment are covered consistently across years. A cluster of learner expectations is a group of expectations related to a common topic or activity, such as interpreting explicit information in a passage of text or adding and subtracting two-digit numbers. Not all of the expectations in a cluster can be measured in any one assessment, because it would take too long. However, across two or three years, all learner expectations in a cluster are measured, except for learner expectations that cannot be measured with a paper-and-pencil assessment (e.g., reading out loud, speaking, collecting data for a survey).
Primary and Junior Specifications

**Reading items** measure three knowledge and skills categories specified in the language curriculum for the primary and junior divisions: *explicit information, implicit information,* and *connections.*

- Explicit information items require students to detect and understand information and ideas stated explicitly in reading selections that represent a variety of text types (e.g., stories, poems, graphs, diagrams).
- Implicit information items probe students’ understanding of implicitly stated information and ideas in reading selections that represent a variety of text types.
- Connections items require students to demonstrate their understanding of reading selections by comparing and contrasting the ideas presented in texts and drawing upon their own knowledge and experience gained from other texts they have read and the world around them.

**Writing items** measure use of *conventions and topic development* for the different forms of writing specified in the language curriculum for the primary and junior divisions:

- Conventions items measure students’ ability to recognize and use correct spelling, subject-verb agreement, capitalization, and punctuation.
- Topic-development items measure students’ ability to organize ideas and provide supporting detail and evidence.

**Mathematics items** are referenced to five content areas, or strands, in the mathematics curriculum for the primary and junior divisions: *number sense and numeration, measurement, geometry and spatial sense, patterning and algebra,* and *data management and probability.*

- Measurement items require students to use area relationships, understand the dimensions of the shapes needed to calculate their areas, and work with metric area units.
- Geometry and spatial sense items require students to identify, describe, and perform geometrical transformations; identify angles; and use rulers and protractors accurately.
- Patterning and algebra items require students to complete or “grow” patterns and to use diagrams, tables, and number sequences to represent the stages of patterns.
- Data management and probability items require students to represent and predict the probability of an outcome; compare probabilities using common representations (e.g., common denominators, percents, or decimals); and interpret graphs.

Mathematics items are also referenced to three knowledge and process skills in the provincial mathematics curriculum for the primary and junior divisions: *knowledge and understanding, application,* and *thinking.*

- Knowledge and understanding items require students to demonstrate subject specific content (knowledge) and comprehend its meaning and significance (understanding).
- Application items require students to select an appropriate mathematical tool and apply the appropriate information.
- Thinking items require students to select and sequence a variety of tools to solve a problem and demonstrate a critical-thinking process. That is, to answer these items, students first need to make a plan.
Grade 9 Specifications

Two mathematics courses—academic and applied—are offered in Grade 9. Separate assessments are developed for these two courses.

Mathematics items for the academic course are referenced to four content areas, or strands, specified in the academic mathematics curriculum for Grade 9: number sense and algebra, linear relations, analytic geometry, and measurement and geometry.

- Number sense and algebra items require students to operate with exponents, manipulate expressions, and solve equations.
- Linear relations items require students to use data management procedures to investigate relationships, understand the characteristics of linear relations, and connect various representations of linear relations.
- Analytic geometry items require students to investigate the relationships between the equation of a relationship and the shape of its graph and the properties of slope, and to use the properties of linear relations to solve problems.
- Measurement and geometry items require students to determine the optimal values of measurement; solve problems involving perimeter, area, surface area, and volume; and investigate and apply geometric relationships.

Mathematics items for the applied course are referenced to three content areas: number sense and algebra, linear relations, and measurement and geometry.

- Number sense and algebra items require students to solve problems involving proportional reasoning, simplify expressions, and solve equations.
- Linear relations items require students to determine characteristics of linear relations; investigate constant rate of change; connect various representations of linear relations; and solve problems using the representations.
- Measurement and geometry items require students to determine the optimal measures of rectangles for a given condition; solve problems involving perimeter, area, and volume; and investigate and apply geometric relationships.

Items in both courses are referenced to the same three knowledge and process skills as are those for primary and junior mathematics: knowledge and understanding, application, and thinking.

OSSLT Specifications

The reading and writing expectations for the Grade 10 OSSLT are the same as the reading and writing expectations for the primary and junior reading and writing assessments, but are referenced to reading and writing learner expectations across all subjects in The Ontario Curriculum up to the end of Grade 9. Reading items measure students’ understanding of explicit and implicit information and ideas in a variety of text types required by the curriculum and students’ ability to make connections between what they read and their own personal knowledge and experience. Writing items require students to organize main ideas and to provide supporting details and evidence and use correct spelling, grammar, and punctuation in the different forms of written communication required by the curriculum.
How Are Items Developed?
How Are Items Developed?

Separate assessments are developed for English- and French-language students. New items for each assessment are carefully developed and reviewed using the following eight steps for both languages:

1. An item-writing committee is established for each assessment. Each committee has 10 to 20 teachers and principals who are selected because of their strong knowledge about the students to be assessed, familiarity with *The Ontario Curriculum*, and experience in writing instructional or assessment materials for students.

2. Before each writing session, committee members take a training workshop on how to develop stand-alone multiple-choice items and multiple-choice items that refer to a reading selection, a chart, or a data table. They also learn how to develop prompts or tasks for open-response items as well as scoring rubrics for scoring students’ responses to these items. When developing items, writers must follow these rules:
   - i. Items must be based on *The Ontario Curriculum* and clearly refer to the learning expectations for the assessment.
   - ii. Items must be evidence-based:
     - Given the purpose of the assessment, what content knowledge and skills are both useful and interesting to claim about examinees?
     - What is the reasonable and observable evidence in student work or performance required to support the claims?
   - iii. Items must be age and grade appropriate and contain accurate information that does not rely on students possessing subject-specific content knowledge beyond the grade level for the assessment or on students possessing knowledge beyond a reasonable level of personal experience.
   - iv. Items must allow the widest range of students with special education needs and students whose first language is neither English nor French to respond in meaningful ways. For example, unfamiliar words, idioms, and idiomatic expressions should be avoided unless they are an integral part of the item. To the greatest extent possible, items should be in a form that allows accommodations to be made, as needed, for students with special education needs and English or French learners.
   - v. Distractors (wrong answers) for multiple-choice items must be plausible; writing prompts must allow a range of responses; scoring rubrics for writing prompts for open-response items must be clearly related to the writing prompts and effectively capture the range of student responses.
   - vi. Pictures, graphs, and tables must be appropriate for the item and assist students who may be unfamiliar with the topic measured by the item.
   - vii. Standard Canadian spelling must be used; sentence structure must be age and grade appropriate. Technical terms, if used, must be explained in the context of the item.
   - viii. Items must be free of content that may offend or unfairly penalize students on the basis of gender, where they live, ethnicity, culture, religion, or socio-economic status. For example, items, reading passages, diagrams, and charts must avoid favouring one gender; avoid stereotypical gender roles, specific celebrations and religious references; and avoid suggesting that belonging to one socio-economic group is more advantageous than belonging to another.
After training, item writers are assigned to specific areas of the curriculum so that the test specifications for the assessment are fully covered. They write items with varying difficulty and that discriminate less able students from more able students for the areas to which they are assigned.

