

EQAO's 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*



CHAPTER 1: PURPOSE AND OVERVIEW OF THIS REPORT.....	1
THE EQAO ASSESSMENT PROGRAM: PRIMARY (GRADES 1–3), JUNIOR (GRADES 4–6), GRADE 9 AND ONTARIO SECONDARY SCHOOL LITERACY TEST	1
CHAPTER 2: ASSESSMENT DESIGN AND DEVELOPMENT.....	3
ASSESSMENT FRAMEWORKS	3
ASSESSMENT BLUEPRINTS.....	3
ITEM DEVELOPMENT	3
Item Developers	3
Training for Item Developers	4
EQAO Education Officer Review	4
Item Tryouts	5
ASSESSMENT DEVELOPMENT AND SENSITIVITY REVIEW COMMITTEES	5
EQAO Assessment Development Committees.....	5
EQAO Sensitivity Committee	6
FIELD TESTING	6
TEST CONSTRUCTION: SELECTING ITEMS FOR THE OPERATIONAL FORM	7
QUESTIONNAIRES	7
CHAPTER 3: TEST ADMINISTRATION AND PARTICIPATION.....	9
ASSESSMENT ADMINISTRATION	9
The Administration Guides	9
Support for Students with Special Education Needs and English Language Learners: The Guides for Accommodations.....	9
QUALITY ASSURANCE	11
CHAPTER 4: SCORING.....	13
THE RANGE-FINDING PROCESS	13
Pre-Range Finding.....	13
Range Finding	14
Overview of the Range-Finding Process	15
PREPARING TRAINING MATERIALS FOR SCORING: ANCHORS, TRAINING PAPERS, QUALIFYING TESTS, VALIDITY PAPERS	15
FIELD-TEST SCORING	16
Scoring Open-Response Field-Test Responses	16
Training Field-Test Scoring Leaders and Scorers	16
Developing Additional Scoring Training Materials Before Scoring Operational Items.....	17
SCORING OPEN-RESPONSE OPERATIONAL ITEMS	17
Scoring Rooms for Scoring Open-Response Operational Items.....	17
Training for Scoring Open-Response Operational Items.....	18
Training of Scoring Leaders and Scoring Supervisors for Scoring Open-Response Operational Items	18
Training of Scorers for Scoring Open-Response Operational Items	19
PROCEDURES AT THE SCORING SITE.....	20
Students at Risk	20
Inappropriate Content, Cheating and Other Issues	21
Ongoing Daily Training	21
Daily Scoring Centre Reports for Monitoring the Quality of Open-Response Item Scoring	21
Required Actions: Consequences of the Review and Analysis of Daily Scoring-Centre Data Reports	24
Audit.....	24
SCORER RELIABILITY AND VALIDITY	24
Scorer Reliability	24
Scoring Validity	36

CHAPTER 5: EQUATING.....	49
IRT MODELS	49
EQUATING DESIGN	50
EQUATING SAMPLES.....	50
CALIBRATION	51
ASSESSMENTS OF READING, WRITING AND MATHEMATICS: PRIMARY AND JUNIOR DIVISIONS	51
Description of the IRT Model	51
Equating Sample: Exclusion Rules.....	51
Equating Steps.....	52
Eliminating Items and Collapsing of Score Categories	52
Equating Results.....	53
GRADE 9 ASSESSMENT OF MATHEMATICS	58
Description of the IRT Model	58
Equating Sample.....	59
Equating Steps.....	59
Eliminating Items and Collapsing of Score Categories	60
Equating Results.....	60
ONTARIO SECONDARY SCHOOL LITERACY TEST (OSSLT)	61
Description of the IRT Model	61
Equating Sample.....	61
Equating Steps.....	62
Scale Score	62
Eliminating Items and Collapsing of Score Categories	62
Equating Results.....	63
REFERENCES	63
CHAPTER 6: REPORTING RESULTS	64
REPORTING THE RESULTS OF THE ASSESSMENTS OF READING, WRITING AND MATHEMATICS, PRIMARY AND JUNIOR DIVISIONS.....	64
REPORTING THE RESULTS OF THE GRADE 9 ASSESSMENT OF MATHEMATICS	66
REPORTING THE RESULTS OF THE OSSLT	66
INTERPRETATION GUIDES.....	67
CHAPTER 7: STATISTICAL AND PSYCHOMETRIC SUMMARIES.....	68
ASSESSMENTS OF READING, WRITING AND MATHEMATICS, PRIMARY AND JUNIOR DIVISIONS	68
Classical Test Theory (CTT) Analyses.....	68
Item Response Theory (IRT) Analyses	69
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)	83
GRADE 9 ASSESSMENT OF MATHEMATICS	85
Classical Test Theory (CTT) Analyses.....	85
Item Response Theory (IRT) Analyses	86
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)	92
ONTARIO SECONDARY SCHOOL LITERACY TEST (OSSLT)	94
Classical Test Theory (CTT) Analyses.....	94
Item Response Theory (IRT) Analysis.....	94
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)	97
DIFFERENTIAL ITEM FUNCTIONING (DIF)	98
Primary- and Junior-Division Assessments.....	100
Grade 9 Mathematics.....	102
OSSLT.....	104
DECISION ACCURACY AND CONSISTENCY	105
Accuracy.....	105
Consistency	106
Estimation from One Test Form.....	106
Primary and Junior Assessments	108

Grade 9 Assessment of Mathematics.....	109
OSSLT.....	109
REFERENCES	110
APPENDIX 7.1: SCORE DISTRIBUTIONS AND ITEM STATISTICS FOR THE 2009 EQAO ASSESSMENTS.....	111

CHAPTER 1: PURPOSE AND OVERVIEW OF THIS REPORT

The EQAO Assessment Program: Primary (Grades 1–3), Junior (Grades 4–6), Grade 9 and Ontario Secondary School Literacy Test

In order to fulfill its mandate, EQAO conducts four province-wide assessments: the Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions; the Grade 9 Assessment of Mathematics (academic and applied); and the Ontario Secondary School Literacy Test (OSSLT). All four assessments are conducted annually and involve all students in the specified grades in all publicly funded schools in Ontario, as well as a number of students in private schools that use *The Ontario Curriculum*.

EQAO assessments are developed in keeping with the *Principles for Fair Student Assessment Practices for Education in Canada* (1993), a document widely endorsed by Canada's psychometric and education communities. The assessments measure how well students are achieving selected expectations outlined in *The Ontario Curriculum*. The assessments contain performance-based activities requiring written responses to open-response questions as well as multiple-choice questions through which students demonstrate what they know and can do in relation to the curriculum expectations measured. One version of each assessment is developed for English-language students and another version is developed for French-language students. Both versions have the same number of items and kinds of tasks, but reflect variations in the curricula for the two languages.

The assessments provide individual student, school, school board and province-wide results on student achievement of selected *Ontario Curriculum* expectations. Every year, EQAO posts school and board results on its Web site (www.eqao.com) for public access. EQAO publishes annual provincial reports in English and in French for education stakeholders and the general public on the EQAO Web site. The assessment results provide valuable information that support improvement planning by schools, school boards and the Ontario Ministry of Education.

The annual Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions, measure how well elementary-school students have met the reading, writing and mathematics curriculum expectations assessed by EQAO and outlined in *The Ontario Curriculum, Grades 1–8: Language* (revised 2006) and *The Ontario Curriculum, Grades 1–8: Mathematics* (revised 2005). The reading component assesses students on their skill at understanding explicit and implicit information and ideas in a variety of text types required by the curriculum. The reading component also requires students to make connections between what they read and their own personal knowledge and experience. The writing component assesses students on their skill at organizing main ideas and supporting details using correct spelling, grammar and punctuation in a variety of written communication forms required by the curriculum. The mathematics component assesses students on their knowledge and skill across the five mathematical strands in the curriculum: number sense and numeration, measurement, geometry and spatial sense, patterning and algebra, and data management and probability.

The Grade 9 Assessment of Mathematics measures how well students have met the expectations

outlined in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005). The Grade 9 assessment is conducted twice annually to accommodate secondary schools with semestered programs. EQAO develops separate versions of the Grade 9 Assessment of Mathematics for students in academic and applied courses. Students in Grade 9 academic mathematics are assessed on their knowledge and skill across the four mathematical strands in the curriculum: number sense and algebra, linear relations, analytic geometry, and measurement and geometry. Students in Grade 9 applied mathematics are assessed on their knowledge and skill across the three mathematical strands in the curriculum: number sense and algebra, linear relations, and measurement and geometry. Two parallel forms of the Grade 9 Assessment of Mathematics are developed for each course (academic and applied), with one form administered toward the end of the first semester and the second form administered toward the end of the second semester.

The OSSLT is administered annually and assesses Grade 10 students' literacy skills based on reading and writing curriculum expectations across all subjects in *The Ontario Curriculum* up to the end of Grade 9. The reading component assesses students on their skill at understanding explicit and implicit information and ideas in a variety of text types required by the curriculum. It also assesses students on their ability to make connections between what they read and their own personal knowledge and experience. The writing component assesses students on their skill at organizing main ideas and supporting details using correct spelling, grammar and punctuation for communication in written forms required by the curriculum. Successful completion of the OSSLT is one of 32 requirements for the Ontario Secondary School Diploma (OSSD).

EQAO education officers involve educators across the province in most aspects of EQAO assessments, including

- design,
- development of items and item-specific scoring rubrics,
- review of items for curriculum content and sensitivity,
- administration of the assessments in schools,
- scoring student responses to open-response items and
- reporting assessment results.

Educators are selected to participate in EQAO activities based on the following criteria:

- geographic location (balance between northern, southern, eastern and western parts of the province);
- rural and urban locations;
- current elementary and secondary experience (teachers, administrators, subject experts and consultants);
- expertise in assessment, evaluation and large-scale assessments and
- previous experience working with EQAO.

Educators interested in taking part in EQAO assessments can apply by visiting the EQAO Web site, www.eqao.com, and clicking "Careers and Educator Opportunities."

CHAPTER 2: ASSESSMENT DESIGN AND DEVELOPMENT

Assessment Frameworks

EQAO posts the current framework for each large-scale assessment on its Web site to provide educators, students, parents and the general public with a detailed description of the assessment, including an explanation of how the assessment relates to *Ontario Curriculum* expectations. The English-language and French-language frameworks for the four EQAO assessments can be found on the EQAO Web site, www.eqao.com.

Assessment Blueprints

EQAO assessment blueprints are used to develop multiple-choice and open-response items for each assessment so that each year the assessment has the same characteristics. This consistency in assessment design ensures that the number and types of items, the coverage of *Ontario Curriculum* expectations and the difficulty of the assessments are comparable each year. It should be noted that not all expectations can be measured in a large-scale assessment. Measurable curriculum expectations are clustered by topic, and items are then mapped to these clusters of expectations. Not all of the measurable expectations in a cluster are measured in any one assessment; however, over time all measurable expectations in a cluster are assessed.

The blueprints can be found in EQAO's assessment frameworks. A more detailed version of the blueprints is provided to item developers.

Item Development

New items are developed and field tested each year before becoming operational items in future assessments. Educators from across the province assist EQAO with all aspects of the development of the assessments, including

- finding or developing reading selections appropriate for the grade level of students;
- developing multiple-choice and open-response reading, writing and mathematics items and item-specific scoring rubrics for open-response items;
- trying out items as they are being developed and
- reviewing reading selections, items and item-specific scoring rubrics for curriculum content and possible bias against subgroups of students (e.g., students with special education needs; English language learners; students of a particular gender or ethnic or racial background).

Item Developers

EQAO recruits and trains experienced educators in the areas of English- and French-language (reading and writing) and mathematics to participate in its item-writing committees. The item-writing committee for each assessment comprises 10 to 20 educators who serve for terms of one to five years. Committee members meet twice a year to write and revise items, discuss results of item tryouts and review items that will be considered for use in subsequent operational assessments.

Item developers construct multiple-choice items in reading, writing and mathematics; open-response items in reading and mathematics; and open-response writing prompts for short- and long-writing tasks. All are referenced to *Ontario Curriculum* expectations and matched to the blueprints for the individual assessments. Item developers are provided with a copy of the *Development Specifications Guide for EQAO Assessments* to assist them in the development of multiple-choice and open-response items and writing prompts.

Item writers for EQAO assessments are selected based on their

- expert knowledge and recent classroom experience in at least one of the following:
 - literacy education (language, English, language across the curriculum) and
 - mathematics education;
- familiarity with and knowledge about the elementary and/or secondary school curricula in Ontario (especially in language, English and mathematics);
- familiarity with the cross-curricular literacy requirements of the elementary and secondary curricula in Ontario (OSSLT);
- expertise and experience in the application of rubrics based on the achievement charts in *The Ontario Curriculum* (to identify varying levels of student performance, especially in elementary language and elementary and secondary mathematics);
- excellent written communication skills;
- comfort using computer word-processing programs (and, for writers of mathematics items, mathematics software);
- experience in writing instructional or assessment materials for students;
- proven track record of working collaboratively with others and accepting instruction and feedback and
- access to grade and subject classrooms to conduct item tryouts.

Training for Item Developers

The field-test material for 2008–2009 was developed by EQAO in partnership with Educational Testing Service (ETS) Canada. EQAO and ETS staff led a four-day workshop and development session for item writers. They spent approximately half a day introducing item writers to the criteria for item writing. EQAO provided an overview of the assessments, including a description of the frameworks, and then ETS provided details on the elements of effective item writing. The remaining time involved a guided item-writing session structured by EQAO, in consultation with ETS.

In general, each item writer was assigned items to write based on the EQAO blueprint for the assessment. At the end of the item-writing session, the draft material was examined by EQAO and ETS staff. ETS further refined the items prior to their presentation to the Assessment Development and Sensitivity review committees. Further revisions took place following the committee reviews to prepare the items for EQAO to select for embedding in the matrix field test.

EQAO Education Officer Review

When the first draft of the items and item-specific scoring rubrics are developed by the item writers, they are reviewed by EQAO education officers. The education officers ensure that each

item is referenced correctly to curriculum expectations and difficulty levels. For the multiple-choice items, the education officers consider the clarity and completeness of the stem, the correctness of the correct answer and the plausibility of the three incorrect options. For the open-response items, the education officers consider the correspondence between the items and their scoring rubrics to determine if the items will elicit the range of responses expected.

Item Tryouts

After the initial review of first-draft items by the education officers, item writers use cognitive labs in their own classes and schools to try out the items they developed with students. These item tryouts allow item writers to see if their items are working as intended. The student responses are used to inform the editing and refining of stems of multiple-choice items, multiple-choice options, open-response items and item-specific scoring rubrics for open-response items. The results of these item tryouts are provided to EQAO education officers to help them review, revise and edit the items.

Assessment Development and Sensitivity Review Committees

EQAO recruits and trains Ontario educators with expertise in the areas of English- and French-language (reading and writing), mathematics and equity issues to participate in its Assessment Development and Sensitivity committees. All field-test and operational assessment materials that appear on EQAO assessments are reviewed by these committees.

