Education Quality and Accountability Office

EQAO

EQAO's Executive Summary of the Technical Report

for the 2008–2009 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

EQAO, 2 Carlton Street, Suite 1200, Toronto ON M5B 2M9 • Telephone: 1-888-327-7377 • Web site: www.eqao.com • © 2010 Queen's Printer for Ontario

Introduction

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

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

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

Upon completion, the assessments yield individual student, school, school board and provincial 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 report is two-fold: to describe the technical quality of the 2009 EQAO assessments and to outline some of the procedures and professional expertise that were used to ensure the accuracy, validity, reliability and psychometric integrity of the following assessments administered in 2008–2009:

- 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

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

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 detailed blueprints. EQAO recruits and trains educators with expertise 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 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, content and 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-response 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 developing items that assess the curriculum expectations and are appropriate, fair and accessible to the broadest range of students in Ontario.

The Assessment Development Committee reviews the items to ensure that they are appropriate to the age and grade of the students, the curriculum expectations being measured and the purpose of the assessment. The Sensitivity Review Committee focuses on issues of fairness and accessibility to the broadest range of students in Ontario by considering equity issues in education (such as gender and 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.

Field Testing

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

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 circumvents the tendency for students to lack motivation 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 over time can be validly compared. The quality of year-to-year comparisons of data depends on this equating.

Test Construction

A number of guidelines must be adhered to when tests are constructed.

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 representation and aspects of provincial demographics (urban/rural, north/south).
- Since items are mapped to clusters of curriculum expectations, and not all expectations in a cluster will be assessed on any one assessment, it is important that all measurable expectations in a cluster be included on at least one assessment over a period of years.

Questionnaires

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

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 are related to a number of factors associated with student achievement.

Test Administration

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

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 analyzes the student response data files to identify unusual response patterns. If unusual response patterns are identified, the school is contacted for additional information. EQAO examines overall school results over time to identify large increases in achievement results in the current year. School superintendents are asked to document explanations for these large changes.

Scoring

Rigorous scoring procedures ensure the reliability of the assessment results.

Multiple-choice items are machine scored. Written responses to open-response reading and mathematics items and to writing prompts are carefully and systematically scored by qualified scorers, who are mostly 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 scoring 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 scoring and the Ontario education system. Committee members reach agreement on the student responses that represent each score point in a specific rubric. This coding of student responses is used to identify anchor papers for the rubric and to train scorers who will score student responses.

Training

Extensive training and monitoring procedures maintain high 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, scoring supervisors and scoring leaders;
- the supervision of scorers by scoring leaders and supervisors;
- daily review and calibration activities;
- double blind scoring for all student OSSLT responses;
- procedures for reporting evidence of students at risk, inappropriate content and teacher interference;
- daily checks against expertly scored student responses 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 any given student response will receive the same score regardless of which scorer scores it.

Maintaining Accuracy and Consistency During Scoring

EQAO monitors the accuracy and consistency of scoring by checking the scores assigned to selected student responses by scorers against those assigned by experts and by checking for agreement between pairs of scores on randomly assigned student responses.

In preparation to monitor scoring accuracy, several student responses are scored by expert scorers prior to the scoring session. During scoring, the scores assigned to a given student response are compared with those assigned by the experts to see if they are exactly the same, separated by one score point (adjacent) or separated by more than one score point (non-adjacent). Scorers who do not have high levels of exact agreement with the expert scores are given additional training. EQAO achieves 95% exact-plus-adjacent agreement for nearly all questions across all assessments (156 out of 168 items).

EQAO monitors interrater reliability (consistency) by comparing the scores assigned to randomly selected student responses by pairs of scorers. At least two scorers score each OSSLT student response. If the two scores are not identical or adjacent, an expert scorer adjudicates the score.

Field-Test Scoring

The scoring of open-response 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 places student scores in consecutive years on a common scale so that student achievement results across years can be compared validly.

Item Response Theory (IRT) defines the relationship between student ability with respect to a defined construct and the probability of his or her correctly answering a given item that measures the construct.

Different IRT models are used for the OSSLT and for the primary, junior and Grade 9 assessments.

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 valid comparisons can be made across years and 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 IRT for item calibration, scoring and equating. The general IRT models EQAO uses for the primary, junior and Grade 9 assessments are a modified two-parameter logistic (2PL) model and a modified Rasch model for the OSSLT. Both models are modified by the addition of a fixed guessing parameter set at 0.20 for multiple-choice questions. EQAO uses the generalized partial credit model (GPCM) for open-response items in all assessments.

The forward-fixed parameter common-item non-equivalent group design was used to equate the 2008–2009 and 2007–2008 EQAO tests. The 2008–2009 operational tests were created from items field tested in 2007–2008 and these items were used to link the tests. The 2008–2009 operational tests were calibrated, and the item parameters were fixed on those items that had been field tested on the 2007–2008 tests. The 2007–2008 operational tests were then calibrated with the field-test items to place the 2007–2008 and 2008–2009 tests on a common scale. The item parameters from this calibration were then used to rescore the 2007–2008 student responses. Finally, theta values corresponding to the percentage of students at each achievement level in 2007–2008 were identified and applied to the 2008–2009 student scores to assign students to achievement

levels and to determine the percentage of students at each achievement level in 2008–2009.

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.

Reliability and Accuracy of Assessment Results

Once the scoring was completed, statistical and psychometric analyses using Classical Test Theory and IRT were conducted to provide data on the technical quality of the individual test items and on the overall reliability of the student scores.

To assess the level of precision of the scores for the 2009 EQAO assessments, a variety of test statistics were computed, including Cronbach's alpha reliability coefficient, standard error of measurement, test characteristic curves, test information functions, differential item functioning statistics and classification accuracy and consistency. Overall, the results of these measures indicate that satisfactory levels of precision were obtained. The reliability coefficients ranged from 0.78 to 0.89 for the primary and junior assessments, 0.81 to 0.85 for Grade 9 mathematics, and 0.86 to 0.88 for the OSSLT. The classification accuracy for students at or above the provincial standard or successful on the OSSLT ranged from 0.86 to 0.92, indicating that about 90% of students were correctly classified.

A number of factors contributed to the level of precision: the quality of the individual assessment items, the accuracy and consistency of scoring and the interrelationships among the items. All items on the EQAO assessments are directly linked to expectations in the curriculum. For the operational assessments, EQAO selects items of an appropriate range of difficulty and that discriminate between students with high and low levels of achievement. As described above, a number of practices maintain and improve accuracy and consistency in scoring. Statistics related to these factors for individual items are provided in the full *Technical Report* for all assessments at www.eqao.com.

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.

Parents and guardians receive Individual Student Reports, which provide the overall results for each student.

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

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. Data files and detailed reports that include results suppressed in the public reports and results for subgroups of students, such as English language learners and students with special education needs, are posted on the secure portion of EQAO's Web site, which is available to board and school personnel with user identification numbers and passwords.

Reports include

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

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

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.