EQAO assessment specialists carefully review the newly constructed items and scoring rubrics to ensure that each item is stated clearly and completely, there is one correct answer for each multiple-choice item, each scoring rubric clearly reflects the range of student responses to the corresponding prompt, and each item is clearly referenced to the correct expectation. Revisions are made as needed.

Members of the item-writing committee try out revised versions of their own items with their own students in their classes. They work with their students individually and have each student think out loud so that they can determine whether the student understands each item and how the student got his or her answer.

EQAO assessment specialists revise the items as needed using the results from the item tryouts.

The items for each assessment are then reviewed by an external reviewer and two committees. Additional Ontario teachers and principals serve on the Assessment Development Committee for each assessment (e.g., one committee for Primary Reading and Writing and a second committee for Primary Mathematics) and a Sensitivity Committee, which reviews all of the assessments.

EQAO assessment specialists revise the items using the Assessment Development and Sensitivity Committees’ suggestions.
Each Assessment Development Committee has between 8 and 12 members who have the same qualifications as the item writers. The committee members first independently review all newly developed items, reading passages, writing prompts, scoring rubrics, and mathematical tables and figures to confirm that each clearly meets item-writing criteria i–iii and v–vii, presented at the beginning of this section. The Assessment Development Committee then meets as a whole to discuss the reviews and make recommendations for including, excluding, or revising each item.

The Sensitivity Committee has between 8 and 12 members who are selected for their expertise in the areas of diversity, culture, special education, and second-language acquisition, and the members consider all the assessments. They independently review all multiple-choice items, reading passages, writing prompts, and mathematical tables and figures to confirm that each clearly meets item-writing criteria iv, vi, and vii, presented at the beginning of this section. The Sensitivity Committee then discusses these reviews and suggests including, excluding, or revising each item.

Clearly, EQAO employs current best practice for item development. Evidence-based design and universality are stressed, which allows assessment of the widest possible range of students. EQAO’s detailed procedures for item construction, item tryouts, and revising and reviewing items produce the most effective items possible, which is clearly commendable.
How Is Field Testing Conducted?
How Is Field Testing Conducted?

The final versions of the new items are field tested on samples of students. Except for new writing prompts that require long answers, these items are field tested each year as part of the regular, or operational, assessments. Different sets of field-test items are inserted among the operational items in the current year’s assessment booklets. Usually, there are about 20 sets of field-test items for each assessment, resulting in about 20 different student booklets for each assessment. Each booklet contains the same operational items but a different set of field-test items. The field-test items in each booklet take about 20 percent of the administration time to complete. The operational items, which were field tested the year before, match the test specifications and have desirable measurement properties. Student scores are determined only from the operational items and not from the field-test items.

Since the field-test items are similar to the operational items, students do not know which are operational items and which are field-test items. This solves the problem of students not trying hard enough to answer the field-test items, which is often observed in some field-testing models.

Student booklets for an assessment are distributed in a way that ensures the sample for each booklet scientifically represents the entire student population in Ontario. Thus the samples for the field-test items contained in the booklets also represent the entire provincial student population. Therefore, the analysis of the student responses to the field-test items provides trustworthy results.

As mentioned above, new writing prompts that require long answers are not field tested in the administration of the operational assessments, due to the time needed for the students to formulate and write their responses. Instead they are pilot tested separately. (Pilot tests are similar to field tests, but are less formal.) The purpose of these pilot tests is to ensure that students with varying backgrounds can respond to the long writing prompts and that the prompts generate responses at each score point in the scoring rubrics for these items. While the number of pilot-test schools is small, students in the pilot-test sample clearly represent the range of student responses in the province.

EQAO’s field testing procedure is unique in Canada and reflects best practice that allows equating, the strongest procedure for linking assessments across years. The process of equating is discussed later in this document.
How Are Operational Items Selected?
How Are Operational Items Selected?

The data for the field tested items are used to answer the following questions:

- Are students attracted to all options for each multiple-choice item?
- For the open response items, do students respond to the prompts as expected and do their responses fall into each of the open-response scoring-rubric categories?
- Is there a range of difficulty (i.e., a spread from easy items to more difficult items)?
- Does each item have adequate discrimination (i.e., separate students who perform well on the assessment from students who perform less well)?
- Does each item provide adequate information about student performance, thereby contributing to the reliability of the assessment?

The answers to these five questions are then used to select the final items to be used in the following year’s operational assessments. This means that the operational items on the following year’s assessments have statistical properties that meet the criteria for good test items. It also means that there is a way to equate next year’s assessments to the current year’s assessments.

In addition, the wording and statistical properties of items selected for the next year’s assessment should be identical in the field-test version and the operational version, in order to allow equating.
Stand-alone items (e.g., a single mathematics item), items that refer to reading selections, and writing prompts for the new operational test form for each assessment are selected in the following back-and-forth manner between EQAO assessment specialists and analysts.

1. The assessment specialist for each assessment considers the information provided by the analysts on each field-test item (how difficult the item is, how well the item distinguishes stronger students from less strong students, the range of difficulty) and the target test information function (which characterizes how reliable the scores are at different points in the distribution of scores). The target test information function is the same across years so as to be able to measure change in student performance between one year and the next. Using the target test information function also ensures a high degree of accuracy and precision when placing students in the achievement categories (e.g., Level 1, 2, 3, or 4 for the primary, junior, and Grade 9 assessments; pass/do not pass for the OSSLT). The assessment specialist constructs an operational form that best fits the target test information function subject to four conditions:

   - the items in the operational form best represent the set of learner expectations for the assessment,
   - the items distinguish well between stronger and less strong students,
   - there is a good range of item difficulty, and
   - the difficulty of the new operational form is close to the difficulty of the operational form administered in the previous year.