The goal of these committees is to ensure that items on the Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions, the Grade 9 Assessment of Mathematics and the OSSLT assess literacy and mathematics standards based on *Ontario Curriculum* expectations and that these items are appropriate, fair and accessible to the broadest range of students in Ontario.

Assessment Development and Sensitivity committee members are selected based on their

- knowledge and experience with the elementary and/or secondary school curricula in Ontario;
- expert knowledge and recent classroom experience in at least one of the following:
 - literacy education (language, English or language across the curriculum);
 - mathematics education and
 - equity issues in education (such as multicultural issues, those relating to students with special education needs and English language learners) to ensure that no particular group of students is advantaged or disadvantaged by any item on an assessment (e.g., reading selections, writing prompts or mathematics contexts) and
- experience with large-scale assessments (such as previous work with EQAO) or related expert knowledge.

EQAO Assessment Development Committees

The Assessment Development Committee for each subject in each assessment comprises 10 to 12 Ontario educators who serve for terms of one to five years. Members meet once a year to provide expert advice from a specialized content and assessment perspective on the quality and fairness of materials being proposed for EQAO assessments and to ensure that all field-test and

operational items appropriately assess standards of literacy and mathematics based on *Ontario Curriculum* expectations.

The members of the Assessment Development Committee possess expertise in and current experience with the curriculum and students in at least one of the subjects in the grade being assessed:

- primary assessment: language and/or mathematics in the primary division (Grades 1 to 3)
- junior assessment: language and/or mathematics in the junior division (Grades 4 to 6)
- Grade 9 assessment: mathematics in the intermediate division (Grades 7 to 10)
- OSSLT: literacy across the curriculum in the intermediate division (Grades 7 to 10)

The members of the Assessment Development Committee work collaboratively under the guidance of EQAO education officers to ensure that the materials (e.g., reading selections; reading, writing and mathematics items and writing prompts) for a particular assessment are appropriate to the age and grade of the students, the curriculum expectations being measured and the purpose of the assessment. They make suggestions for the inclusion, exclusion or possible revision of items.

EQAO Sensitivity Committee

The Sensitivity Committee, which considers all four EQAO assessments, comprises 20 to 24 Ontario educators who serve for terms of one to five years. Members meet in focused subgroups once a year to make recommendations that will assist EQAO in ensuring the fairness of all field-test and operational items being proposed for its assessments. They provide expert advice from a specialized equity perspective to ensure that assessment materials are fair for a wide range of students. The members of the Sensitivity Committee possess expertise in and current experience with equity issues in education (issues related to the diversity of Ontario students, students with special education needs and English language learners).

The members of the Sensitivity Committee work collaboratively under the guidance of EQAO education officers to review assessment materials (e.g., reading selections, items) in various stages of development to ensure that no particular group of students is unfairly advantaged or disadvantaged on any item. They make suggestions for the inclusion, exclusion or possible revision of items.

Field Testing

Field testing of assessment materials ensures that assessment items selected for future operational assessments are psychometrically sound and fair for all students. Field testing also provides data to equate each year's assessment with the previous year's assessment, so that assessment results can be validly compared over time. Only items found to be acceptable based on field-test results are used operationally in EQAO assessments.

EQAO uses a matrix-sample design in which newly developed items are embedded as field-test items in each operational form. The field-test items are placed in the booklets according to psychometric principles to ensure that valid and reliable data are obtained for each field-test item. The field-test items are divided into subsets that are inserted into the operational

assessment to guarantee that they are attempted by a representative sample of students. Since the field-test items are like the operational items, the students do not know whether they are responding to a field-test item or an operational item. This likeness can counter the low motivation that may occur among students when they know which items are field-test items and do not count toward their score. The number of field-test items in each subset ensures that no more than 20% of the assessment items given to a student are field-test items. Scores on the field-test items are not used in determining student, school, school board or provincial results.

All items, except for the long-writing prompts on the primary- and junior-division assessments and the OSSLT, are field tested this way. Long-writing prompts require too much student time to be field tested during an operational assessment. Long-writing prompts go through a rigorous process of committee reviews and field trials to ensure their appropriateness. Long-writing prompts are not used for equating.

Test Construction: Selecting Items for the Operational Form

Operational items are selected from the items that were field tested in previous assessments. Several important factors are taken into consideration when these items are selected for an operational form:

- *Data:* The data for individual items, groups of items and test characteristic curves (based on selected items) need to indicate that the assessment items are fair and comparable in difficulty to those on previous assessments.
- *Educator perspective:* The items selected for an assessment are reviewed to ensure that they reflect the blueprint for the assessment and are balanced for aspects such as subject content, gender representations and provincial demographics (e.g., urban or rural, north or south).
- *Curriculum coverage:* As mentioned earlier, it is important to note that while items are mapped to clusters of curriculum expectations, and not all expectations within a cluster are measured in any one assessment, over time, all measurable expectations in a cluster are included on an assessment.

Sample assessments are available on the EQAO Web site, www.eqao.com.

Questionnaires

EQAO constructs student, teacher and principal questionnaires to collect information on instructional practices and attitudes related to literacy and mathematics so that EQAO results can be used to make recommendations to improve student learning. Some of the contextual data derived from these questionnaires are included in EQAO's annual provincial reports and are used to inform the recommendations in these reports.

The Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions, include student, teacher and principal questionnaires. Students are asked about their age; language spoken at home; attitudes toward reading, writing and mathematics; home help with reading, writing and mathematics; use of computers at school for reading, writing and mathematics learning; and use of calculators at school for mathematics learning. Teachers are asked about the size and makeup of their class (e.g., single, combined or multi-grade); the usefulness of the

resources in their classroom; the accessibility of classroom resources and the frequency with which they are used; collaboration with colleagues to plan and discuss curriculum and assessment; their use of assessment-related materials from the EQAO Web site; and their gender, experience, qualifications and participation in professional development. Principals are asked about the school's characteristics (e.g., grades taught, enrolment, student absenteeism and number of staff); their gender; education; participation in professional development; knowledge about instruction and assessment, and use of data for school improvement planning; communication with parents and staff about EQAO testing; and use of the *EQAO Guide to School and Board Improvement Planning* (2005). Copies of this guide can be obtained from the EQAO Web site, www.eqao.com.

The Grade 9 Assessment of Mathematics also includes student and teacher questionnaires. The student questionnaires include questions on students' attitudes toward mathematics; access to a computer, calculator and graphing calculator at home; time spent on mathematics homework; age; language spoken at home; number of times late for and absent from mathematics class; and number of elementary schools attended. The teacher questionnaires include questions about the availability and use of computers and graphing calculators by Grade 9 mathematics students in the school; instructional and assessment practices used in mathematics classrooms in the school; their collaboration with colleagues to plan and discuss curriculum and assessment; their use of assessment-related materials from the EQAO Web site; and their gender, experience, mathematics qualifications and participation in professional development.

The OSSLT includes a student questionnaire that asks students about their access to a computer at home; amount of time spent reading different types of reading materials outside school; access to reading materials at home; language spoken at home and time spent writing in different kinds of written forms outside school.

CHAPTER 3: TEST ADMINISTRATION AND PARTICIPATION

Assessment Administration

To ensure consistent and fair practice across the province in the administration of the assessments, EQAO annually publishes an administration guide and a guide for accommodations for each assessment. They can be found on the EQAO Web site, www.eqao.com.

The Administration Guides

The administration guide for each EQAO assessment describes in detail the administration procedures that principals and teachers must follow to ensure that the administration of the assessment is consistent and fair for all students in the province. Each school is sent copies of the English- or French-language administration guide for training teachers to administer the assessment. The guide outlines in detail what is expected of educators involved in the administration, including

- the procedures to follow (e.g., preparation of materials for distribution to students, do's and don'ts during the assessment),
- what to say to students (e.g., instructions for presenting the assessment) and
- the professional responsibilities of all school staff involved in the assessment.

During the assessment, students answer multiple-choice questions and write their responses to open-response items. Students must work independently in a quiet environment and be supervised at all times.

Support for Students with Special Education Needs and English Language Learners: The Guides for Accommodations

The guide for accommodations for each assessment provides information and directions to assist principals and teachers in making decisions about

- accommodations for students with special education needs,
- special provisions for English language learners and
- the exemption (primary, junior and OSSLT only) or deferral (OSSLT only) of students.

Students with special education needs are allowed accommodations, and English language learners are provided with special provisions to ensure that they can participate in the assessment and demonstrate the full extent of their skills. In cases where the accommodations or special provisions do not address a student's needs, exemption from participation in the assessment is allowed (primary and junior only). Each year, EQAO reviews and updates these accommodations and provisions to ensure that they reflect new developments in supports available for students.

The guides for accommodations and special provisions also clarify the expectations for the documentation of accommodations, special provisions, exemptions and deferrals for students receiving them. The guidelines are based on two Ontario Ministry of Education policy documents: *Individual Education Plans: Standards for Development, Program Planning, and Implementation* and *Policy/Program Memorandum 127* (Grade 9 and OSSLT only), both available at www.edu.gov.on.ca.

Definition of “Accommodations”

Accommodations are defined in the accommodations guides (modified from *Policy/Program Memorandum 127*) as follows:

“Accommodations” are supports and services that enable students with special education needs to demonstrate their competencies in the skills being measured by the assessment. 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. It is expected that accommodations will not alter the content of the assessment or affect the validity or reliability.

On the other hand, “modifications,” which are not allowed, are changes to content and to performance criteria. Modifications are not permitted, because they would affect the validity and reliability of the assessment results.

Clarification of instructions for all students is permitted prior to the assessment. Clarification of questions during the assessment (e.g., rewording or explaining) is not allowed.

Special Version Assessments for Accommodated Students

EQAO provides the following special versions of the assessments to accommodate the special education needs of students:

- contracted Braille version, plus a set of regular-print booklets for the scribe’s use
- uncontracted Braille (OSSLT only), plus a set of regular-print booklets for the scribe’s use
- large-print version—white paper
- large-print version—blue, green or yellow paper
- regular-print version—blue, green or yellow paper
- audio CD version, plus a set of regular-print booklets
- audio CD version, plus a set of large-print student booklets
- one single-sided hard copy to be scanned for use with assistive devices and technology resources, such as text-to-speech software for the school, plus the required sets of regular-print student booklets

Special Provisions for English Language Learners

“Special provisions” are adjustments to the setting and/or timing of the assessment for English language learners. These provisions do not affect the validity or reliability of the assessment results for these students.

Exemptions (Primary, Junior and OSSLT only)

If a Grade 3 or 6 student is unable to participate in all or part of the assessment, even with accommodations or special provisions, the student may be exempted. A Grade 3 or 6 student must be exempted from

- reading, if the student has to be read to by a teacher or another adult, and
- mathematics, if mathematics terms have to be defined.

All students working toward a Grade 9 academic- or applied-level mathematics credit must participate.

If a student's Individual Education Plan (IEP) states that he or she is not working toward an Ontario Secondary School Diploma (OSSD), the student may be exempted from the OSSLT.

Deferrals (OSSLT only)

All Ontario secondary-school students must write the OSSLT in their Grade 10 year. However, this requirement can be deferred for a year (one or more times), even though the student is working toward an OSSD, if one of the following applies:

- the student has been identified as exceptional by an Identification, Placement and Review Committee (IPRC) and is not able to participate in the assessment, even with permitted accommodations;
- the student has not yet acquired the reading and writing skills appropriate for Grade 9;
- the student is an English language learner and has not yet acquired a level of proficiency sufficient to participate in the test or
- the student is new to the board, requires accommodations and appropriate accommodations cannot be provided.

All deferred students must eventually complete the literacy requirement prior to graduation.

Quality Assurance

EQAO has established quality assurance procedures to help ensure that its assessments are administered consistently and fairly across the province and that the data produced is valid and reliable. EQAO follows a number of procedures to ensure that parents, educators and the public have confidence in the validity and reliability of the results reported.

- *Quality assurance monitors:* EQAO contracts quality assurance monitors to visit and observe the administration of the assessments (in a random sample of schools), to determine the extent to which EQAO guidelines are being followed.
- *Database analyses:* EQAO conducts two types of statistical analyses of student response data. The first analysis identifies student response patterns that suggest the possibility of collusion between two or more students. The second examines unusual changes in the proportion of primary, junior and Grade 9 students in a school performing at or above the provincial standard (Level 3) over time and overall patterns of school results for open-response items. In the case of the OSSLT, the data are analyzed for unusual changes in the passing rate of a school.

- *Examination of test materials:* EQAO looks for evidence of possible irregularities in the administration of assessments in schools, following each assessment. This is done through an examination of test materials from a random sample of schools prior to scoring.

CHAPTER 4: SCORING

EQAO follows rigorous scoring procedures to ensure that its assessment results are valid and reliable. All responses to open-response field-test and operational reading and mathematics items, as well as writing prompts, are scored by trained scorers. The multiple-choice items, except on the primary assessment, are captured by a scanner. On the primary assessment, students circle the correct option to multiple-choice items in the booklets and their choices are keyed in manually.

Item-specific scoring rubrics and anchors are the key tools used for scoring open-response reading, writing and mathematics items. Anchors illustrate the descriptors for each code in the rubrics. In order to maintain consistency across items and years, item-specific rubrics for open-response items are based on generic rubrics. Each student response to an open-response item is scored according to its best match with one of the code descriptors in the rubric for the item and its anchors. EQAO scoring rubrics describe work at different codes or score points; each code represents a different quality of student performance. The anchors are chosen and validated by educators from across the province during the range-finding process. Scorers are trained to refer constantly to the anchors to ensure consistent scoring. The rubric codes are related to, but do not correspond to, the levels of achievement outlined in the Achievement Charts in the Ministry of Education curriculum documents.

The generic rubrics used to create item-specific rubrics are included in each framework document for each assessment on the EQAO Web site, www.eqao.com.

The main stages of the scoring process are outlined below.

The Range-Finding Process

Range finding is used to define the range of acceptable performances for each code or score point in each scoring rubric, which is used to train scorers. The process is completed in two stages: pre-range finding and range finding.

Range finding for open-response reading and mathematics items and short-writing prompts uses student field-test responses and occurs prior to field-test scoring. Field-test scoring follows operational scoring for the primary, junior and Grade 9 assessments. Field-test scoring for the OSSLT occurs during the summer, after operational scoring has finished.

The long-writing prompts on the primary- and junior-division assessments and on the OSSLT are field trialed with a limited number of students. As a result, range finding for long-writing prompts uses student responses to operational items and occurs just prior to operational scoring.

