

EQAO's Executive Summary of the Technical Report

for the 2006–2007 Assessments

*Assessments of Reading, Writing and Mathematics,
Primary Division, (Grades 1-3) and Junior Division (Grades 4-6),
Grade 9 Assessment of Mathematics, and
Ontario Secondary School Literacy Test*



In order to fulfill its mandate, EQAO conducts four annual province-wide assessments for both English- and French-speaking students.

Large-scale assessments are designed to measure student performance at one point in time under standardized conditions.

EQAO assessment blueprints are used to develop the multiple-choice and open-ended items for each assessment, so that the assessment has the same characteristics each year.

Introduction

EQAO's large-scale assessments measure how well students are achieving selected expectations outlined in *The Ontario Curriculum*. The assessments contain open-ended questions requiring written responses as well as multiple-choice questions. Every year, a framework is prepared for each assessment to describe in detail how the assessment will be constructed.

Upon completion, the assessments yield individual student, school and school board results. The assessment results provide valuable information to support improvement planning by schools, school boards and the Ontario Ministry of Education.

The purpose of this summary paper is to outline some of the technical features and professional expertise that were used to ensure the accuracy, validity and psychometric integrity of the following assessments administered in 2006–2007:

- the Assessments of Reading, Writing and Mathematics, Primary Division (Grades 1–3) and Junior Division (Grades 4–6);
- the Grade 9 Assessment of Mathematics and
- the Ontario Secondary School Literacy Test (OSSLT).

Steps Taken to Produce EQAO Assessments

Item Development

It is critical to ensure that the coverage of *Ontario Curriculum* expectations and the difficulty of the assessments are similar each year. To this end, all new items are based on a detailed version of the blueprints. EQAO recruits and trains educator experts in the areas of literacy (English- and French-language reading and writing) and mathematics from across the province to apply the blueprints in their work on the item-writing committees.

Each year, new items are developed and field tested to become operational items in the next year's assessment.

There are strict criteria, item tryouts, sensitivity reviews, revisions and editing before an item is considered for inclusion in an assessment. For the multiple-choice items, teams examine the clarity and completeness of the stem, the correctness of the keyed response and the plausibility of the three incorrect options. For the open-ended reading and mathematics items and the writing prompts, teams consider the correspondence of the items to their scoring rubrics to determine if the items will elicit the range of responses expected.

A great deal of attention is given to selecting items that are appropriate to the students' age and grade level and the curriculum expectations.

Members of the Assessment Development and Sensitivity Review committees focus on issues of fairness and accessibility to the broadest range of students in Ontario by considering equity issues in education (such as multicultural issues and those affecting English-language learners and students with special education needs) to ensure that no particular group of students are advantaged or disadvantaged by any item on an assessment.

The field-test items are embedded in each year's operational assessments.

Field Testing

Since field-test items have the same format as operational items, students do not know whether they are responding to a field-test item or an operational item. This technique avoids the lack of motivation that occurs in students when new items are field tested outside the operational assessment. Scores on the field-test items are not used in determining student, school, school board or provincial results.

The field testing of assessment materials ensures that the assessment items selected for future operational assessments are psychometrically sound and fair for all students. Field testing also provides data that are used to equate each year's assessment with the previous year's so that results can be validly compared. The quality of year-to-year comparisons of data depends on this equating.

A number of guidelines must be adhered to.

Test Construction

EQAO takes the following guidelines into consideration when selecting items for its operational assessments:

- The items must be fair, and the level of difficulty of the assessment must be comparable to that of previous assessments.
- The items must reflect the blueprint and be balanced for subject content, gender representations and aspects of provincial demographics (urban/rural, north/south).
- Over time all measurable expectations in a cluster must be included on an assessment.

Contextual information is used to analyze data by group and to focus recommendations about improving student learning.

Questionnaire Construction

EQAO develops student, teacher and principal questionnaires to collect contextual data as well as information on instructional practices and attitudes related to literacy and mathematics. The questions in them are related to a wide range of factors associated with student achievement.

The administration guides outline procedures that ensure the assessments are consistent and fair for all students.

Test Administration

Each assessment has an administration guide that outlines teachers' roles and responsibilities and gives instructions about preparing students for the test, what to say to students and student participation (i.e., accommodations for students with special education needs, special provisions for English-language learners, exemptions and deferrals).

Another way in which EQAO ensures the integrity of the assessment procedures is by posting monitors in randomly selected schools during testing. In addition, the agency statistically analyses its student response data files to identify unusual response patterns.

Rigorous scoring procedures ensure the reliability of the assessment results.