2. The analyst for each assessment reviews the initial operational form and computes a number of statistics to evaluate its psychometric properties. The analyst also constructs a test information function to determine if the scores around each cut score (the score used to define the low end of an achievement category) are sufficiently precise to allow accurate and precise placement of students in the achievement categories. Based on a review of these analyses, the analyst may suggest changes to improve the fit of the test information function while maintaining proper coverage of the learner expectations.

3. The assessment specialist and analyst discuss the suggested changes and revise the form.

4. The second and third steps are repeated until the final operational form is created.

Once there is agreement on the final operational form, the Psychometric Expert Panel carefully compares the difficulty, discrimination, and test information function for the new operational form and the previous year’s operational form for each assessment to ensure that the new form satisfies the conditions for equating. This selection process is the same for all EQAO assessments.

The type and number of items in each operational form are the same for both the English-language and French-language assessments (Table 1). Once the operational form has been fixed and the field-test items have been inserted into each student booklet, directions are added that are clear, complete, and age- and grade-level appropriate. A unique bar code that identifies each student is added to each booklet so that students’ responses can be accurately tracked through administration, scoring, analysis, and reporting.
**Table 1: Type and Number of Operational Items in EQAO Assessments**

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Type of item</th>
<th>Multiple-choice</th>
<th>Short answer</th>
<th>Long answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Open-response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td>26</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td>8</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td>28</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Junior</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td>26</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td>8</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td>28</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Grade 9 Mathematics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td></td>
<td>28</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Applied</td>
<td></td>
<td>28</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Grade 10 OSSLT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td>31</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td>8</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

Operational items that have been used in previous assessments are available at [www.eqao.com](http://www.eqao.com) for parents to see the kinds of items administered to their children. These are also available for teachers to use in their classrooms.
How Are EQAO’s Questionnaires Constructed?
How Are EQAO’s Questionnaires Constructed?

EQAO develops and administers Student Questionnaires for all of its assessments, Teacher Questionnaires for all but the OSSLT, and Principal Questionnaires for the primary and junior assessments. The information obtained from these questionnaires is used to provide a context in which to interpret the assessment results and is used in research designed to identify factors that are related to student learning. EQAO is one of the few testing agencies that administers questionnaires and that has an active, ongoing research program designed to investigate better ways to improve student learning.

The following nine-stage procedure is closely followed to develop each questionnaire:

1. EQAO staff conducts a literature review to identify factors that have been shown to influence student learning.

2. A committee of EQAO assessment specialists and analysts choose factors to be considered, taking into account the amount of time available for students, teachers, or principals to respond to their questionnaires.

3. EQAO’s Research Priorities Committee reviews the factors and makes modifications as needed.

4. The committee of EQAO assessment specialists and analysts identifies items from other questionnaires and develops new items for the selected factors and assembles the items together to form the questionnaire.

5. Two to three outside researchers in the area of school improvement review the questionnaire to determine the breadth of coverage and item clarity.

6. A committee of teachers and principals reviews the questionnaire for clarity and to be sure that students, teachers, or principals will have the information needed to respond.

7. The questionnaire is pilot tested on samples of students, teachers, or principals.

8. The committee of EQAO assessment specialists and analysts revises the questionnaire based on the results of the two reviews and the pilot tests.

9. EQAO’s Research Priorities Committee reviews the revised questionnaire.
The questions included in the questionnaires differ according to the nature of the curriculum for the students at different stages in school. What is measured by the different questionnaires is described below, beginning with the primary and junior questionnaires.

**Primary and Junior Questionnaires**

- The Student Questionnaire includes questions about students’ attitudes toward reading, writing, and mathematics; their perceptions about their abilities in reading, writing, and mathematics; and behaviours in and out of schools.

- The Teacher Questionnaire includes questions about classroom characteristics; access to and use of resources for teaching reading, writing, and mathematics, and assessing student performance in these subjects; teacher collaboration; parental engagement; use of EQAO resources and data; and teacher characteristics (e.g., background, experience, qualifications, professional development).

- The Principal Questionnaire includes questions about the school’s characteristics; school improvement planning, staffing and support; staff collaboration; student attendance; parental engagement; use of EQAO resources and data; and principal characteristics (e.g., gender, experience, qualifications, professional development).

**Grade 9 Questionnaires**

- The Student Questionnaire includes questions about student attitudes toward mathematics and their perceptions about their performance in mathematics, use of technology in the classroom, student attendance in mathematics classes, time spent on mathematics homework and on extracurricular activities, parental engagement, future expectations, use of EQAO mathematics results as part of students’ final course marks, and student characteristics (e.g., age, language spoken, number of schools attended).

- The Teacher Questionnaire includes questions about classroom characteristics, access to and use of resources for teaching mathematics and assessing student performance in mathematics, instructional practices, teacher collaboration, school improvement planning, parental engagement, use of EQAO resources and data, use of EQAO Grade 9 mathematics results as part of students’ course marks, and teacher information (e.g., background, experience, professional development).
OSSLT Questionnaire

- The Student Questionnaire asks students about access to a computer at home, the amount of time spent reading in English or French outside school and the different types of materials read outside school, access to reading materials, time spent writing in English or French outside school, and the language spoken at home.
How Are Test Booklets and Questionnaires Administered?
How Are Test Booklets and Questionnaires Administered?

Administration Schedule

The primary and junior assessments of reading, writing, and mathematics are administered over a two-week period at the end of May and/or the beginning of June. Each assessment is divided into two parts, and one hour is allotted for each part. Additional time may be provided immediately following the allotted hour for students who need more time. The Student Questionnaire is administered separately, after all the assessments have been completed.

The Grade 9 academic and applied mathematics assessments are administered over a two-week period toward the end of the first semester (January) and toward the end the second semester (June). Semestered students write it toward the end of the course, either in January or June. Students in a full-year course write it at the end of the second semester with the second semester students. The academic and applied versions are divided into two parts. Each part is administered on a separate day. Even though the estimated time to complete each part is 40 minutes, students are given one hour to complete each part to accommodate students who may require extra time. The Student Questionnaire is administered after students have completed the second part of the assessment or on a separate day.