Pre-Range Finding

During pre-range finding, EQAO selects responses that represent the full range of codes or score points for each item or prompt. These responses are used by the range-finding committee. An overview of the process is provided below. A few minor variations of this process occur across

assessments and between field-test and operational range finding. However, the following points describe the process:

1. EQAO education officers are responsible for pre-range finding.
2. Once student booklets arrive at EQAO from schools, a purposeful, demographically representative sample of about 500 student responses for each open-response field-test reading or mathematics item, short-writing prompt and operational long-writing prompt is set aside for pre-range finding.
3. Education officers read through 25 booklets, or more if necessary, to see if there is a range of responses and if the item or prompt worked with students. The pre-range finding process for items or prompts does not proceed unless there is a range of responses.
4. Typically, booklets are sorted into four piles based on the range of responses: approximately 20 low, 20 medium, 20 high and 25 mixed. The booklets chosen for the piles represent the full range of student responses, including off-topic, incorrect, typical and unusual responses (but not blanks). The mixed pile is constructed after the other three piles have been determined.
5. Responses that are blank or difficult to read, due to poor handwriting or light ink, are not selected.
6. Four cover sheets (showing the item and booklet numbers) for each item or prompt are printed and labelled as high, medium, low and mixed.

Range Finding

During the range-finding process, subject experts from the Ontario education system meet to prepare high-quality scoring tools and training materials for scorers, in order to ensure the accurate, reliable and consistent scoring of open-response items on EQAO assessments. These experts select representative samples of student responses to define and illustrate the range of student performances within the scoring rubric codes and provide consensus on the coding of student responses used to train scorers of open-response items.

Range-finding committees, consisting of 8–25 Ontario educators, meet up to three times a year to make recommendations about student responses that will be used as anchors during scoring. They also discuss other possible responses to be used as training materials for scorers (e.g., training papers, qualifying test responses and possible responses for calibration activities during scoring).

The qualifications for range-finding committee members include

- expertise and experience in the application of rubrics based on the achievement charts in *The Ontario Curriculum* (to identify varying levels of student performance in language and mathematics);
- the ability to explain clearly and concisely the reasons why a student response is at one of the codes in a rubric and
- expertise in and current experience with the curriculum and the grades being assessed:
 - primary-division assessment: language and mathematics in the primary division (Grades 1 to 3),
 - junior-division assessment: language and mathematics in the junior division (Grades 4 to 6),

- Grade 9 mathematics assessment: mathematics in the intermediate division (Grades 7 to 10) and
- OSSLT: literacy across the curriculum in the intermediate division (Grades 7 to 10).

Members of the range-finding committees

- use their scoring expertise to assign the appropriate generic rubric and/or item-specific rubric codes to a set of student responses for each group of assessment items;
- share the codes they have assigned with other members of the committee;
- work collaboratively with other members of the committee, under the guidance of an EQAO education officer, to reach consensus on appropriate codes for each student response used to train scorers;
- make recommendations for refinements to the item-specific rubrics and
- suggest wording for the annotations explaining the codes assigned.

Overview of the Range-Finding Process

1. Range-finding committee members (including subject experts and current classroom teachers) are recruited and selected for each assessment.
2. Range-finding committee meetings are facilitated by EQAO education officers. After thorough training, the committee is often divided into groups of three or four members.
3. Each group discusses a set of items, prompts and associated item-specific scoring rubrics and recommends appropriate responses to be used as anchors, training papers and qualifying test items to train scorers for each task.
4. The discussions focus on
 - the content and requirements of each item or task;
 - consensus scores by the group for student responses and
 - scoring rules, as required, to ensure consistent scoring of each item or task.

Preparing Training Materials for Scoring: Anchors, Training Papers, Qualifying Tests, Validity Papers

EQAO education officers prepare materials to train scorers for both field-test and operational open-response items. They consider all recommendations and scoring decisions reached during the range-finding process and make final decisions about which student responses will be used for anchors, scorer training, qualifying tests and monitoring the validity (accuracy) and reliability (consistency) of scoring. A full range of training materials is not required for field-test scoring.

Training materials include

- generic or item-specific rubrics,
- anchors that solidly represent the codes in scoring rubrics,
- training papers that represent both solid score-point responses and unusual or less-solid responses (e.g., shorter than normal, atypical approaches, mix of very low and very high attributes),
- annotations for each anchor and training paper used,
- solid score-point responses for one or more qualifying tests,

- possible responses used for ongoing training during the daily calibration activity (operational scoring only) and
- responses used for monitoring validity (operational scoring only).

Field-Test Scoring

Field-test scoring occurs following operational scoring. Since these items will be used in future assessments, field-test items are scored according to the same high standards applied to the scoring of operational items. The most reliable and productive scoring leaders and scorers of operational items are selected to score items similar to those they scored on the operational assessment, to ensure consistency of year-to-year scoring and to reduce the time required to train scorers. Education officers arrange for sufficient copies of materials to train scorers of field-test items.

Scoring Open-Response Field-Test Responses

A sample of approximately 1200 demographically representative English- and 500 French-language student responses for each field-test item or prompt is scored. Student confidentiality is maintained at all times, and all training materials are kept secure.

In-depth training for the first item and prompt are provided to scorers by their scoring leader. When field-test scoring does not immediately follow operational scoring, scorers write a qualifying test on the first item or prompt before scoring begins. Scorers are trained on each item and complete scoring of one item before proceeding to the next. There are no qualifying tests for subsequent items and prompts.

Item-analysis statistical reports are prepared following field-test scoring. These reports, together with scorer comments related to field-test item performance, are used to inform test construction.

Training Field-Test Scoring Leaders and Scorers

Field-test scorers and leaders are trained on the scoring requirements of field-test items, prompts and item-specific rubrics in order to produce valid and reliable item- and prompt-specific data for operational test construction.

Scoring leaders for each scoring room (open-response reading and mathematics items and short-writing prompts) are trained by EQAO education officers. These scoring leaders then train scorers. Training includes

- an introduction to the purpose of field-test scoring,
- an explanation of the need to report suspected abuse to the Children’s Aid Society,
- a grounding in field-test scoring procedures (using the first item or prompt and its scoring rubric, anchors and training papers),
- a qualifying test on the first item or prompt (when field-test scoring does not immediately follow operational scoring) and
- an introduction to subsequent items and prompts and their scoring rubric, anchors and training papers prior to the scoring of each item.

Standards for passing the qualifying test are the same as those for scoring operational items.

Developing Additional Scoring Training Materials Before Scoring Operational Items

When a full range of training materials has not been used for field-test scoring of open-response reading and mathematics items or writing prompts, EQAO develops additional scoring materials using the original range-finding data or field-test scoring data. In the latter case, education officers bundle student responses so that range finders can select additional scorer-training materials, such as anchors, training papers or qualifying tests, for operational scoring.

Education officers are responsible for arranging all of the materials required to train the scorers who will score operational items.

Scoring Open-Response Operational Items

EQAO has rigorous policies and procedures for the scoring of operational assessment items and prompts to ensure the reliability of assessment results.

The primary, junior and Grade 9 assessments are scored by qualified Ontario educators. The primary- and junior-division assessments are scored by educators representing all primary and junior grades. The Grade 9 Assessment of Mathematics is scored by educators with expertise in mathematics and experience working with Grade 9 students. Scoring provides teachers with valuable professional development in the area of understanding curriculum expectations and assessing student achievement.

The OSSLT is scored during the school year. EQAO recruits as many teacher scorers as possible (members of the Ontario College of Teachers) and completes the complement of required scorers with other-degree personnel. As part of the initial screening process administered by the contractor that recruits the other-degree scorers, applicants write a test to ensure that they have sufficient proficiency in English or French in order to effectively score the test.

Scoring Rooms for Scoring Open-Response Operational Items

A limited set of the operational assessment items is scored in each scoring room under the leadership of a scoring leader, who trains all scoring supervisors and scorers in the room. Scoring leaders, with the assistance of the scoring supervisors, manage the training, scoring and retraining of the scorers. All scorers are trained to use the EQAO scoring guide (rubrics and anchors) for each item they are scoring. Following training, scorers must pass a qualifying test. The validity (accuracy) and reliability (consistency) of scoring is tracked daily at the scoring site, and retraining occurs when required. All scoring procedures are conducted under the supervision of EQAO's program managers and education officers.

Scorers sit alone at tables and score individually. Scorers can discuss anomalous responses with their scoring leader or supervisor.

Operational open-response reading, writing and mathematics items for the primary- and junior-division assessments and operational mathematics items for the Grade 9 assessment are single scored. To measure overall scorer reliability (consistency), scores from the sets of validity booklets are used for those items that are not part of the validity process. More information about scoring validity is provided later in this chapter.

Each open-response reading item and writing prompt on the OSSLT is scored independently by two trained scorers using the same rubric. A “blind scoring” model is used; that is, scorers do not know what score has been assigned by the other scorer. The routing system automatically ensures that responses are read by two different scorers. If the two scores are in exact agreement, that score is assigned to the student. If the two scores are adjacent, the higher score (for reading and short-writing responses) or the average of the two scores (for news reports and paragraphs expressing an opinion) is assigned to the student. If the two scores are non-adjacent, the response is scored again by an expert scorer to determine the correct score for the student. This rigour ensures that parents, students and teachers can be confident that all students have received valid scores.

“Exact agreement” means that the code or score point assigned to an open-response item by a pair of scorers is exactly the same. “Adjacent” means that there is a difference of one score point between the codes assigned to an open-response item by a pair of scorers. “Non-adjacent” means that there is a difference of more than one score point between the codes assigned to an open-response item by a pair of scorers.

Training for Scoring Open-Response Operational Items

The purpose of training is 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, resulting in valid (accurate) and reliable (consistent) student scores.

Training of Scoring Leaders and Scoring Supervisors for Scoring Open-Response Operational Items

Scoring leaders must have subject expertise and be, first and foremost, effective teachers of adults. They must encourage scorers to abandon preconceived notions about scoring procedures and align their thinking and judgment to the procedures and scoring materials for the items being scored. The responsibilities of scoring leaders include

- training all scoring supervisors and scorers in the room;
- overseeing the scoring of items;
- ensuring that scoring materials are applied consistently and
- resolving issues that arise during scoring.

Scoring leaders are also responsible for reviewing and analyzing daily data reports to ensure that a high quality of scoring occurs in their scoring rooms.

Scoring supervisors are selected from a pool of experienced and proficient EQAO scorers. Scoring supervisors assist scoring leaders and ensure that their assigned scorers are qualified and are scoring accurately. Scoring supervisors may also be asked to retrain individual scorers when necessary.

The training for scoring leaders and scoring supervisors is conducted before scoring begins. EQAO education officers train scoring leaders and oversee the training of scoring supervisors. Scoring leaders then train scoring supervisors. Supervisor training is substantially similar to the training and qualifying for scorers. The only difference is that supervisors receive additional

training regarding scoring materials, room management problems and issues that may arise during scoring. For Grade 9 scoring, an EQAO education officer trains the scoring leaders and supervisors for a room at the same time.

Following training and prior to scoring, scoring leaders and scoring supervisors must pass a qualifying test which involves scoring 14 to 20 student responses for the item(s) to be scored in their rooms. The items included in the qualifying tests are selected during the range-finding process. Scoring leaders and supervisors must attain at least an 80% exact and an 100% exact-plus-adjacent match with the expertly assigned scores. Scoring leaders or supervisors who fail the qualifying test may not continue in the role of leader or supervisor.

Training of Scorers for Scoring Open-Response Operational Items

The purpose of training for open-response operational items is to ensure that all scorers become experts in scoring specific items or subsets of items. All operational items require a complete set of scoring materials: item-specific rubrics, anchors (actual student responses illustrating work at each code on the rubric) and their annotations, training papers, a qualifying test, validity booklets and items for the daily calibration activity.

To obtain high levels of validity (accuracy) and reliability (consistency) during scoring, EQAO adheres to stringent criteria for selecting, training and qualifying scorers. Various other quality control procedures (outlined below) are used during the scoring process to identify scorers who need to be retrained or possibly dismissed from scoring.

EQAO scoring procedures and processes have been developed to ensure validity and reliability in scoring. The purpose of training is 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, resulting in accurate and consistent scoring of student responses. All scorers in a room are trained to score using the same scoring materials. These scoring materials are approved by EQAO and cannot be altered. During training, scorers are told they may have to adjust their thinking about scoring student performance in a classroom setting in order to accept EQAO's standards and practices for scoring assessments.

Training for scorers on the limited number of open-response items scored in a room takes approximately half a day and includes the following:

- a brief overview of the assessment being scored (including the purpose and goal of the assessment and the unique features of the assessment and assessment population);
- general instructions about the security, confidentiality and suitability of the scoring materials;
- instruction on entering scores into the Personal Digital Assistant (PDA) used to collect scoring data:
 - prior to entering the scores, scorers scan the unique student booklet barcodes into the PDA (which has a built-in barcode scanner) in order to link student names to their corresponding scores and
 - scorers enter their scores for student responses into the PDA, then synchronize the PDA in a cradle connected to a laptop, which uploads the data to the server;

- a thorough review and discussion of the scoring materials (item, generic and/or item-specific rubric, anchors and their annotations) for each item to be scored:
 - emphasis is placed on the scorer's understanding of how the responses differ in incremental quality and how each response reflects the description of its code on the rubric;
 - the anchors consist of responses that are typical (rather than unusual or uncommon), solid (rather than controversial or borderline) and true (meaning they cannot be changed by anyone);
- scoring of a series of training papers, consisting of selected expertly scored student responses:
 - training papers typically contain responses that are solid as well as responses that are unusual and/or less solid (e.g., are shorter than normal, employ atypical approaches, contain both very low and very high attributes);
 - scorers will first score the responses then compare their scores with the annotated correct scores provided;
 - scorers will then discuss the annotation and characteristics of each correct response with their scoring leader and supervisor and
 - scorers will internalize the rubric during this process and adjust their individual scoring to conform to it.

Scorers are also trained to

- read responses in their entirety prior to making any scoring decisions;
- view responses as a whole without focusing on particular details (such as spelling);
- remain objective and fair, and view the whole response through the filter of the rubric and
- score all responses in the same way so that they do not adjust the scoring to take into account a particular characteristic they think they know about the student (e.g., special education needs, English language learner).

Following training and prior to scoring, scorers must pass a qualifying test consisting of 14 to 20 student responses to all the items to be scored in the room. These items are selected during the range-finding process as examples of solid score points for rubrics. Scorers must attain at least a 70% exact match with the expertly assigned score. This ensures that scorers have understood and can apply the information they received during training. Scorers who fail the qualifying test the first time may undergo further training and write the test a second time. Scorers who fail to pass the qualifying test a second time are dismissed.

Procedures at the Scoring Site

Students at Risk

On occasion, a student's response to an open-response question will contain evidence that he or she may be at risk (e.g., content that states or implies threats of violence to self or others, or possible abuse or neglect). Copies of student responses that raise concerns are sent to the student's local Children's Aid Society. It is the legal responsibility and duty of scorers, in consultation with their site managers, to inform the Children's Aid Society of such cases.

Inappropriate Content, Cheating and Other Issues

Student responses to open-response questions occasionally contain inappropriate content or evidence of possible teacher interference or other issues. Booklets containing these issues are sent to the exceptions room for consideration.

Offensive Content: Obscene, racist or sexist content in student response booklets are reviewed by EQAO staff to determine whether the school should be contacted. If the offensive content warrants the notification of the school, EQAO will contact the school.