Scoring

Multiple-choice items are machine scored. Written responses to open-ended reading and mathematics items and to writing prompts are carefully and systematically scored by Ontario educators. A generic rubric for each type of open-response item describes work at the different levels of performance and maintains consistency across items and years. An item-specific rubric for each open-response item in each subject is based on these generic rubrics, and anchors illustrate the descriptors for the score points in the rubric.

Range Finding

The purpose of range finding, which is the first step in the anchor selection process, is to define and illustrate the range of performances within each code of the scoring rubrics. This task is accomplished by a committee of educators with expertise in the Ontario education system, who agree by consensus on the student responses that represent each score point in a specific rubric. This coding of student responses is then used to select anchor papers for the rubric and to train scorers who will score student responses.

Extensive training and monitoring procedures yield results with the highest levels of validity.

Every scorer participates in extensive training to develop a clear and common understanding of the scoring materials, so that each scoring leader, scoring supervisor and scorer interprets and applies the scoring materials in the same way.

Other safeguards and quality control procedures include

- qualifying tests for scorers;
- the supervision of scorers by scoring leaders;
- daily review and calibration activities;
- double blind scoring;
- procedures for reporting evidence of students at risk, inappropriate content and teacher interference;
- daily checks against validity papers to protect against drifts in scoring;
- monitoring of scorers' accuracy rates;
- standards for productivity;
- auditing for scoring inconsistencies and
- performance standards for scoring leaders and supervisors.

All these measures are designed to ensure that students' work is scored consistently regardless of which scorer scores it.

Measuring Marker Reliability and Validity

Interrater reliability and scoring validity measures are used to gauge the consistency of markers' performance. Scores assigned to a given student response are compared to see if they are exactly the same, separated by one score point (adjacent) or separated by more than one score point. At least two markers score each OSSLT student response. If the two scores are not identical or adjacent, an expert marker adjudicates the score.

Interrater Reliability

Interrater reliability is determined by comparing the scores assigned by pairs of markers on randomly selected student responses.

For both reliability and validity, EQAO has set the following industry standards as targets:

- **60% exact agreement for six-point rubrics**
- **70% exact agreement for four-point rubrics**
- **75% exact agreement for three-point rubrics**
- **95% exact-plus-adjacent agreement**

For the primary and junior assessments, the EQAO target for exact agreement was met for 85% of the English and 70% of the French reading items, 92% of the English and none of the French writing rubrics, and 100% of the English and 75% of the French mathematics items. The EQAO target for exact-plus-adjacent agreement was met for all reading items, 75% of the English and 58% of the French writing rubrics, and all English and 63% of the French mathematics items.

For Grade 9 mathematics, the percentages of exact agreement and exact-plus-adjacent agreement exceeded the EQAO targets for all items in both languages, with the exception of exact agreement on one French item on the applied version of the mathematics assessment.

For the OSSLT reading items, the percentages of exact agreement did not meet the EQAO target on any of the items. The percentages of exact-plus-adjacent agreement met the EQAO target on 25% of the English and 75% of the French items. For the OSSLT writing prompts, the percentages of exact agreement met the EQAO target only for the English short-writing prompts. The percentages of exact-plus-adjacent agreement met the EQAO target on 67% of the English and 50% of the French writing rubrics.

The validity of the marking is measured by comparing the level of agreement between scores assigned by the markers and those assigned by an expert panel.

Validity

For the primary and junior assessments, the EQAO target for exact agreement was met for 20% of the English and 55% of the French reading items, none of the English and all of the French writing rubrics, and all mathematics items except one in French. The EQAO target for exact-plus-adjacent agreement was met for 70% of English and all French reading items, half of the English and 83% of the French writing rubrics, and all mathematics items.

For Grade 9 mathematics, the percentages of exact agreement and exact-plus-adjacent agreement exceeded the EQAO targets for all items on the assessment in both languages.

For the OSSLT, the percentages of exact agreement met the EQAO target for none of the English and half of the French reading items, and only for the short-writing prompts. The percentages of exact-plus-adjacent agreement met the EQAO target for half of the reading items and 67% of the writing rubrics.

When markers did not agree with the scores assigned by the panel, the tendency for them to assign higher and lower marks than those assigned by the panel was similar, with some variation among items.

Field-Test Scoring

The scoring of open-ended field-test items is conducted after operational scoring. The most reliable and productive scorers from operational scoring score the field-test items to ensure consistency year to year and to reduce the time required to train scorers. All field-test scoring staff are trained on field-test items and prompts, item-specific rubrics and scoring requirements in order to produce valid and reliable item- and prompt-specific data for operational test construction and equating.

Equating and Analysis of Student Scores

Equating is a process that converts measurements from one scale to another, much as one might convert temperature measures from Fahrenheit to Celsius.