The OSSLT is administered to all students during a single morning at the end of March or the beginning of April. The date and time are the same for all students. The OSSLT is in two parts. Students are given 75 minutes to complete each part, with a 15-minute break between the two parts. The Student Questionnaire is administered after students have completed the second part.

Teachers and principals complete their questionnaires either while the students are writing (if they are not supervising the students) or at another time before the materials for the school are returned to EQAO.

Administration Procedure

A very detailed procedure is followed by EQAO, principals, and teachers to ensure that the assessments are administered in the same way across all the schools in the province. This ensures that all students have the same opportunity to show what they know and can do.

EQAO provides a Teacher Bulletin in early fall that outlines what will be assessed, what resource materials are available on EQAO’s Web site (www.eqao.com), and the steps principals and teachers must follow to ensure a fair and complete administration.
EQAO also provides two more documents to each school: the *Administration Guide* and the *Guide for Accommodations, Special Provisions, and Exemptions* for English- and French-language learners. Each school receives a copy of each guide for the current year well before the assessments so that teachers and principal have sufficient time to complete the required activities before each assessment.

The *Administration Guide* outlines in detail what principals and teachers must do before, during, and after the assessment is administered. The things they must do before the assessment is administered include the following:

- Order the student booklets, the appropriate questionnaires, and any materials needed for students with special education needs.
- Check they have received the correct number and type of booklets, questionnaires, and special needs materials.
- Check that the unique student barcode on each booklet corresponds to the number attached to the appropriate student in the school.
- Make sure the assessment materials are securely stored before the assessment dates.
- Ensure that students’ accommodations are consistent with the regular classroom assessment practice outlined in the students’ Individual Education Plans (IEPs) and that, if needed, proper arrangements have been made (e.g., scheduling a scribe).
- Schedule the assessment date or dates in the school.
- Train all school staff involved in the assessment.

The *Administration Guide* also outlines what principals and teachers are responsible for during the test administration. This includes the following:

- Before the assessment begins, ensure that all students have a dictionary and thesaurus for the short- and long-writing tasks and a calculator, ruler, protractor, and other manipulatives they use in class for mathematics; and remove materials posted in the classroom related to the assessment.
- Ensure that students understand how to select and record answers for the multiple-choice items and write responses to the open-response items, and that they must work independently.
- Make sure students have no cellphones or other text-messaging devices and do not receive explanations, definitions, translations, or examples of reading, writing, or mathematics terminology.
- Make sure that students do not receive any encouragement or influence to alter or revise their responses, and are supervised at all times.

Finally, the *Administration Guide* outlines what staff should do after the test has been administered, including the following:

- Ensure that student responses are not rewritten, edited, or corrected in any way.
- Prepare an assessment report that outlines all breaches or unanticipated circumstances that interfered with or otherwise compromised the collection of assessment information.
- Return completed booklets, unused booklets, and the assessment report to EQAO by the required date.

**Accommodations, Special Provisions, and Exemptions**

The *Guide for Accommodations, Special Provisions, and Exemptions* provides information and directions to help principals and teachers make decisions about accommodations for students with special education needs, special provisions for students whose proficiency in the language of instruction is inadequate for them to respond in the anticipated manner, exemptions (primary, junior, and OSSLT only) and deferrals (OSSLT only).
EQAO employs a number of accommodations to enable students with special education needs to demonstrate their achievement of the knowledge and skills being measured by the assessment. These accommodations increase the universality of the assessments and take into account the backgrounds and prior experiences of students. Accommodations change only the way in which the assessment is administered or the way in which a student responds to the components of the assessment. The accommodations used must match the accommodations outlined in the student’s IEP. Permissible accommodations include regular- and large-print versions on white, blue, green, or yellow paper; an audio CD version plus a set of regular- or large-print booklets; regular- or large-print booklets plus an electronic file for use with assistive technology; and contracted or un-contracted Braille versions of the assessment plus a set of regular-print booklets for the scribe to use. Accommodations do not alter the assessment’s content or reliability and they do not affect the validity of the interpretation of the student’s results. In contrast, modifications, which involve changes to content and to performance criteria, are not permitted.

Special provisions for English- and French-language learners are adjustments to the setting or timing of an assessment. These provisions do not affect the reliability or validity of the interpretation of the results for these students.

One of the 32 requirements for acquiring the Ontario Secondary School Diploma is to successfully complete the OSSLT. Given this, students may be deferred in the first year they are expected to take the OSSLT if they have been identified as exceptional by an Identification, Placement, and Review Committee (IPRC) and are not able to participate in the assessment, even with permitted accommodations;

- have not yet acquired the reading and writing skills expected by the end of Grade 9;
- are an English- or French-language learner and have not yet acquired a level of proficiency sufficient to participate in the test; or
- are new to the school and require accommodations that cannot be provided in time for the administration.

All deferred students who wish to obtain an Ontario Secondary School Diploma must eventually complete the literacy requirement. If a student has attempted the OSSLT and has been unsuccessful at least once, the principal has the discretion to allow the student to take the Ontario Secondary School Literacy Course (OSSLC).

In cases where the available accommodations and special provisions do not address a student’s need, exemption from participation in an assessment is allowed in Grades 3 and 6. Students must be exempted for reading if a teacher or another adult must read to them and for mathematics if mathematical terms have to be defined for them. For the OSSLT, students are exempted if they are not working toward the Ontario Secondary School Diploma. However, all students enrolled in a Grade 9 academic or applied mathematics course must participate in the Grade 9 assessment.

The documents referred to in this section may be obtained by going to www.eqao.com.
What Is the Quality of EQAO Assessments? Excellent!

How Are Student Responses Scored?
How Are Student Responses Scored?

Student responses to multiple-choice items for the primary assessment are manually entered into a computer file and are completely verified. Students’ responses to the multiple-choice items on the other assessments are read into the computer using an optical reader. EQAO conducts a quality assurance check to ensure that student responses are captured by the optical reader with 99.5 percent accuracy.

Each year, the total number of scorers for the open-response items is about 3700: 1600 for the primary and junior assessments, 300 for the Grade 9 Assessment of Mathematics, and 1800 for the OSSLT. The scoring takes place in a large building with movable walls to create multiple scoring rooms. The open-response items for the OSSLT are scored over a three-week period in April; the open-response items for the remaining assessments are scored during the first two weeks in July.

Because of the subjectivity of scoring student responses to open-response items, EQAO adheres to a rigorous scoring procedure to ensure that the scorers judge student responses consistently and accurately.