Cheating: When there is any evidence in a booklet that may indicate some form of irregularity (e.g., many changed answers, evidence of possible teacher interference), the booklets are reviewed by EQAO staff to determine whether the school should be notified.

Damaged or Misprinted Booklets: In very few cases, booklets given to students are torn or stapled incorrectly, or have missing pages or a defaced barcode that cannot be scanned. In such cases, students are not penalized. These damaged booklets are further reviewed by EQAO staff to determine whether the results of these booklets should be pro-rated based on the results of responses unaffected by such problems.

Ongoing Daily Training

Scoring leaders provide clarification on scoring of specific items and key elements of item-specific rubrics in their scoring rooms. EQAO conducts morning and afternoon training to refresh scorers' understanding of the scoring materials and to ensure that they apply the scoring materials accurately and consistently from one day to the next, and before and after lunch breaks.

Daily Morning Review of Anchors

Scoring leaders begin each day with a review of all or a portion of the rubric and anchors. The purpose of the review is to refocus scorers and highlight any section of the rubric that requires attention. This review is more comprehensive after a weekend break (or following any extended break).

Daily Afternoon Calibration Activity

Scorers begin each afternoon by scoring one or more of the selected calibration items (expertly scored student responses that were challenging to score). Calibration items facilitate the review of and response to scoring issues raised by scorers or the daily scoring data reports. Scorers score and record the scores for the calibration items. Scoring leaders review the calibration item scores and provide scorers with an explanation of the issues raised and clear information and guidance on the correct way to score these responses. Individual scorers or groups of scorers that encounter difficulty with daily afternoon calibration items can address their issues with their scoring leader or scoring supervisor.

Daily Scoring Centre Reports for Monitoring the Quality of Open-Response Item Scoring

Scoring leaders and supervisors receive daily data reports showing daily and cumulative validity, reliability and productivity data for individual scorers and for groups of scorers in their room. These data reports are described below.

Daily and Cumulative Validity

During scoring, EQAO tracks the validity (accuracy) of scorers through the use of expertly scored validity booklets. Scorers score up to 10 validity booklets a day. The validity booklets ensure that scorers are giving correct and accurate scores by comparing their scores with those assigned during the range-finding process. Scoring leaders and supervisors use the report data to determine whether scorers are drifting from the scoring standards (provided by scorer training and training materials) and whether any retraining is required. During scoring, all scorers must maintain a minimum accuracy rate on the validity booklets. The accuracy rates are as follows: 75% exact and 95% adjacent for three-point rubrics, 70% exact and 95% exact-plus-adjacent agreement for four-point rubrics and an accuracy rate of 60% exact and 95% exact-plus-adjacent agreement for six-point rubrics.

“Exact agreement” means that the code or score point assigned to an open-response item by a pair of scorers is exactly the same. “Adjacent” means that there is a difference of one score point between the codes assigned to an open-response item by a pair of scorers. “Non-adjacent” means that there is a difference of more than one score point between the codes assigned to an open-response item by a pair of scorers.

The reports summarize daily and cumulative level-of-agreement data (exact, adjacent and high or low non-adjacent agreement) on validity booklets with preset scores. Adjacent-low means that the score assigned to a certain response by a scorer is one point below the score assigned by the expert panel. Adjacent-high means that the score is one point above the score given by the expert panel, and non-adjacent means that the difference between the scores assigned by the scorer and the expert panel is greater than one score point.

The reports also include a cumulative-trend review and are summarized by item or item set, rubric, room, group and scorer. Scorers are listed from low to high validity. Scorers not meeting the exact-agreement requirement are highlighted in the report.

Accuracy is measured primarily by the use of validity metrics. The daily data reports for scorers who pass the qualifying test after retraining are carefully monitored to ensure that they continue to meet standards. If, after a minimum of 10 validity items, a scorer falls below the required exact-plus-adjacent-agreement standards, the scorer receives retraining (including a careful review of the anchors). If retraining does not correct the situation, the scorer may be dismissed. The scores of dismissed scorers are audited and, if necessary, rescored.

Daily and Cumulative Mean Score and Score-Point Distribution

Daily and cumulative mean-score and score-point-distribution data are used to monitor room and individual-scorer drift. These data reports guide ongoing training (calibration) at both the individual and room levels. These data are used to confirm validity results and group-level interventions and are not intended to identify scorers in need of retraining.

These reports identify and summarize (by item or item set, room, group, rubric and scorer) the daily and cumulative mean score and the distribution of assigned score points.

Daily and Cumulative Reliability Reports

Open-response primary- and junior-division and Grade 9 assessment items in reading, writing and mathematics are single scored. To measure overall scorer reliability (consistency) during scoring, scores from the sets of validity booklets are used for items that are not part of the validity process. All open-response OSSLT items are routed for a second scoring, which is used to monitor interrater reliability. Scoring leaders and supervisors use this secondary source of data, in conjunction with validity data, to determine whether individual scorers or groups of scorers require retraining.

The reports identify and summarize daily and cumulative levels of interrater agreement on two independent scorings, including exact, adjacent and high and low non-adjacent agreement. The reports are summarized by item or item set, room, group, rubric and scorer, and scorers are listed from low to high reliability. Scorers not meeting the exact-agreement requirement are highlighted in the report.

Daily and Cumulative Productivity

During scoring, EQAO tracks scoring-centre productivity and monitors progress through daily productivity reports to ensure that all scoring will be completed during the scoring session. The reports show the number and percentage of responses for which the scoring is complete. These reports, which are provided to scoring leaders and supervisors, report daily and cumulative productivity. The reports also track the productivity of each scorer to ensure that daily targets and minimums are met. Productivity targets and minimums are set for each room, taking into consideration the subset of items being scored. Productivity targets for scoring supervisors are 50% to 60% of scorer targets.

The reports are summarized by room, group and individual scorer and include the daily and cumulative number of student responses scored and a cumulative-trend review. The reports list scorers from low to high productivity. Scorers not meeting the minimum productivity rate for the room in which they are scoring are highlighted in the report. Scoring leaders and supervisors review the data highlighted in this report to determine whether retraining is required for any scorer.

Scoring completion reports also compare, on a daily and cumulative basis, the number of scorings completed with completion targets for the scoring room. The report is used to ensure that all scoring is completed during the scoring session.

Aggregated Daily and Cumulative Individual Scorer Data

These reports combine validity data with secondary data for each scorer. The reports include not only daily and cumulative validity data but also daily and cumulative reliability, mean score and productivity data. The reports list scorers from low to high validity. Scorers not meeting the exact-agreement requirement of 75% on three-point rubrics, 70% on four-point rubrics or 60%

on six-point rubrics are highlighted in this report. As such, this report assists the scoring leaders in identifying the scorers that require retraining.

Required Actions: Consequences of the Review and Analysis of Daily Scoring-Centre Data Reports

Scoring leaders are responsible for the daily review and analysis of all scoring-centre data reports to ensure the quality of the scoring in their scoring room. EQAO personnel (the chief assessment officer, director of assessment and reporting and education officers) also review the daily reports and work with scoring leaders to identify

- individual scorers who need retraining,
- groups of scorers who need retraining,
- calibration items that will ensure quality scoring,
- issues arising that require additional training for an entire room and
- productivity issues.

Scoring leaders share the data and data-related issues with the appropriate scoring supervisors so that interventions can be planned to ensure the quality of individual and group scoring. The following occurs when a scorer is not meeting the validity metrics:

- the scorer is retrained and re-qualified if the exact-plus-adjacent standard is not met or
- the scorer is retrained and participates in recalibration if the exact-agreement requirement is not met.

Scorers, as well as their leaders and supervisors, are required to demonstrate their ability to score student responses accurately and consistently throughout training, qualification and scoring. Scoring supervisors and scorers must meet EQAO standards for validity, reliability and productivity in order to continue. If a scoring supervisor or scorer does not meet one or more of these criteria standards, he or she will receive retraining. If his or her scoring does not improve, the scoring supervisor or scorer may be dismissed. All retraining is documented by scoring leaders and supervisors.

Audit

EQAO audits individual student score sheets for inconsistencies that may indicate incomplete scoring. Any booklet scored entirely blank is rerouted for scoring a second time.

Scorer Reliability and Validity

The procedures used for estimating the reliability and validity of EQAO assessments are summarized below. The estimates of interrater reliability are presented in Tables 4.1 to 4.21, and those of validity are presented in Tables 4.22 to 4.42. Two sets of results are reported for each writing prompt: one for topic development and one for conventions.

Scorer Reliability

Test reliability is affected by different sources of measurement error. In the case of open-response items, inconsistency in scoring is the source of error. As part of its attempt to minimize

this type of error, EQAO tracks reliability using scores assigned to items in the pre-selected validity booklets that are not part of the validity process.

The percentage of agreement between the scores awarded by a pair of scorers is known as the interrater reliability. Four indices are used to identify interrater reliability: percentage of exact agreement, percentage of exact-plus-adjacent agreement, percentage of adjacent agreement and percentage of non-adjacent scores.

“Exact agreement” means that the code or score point assigned to an open-response item by a pair of scorers is exactly the same. “Adjacent agreement” means that there is a difference of one score point between the codes assigned to an open-response item by a pair of scorers. “Non-adjacent agreement” means that there is a difference of more than one score point between the codes assigned to an open-response item by a pair of scorers.

The process for determining interrater reliability for the OSSLT does not require the reinsertion of selected booklets because all student responses to open-response items are scored by at least two scorers. OSSLT interrater reliability is reflected using the scores from the two scorers and is measured by the same four indices described above.

Scoring reliability estimates for the primary, junior and Grade 9 assessments and the OSSLT are presented in the following sections.

Assessment of Reading, Writing and Mathematics: Primary Division

There are 10, three and eight open-response items for the reading, writing and mathematics components of the assessment, respectively. Four-point scoring rubrics are used for reading and mathematics. For the writing component, there are two short-writing items and one long-writing item that are scored for topic development and use of conventions. A four-point scoring rubric is used for topic development and a three-point scoring rubric for conventions. The scoring reliability estimates for reading, writing and mathematics for the primary division are displayed in Tables 4.1 to 4.7. The statistics are provided for each item and for the aggregate of the items for each assessment. For writing, the aggregate statistics for short-writing prompts, long-writing prompts and all prompts are provided.

For the reading component, the percentage of exact agreement for English-language individual items ranged from 67.5% to 85.7%, and from 56.8% to 88.6% for French-language items (Tables 4.1 and 4.5). The percentage of exact-plus-adjacent agreement was larger and less variable, ranging from 96.8% to 99.8% for English-language and 94.1% to 100.0% for French-language items. The corresponding aggregate values were 72.6% and 98.3% for the English-language version of the assessment and 74.7% and 98.6% for the French-language version of the assessment. With the exception of four items on the English-language assessments and three items on the French-language assessment, these indices met EQAO’s target of 70% exact agreement. All English-language items met the EQAO target of 95% for exact-plus-adjacent agreement, and all but one French-language item met this target.

The statistics for the writing component of the assessment are displayed in Tables 4.2 and 4.6. The aggregate percentages of exact agreement were 71.1% for the English-language version of

the assessment and 60.2% for the French-language version of the assessment. However, as with the reading component, the percentages of exact-plus-adjacent agreement were larger. They ranged from 97.3% to 99.5% for the English-language version and from 87.7% to 100.0% for the French-language version, meeting the EQAO target for all rubrics. The percentages of exact-agreement were below the EQAO targets (70% for the four-point rubrics for topic development and 75% for the three-point rubrics for conventions) for the French-language assessment but were met for all but two rubrics for the English-language assessment. The target of 95% for exact-plus-adjacent agreement was met for all rubrics in the English-language assessment, and only two rubrics in the French-language assessment did not meet the target.

For the mathematics component of the primary-division assessment, the mathematics items on the English-language assessment are translated into French and administered to students in the French Immersion program in some school boards. Since the booklets for these French Immersion students are scored by a separate group of scorers, reliability results are reported separately. The percentages of exact- and exact-plus-adjacent agreement for the mathematics items were large (Tables 4.3, 4.4 and 4.7). The aggregate percentages of agreement were, respectively, 77.9% and 97.3% for English-language items, 73.1% and 96.4% for French Immersion items and 80.9% and 97.9% for French-language items. The indices for all items met the EQAO target of 70% exact-agreement, and the indices for most items met the 95% target for exact-plus-adjacent agreement, except for one French-language item and one French Immersion item.

Table 4.1 Interrater Reliability Estimates for Reading: Primary Division (English)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
EQRE20994.083	1	5	587	99.8	85.7	14.1	0.2
EQRE20999.083	1	6	730	99.5	73.7	25.8	0.5
EQRE20988.082	1	11	762	98.3	73.2	25.1	1.7
EQRE20983.082	1	12	738	98.0	69.2	28.7	2.0
EQRE20827.070	2	11	588	99.5	78.9	20.6	0.5
EQRE20821.070	2	12	541	98.2	71.3	26.8	1.8
EQRE20939.078	2	21	724	98.1	68.6	29.4	1.9
EQRE20945.078	2	22	561	96.8	72.7	24.1	3.2
EQRE21034.086	2	28	834	98.0	68.6	29.4	2.0
EQRE21030.086	2	29	633	97.3	67.5	29.9	2.7
Aggregate			6698	98.3	72.6	25.7	1.7

Table 4.2 Interrater Reliability Estimates for Writing: Primary Division (English)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
EQWE30434T	2	23	791	97.3	63.7	33.6	2.7
EQWE30434V	2	23	931	99.5	71.3	28.1	0.5
Aggregate Long Writing			1722	98.4	67.5	30.9	1.6
EQWE20250T	1	13	565	98.2	76.1	22.1	1.8
EQWE20250V	1	13	784	99.1	77.3	21.8	0.9
EQWE20245T	1	14	539	96.8	70.9	26.0	3.2
EQWE20245V	1	14	735	98.6	67.1	31.6	1.4
Aggregate Short Writing			2623	98.2	72.8	25.4	1.8
Aggregate Across All Items			4345	98.3	71.1	27.2	1.7

Table 4.3 Interrater Reliability Estimates for Mathematics: Primary Division (English)

Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
EQME20944	7	1216	98.0	77.1	20.9	2.0
EQME20880	8	1208	95.1	74.8	20.3	4.9
EQME20811	9	1216	97.9	76.6	21.2	2.1
EQME20822	10	1262	98.0	85.8	12.2	2.0
EQME10170	27	1237	98.1	78.7	19.4	1.9
EQME20825	28	1238	97.1	75.9	21.2	2.9
EQME20878	29	1153	97.1	77.4	19.8	2.9
EQME20941	30	1232	97.4	77.0	20.4	2.6
Aggregate		9762	97.3	77.9	19.4	2.7

Table 4.4 Interrater Reliability Estimates for Mathematics: Primary Division (French Immersion)

Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
EQME20944	6	158	95.6	74.1	21.5	4.4
EQME20880	7	149	95.3	53.7	41.6	4.7
EQME20811	8	147	100.0	78.9	21.1	0.0
EQME20822	9	146	96.6	74.7	21.9	3.4
EQME10170	26	166	93.4	71.1	22.3	6.6
EQME20825	27	177	97.2	70.1	27.1	2.8
EQME20878	28	156	98.1	82.1	16.0	1.9
EQME20941	29	167	95.2	80.2	15.0	4.8
Aggregate		1266	96.4	73.1	23.3	3.6