Item Response Theory defines the relationship between a student with a given ability level and the probability of the student answering a given item correctly.

There are many similarities but some differences among the primary, junior, Grade 9 and OSSLT equating procedures.

EQAO employs various statistical models, designs and analyses to ensure the comparability of the assessment results from one administration to the next. The equating process places student scores in two adjacent years on a common scale. This controls for small differences in difficulty among the tests from year to year and ensures that students in one year are not given an unfair advantage over students in another.

Cut scores are then determined to mark the levels of performance used in reporting EQAO results (four achievement levels for primary, junior and Grade 9 and successful or unsuccessful for the OSSLT). Students assigned a given result in the current year will have demonstrated the same level of skill and knowledge as those students assigned that level in previous years.

EQAO uses Item Response Theory (IRT) for item calibration, scoring and equating. The general IRT models EQAO uses to estimate the item and ability parameters for multiple-choice and open-response items are the three-parameter logistic (3PL) model and the generalized partial credit model (GPCM), respectively. The 3PL model is modified by fixing the guessing parameter at 0.20 for all multiple-choice items. For the OSSLT, EQAO uses a modified Rasch model with a fixed guessing parameter of 0.20.

The forward-fixed parameter common-item non-equivalent group design was used to equate the 2006–2007 and 2005–2006 EQAO tests. The 2006–2007 operational tests were created from items field tested in 2005–2006, and these items were used to link the tests. The items in both years were calibrated to the scale for each 2006–2007 operation test, and these item parameters were then used to rescore the 2005–2006 student responses. Finally, theta values corresponding to the percentage of students at each achievement level in 2005–2006 were identified and applied to the 2006–2007 distributions to determine the percentage of students at each achievement level in 2006–2007.

Equating procedures and cut-score determinations were conducted separately for the English and French versions of the assessments.

All analyses were performed twice: once by EQAO staff and once by an external contractor. The results were then compared and differences were resolved, thereby ensuring the accuracy of the results.

Once the scoring was completed, statistical and psychometric analyses using Classical Test Theory and Item Response Theory were conducted.

A variety of test statistics were computed, including Cronbach's alpha reliability coefficient, standard errors of measurement, test characteristic curves and test information functions, to assess the level of precision of the scores for the 2007 EQAO assessments. Overall, the results of these measures indicate that satisfactory levels of precision were obtained. The reliability coefficients ranged from 0.77 to 0.89 for the primary and junior assessments, 0.81 to 0.86 for Grade 9 mathematics, and 0.85 to 0.87 for the OSSLT.

Reporting Results

The results describe student achievement at critical stages in a student's education.

Assessment results are reported at the student, school, school board and provincial levels. They provide valuable information to support improvement planning at all of these levels.

For the primary, junior and Grade 9 assessments, a four-level scale is used to report student achievement in reading, writing and mathematics.

The achievement levels used to report results for the primary, junior and Grade 9 assessments are taken from *The Ontario Curriculum*, which sets Level 3 as the provincial standard. Levels 1 and 2 indicate achievement below the provincial standard, while Level 4 indicates achievement beyond the standard.

For the OSSLT, two levels are reported: successful or unsuccessful.

For the OSSLT, students are classified as successful or unsuccessful. Successful students will have met the level of literacy competency expected at the end of Grade 9 and will thereby have satisfied the literacy requirement for graduation. Feedback is provided to unsuccessful students to assist them in working to meet the standard.

All reports are designed with the utmost attention to accuracy, clarity and effectiveness.

Reports include

- **tables of results**
- **graphs of results**
- **guiding information for interpreting results**

Reports for boards and schools are posted on the public EQAO Web site, www.eqao.com. If the number of students for a given school or board is fewer than 15, in order to prevent the possibility of identification of individual student results, results are not reported. Detailed reports that include results for all students and data files are posted on the secure portion of EQAO's site, which is available to board and school personnel with identification numbers and passwords.

Parents and guardians receive Individual Student Reports, which provide the overall results for the primary and junior division reading, writing and mathematics assessments, and the applied or academic version of the Grade 9 mathematics assessment. Provincial, board and school reports provide contextual information; results for each subject for the school, board and province; results by gender and English-language learner and special needs status; results over time; areas of strength and areas for improvement with reference to the curriculum; item information results for each student, school and board and the province; and results from the student, teacher and principal questionnaires. Results from the teacher and principal questionnaires are not reported at the school level.

Similarly, for the OSSLT, individual student results are reported to students, as are the results of first-time eligible and previously eligible students at the school, board and provincial levels to schools and boards.

In addition, EQAO publishes a provincial report for education stakeholders and the general public for the English and French versions of each assessment. These are available on the public EQAO Web site.