Scoring Materials

The key tools for scoring open-response items are item-specific scoring rubrics and anchor papers. Item-specific scoring rubrics are developed from generic rubrics, which are constant across the years to meet the requirements for equating. The scoring rubrics generally have three, four, or six ordered score codes; the highest code represents the best performance and the lowest code represents the poorest performance. As indicated earlier, the rubrics match what students are expected to do when they respond to the writing prompt or mathematics task. Anchor papers are exemplar papers used to clarify or “anchor” each of the scoring categories in a scoring rubric.

The anchor papers, along with training papers, validity papers, and papers for the qualifying test are selected in two stages. First, a Pre-Range-Finding Committee, consisting of scoring leaders (see below) and EQAO assessment specialists, selects student responses to each open-response item that represent the full range of score points in the scoring rubric for the item. Once student booklets arrive at EQAO from schools, a demographically representative sample of about 500 student responses for each open-response reading or mathematics item and each short-writing and each long-writing item is provided to the committee. The committee members then select a number of booklets and sort them into four piles based on the quality of the responses: approximately 30 low quality, 30 medium quality, 30 high quality, and 40 mixed (low, medium, high) quality.
Each pile contains a range of students’ written responses that correspond to that level of quality. Items left unanswered (“blanks”) or responses that are difficult to read due to poor handwriting or light ink are not selected for pre-range finding.

At the second stage, a Range-Finding Committee, consisting of eight to 25 Ontario principals and teachers, identifies potential anchor papers, validity papers, training papers, and qualifying papers for the qualifying test from the set of sorted pre-range finding papers for each open-response item. As indicated above, anchor papers are used to clarify each of the score codes in a scoring rubric. Validity papers are used during scoring to ensure that the scorers are accurately scoring the student responses. Training papers are used to train the scorers. The qualifying test consists of 14 to 20 qualifying papers. This test is used to assess scoring leaders, scoring supervisors, and scorers to see if they are properly prepared to begin scoring. Once all the papers have been identified, EQAO assessment specialists then decide which of the identified papers will be used for each purpose.

Scoring Leaders and Scorers

Scoring leaders have subject expertise, are effective teachers of adults, and have previously been scorers. They are trained by the assessment specialists on the items to be scored in the scoring room to which the leaders are assigned. Scoring leaders train the scoring supervisors and scorers in their rooms so that the supervisors and scorers align their thinking and judgment to the procedures and scoring materials for the open-response item or items being scored in their room. During scoring, scoring leaders ensure that the scoring materials are applied consistently, resolve issues that arise during scoring, and review and analyze daily scoring reports to ensure that the scoring in their room is accurate and precise.

Scoring supervisors are selected from a pool of experienced and proficient EQAO scorers. They assist the scoring leaders during scoring by supervising a “pod” of scorers in a scoring room. They ensure that the scorers in their pod are scoring accurately and precisely. Scoring supervisors may also be asked to retrain individual scorers when necessary.

Scoring leaders and scoring supervisors must take the qualifying test and attain at least an 80 percent exact and a 100 percent exact-plus-adjacent agreement with the assigned scores for the qualifying papers included in the test. “Exact agreement” means that the scores assigned to a student’s response by two scorers are exactly the same. “Exact plus adjacent” means that there is a difference of one point between the scores assigned to the student’s response by two scorers. Scoring leaders or supervisors who fail the qualifying test may not continue in the role of leader or supervisor.

Scorers are current, retired, or prospective teachers and principals and, in the case of the OSSLT, non-educators who all have at least one university degree. They are carefully trained by the scoring leader in the room to which they have been assigned. The scoring leader uses the anchor and training papers for the open-response items to be scored in the room to train the scorers. Following training, the scorers take the qualifying test. They must attain at least a 70 percent exact match with the expertly assigned score. Scorers who pass the test begin scoring. Scorers who fail the test the first time undergo further training and write a second qualifying test. Scorers who fail to pass the second qualifying test are dismissed.
Scoring Procedure

One long-writing item, two short reading or two short writing items, or four open-response mathematics items are scored in a scoring room under the leadership of the scoring leader, assisted by one or more scoring supervisors. The student booklets containing the open-response items are bundled together in lots of 20. These bundles are formed in such a way that booklets from the same school do not appear in a bundle. Each scorer takes a bundle and scores the open-response items assigned to the scoring room. After scoring each student’s responses, scorers scan the booklet barcode and enter the student’s score into EQAO’s computer for analysis.

Student responses to the open-response OSSLT items are scored independently by two different scorers. If the two scores agree exactly, that score is assigned to the student. If the two scores are adjacent, the higher score for reading and short writing items or the average of the two scores for long-writing tasks is assigned to the student. If the two scores are non-adjacent, the response is rescored by an expert scorer to determine the correct score for the student.

Student responses to the open-response items in the primary, junior, and Grade 9 assessments are scored only once. The reason for this difference in scoring procedures for the different assessments is that passing the OSSLT is a graduation requirement, while student scores on the other assessments do not affect promotion from one grade to the next or high school graduation.

To eliminate any possible changes in scoring across years that might confound the measurement of change from one year to the next, the responses to the open-response field-test items from the previous year that are to be used to equate the current and previous years’ operational forms are rescored during the scoring of the current operational responses.
Monitoring Scoring

During scoring, calibration papers are scored to ensure that the scorers are scoring in the same way. Calibration papers are selected by the scoring leaders and scoring supervisors from the papers scored on the previous day. These papers contain student responses that have been brought to the attention of the leaders by a scorer in the scoring room due to an issue in scoring or to an issue identified in the daily score reports for the room. All scorers score the response to the calibration paper at the same time and record their scores on a sheet provided. After collecting the sheets, the scoring leader explains the issues raised by the calibration paper and gives clear and accurate information on the correct scoring of the paper. If an individual scorer or group of scorers experienced difficulty scoring the calibration paper, the scoring leader or a scoring supervisor addresses the issue with the individual scorer or a group of scorers. There is usually one calibration paper scored each day, but there may be more, if more scoring issues arise.

Also during scoring, each scorer will score sets of validity papers each day to evaluate scoring accuracy. Scoring accuracy is assessed by computing the agreement between the scores assigned by the scorers and the scores assigned during range finding.

Scoring consistency or interrater reliability for the OSSLT is assessed by examining the agreement between the scores assigned by the two independent scorers. Scoring consistency for the primary, junior, and Grade 9 assessments is assessed by examining the agreement between the scores assigned by the scorers to the other open-response items in the booklets containing the validity papers.