Table 4.5 Interrater Reliability Estimates for Reading: Primary Division (French)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
OQLF20486.063	1	5	74	98.6	56.8	41.9	1.4
OQLF20489.063	1	6	42	100.0	71.4	28.6	0.0
OQLF20518.066	1	11	99	97.0	81.8	15.2	3.0
OQLF20519.066	1	12	57	98.2	66.7	31.6	1.8
OQLF20466.061	2	11	48	100.0	87.5	12.5	0.0
OQLF20467.061	2	12	49	98.0	77.6	20.4	2.0
OQLF20509.065	2	21	35	100.0	88.6	11.4	0.0
OQLF20506.065	2	22	60	100.0	75.0	25.0	0.0
OQLF20538.068	2	28	67	100.0	82.1	17.9	0.0
OQLF20539.068	2	29	34	94.1	58.8	35.3	5.9
Aggregate			565	98.6	74.7	23.9	1.4

Table 4.6 Interrater Reliability Estimates for Writing: Primary Division (French)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
OQEF20230T	2	23	87	100.0	57.5	42.5	0.0
OQEF20230V	2	23	129	100.0	66.7	33.3	0.0
Aggregate Long Writing			216	100.0	62.1	37.9	0.0
OQEF20220T	1	13	65	87.7	55.4	32.3	12.3
OQEF20220V	1	13	63	96.8	66.7	30.2	3.2
OQEF20226T	1	14	37	94.6	45.9	48.6	5.4
OQEF20226V	1	14	58	96.6	69.0	27.6	3.4
Aggregate Short Writing			223	93.9	59.2	34.7	6.1
Aggregate Across All Items			439	95.9	60.2	35.8	4.1

Table 4.7 Interrater Reliability Estimates for Mathematics: Primary Division (French)

Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
OQMF20510	7	200	99.5	83.5	16.0	0.5
OQMF20640	8	170	99.4	81.8	17.6	0.6
OQMF20513	9	190	93.7	82.1	11.6	6.3
OQMF10044	10	187	97.9	79.7	18.2	2.1
OQMF20639	27	165	100.0	74.5	25.5	0.0
OQMF20504	28	171	97.1	85.4	11.7	2.9
OQMF20526	29	176	96.0	77.3	18.8	4.0
OQMF20636	30	178	100.0	82.6	17.4	0.0
Aggregate		1437	97.9	80.9	17.1	2.1

Assessment of Reading, Writing and Mathematics: Junior Division

The number of open-response items and scoring rubrics used to score the items are the same for the primary- and junior-division assessments of reading, writing and mathematics. The scoring reliability estimates for reading, writing and mathematics for the junior division are reported in Tables 4.8 to 4.13. The patterns of percentage agreement for the junior-division assessment were similar to the patterns observed for the primary-division assessment.

The aggregate percentages of exact agreement for English-language reading, writing and mathematics were 61.8%, 71.3% and 76.7%, respectively. The percentages for French-language reading, writing and mathematics were 68.9%, 62.5% and 79.2%, respectively. The percentages of exact-plus-adjacent agreement for each subject in both languages were large and less variable; the aggregate percentages ranged from 95.9% to 98.6%.

The EQAO target of 70% exact agreement was not met for any item on the English-language reading component, and it was not met for six items on the French-language reading component. The EQAO target of 95% exact-plus-adjacent agreement was met for six items on the English-language version of the assessment and for all items on the French-language version of the assessment.

The EQAO targets for the percentages of exact agreement (70% for topic development and 75% for conventions) were met for only one rubric for the English-language writing component and were not met for any rubric for the French-language writing component.

The EQAO target of 70% exact agreement was met for most mathematics items, except for three English-language items and one French-language item. The EQAO target of 95% exact-plus-adjacent agreement was met for most mathematics items, except for one French-language item.

Table 4.8 Interrater Reliability Estimates for Reading: Junior Division (English)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
EQRE21160.094	1	5	709	94.9	57.5	37.4	5.1
EQRE21165.094	1	6	881	96.5	62.2	34.3	3.5
EQRE21274.104	1	11	699	96.0	62.5	33.5	4.0
EQRE21275.104	1	12	666	95.5	58.7	36.8	4.5
EQRE21102.091	2	11	787	94.9	59.6	35.3	5.1
EQRE21114.091	2	12	785	96.1	62.4	33.6	3.9
EQRE21243.101	2	21	720	94.7	66.8	27.9	5.3
EQRE21242.101	2	22	722	97.5	58.2	39.3	2.5
EQRE31965.172	2	28	783	98.9	67.3	31.5	1.1
EQRE31964.172	2	29	700	94.4	62.4	32.0	5.6
Aggregate			7452	96.0	61.8	34.1	4.0

Table 4.9 Interrater Reliability Estimates for Writing: Junior Division (English)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
EQWE20322T	2	23	1383	96.5	65.7	30.8	3.5
EQWE20322V	2	23	1247	99.6	71.8	27.8	0.4
Aggregate Long Writing			2630	98.0	68.7	29.3	2.0
EQWE20315T	1	13	967	98.8	67.0	31.7	1.2
EQWE20315V	1	13	938	99.9	74.1	25.8	0.1
EQWE20304T	1	14	994	99.1	72.6	26.5	0.9
EQWE20304V	1	14	999	99.2	76.5	22.7	0.8
Aggregate Short Writing			3898	99.2	72.6	26.7	0.8
Aggregate Across All Items			6528	98.8	71.3	27.6	1.2

Table 4.10 Interrater Reliability Estimates for Mathematics: Junior Division (English)

Item Code	Sequence	No. of Pairs	% Exact Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
EQME21089	7	1264	99.4	83.5	16.0	0.6
EQME21119	8	1121	95.2	68.2	26.9	4.8
EQME21138	9	1170	98.7	79.9	18.8	1.3
EQME21043	10	1218	99.3	88.2	11.2	0.7
EQME21104	27	1247	98.9	68.6	30.3	1.1
EQME10229	28	1200	97.0	75.1	21.9	3.0
EQME21137	29	1198	95.5	64.4	31.1	4.5
EQME21135	30	1200	99.7	85.8	13.9	0.3
Aggregate		9618	98.0	76.7	21.3	2.0

Table 4.11 Interrater Reliability Estimates for Reading: Junior Division (French)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
OQLF20595.072	1	5	68	100.0	80.9	19.1	0.0
OQLF20596.072	1	6	31	96.8	54.8	41.9	3.2
OQLF20637.076	1	11	38	100.0	76.3	23.7	0.0
OQLF20635.076	1	12	71	98.6	54.9	43.7	1.4
OQLF20566.070	2	11	78	100.0	85.9	14.1	0.0
OQLF20568.070	2	12	63	100.0	65.1	34.9	0.0
OQLF20617.074	2	21	53	98.1	66.0	32.1	1.9
OQLF20615.074	2	22	27	100.0	66.7	33.3	0.0
OQLF20667.079	2	28	44	100.0	79.5	20.5	0.0
OQLF20669.079	2	29	54	98.1	50.0	48.1	1.9
Aggregate			527	99.2	68.9	30.4	0.8

Table 4.12 Interrater Reliability Estimates for Writing: Junior Division (French)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
OQEF30611T	2	23	87	98.9	57.5	41.4	1.1
OQEF30611V	2	23	108	100.0	60.2	39.8	0.0
Aggregate Long Writing			195	99.4	58.8	40.6	0.6
OQEF20268T	1	13	45	95.6	60.0	35.6	4.4
OQEF20268V	1	13	73	100.0	67.1	32.9	0.0
OQEF20278T	1	14	49	95.9	59.2	36.7	4.1
OQEF20278V	1	14	65	100.0	70.8	29.2	0.0
Aggregate Short Writing			232	97.9	64.3	33.6	2.1
Aggregate Across All Items			427	98.4	62.5	35.9	1.6

Table 4.13 Interrater Reliability Estimates for Mathematics: Junior Division (French)

Item Code	Sequence	No. of Pairs	% Exact Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
OQMF10131	7	232	94.4	82.8	11.6	5.6
OQMF20796	8	254	95.7	76.4	19.3	4.3
OQMF20718	9	228	99.1	85.1	14.0	0.9
OQMF20844	10	285	94.7	72.6	22.1	5.3
OQMF20710	27	161	97.5	83.2	14.3	2.5
OQMF20747	28	173	97.1	79.8	17.3	2.9
OQMF10203	29	179	97.2	62.0	35.2	2.8
OQMF10127	30	162	98.8	92.0	6.8	1.2
Aggregate		1674	96.8	79.2	17.6	3.2

Grade 9 Assessment of Mathematics: Academic and Applied

The Grade 9 Assessment of Mathematics has separate English-language and French-language versions for students enrolled in academic and applied courses. The assessment is administered in January for students completing their courses in the first semester and in June for students in second-semester and full-year courses. The scoring reliability estimates for the Grade 9 Assessment of Mathematics are presented in Tables 4.14 to 4.17 for both administrations for each course. The tables present statistics for each open-response item and the aggregate for open-response items for each administration. They also include aggregate statistics across the winter and spring administrations because the student responses from both administrations were scored during the same scoring session.

As shown in these tables, the scoring was highly consistent. For example, the aggregate percentages of exact agreement among pairs of scorers ranged from 75.2% to 91.7% across all languages, courses and administrations. The aggregate percentages of exact-plus-adjacent agreement ranged from 95.7% to 99.2%. The percentages of exact agreement exceeded the EQAO targets of 70%, except for three items on the English-language version of the assessment, which were slightly below the target. The percentages of exact-plus-adjacent agreement exceeded the EQAO targets of 95% for most of the items on the assessments of both languages, except six items on the English-language version of the test and two items on the French-language version of the test.

Table 4.14 Interrater Reliability Estimates for Grade 9 Applied Mathematics (English)

Administration	Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
Winter	EQME10643	NR	261	98.9	88.1	10.7	1.1
	EQME21498	15	219	94.5	71.2	23.3	5.5
	EQME21413	NR	232	96.6	62.5	34.1	3.4
	EQME21492	21	242	93.4	76.9	16.5	6.6
	EQME21472	16	224	98.7	74.6	24.1	1.3
	EQME21454	NR	242	97.1	80.2	16.9	2.9
	EQME21429	6	234	97.9	90.2	7.7	2.1
Aggregate			1654	96.7	77.7	19.0	3.3
Spring	EQME21441	NR	275	92.4	80.7	11.6	7.6
	EQME21494	16	266	93.2	68.8	24.4	6.8
	EQME21427	NR	256	98.4	85.2	13.3	1.6
	EQME21425	NR	244	98.8	80.7	18.0	1.2
	EQME10631	15	265	99.2	87.2	12.1	0.8
	EQME21474	21	259	98.0	79.9	18.1	2.0
	EQME21426	5	270	94.0	80.4	13.6	6.0
Aggregate			1835	96.3	80.4	15.9	3.7
Aggregate Across Administrations			3489	96.5	79.0	17.5	3.5

Table 4.15 Interrater Reliability Estimates for Grade 9 Academic Mathematics (English)

Administration	Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
Winter	EQME21256	5	524	95.4	82.6	12.8	4.6
	EQME21275	NR	538	97.2	84.8	12.5	2.8
	EQME21232	NR	551	98.0	84.8	13.2	2.0
	EQME21282	21	529	97.4	85.1	12.3	2.6
	EQME21225	10	499	97.6	74.3	23.2	2.4
	EQME10374	16	495	98.0	74.5	23.4	2.0
	EQME21284	NR	496	97.8	82.1	15.7	2.2
Aggregate			3632	97.3	81.2	16.2	2.7
Spring	EQME21213	4	449	95.8	79.5	16.3	4.2
	EQME21259	NR	458	90.8	69.2	21.6	9.2
	EQME21239	NR	430	96.7	72.1	24.7	3.3
	EQME21241	21	451	95.6	76.3	19.3	4.4
	EQME21219	9	616	96.3	74.0	22.2	3.7
	EQME21276	15	638	96.2	71.9	24.3	3.8
	EQME10476	NR	614	98.5	83.6	15.0	1.5
Aggregate			3656	95.7	75.2	20.5	4.3
Aggregate Across Administrations			7288	96.5	78.2	18.3	3.5

Table 4.16 Interrater Reliability Estimates for Grade 9 Applied Mathematics (French)

Administration	Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
Winter	OQMF21364	NR	37	97.3	91.9	5.4	2.7
	OQMF10269	6	46	97.8	89.1	8.7	2.2
	OQMF21420	NR	41	100.0	92.7	7.3	0.0
	OQMF10313	NR	50	98.0	90.0	8.0	2.0
	OQMF21333	NR	14	100.0	92.9	7.1	0.0
	OQMF10296	13	16	100.0	75.0	25.0	0.0
	OQMF21390	20	12	100.0	100.0	0.0	0.0
Aggregate			216	99.0	90.2	8.8	1.0
Spring	OQMF21392	NR	45	100.0	97.8	2.2	0.0
	OQMF10269	6	44	97.7	88.6	9.1	2.3
	OQMF21365	20	44	100.0	93.2	6.8	0.0
	OQMF21363	NR	41	100.0	95.1	4.9	0.0
	OQMF21333	NR	30	100.0	86.7	13.3	0.0
	OQMF21038	NR	32	96.9	84.4	12.5	3.1
	13536	13	26	100.0	96.2	3.8	0.0
Aggregate			262	99.2	91.7	7.5	0.8
Aggregate Across Administrations			478	99.1	91.0	8.2	0.9

Table 4.17 Interrater Reliability Estimates for Grade 9 Academic Mathematics (French)

Administration	Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
Winter	OQMF20924	15	58	87.9	77.6	10.3	12.1
	OQMF20919	11	58	96.6	87.9	8.6	3.4
	OQMF10397	6	54	98.1	83.3	14.8	1.9
	OQMF10370	21	55	92.7	83.6	9.1	7.3
	OQMF20905	NR	75	98.7	90.7	8.0	1.3
	OQMF20926	NR	79	98.7	86.1	12.7	1.3
	OQMF20913	NR	76	100.0	82.9	17.1	0.0
Aggregate			455	96.1	84.6	11.5	3.9
Spring	OQMF10367	NR	103	97.1	90.3	6.8	2.9
	OQMF20925	15	97	99.0	90.7	8.2	1.0
	OQMF10380	5	93	96.8	83.9	12.9	3.2
	OQMF20920	NR	109	96.3	79.8	16.5	3.7
	OQMF20912	NR	112	100.0	92.9	7.1	0.0
	OQMF10353	21	140	99.3	84.3	15.0	0.7
	OQMF20918	10	142	95.8	85.9	9.9	4.2
Aggregate			796	97.7	86.8	10.9	2.3
Aggregate Across Administrations			1251	96.9	85.7	11.2	3.1

Ontario Secondary School Literacy Test (OSSLT)

The scoring reliability estimates for the OSSLT are presented in Tables 4.18 to 4.21. All of the open-response items included in both language versions of the OSSLT are scored twice. The percentages of exact agreement for reading ranged from 58.7% to 70.9% for the English-language version and from 55.8% to 64.9% for the French-language version. As with the other assessments, the percentages of exact-plus-adjacent agreement were larger and less variable. The aggregate percentages were 96.1% for the English-language version and 95.2% for the French-language version. The percentages of exact agreement for all reading items in both English and French were below the EQAO target of 75% for three-point rubrics. The percentage of exact-plus-adjacent agreement met the EQAO target of 95% for three items in English and three items in French.