EQAO directors, managers, and assessment specialists meet each morning to review scorer validity, reliability, and productivity for the previous day. Problem areas are identified. The assessment specialists then share these problems with the scoring leaders. The scoring leaders share the problems with the scorers in their rooms after the morning break and suggest ways to address the problems. Some scorers will receive additional training based on the results from scoring the validity papers.
Responses Needing Special Attention

Scorers may find student responses to an open-response item contain evidence that the student may be at risk (e.g., the response’s content states or implies threats of violence to him- or herself or others, or possible abuse or neglect). In such cases, scorers, in consultation with the scoring site manager, are legally required to inform the Children’s Aid Society and provide the Society with a copy of the student’s response.

Student responses to open-response items may be inappropriate. Obscene, racist, sexist, or otherwise offensive content in a student’s response is reviewed by EQAO staff to determine whether or not the school or school board should be contacted.

When there is evidence in a completed assessment that may indicate some form of irregularity (e.g., many changed answers, teacher interference), EQAO staff review the booklet to determine whether or not the school should be notified. If cheating is confirmed, the student gets no score on the assessment.

How Accurate and Consistent Is the Scoring?

With such a rigorous scoring process, how good is the scoring? EQAO has adopted targets for exact agreement and exact-plus-adjacent agreement for scores on the validity papers (accuracy) and interrater reliability (consistency). These targets are generally considered acceptable in large-scale testing. The accuracy and consistency targets for exact agreement are 75 percent for three-point rubrics, 70 percent for four-point rubrics, 65 percent for five-point rubrics and 60 percent for six-point rubrics. The target for exact-plus-adjacent agreement is 95 percent for all rubrics. Scorers failing to meet accuracy expectations receive additional training and, if they continue to fail to meet accuracy expectations, they are dismissed.

The accuracy and consistency targets for exact agreement are
- 75% for three-point rubrics,
- 70% for four-point rubrics,
- 65% for five-point rubrics, and
- 60% for six-point rubrics.

The target for exact-plus-adjacent agreement is
- 95% for all rubrics.

The exact-plus-adjacent agreement targets for accuracy were exceeded for all the open-response items scored in 2012 and were generally met in previous years (visit www.eqao.com to find the Technical Report for each assessment year). The exact targets were exceeded for all but five items in 2012. These results reflect the rigour of the scoring process and indicate that parents, students, and teachers can be confident that all students have received accurate and precise scores.
How Are Scored Student Responses Analyzed?
How Are Scored Student Responses Analyzed?

Once all the items have been scored, the next phase in the assessment process is to analyze the student data. EQAO uses two analysis models. The older model is called the classical test score model (CTSM). The newer model is called the item response model (IRM). Each model has two main statistics: item difficulty, which describes how easy or difficult an item is, and item discrimination, which describes how well the item separates students who perform well from students who perform less well.

EQAO does the analysis in three sequential stages, which is common in large-scale testing programs.

1 First, CTSM is used to conduct an item analysis of students’ responses to the items in the current operational form. Item analysis determines the difficulty and discrimination of each operational item. The percentage of students who selected the correct answer denotes the difficulty of a multiple-choice item. The average mean score denotes the difficulty of an open-response item. Discrimination is denoted by the correlation between the item responses and the total test scores and indicates how well the item separates students who performed well on the test from students who performed less well on the test. Items should range in difficulty from easy to hard in an assessment and each item’s difficulty should differ very little between the operational form and when the item was field tested. An item’s discrimination should be at least 0.20.

2 Second, CTSM is used to examine the characteristics of the operational form. These characteristics include the average or mean score, the spread or variability of scores, and the reliability of the test. Reliability is an indication of how consistently the items in the operational form measure what they were designed to measure, such as student achievement of the Grade 3 mathematics learning expectations. EQAO uses the most commonly used reliability measure, Cronbach’s alpha. Theoretically, this coefficient has a maximum value of 1.0, but it is impossible to construct a test with a reliability of 1.0. Higher reliability coefficients indicate lower random measurement error and higher precision. The reliabilities of EQAO’s assessments are all acceptably high. For example, in 2012, the reliability varied from 0.86 to 0.90 for the primary and junior reading and mathematics components, from 0.81 to 0.83 for the primary and junior writing components, from 0.82 to 0.87 for the academic and applied versions of the Grade 9 Assessment of Mathematics, and from 0.85 to 0.89 for the OSSLT. Similar values have been obtained in previous years for each of these assessments. The reliabilities obtained for all of EQAO’s assessments indicate that the test scores from these assessments provide a clearly acceptable level of precision.

3 Third, the scores on the current and the previous year’s assessments are equated to provide a clear measure of change. While EQAO constructs the assessments to be as similar as possible in content and difficulty, the difficulty of the two assessments may differ somewhat. Differences in difficulty confound actual differences in student performance between two years. Equating adjusts for differences in difficulty, but not actual differences in student performance. The purpose of equating is to ensure that EQAO can make valid comparisons of results over years.
In order to equate the achievement results for two groups of students (e.g., primary students in 2011 and in 2012) it is necessary to administer some test items to both groups. With the exception of the long-response writing items, all the items in the 2012 form were field tested in 2011. Therefore, all the primary students in 2012 and a provincially representative sample of primary students in 2011 responded to these items. Since students responded to all items except the long-response writing items in both years, EQAO’s equating is very strong. In most equating designs, roughly 20 percent of the items are common to two tests. For the primary and junior writing components, all items but one are common; for the mathematics components, all items are common; and for the OSSLT, all but two items are common. The greater the number of common items, the greater the confidence we can have in the equated results. Consequently, the interpretation of change between two years is trustworthy. Thus, any changes in student assessment results from one year to the next are attributed to changes in student learning and not to changes in difficulty in the operational form.

**Equating Process**

**The first step** in the equating process is to complete a *calibration* of the current operational form, but this time the IRM is used to obtain the difficulty and discrimination for each operational item. These are called item parameters and are related to the difficulty and discrimination in the CTSM. The IRM is used because it is particularly useful for equating.

**The second step** is to generate item parameters on the same scale for the operational items on both the current and previous years’ assessments, which requires a recalibration of the operational form for the previous year. This is done by calibrating the operational items from the previous year with the field-test items that were carried forward to be operational in the current year. In this calibration, the parameters from the first equating step are used. These parameters are fixed on the field-test items in the previous year and the operational items in the previous year are then calibrated to their scale. The result is a set of item parameters on a common scale for the operational items in the previous and current years.

**The third step** is to obtain student scores for both the current and previous year. Given that the item parameters for the operational items in the current and previous years’ assessments are on the same scale, students’ scores for both years are on the same scale. Therefore, students in the current year and students in the previous year who have identical scores have the same ability as measured by the two assessments. This is comparable to measuring temperature with a thermometer. A weatherperson in New York measures the temperature in Fahrenheit, while a weatherperson in Toronto measures the temperature in Celsius. Thus one scale needs to be converted to the other before we can compare temperatures in the two cities. If the Fahrenheit scale is converted to the Celsius scale, then the two “scores” will be on the same scale. If the score on the recalibrated scale in New York is 18 degrees
and the score on the scale in Toronto is 18 degrees, we can say that the temperature in New York is the same as in Toronto.

The final step in equating is to assign students to achievement categories using cut scores. There are four cut scores for the primary, junior, and Grade 9 assessments and one cut score for the OSSLT. The cut score that separates students at Level 3 from students at Level 4 on the primary reading component is the lowest student score that would be assigned a Level 4. All students with this score or higher are classified as Level 4. We need to be sure that the cut scores in the current year are the same as the cut scores in the previous year. To do this, the percentages of students at the various levels in the previous year are used to find new cut scores in the recalibrated scale for the previous year with the same percentages of students at each level. These new cut scores are the cut scores for the current year. For example, suppose the percentage of students at Level 4 in the previous year was 20.5 percent. Using this percentage, we find the new cut score in the recalibrated scale for the previous year that gives the same percentage of students at Level 4, 20.5 percent. We can use this new cut score for the current year, because the recalibrated scale for the previous year and the current scale have been equated. Suppose we find that 21.5 percent of students in the current year are at or above this cut score for Level 4. Then there are 1.0 percent more students at Level 4 in the current year than in the previous year.

In addition to the equating described above, the results for Grade 9 students who take the mathematics assessments in January and the results for students who take the mathematics assessments in June need to be equated to make sure that there is no advantage in taking the assessment at one time over the other time. Half of the items on the January assessment appear on the June assessments. These common items are used to equate the January and June results. Once the equating process is completed, any changes in student learning between the two semesters can be detected.
What Is the Quality of EQAO Assessments? Excellent!

How Are Cut Scores Established?
How Are Cut Scores Established?

EQAO ensures that the performance standards and cut scores for each assessment are carefully developed. A performance standard is a clear description of what examinees know and can do at each level of performance. They provide qualitative descriptions of the knowledge and skills that individuals at each level possess.

As indicated above, cut scores are used to delineate different levels of performance as identified in the performance standards. Cut scores are set on a score scale. Students who score below the cut score should not be able to correctly answer most of the items associated with performance above the cut score. Performance standards are established and cut scores are set when an assessment program first begins and they are reestablished and reset each time there is a change in the curriculum or a change in the way the assessment is conducted. EQAO also reviews the performance standards periodically to ensure that they are accurate descriptions of what students know and can do and that the cut scores correspond to what students know and can do.

Ontario teachers and principals establish performance standards and set cut scores for the assessments. One panel establishes the standards and a second panel sets the cut scores, with some overlap between the two panels. The teachers and principals on these panels are selected because of their strong knowledge about the students to be assessed, familiarity with The Ontario Curriculum, and experience in writing instructional or assessment materials for students. Each panel is led by an outside expert and the assessment specialist for the assessment.

Establishing Performance Standards

To establish the performance standards, the five to 10 teachers and principals that make up the panel review a sample of 20 student booklets selected from the population of booklets completed by Ontario students in the most recent assessment. No scoring marks appear in the booklets. The panel members are trained to draw distinctions among responses that have “qualitative differences in performance.” After a training session, each panellist independently sorts the booklets into two piles, which reflect differences in performance. For the OSSLT, only two piles are necessary because it is a pass/fail test. For the primary, junior, and Grade 9 assessments, the panellists then sort each of their piles into two piles for a total of four piles, since there are four achievement levels for these assessments. They review their final piles to see if they wish to make any changes. At the end, each panellist has piles that represent their judgments about
qualitative differences in student performance on the assessment. Two or three papers that all the panellists placed in the same pile are then selected. Each panellist then independently describes what qualities are evident in these papers. The panellists are instructed to describe the performance rather than evaluate it, stress what students displayed in their work (not what they could not display), use qualitative language in their descriptions (rather than quantitative language), and avoid content and technical terms where possible. This process is repeated for all the piles for the assessment. These descriptions are summarized by the experts under the following three headings: Level of Accuracy, Degree of Complexity, and Depth of Analysis. Examples of performance standards can be found by visiting www.eqao.com.

Setting Cut Scores

A panel of approximately 20 teachers and principals set the cut scores that separate the achievement categories identified in the performance standards for the assessment. There are several different procedures and modifications of procedures for setting cut scores. The decision about what procedure to use is usually made by an expert in standard setting who works with the panel to set the cut scores.

The session begins with training. First, the panel members independently take and score the test. They then discuss their experiences in taking and marking the test. Doing these first three activities ensures that the panel members are very familiar with the assessment’s items. The panel members then carefully review the performance standards and pay particular attention to the qualitative differences between the adjacent standards, and the assessment specialist leads a discussion about what the panellists found in their reviews. The standard-setting expert then presents the procedure the panellists will use to set the cut scores.

The panel members then set their initial cut scores. There is typically a lot of variation among the members’ cut scores, so the members then discuss why they set the scores they did. After the discussion, members then independently set the cut scores a second time. Typically there is still variation among the panellists’ cut scores. Therefore, these scores are used to produce impact data, which shows the percentages of students that would be in each achievement category if these cut scores were used. The panel discusses these results and then sets cut scores a third time. While these cut scores are shared with the panel, they are not discussed, since the amount of agreement among the panel members’ cut scores at this point is usually acceptable.
How Accurate and Precise Is the Placement of Students in Achievement Categories?

Each year, EQAO assesses how accurately and consistently students are placed in each of the achievement categories. The values for both accuracy and consistency are quite high, ranging from 0.89 to 0.99 across the cut scores for the nine assessments. The differences between these values and the maximum value, which is 1.00, are attributable to random error of measurement.
What Steps Does EQAO Take to Ensure the Quality of Student Results?
What Steps Does EQAO Take to Ensure the Quality of Student Results?