For writing, each open-response item is scored for content and use of conventions. For the long-writing items, a six-point rubric is used for content and a four-point rubric is used for conventions. For the short-writing items, the content and conventions rubrics were combined into a five-point scale. For long-writing items on both language versions of the test, the percentages of exact agreement ranged from 42.0% to 49.6% for content and from 50.9% to 58.6% for conventions. For short-writing items, the percentages of exact agreement ranged from 54.9% to 63.7%. The percentages of exact-plus-adjacent agreement ranged from 86.9% to 98.4% across all prompts.

The percentages of exact agreement did not meet the EQAO targets of 60% for content and 70% for conventions for any of the long-writing prompts. Also, the percentages of exact agreement

did not meet EQAO targets of 65% for the short-writing prompts. The percentages of exact-plus-adjacent agreement met the EQAO target of 95% for three out of six writing rubrics in English and four out of six writing rubrics in French.

Table 4.18 Interrater Reliability Estimates for Reading: OSSLT (English)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
EQRE21427.116	I	6	185 385	96.8	58.7	38.2	3.2
EQRE21398.113	V	7	185 397	91.7	69.7	22.0	8.3
EQRE21474.120	IX	6	185 473	96.9	67.1	29.8	3.1
EQRE21475.120	IX	7	185 588	99.1	70.9	28.3	0.9
Aggregate			741 843	96.1	66.6	29.6	3.9

Table 4.19 Interrater Reliability Estimates for Writing: OSSLT (English)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
1W3_057T	IV	1	185 381	89.6	44.2	45.5	10.4
1W3_057V	IV	1	185 381	98.3	58.5	39.7	1.7
1W2_054T	VI	1	185 481	94.2	49.6	44.5	5.8
1W2_054V	VI	1	185 481	97.8	58.1	39.6	2.2
Aggregate Long Writing			741 724	95.0	52.6	42.3	5.1
EQWE20375T & V	III	1	185 513	94.7	54.9	39.8	5.3
EQWE20381T & V	X	1	185 612	95.5	59.4	36.1	4.5
Aggregate Short Writing			371 125	95.1	57.1	38.0	4.9
Aggregate Across All Items			1 112 849	95.0	54.1	40.9	5.0

Table 4.20 Interrater Reliability Estimates for Reading: OSSLT (French)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
OQLF20722.083	I	6	6 519	96.0	59.6	36.4	4.0
OQLF20681.080	V	7	6 519	95.0	64.9	30.1	5.0
OQLF20751.085	IX	6	6 521	93.0	55.8	37.1	7.1
OQLF20754.085	IX	7	6 521	97.0	59.5	37.5	3.1
Aggregate			26 080	95.2	60.0	35.3	4.8

Table 4.21 Interrater Reliability Estimates for Writing: OSSLT (French)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
2É3_024T	IV	1	6 508	86.9	42.0	44.9	13.1
2É3_024V	IV	1	6 508	98.3	58.6	39.7	1.8
2É2_025T	VI	1	6 500	87.2	43.8	43.4	12.8
2É2_025V	VI	1	6 500	95.2	50.9	44.3	4.8
Aggregate Long Writing			26 016	91.9	48.8	43.1	8.1
OQEF20359T & V	III	1	6 515	97.2	55.5	41.7	2.8
OQEF20379T & V	X	1	6 500	98.4	63.7	34.8	1.6
Aggregate Short Writing			13 015	97.8	59.6	38.2	2.2
Aggregate Across All Items			39 031	93.9	52.4	41.5	6.2

It is important to note that all student responses on the OSSLT were scored independently by two scorers and that all student responses with non-adjacent scores were adjudicated by an expert scorer. Scores used to compute student results had 100% exact-plus-adjacent agreement for one pair of scorers.

Scoring Validity

As described earlier in this chapter, scoring validity is assessed by having scorers score blind validity booklets. A set of validity booklets that contain scores to student responses assigned by an expert panel is prepared for each group of scorers. These blind validity booklets are circulated as frequently as possible, so that most scorers score one validity booklet per day. Validity is assessed by examining the agreement between the scores assigned by the scorers and those assigned by the expert panel. The following six indices are computed: percentage of exact agreement, percentage of exact-plus-adjacent agreement, percentage of adjacent agreement, percentage of adjacent-low agreement, percentage of adjacent-high agreement and percentage of non-adjacent agreement.

Adjacent-low means that the score assigned to a certain response by a scorer is one point below the score assigned by the expert panel. Adjacent-high means that the score is one point above the score given by the expert panel; and non-adjacent means that the difference between the scores assigned by the scorer and the expert panel is greater than one score point.

Assessment of Reading, Writing and Mathematics: Primary Division

The validity estimates for reading, writing and mathematics for the primary assessment are provided in Tables 4.22 to 4.28. For reading, the percentages of exact agreement ranged from 70.6% to 93.4% for the English-language version and from 69.5% to 97.3% for the French-language version. Both English- and French-language scorers had the tendency to score lower than the expert panel: The aggregates were 9.8% and 5.8%, respectively, for the English-language assessment and 8.6% and 4.6%, respectively, for the French-language assessment, as shown in Tables 4.22 and 4.26. The percentages of exact-plus-adjacent agreement were larger

and less variable, ranging from 96.9% to 99.8% for the English-language version and 95.3% to 100% for the French-language version of the assessment. The percentages of exact agreement met the EQAO target of 70% for four-point rubrics for all English- and French-language items. The percentage of exact-plus-adjacent agreement met the EQAO target of 95% for all English- and French-language items as well.

The percentages of exact agreement for writing were smaller than those for reading (see Tables 4.23 and 4.27). They ranged from 63.8% to 85.1% for English-language items and from 67.1% to 87.3% for French-language items. English-language scorers had a greater tendency to score lower (18.6% lower and 11.0% higher) than the expert panel for the long-writing items. French-language scorers had a greater tendency to score higher than the expert panel for long-writing items (15.3 % higher and 4.8% lower) and for short-writing items (14.7 % higher and 7.8% lower). The percentages of exact-plus-adjacent agreement were large for both long- and short-writing items. The aggregates were 98.9% and 98.4%, respectively, for English-language items, and 99.6% and 99.4%, respectively, for French-language items. The percentage of exact agreement for four rubrics in English met the EQAO targets (70% for topic development and 75% for conventions). These targets were met for five of the six rubrics in French. The percentage of exact-plus-adjacent agreement met the EQAO target of 95% for all English- and French-language items.

The percentages of exact agreement were larger for mathematics items than for reading and writing items on the English-language version of the assessment (see Tables 4.24). The aggregate percentages of exact agreement were 88.5% for the English-language, 83.1% for the French Immersion and 83.4% for the French-language version of the assessment. The aggregate percentages of exact-plus-adjacent agreement were 98.7%, 98.8% and 99.1%, respectively. The aggregate adjacent-high percentage was larger than the aggregate adjacent-low percentage for the scoring of French-language responses, but the aggregate adjacent-high percentage was smaller than the aggregate adjacent-low percentage for the French Immersion and the English-language responses. The percentage of exact agreement met the EQAO target of 70% for four-point rubrics for all English-language items and for four of the eight items on the French-language version of the assessment. The percentages of exact-plus-adjacent agreement for all items on the English- and French-language versions of the assessment met the EQAO target of 95%. One French Immersion item did not meet this target.

Table 4.22 Validity Estimates for Reading: Primary Division (English)

Item Code	Booklet	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
EQRE20994.083	1	5	1 506	99.1	93.4	5.7	4.3	1.4	0.9
EQRE20999.083	1	6	1 221	99.8	92.1	7.7	5.3	2.4	0.3
EQRE20988.082	1	11	1 524	99.2	88.8	10.4	5.4	5.00	0.9
EQRE20983.082	1	12	1 573	99.4	82.5	16.9	6.9	10.0	0.6
EQRE20827.070	2	11	1 131	99.2	70.6	28.7	17.1	11.6	0.8
EQRE20821.070	2	12	1 134	96.9	72.6	24.3	19.1	5.2	3.1
EQRE20939.078	2	21	1 170	98.6	82.2	16.4	9.7	6.8	1.4
EQRE20945.078	2	22	1 368	99.4	80.5	18.9	14.5	4.5	0.6
EQRE21034.086	2	28	1 602	99.6	87.9	11.7	4.9	6.8	0.4
Aggregate			13 730	98.8	83.2	15.6	9.8	5.8	1.3

Table 4.23 Validity Estimates for Writing: Primary Division (English)

Item Code	Booklet	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
EQWE30434T	2	23	1919	98.6	63.8	34.8	21.2	13.5	1.3
EQWE30434V	2	23	1637	99.3	75.7	23.6	15.6	8.0	0.6
Aggregate Long Writing			3556	98.9	69.3	29.6	18.6	11.0	1.0
EQWE20250T	1	13	1676	98.2	74.0	24.2	11.5	12.6	1.7
EQWE20250V	1	13	1185	99.6	85.1	14.5	6.6	7.8	0.3
EQWE20245T	1	14	1542	97.8	69.4	28.3	14.9	13.3	2.2
EQWE20245V	1	14	1137	98.4	72.4	25.9	10.8	15.1	1.5
Aggregate Short Writing			5540	98.4	74.8	23.6	11.3	12.3	1.5
Aggregate All Items			9096	98.6	72.6	26.0	14.1	11.8	1.3

Table 4.24 Validity Estimates for Mathematics: Primary Division (English)

Item Code	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
EQME20944	7	855	100.0	85.7	14.3	7.4	6.9	0.0
EQME20880	8	870	97.0	87.5	9.5	7.0	2.5	3.0
EQME20811	9	856	99.7	89.5	10.2	6.7	3.5	0.4
EQME20822	10	761	99.1	85.8	13.3	9.6	3.7	0.9
EQME10170	27	784	99.9	93.0	6.9	5.0	1.9	0.1
EQME20825	28	782	100.0	95.8	4.2	3.1	1.2	0.0
EQME20878	29	952	96.7	84.0	12.7	5.2	7.6	3.3
EQME20941	30	792	97.7	88.1	9.6	2.0	7.6	2.3
Aggregate		6652	98.7	88.5	10.2	5.7	4.4	1.3

Table 4.25 Validity Estimates for Mathematics: Primary Division (French Immersion)

Item Code	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
EQME20944	6	97	100.0	87.6	12.4	9.3	3.1	0.0
EQME20880	7	114	99.1	77.2	21.9	13.2	8.8	0.9
EQME20811	8	118	99.2	92.4	6.8	4.2	2.5	0.9
EQME20822	9	120	100.0	80.8	19.2	5.8	13.3	0.0
EQME10170	26	123	99.2	87.8	11.4	10.6	0.8	0.8
EQME20825	27	100	96.0	71.0	25.0	23.0	2.0	4.0
EQME20878	28	142	97.2	80.3	16.9	2.8	14.1	2.8
EQME20941	29	123	100.0	87.0	13.0	6.5	6.5	0.0
Aggregate		937	98.8	83.1	15.7	9.0	6.7	1.2

Table 4.26 Validity Estimates for Reading: Primary Division (French)

Item Code	Booklet	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
OQLF20486.063	1	5	98	100	88.8	11.2	5.1	6.1	0.0
OQLF20489.063	1	6	160	100	90.0	10.0	5.0	5.0	0.0
OQLF20518.066	1	11	121	98.4	90.1	8.3	1.7	6.6	1.7
OQLF20519.066	1	12	199	99.5	82.9	16.6	9.1	7.5	0.5
OQLF20466.061	2	11	112	100	87.5	12.5	9.8	2.7	0.0
OQLF20467.061	2	12	109	100	85.3	14.7	11.9	2.8	0.0
OQLF20509.065	2	21	111	100	97.3	2.7	1.8	0.9	0.0
OQLF20506.065	2	22	66	100	97.0	3.0	3.0	0.0	0.0
OQLF20538.068	2	28	128	95.3	69.5	25.8	19.5	6.3	4.7
Aggregate			1164	99.1	85.9	13.2	8.6	4.6	0.9

Table 4.27 Validity Estimates for Writing: Primary Division (French)

Item Code	Booklet	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
OQEF20230T	2	23	302	99.3	73.8	25.5	5.9	19.5	0.6
OQEF20230V	2	23	213	100.0	87.3	12.6	3.2	9.3	0.0
Aggregate Long Writing			515	99.6	79.4	20.1	4.8	15.3	0.3
OQEF20220T	1	13	149	100.0	77.8	22.1	6.7	15.4	0.0
OQEF20220V	1	13	152	99.3	84.8	14.4	10.5	3.9	0.6
OQEF20226T	1	14	131	98.4	67.1	31.3	9.1	22.1	1.5
OQEF20226V	1	14	89	100.0	75.2	24.7	3.3	21.3	0.0
Aggregate Short Writing			521	99.4	76.7	22.6	7.8	14.7	0.5
Aggregate All Items			1036	99.5	78.0	21.4	6.3	15.0	0.4

Table 4.28 Validity Estimates for Mathematics: Primary Division (French)

Item Code	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
OQMF20510	7	115	99.1	83.5	15.7	4.4	11.3	0.9
OQMF20640	8	181	99.5	86.7	12.7	1.1	11.6	0.6
OQMF20513	9	141	97.9	83.0	14.9	1.4	13.5	2.1
OQMF10044	10	149	98.7	72.5	26.2	10.7	15.4	1.3
OQMF20639	27	151	100.0	91.4	8.6	4.6	4.0	0.0
OQMF20504	28	142	100.0	90.1	9.9	2.8	7.0	0.0
OQMF20526	29	122	100.0	86.9	13.1	1.6	11.5	0.0
OQMF20636	30	126	97.6	71.4	26.2	17.5	8.7	2.4
Aggregate		1127	99.1	83.4	15.7	5.3	10.4	0.9

Assessment of Reading, Writing and Mathematics: Junior Division

The pattern of exact-plus-adjacent percentage agreement for the junior-division assessment was similar to that of the primary-division assessment, but the percentages of exact agreement tended to be smaller for the junior-division assessment (see Tables 4.29 to 4.34). The percentages of exact agreement for the junior-division assessment were larger for mathematics than for reading and writing.

For reading, the percentages of exact agreement ranged from 64.9% to 80.9% for English-language items and from 65.6% to 96.7% for French-language items (see Tables 4.29 and 4.32). Percentages of exact-plus-adjacent agreement were larger and less variable, ranging from 95.5% to 99.7% for the English-language version and 93.7% to 100% for the French-language version of the assessment. The percentages of exact agreement met the EQAO target of 70% for four-point rubrics for seven out of ten English-language items and eight out of ten French-language items. The percentages of exact-plus-adjacent agreement met the EQAO target of 95% for all English-language items and nine out of 10 French-language items.

The percentages of exact agreement for writing were smaller than those for reading (see Tables 4.30 and 4.33). The percentages of exact agreement ranged from 63.4% to 77.9% for English-language items and from 62.8% to 76.5% for French-language items. Percentages of exact-plus-adjacent agreement were larger and less variable, ranging from 98.4% to 99.9% for English-language items and 98.0% to 100% for French-language items. The percentages of exact agreement met the EQAO targets for half of the rubrics for both French- and English-language versions of the assessment. The percentages of exact-plus-adjacent agreement for all English- and French-language rubrics met the EQAO target of 95%.

For mathematics, the aggregate percentages of exact agreement were 88.8% for English-language items and 83.3% for French-language items (see Tables 4.31 and 4.34). The aggregate percentages of exact-plus-adjacent agreement were 99.7% and 95.6%. The percentages of exact agreement met the EQAO target of 70% for four-point rubrics for all French-language items and all but one English-language item. The percentages of exact-plus-adjacent agreement met the EQAO target of 95% for all items.

At the aggregate level, there was little difference between the adjacent-low percentage and the adjacent-high percentage for English-language and French-language reading items. The adjacent-high percentage was larger than the adjacent-low percentage for both English- and French-language writing and mathematics items. This percentage was particularly high for English-language writing items.

Table 4.29 Validity Estimates for Reading: Junior Division (English)

Item Code	Booklet	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
EQRE21160.094	1	5	1 812	99.7	73.0	26.7	15.2	11.5	0.3
EQRE21165.094	1	6	1 471	98.8	72.3	26.5	14.7	11.8	1.2
EQRE21274.104	1	11	1 384	97.4	68.6	28.8	5.3	23.6	2.6
EQRE21275.104	1	12	1 443	97.8	65.8	32.0	16.2	15.7	2.2
EQRE21102.091	2	11	1 626	98.4	74.7	23.7	16.2	7.5	1.6
EQRE21114.091	2	12	1 626	95.5	64.9	30.6	15.1	15.4	4.6
EQRE21243.101	2	21	1 488	98.1	72.5	25.6	16.3	9.3	1.9
EQRE21242.101	2	22	1 501	99.3	80.9	18.4	7.5	10.9	0.7
EQRE31965.172	2	28	1 557	99.4	74.2	25.2	12.1	13.2	0.6
Aggregate			15 400	98.3	72.0	26.3	13.7	12.6	1.7

Table 4.30 Validity Estimates for Writing: Junior Division (English)

Item Code	Booklet	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
EQWE20322T	2	23	2 802	99.0	63.4	35.5	13.7	21.8	0.9
EQWE20322V	2	23	2 821	99.5	69.5	30.0	9.6	20.0	0.4
Aggregate Long Writing			5 623	99.2	66.5	32.7	11.8	20.9	0.6
EQWE20315T	1	13	1 928	99.2	74.6	24.5	12.0	12.5	0.7
EQWE20315V	1	13	1 992	99.8	77.9	21.8	11.6	10.2	0.1
EQWE20304T	1	14	2 256	98.4	64.6	33.7	13.7	19.9	1.5
EQWE20304V	1	14	2 041	99.9	76.1	23.7	7.8	15.8	0.0
Aggregate Short Writing			8 217	99.3	73.0	26.2	11.3	14.8	0.6
Aggregate All Items			13 840	99.3	70.4	28.9	11.5	17.3	0.6

Table 4.31 Validity Estimates for Mathematics: Junior Division (English)

Item Code	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
EQME21089	7	671	99.9	86.1	13.7	2.4	11.3	0.2
EQME21119	8	962	99.6	81.9	17.7	8.1	9.6	0.4
EQME21138	9	861	99.9	89.2	10.7	8.1	2.6	0.1
EQME21043	10	760	99.9	94.5	5.4	1.6	3.8	0.1
EQME21104	27	754	99.9	82.8	17.1	7.2	10.0	0.1
EQME10229	28	848	99.7	93.2	6.5	5.3	1.2	0.4
EQME21137	29	853	99.2	87.7	11.5	6.5	5.0	0.8
EQME21135	30	849	100.0	95.1	5.0	2.1	2.8	0.0
Aggregate		6558	99.7	88.8	11.0	5.3	5.7	0.3

Table 4.32 Validity Estimates for Reading: Junior Division (French)

Item Code	Booklet	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
OQLF20595.072	1	5	63	93.7	71.4	22.2	9.5	12.7	6.4
OQLF20596.072	1	6	150	98.0	83.3	14.7	10.7	4.0	2.0
OQLF20637.076	1	11	162	99.4	93.2	6.2	4.3	1.9	0.6
OQLF20635.076	1	12	83	100.0	78.3	21.7	13.3	8.4	0.0
OQLF20566.070	2	11	159	97.5	76.1	21.4	9.4	12.0	2.5
OQLF20568.070	2	12	151	100.0	65.6	34.4	6.6	27.8	0.0
OQLF20617.074	2	21	63	100.0	82.5	17.5	11.1	6.4	0.0
OQLF20615.074	2	22	120	99.2	96.7	2.5	0.8	1.7	0.8
OQLF20667.079	2	28	98	99.0	68.4	30.6	21.4	9.2	1.0
Aggregate			1164	98.8	80.5	18.3	9.6	8.7	1.2

Table 4.33 Validity Estimates for Writing: Junior Division (French)

Item Code	Booklet	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
OQEF30611T	2	23	262	98.0	63.7	34.3	16.4	17.9	1.9
OQEF30611V	2	23	215	100.0	75.8	24.1	11.1	13.0	0.0
Aggregate Long Writing			477	98.9	69.1	29.7	14.0	15.7	1.0
OQEF20268T	1	13	164	100.0	74.3	25.6	12.8	12.8	0.0
OQEF20268V	1	13	107	99.0	73.8	25.2	11.2	14.0	0.9
OQEF20278T	1	14	148	99.3	62.8	36.4	17.5	18.9	0.6
OQEF20278V	1	14	118	100.0	76.2	23.7	14.4	9.3	0.0
Aggregate Short Writing			537	99.6	71.5	28.1	14.1	13.9	0.3
Aggregate All Items			1014	99.3	70.4	28.9	14.1	14.7	0.6

Table 4.34 Validity Estimates for Mathematics: Junior Division (French)

Item Code	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
OQMF10131	7	228	96.5	79.8	16.7	0.4	16.2	3.5
OQMF20796	8	180	99.4	85.6	13.9	1.1	12.8	0.6
OQMF20718	9	240	84.6	75.0	9.6	5.0	4.6	15.4
OQMF20844	10	114	95.6	81.6	14.0	12.3	1.8	4.4
OQMF20710	27	150	99.3	91.3	8.0	4.7	3.3	0.7
OQMF20747	28	123	99.2	86.2	13.0	8.9	4.1	0.8
OQMF10203	29	107	100.0	91.6	8.4	0.9	7.5	0.0
OQMF10127	30	148	97.3	84.5	12.8	3.4	9.5	2.7
Aggregate		1290	95.6	83.3	12.3	4.1	8.1	4.4

Grade 9 Assessment of Mathematics

The validity results for the Grade 9 mathematics assessment are reported in Tables 4.35 to 4.38. The percentages of exact agreement between the expert panel and the scorers were high for each administration of the academic and applied versions of the assessment for both languages. The aggregate percentages of exact agreement ranged from 75.8% to 88.4%. The percentages of exact-plus-adjacent agreement were larger, with the aggregate values ranging from 97.3% to 99.4%. The percentages of exact agreement met the EQAO target of 70% for all but one English-language item and four French-language items. The percentages of exact-plus-adjacent agreement met the EQAO target for all but two English-language items and two French-language items. There were few validity papers for each item for the French-language scoring.

Table 4.35 Validity Estimates for Grade 9 Applied Mathematics (English)

Administration	Item Code	Sequence	Item No.	No. of Scores	% Exact-Plus-Adjnt.	% Exact	% Adjnt.	% Adjnt.-Low	% Adjnt.-High	% Non-Adjnt.
Winter	EQME10643	NR	8	134	100.0	85.1	14.9	6.7	8.2	0.0
	EQME21498	15	9	195	100.0	88.7	11.3	8.7	2.6	0.0
	EQME21413	NR	10	187	94.1	62.0	32.1	9.1	23.0	5.9
	EQME21492	21	11	157	97.5	77.7	19.8	14.7	5.1	2.6
	EQME21472	16	26	212	99.5	89.2	10.4	4.7	5.7	0.5
	EQME21454	NR	28	203	100.0	89.2	10.8	5.9	4.9	0.0
	EQME21429	6	29	213	98.1	81.2	16.9	9.4	7.5	1.9
Aggregate				1301	98.5	82.1	16.4	8.3	8.1	1.5
Spring	EQME21441	NR	8	160	97.5	82.5	15.0	8.8	6.3	2.5
	EQME21494	16	9	178	100.0	78.1	21.9	7.3	14.6	0.0
	EQME21427	NR	10	200	99.5	87.5	12.0	5.0	7.0	0.5
	EQME21425	NR	11	209	98.6	78.5	20.1	2.4	17.7	1.4
	EQME10631	15	26	220	100.0	86.8	13.2	3.6	9.6	0.0
	EQME21474	21	28	323	92.0	74.3	17.7	4.0	13.6	8.1
	EQME21426	5	29	227	100.0	80.2	19.8	13.7	6.2	0.0
Aggregate				1517	97.8	80.6	17.1	6.2	10.9	2.2
Aggregate Across All Administrations				2818	98.1	81.3	16.8	7.2	9.6	1.9

Table 4.36 Validity Estimates for Grade 9 Academic Mathematics (English)

Administration	Item Code	Sequence	Item No.	No. of Scores	% Exact-Plus-Adjnt.	% Exact	% Adjnt.	% Adjnt.-Low	% Adjnt.-High	% Non-Adjnt.
Winter	EQME21282	5	8	417	97.1	89.7	7.4	4.8	2.6	2.9
	EQME21256	NR	9	385	98.4	81.8	16.6	12.5	4.2	1.6
	EQME21275	NR	10	361	98.9	86.4	12.5	8.3	4.2	1.1
	EQME21232	21	11	409	100.0	77.0	23.0	9.8	13.2	0.0
	EQME21225	10	26	499	99.8	88.8	11.0	7.0	4.0	0.2
	EQME10374	16	28	512	99.8	85.0	14.8	12.5	2.3	0.2
	EQME21284	NR	29	504	99.2	88.5	10.7	8.7	2.0	0.8
Aggregate				3087	99.1	85.5	13.6	9.1	4.5	0.9
Spring	EQME21213	4	8	364	99.7	83.2	16.5	10.2	6.3	0.3
	EQME21259	NR	9	348	98.6	73.9	24.7	14.7	10.1	1.4
	EQME21239	NR	10	391	98.5	82.4	16.1	9.0	7.2	1.5
	EQME21241	21	11	356	98.6	85.4	13.2	5.1	8.2	1.4
	EQME21219	9	26	648	99.2	81.6	17.6	4.5	13.1	0.8
	EQME21276	15	28	624	98.9	85.7	13.1	6.9	6.3	1.1
	EQME10476	NR	29	656	100.0	93.5	6.6	4.6	2.0	0.0
Aggregate				3387	99.1	84.5	14.6	7.2	7.4	0.9
Aggregate Across All Administrations				6474	99.1	85.0	14.1	8.1	6.0	0.9

Table 4.37 Validity Estimates for Grade 9 Applied Mathematics (French)

Administration	Item Code	Sequence	Item No.	No. of Scores	% Exact-Plus-Adjnt.	% Exact	% Adjnt.	% Adjnt.-Low	% Adjnt.-High	% Non-Adjnt.
Winter	OQMF21364	NR	8	37	94.6	70.3	24.3	18.9	5.4	5.4
	OQMF10269	6	9	23	100.0	100.0	0.0	0.0	0.0	0.0
	OQMF21420	NR	10	28	100.0	92.9	7.1	0.0	7.1	0.0
	OQMF10313	NR	11	16	100.0	50.0	50.0	0.0	50.0	0.0
	OQMF21333	NR	26	16	100.0	93.8	6.3	0.0	6.3	0.0
	OQMF10296	13	28	15	100.0	86.7	13.3	0.0	13.3	0.0
	OQMF21390	20	29	17	100.0	100.0	0.0	0.0	0.0	0.0
Aggregate				152	98.7	84.2	14.5	4.6	9.9	1.3
Spring	OQMF21392	NR	8	20	100.0	90.0	10.0	0.0	10.0	0.0
	OQMF10269	6	9	27	100.0	92.6	7.4	7.4	0.0	0.0
	OQMF21365	20	10	19	100.0	73.7	26.3	0.0	26.3	0.0
	OQMF21363	NR	11	31	96.8	83.9	12.9	3.2	9.7	3.2
	OQMF21333	NR	26	24	100.0	91.7	8.3	0.0	8.3	0.0
	OQMF21038	NR	28	19	100.0	84.2	15.8	5.3	10.5	0.0
	13536	13	29	32	100.0	96.9	3.1	0.0	3.1	0.0
Aggregate				172	99.4	88.4	11.1	2.3	8.7	0.6
Aggregate Across All Administrations				324	99.1	86.4	12.7	3.4	9.3	0.9

Table 4.38 Validity Estimates for Grade 9 Academic Mathematics (French)

Administration	Item Code	Sequence	Item No.	No. of Scores	% Exact-Plus-Adjnt.	% Exact	% Adjnt.	% Adjnt.-Low	% Adjnt.-High	% Non-Adjnt.
Winter	OQMF20924	15	8	43	97.7	74.4	23.3	14.0	9.3	2.3
	OQMF20919	11	9	43	97.7	67.4	30.2	16.3	14.0	2.3
	OQMF10397	6	10	48	100.0	91.7	8.3	2.1	6.3	0.0
	OQMF10370	21	11	51	96.1	76.5	19.6	17.7	2.0	3.9
	OQMF20905	NR	26	48	100.0	81.3	18.8	18.8	0.0	0.0
	OQMF20926	NR	28	49	93.9	65.3	28.6	16.3	12.2	6.1
	OQMF20913	NR	29	48	95.8	72.9	22.9	4.2	18.8	4.2
Aggregate				330	97.3	75.8	21.5	12.7	8.8	2.7
Spring	OQMF10367	NR	8	62	96.8	77.4	19.4	17.7	1.6	3.2
	OQMF20925	15	9	64	100.0	85.9	14.1	14.1	0.0	0.0
	OQMF10380	5	10	78	100.0	82.1	18.0	15.4	2.6	0.0
	OQMF20920	NR	11	51	100.0	72.6	27.5	21.6	5.9	0.0
	OQMF20912	NR	26	113	100.0	91.2	8.9	5.3	3.5	0.0
	OQMF10353	21	28	83	90.4	66.3	24.1	10.8	13.3	9.6
	OQMF20918	10	29	82	100.0	90.2	9.8	1.2	8.5	0.0
Aggregate				533	98.1	81.8	16.3	11.1	5.3	1.9
Aggregate Across All Administrations				863	97.8	79.5	18.3	11.7	6.6	2.2

Ontario Secondary School Literacy Test (OSSLT)

The scoring validity estimates for the OSSLT are reported in Tables 4.39 to 4.42. For reading, the aggregate percentage of exact agreement was 77.8% for English-language items and 67.1% for French-language items. The aggregate percentage of exact-plus-adjacent agreement was 97.7% for the English-language version of the assessment and 93.4% for the French-language version. The percentages of exact agreement met the EQAO target of 75% for three-point rubrics on three out of four English-language items and one out of four French-language items. Among English- and French-language scorers, the aggregate adjacent-high percentage was larger than the aggregate adjacent-low percentage.