EQAO has several quality assurance procedures to ensure that its assessments are administered consistently and fairly across the province and that the assessment results are reliable and can be validly interpreted.

- EQAO contracts an outside organization to visit a random sample of schools to determine whether the test administration guidelines, use of accommodations for students with special needs, special provisions for first-language learners, and exemptions are followed correctly.
- Principals and teachers at every school in the province are required to report to EQAO any breaches or unanticipated circumstances that interfere with or otherwise compromise the collection of assessment information.
- Once the completed assessments from all schools have been received, EQAO selects a random sample of schools and examines student booklets from these schools for possible irregularities in administration (e.g., different writing in the same student booklet, a large number of changed answers).
- All student responses to multiple-choice items are analyzed to see if there are patterns that suggest possible collusion among two or more students.
- To ensure that the equated scale and cut scores are correct, a qualified external contractor independently replicates the equating conducted by EQAO analysts.
- Once the final results have been determined for each school, EQAO examines the proportion of students in a school performing at or above the provincial standard (at or above Level 3) to see if there are any dramatic shifts over time for the primary, junior, and Grade 9 assessments. EQAO also examines the proportion of students in a school at or above the passing score for the OSSLT to see if there are any dramatic shifts over time.
- As part of EQAO’s research program, EQAO analysts conducted two studies, one to evaluate the item response model used and one to evaluate the equating method used. The results of both studies confirmed that both the current item response models used and the equating method used are appropriate and function well (Xie, 2007; Pang, Madera, Radwan, & Zhang, 2010). To view the two research reports and EQAO’s Technical Report, which contains a more detailed description of the data-analysis procedures, visit www.eqao.com.
How Are EQAO Results Reported and Used?
How Are EQAO Results Reported and Used?

Not only does EQAO report the assessment results effectively, but they also provide interpretive guides, helpful hints, videos, yearly conferences, and effective workshops throughout the school year. These make EQAO unique in Canada and represent best reporting practice.

EQAO reports assessment results at the provincial, school board, school, and student levels. There are two main reports at the provincial level:

- **Ontario Student Achievement: EQAO’s Provincial Elementary School Report on the Results of the Assessments of Reading, Writing and Mathematics, Primary Division (Grades 1–3) and Junior Division (Grades 4–6);**

EQAO also produces Highlights of the Provincial Results, which are much shorter versions of the above reports. These reports are publicly available at www.eqao.com.

At the school-board and school levels, the percentages of students meeting the provincial standard (Levels 3 and 4) and the percentages in each achievement category over time are publicly available at www.eqao.com. These reports also include contextual student data and results from the Student Questionnaires. However, EQAO does not publicly release school results when fewer than 15 students responded to the assessment and fewer than six students responded to the questionnaire in order to prevent the identification of individual students. In the case of these small schools, a three-year “rolling” average is reported.

More detailed school-board and school results are posted on the secure section of EQAO’s Web site and only school-board and school personnel can access them. The reports include the results for schools regardless of the number of students. In addition, school boards and schools receive data files with achievement results for each of their students and data files with aggregated results for the province and each school board. School personnel can access an interactive Web application that allows them to work with their data and to compare their school to similar schools. Providing these additional opportunities means school boards and schools can do analyses tailored to their particular needs.
At the student level, students and their parents receive Individual Student Reports (ISRs). The ISR for the primary, junior, and Grade 9 assessments shows the student’s achievement level for each component (reading, writing, and mathematics) as a small black square placed at one of five positions within that level. For example, if a student’s score was close to the cut score for that level, the square would appear at the left-most position. The ISR also includes school, school board, and provincial results to provide a context to help interpret the student’s results. The ISR for the OSSLT shows the student’s scale score and, for those students who have not passed, provides a short description of what they need to do in order to pass. The ISR also shows the median score for the school and for the province.

EQAO provides a number of printed resources for board members, principals, teachers, and parents to help them interpret and use the assessment and questionnaire results and to guide subsequent follow-up activities (visit www.eqao.com). EQAO also has an effective and active workshop program designed to help school personnel use the assessment results to examine and, when needed, alter instruction so as to improve student learning.

As well, EQAO provides all schools in Ontario with school success stories. The stories are a way of suggesting how school personnel can use the assessment results to improve student learning.

Clearly, the assessment results and information provided to the schools are used to benefit students. Ontario principals and teachers value EQAO’s assessment and questionnaire results provided to them each year. For example, more than 3400 elementary school principals, 8500 Grade 3 teachers, and 7300 Grade 6 teachers were surveyed in 2011. Their survey responses indicated

- **96%** of principals use EQAO achievement results and questionnaire data to guide school improvement initiatives for reading, writing, and mathematics;
- **96%** of principals also use EQAO data to identify program strengths and areas for improvement in these subjects;
- **82%** of Grade 3 and 80 percent of Grade 6 teachers use EQAO data to identify areas of program strength and areas for improvement in these subjects; and
- **80%** of Grade 3 and 78% of Grade 6 teachers reported using EQAO data to identify how well students are meeting curriculum expectations.

(For questionnaire results, visit www.eqao.com.)
EQAO plays an important role in Ontario’s education system by providing:

- an independent gauge of student learning in the core areas of reading, writing, and mathematics at four key points in a student’s journey through school,
- an indication of changes in student learning in the core subjects from one year to the next,
- important information that government, school boards, principals, and teachers can use to improve the quality of education of students, and
- a measure of the quality of our publicly funded schools across years.

EQAO’s assessment program is large and complex and EQAO takes its responsibility very seriously. The Assessments of Reading, Writing, and Mathematics, Primary and Junior Divisions; applied and academic versions of the Grade 9 Assessment of Mathematics; and the Ontario Secondary School Literacy Test (OSSLT) are clearly based on The Ontario Curriculum and measure knowledge and skills students are expected to learn. The procedures used to develop and review assessment and questionnaire items, administer the assessments to students, score student responses, analyze and equate the results of the previous and current years, and report the results and ensure they are used are best of class procedures. EQAO’s quality assurance procedures are equally compelling, as is its willingness to examine its own procedures to ensure that the ones in use are current and best of class.

Turning to the future, in light of the rapidly growing use of technology in education and assessment, EQAO is preparing to implement computer-administered assessments systematically and carefully in order to take advantage of best of class technology available for computer-based assessments.
References