For long-writing items, the percentages of exact agreement ranged from 40.6% to 55.7% for content and from 63.2% to 69.4% for conventions. For short-writing items, the percentages of exact agreement ranged from 60.9% to 74.2%. The percentages of exact-plus-adjacent agreement ranged from 87.9% to 99.4%.

The percentages of exact agreement did not meet the EQAO targets of 60% for content and 70% for conventions on the long-writing prompts in the English-language or French-language assessments. The percentages of exact agreement did meet the EQAO target of 65% for one out of two short-writing prompts in the English-language version and for both short-writing prompts in the French-language version. The aggregate percentages for adjacent-high are higher than those for adjacent-low.

It is important to note that all student responses on the OSSLT were scored independently by two scorers and that all student responses with non-adjacent scores were adjudicated by an expert scorer. Scores used to compute student results had a 100% exact-plus-adjacent agreement for one pair of scorers.

Table 4.39 Validity Estimates for Reading: OSSLT (English)

Item Code	Section	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
EQRE21427.116	I	6	10 069	98.9	69.8	29.0	10.7	18.3	1.1
EQRE21398.113	V	7	12 947	94.3	79.6	14.7	4.9	9.9	5.7
EQRE21474.120	IX	6	10 971	98.8	81.2	17.5	9.8	7.7	1.2
EQRE21475.120	IX	7	10 328	99.7	79.7	20.0	8.2	11.8	0.3
Aggregate			44 315	97.7	77.8	19.9	8.2	11.7	2.3

Table 4.40 Validity Estimates for Writing: OSSLT (English)

Item Code	Section	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
1W3_057T	IV	1	17 630	92.4	51.0	41.4	21.3	20.1	7.6
1W3_057V	IV	1	22 429	99.4	69.4	30.0	13.9	16.1	0.6
1W2_054T	VI	1	23 550	96.2	55.7	40.5	19.5	21.0	3.8
1W2_054V	VI	1	23 885	99.4	66.9	32.5	13.5	19.1	0.6
Aggregate Long Writing			87 494	97.1	61.3	35.8	16.8	19.0	2.9
EQWE20375T&V	III	1	4 546	95.9	60.9	35.1	17.3	17.8	4.1
EQWE20381T&V	X	1	3 738	97.2	65.6	31.6	12.7	18.9	2.8
Aggregate Short Writing			8 284	96.5	63.0	33.5	15.3	18.3	3.5
Aggregate Across All Items			95 778	97.1	61.5	35.6	16.7	19.0	2.9

Table 4.41 Validity Estimates for Reading: OSSLT (French)

Item Code	Section	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
OQLF20722.083	I	6	130	99.2	86.9	12.3	7.7	4.6	0.8
OQLF20681.080	V	7	132	90.9	71.2	19.7	6.8	12.9	9.1
OQLF20751.085	IX	6	238	89.5	59.7	29.8	12.2	17.7	10.5
OQLF20754.085	IX	7	303	95.1	62.7	32.3	18.2	14.2	5.0
Aggregate			803	93.4	67.1	26.3	12.8	13.5	6.6

Table 4.42 Validity Estimates for Writing: OSSLT (French)

Item Code	Section	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
2É3_024T	IV	1	413	90.8	44.1	46.7	8.0	38.7	9.2
2É3_024V	IV	1	428	98.4	68.7	29.7	13.1	16.6	1.6
2É2_025T	VI	1	315	87.9	40.6	47.3	24.8	22.5	12.1
2É2_025V	VI	1	272	98.9	63.2	35.7	22.8	12.9	1.1
Aggregate Long Writing			1428	94.0	54.3	39.6	16.0	23.6	6.0
OQEF20359T&V	III	1	504	98.2	74.2	24.0	18.5	5.6	1.8
OQEF20379T&V	X	1	419	99.3	68.3	31.0	9.6	21.5	0.7
Aggregate Short Writing			923	98.7	71.5	27.2	14.4	12.8	1.3
Aggregate Across All Items			2351	95.8	61.1	34.8	15.4	19.4	4.2

CHAPTER 5: EQUATING

For security purposes, EQAO constructs a different test every year while ensuring that content and statistical specifications are similar to those in the tests used in previous years. Despite such efforts to ensure similarity, assessments from year to year may differ somewhat in their difficulty. To account for this, EQAO uses a process called equating, which adjusts for differences in difficulty among the tests from year to year (Kolen & Brennan, 2004). Equating ensures that students in one year are not given an unfair advantage over students in another year and that reported changes in achievement levels are due to differences in student performance, and not to differences in test difficulty.

From time to time, the Ministry of Education makes modifications to *The Ontario Curriculum*. As a result, EQAO assessments are modified in content and length to reflect these curriculum changes. As such, the modified tests differ in content and statistical specifications from tests used in previous years (prior to the curriculum revisions). In such cases, EQAO uses a process called scaling to link the previous years' tests with the following year's modified tests.

The processes used in equating and scaling are similar, but their purposes are different. Equating is used to adjust for differences in difficulty among tests that are similar in content and statistical specifications. Scaling is used to link two tests that are different in content and statistical specifications (Kolen & Brennan, 2004). Since there were no significant changes to the test specifications from 2008 to 2009, only equating procedures were used in 2009.

The following sections describe the Item Response Theory (IRT) models, equating design, equating samples and calibration procedures used during the 2008–2009 school year for the various EQAO assessments.

IRT Models

Item-response models define the relationship between an unobserved construct or ability (θ) and the probability (P) of a student correctly answering a dichotomously scored item. For polytomously scored items, the models define the relationship between an unobserved construct or ability and the probability of a student receiving a particular score on the item. The Three-Parameter Logistic (3PL) model and the Generalized Partial Credit Model (GPCM) are the general models used by EQAO to estimate the item and ability parameters of multiple-choice and open-response items. The Three-Parameter Logistic (3PL) model is given by

$$P_i(\theta) = c_i + (1 - c_i) \frac{\exp^{Da_i(\theta - b_i)}}{1 + \exp^{Da_i(\theta - b_i)}}, \quad (\text{Equation \#1})$$

where $P_i(\theta)$ is the probability of a student (with ability θ) answering item i correctly;
 a_i is the slope parameter for item i ;
 b_i is the difficulty parameter for item i ;
 c_i is the pseudo-guessing parameter for item i and
 D is a scaling parameter equal to 1.7.

The Generalized Partial Credit Model (GPCM) is given by

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^x \alpha_i(\theta - \delta_{ij})\right]}{\sum_{r=0}^{m_i} \exp\left[\alpha_i \sum_{j=0}^r (\theta - \delta_{ij})\right]}, \quad (\text{Equation \#2})$$

where $P_{ix}(\theta)$ is the probability of a student (with ability θ) receiving a particular score, x , on item i ;

α_i is the slope parameter for item i and

δ_{ij} is the item i step difficulty parameter associated with a category score j .

Equating Design

The forward-fixed common-item parameter non-equivalent group design is used to equate EQAO tests over different years. Common items are sets of items that are identical in two tests and are used to create a common scale for all the items in the tests. These common items are selected from the field-test items administered in one year and used as operational items in the next year. There are several steps in the forward-fixed common-item design:

1. Operational item parameters in the current year's test are calibrated.
2. Operational items and field-test items from the previous year are brought forward to the current year's test and are recalibrated.
 - This is done by fixing the parameters of the items common to the two years at the values obtained in the first step.
 - This process places the item parameters from the two years on the same scale.
3. Recalibrated parameters for the operational items from the previous year are then used to rescore the corresponding equating sample.
 - For the OSSLT, the θ -value of the cut point (corresponding to the percentage of successful students in the previous year) is then identified and applied to the current year's test-score distribution to obtain the percentage of successful and unsuccessful students for the current year.
 - For the primary, junior and Grade 9 assessments, the theta values of the cut points (corresponding to the percentage of students at each performance level) are identified and then applied to the current year's test-score distribution to obtain the percentage of students at each performance level for the current year.

Equating Samples

For each assessment, EQAO uses a set of exclusion rules to select a calibration and equating sample. The exclusion rules ensure that the equating sample is representative of the population of students who wrote the test under typical administration conditions. While the exclusion rules are similar for all assessments, there are some differences. Therefore, the exclusion rules are provided below in the description of the equating conducted for each assessment.

Calibration

Calibration is the process of estimating the item parameters that determine the relationship between ability and the probability of answering a multiple-choice item correctly or receiving a particular score on a polytomously scored open-response item. For each assessment, the calibration of the items for the English-language and the French-language populations are conducted separately. The calibrations are conducted using the program PARSCALE 4.1 (Muraki & Bock, 2003). However, before conducting the analyses, the p -values of common items are plotted to identify outliers, which are then eliminated. Also, common items showing large differences in their p -values between the two years are eliminated from the equating process.

Assessments of Reading, Writing and Mathematics: Primary and Junior Divisions

Description of the IRT Model

A modified 3PL model (equation 1) is used for the primary- and junior-division assessments.

The pseudo-guessing parameter is fixed at $0.20 \left(\frac{1}{k+1} \right)$, where k is the number of options) for the multiple-choice items to reflect the possibility of students with a very low ability to correctly answer an item. The GPCM (equation 2) is used for the constructed-response items.

Equating Sample: Exclusion Rules

The following categories of students were excluded from the equating samples for 2007–2008 and 2008–2009:

- a) students who were not attending a publicly funded school;
- b) students who were home-schooled;
- c) French Immersion students (Grade 3 English mathematics only);
- d) students with no work completed (by booklet and item type);
- e) students receiving accommodations;
- f) students who were exempted and
- g) students who did not attempt at least one question in each significant part of the test.

Using these exclusion rules, three student samples were obtained. These were

1. students from the 2007–2008 population who responded to both the 2007–2008 operational-test items and the field-test items that were brought forward to form the 2008–2009 operational assessments and who were not excluded by the rules stated above (calibration sample).
2. students from the 2007–2008 population who wrote the operational assessment and who were not excluded by the rules stated above (equating sample).
3. students from the 2008–2009 population who wrote the operational assessment and who were not excluded by the rules stated below (calibration and equating sample).

The numbers of students in each sample are presented in Table 5.1.

Table 5.1 Number of Students in the Calibration and Equating Samples for the 2007–2008 and 2008–2009 Primary- and Junior-Division Assessments (English and French)

Component	Number of Students		
	2008–2009 Calibration/ Equating Sample	2007–2008 Calibration Sample	2007–2008 Equating Sample
Primary Mathematics (English)	101 992	107 286	107 372
Junior Mathematics (English)	119 484	109 915	125 452
Primary Mathematics (French)	6 217	6 155	6 217
Junior Mathematics (French)	5 900	5 904	5 975
Primary Reading (English)	104 704	34 974	109 634
Junior Reading (English)	117 264	38 254	123 549
Primary Reading (French)	5 967	2 977	6 147
Junior Reading (French)	5 794	2 959	5 919
Primary Writing (English)	105 112	56 440	113 110
Junior Writing (English)	117 136	62 877	125 033
Primary Writing (French)	6 036	4 488	6 134
Junior Writing (French)	5 823	4 671	5 981

Equating Steps

In equating the 2008–2009 tests to the 2007–2008 tests, the EQAO forward-fixed common-item parameter equating procedure was used. The following steps were implemented:

1. The 2008–2009 operational items were calibrated independently to obtain item parameters and student-ability scores for the 2008–2009 equating sample.
2. The 2007–2008 operational items were calibrated together with the field-test items that were brought forward to the 2008–2009 operational assessments. In this calibration, the item parameters of the field-test items were fixed at the values obtained from the 2008–2009 calibration runs (step 1).
3. The 2007–2008 equating sample was scored using the operational item parameters obtained in step 2.
4. The percentage of students at each achievement level was determined for the 2007–2008 equating sample from the levels assigned in 2007–2008. The theta-value cut points that replicated this distribution were identified for each boundary (0/1, 1/2, 2/3 and 3/4).
5. These theta values were then used as the cut points for 2008–2009.
6. The operational item parameters of 2008–2009 obtained in step 1 were used to score the full student population.
7. The cut-score points identified in step 4 were applied to the 2008–2009 student theta values, students were assigned to levels and the percentage of students at each performance level was determined.

Eliminating Items and Collapsing of Score Categories

For the primary- and junior-division assessments, 11 multiple-choice items and four open-response items across the 12 assessments were dropped from the equating link. These items were either modified between field- and operational-test administrations or were not field tested in the

2007–2008 assessment year. The seven open-response items included four long-writing prompts that were not field tested during the administration of the operational tests. The number of items not used in the equating process and the number of dropped items are presented for each test in Table 5.2.

Table 5.2 Number of Items Dropped from the Primary- and Junior-Division Assessments and from the Equating Process (2008–2009)

Assessment Component	No. of Items Not Used in Equating*		No. of Items Dropped from the Assessment
	Multiple-Choice	Open-Response	
Primary Mathematics (English)	2	0	0
Junior Mathematics (English)	0	0	0
Primary Mathematics (French)	0	0	0
Junior Mathematics (French)	1	0	0
Primary Reading (English)	0	0	0
Junior Reading (English)	3	0	0
Primary Reading (French)	1	0	0
Junior Reading (French)	4	0	0
Primary Writing (English)	0	1	0
Junior Writing (English)	0	1	0
Primary Writing (French)	0	1	0
Junior Writing (French)	0	1	0

* Long-writing prompts for the current year are not field tested with the previous year's operational test, so they are never used in the equating link. As such, they have been included in the number of items not used in equating.

Equating Results

The results of the equating process for the reading and mathematics components of the assessments are provided in Tables 5.3 to 5.6. The theta cut scores and the number of students at each achievement level in 2008 and 2009 are reported for both English-language and French-language students. For example, the equated cut scores for the reading component of the English-language primary-division assessment were 1.23 for Levels 3 and 4, –0.34 for Levels 2 and 3, –1.41 for Levels 1 and 2 and –2.39 for NE1 and Level 1.

Since the 2007–2008 and 2008–2009 student thetas are on the same scale, the theta cut scores in the following tables apply to the assessments in both years.

Table 5.3 Equating Results for Reading: Primary Division (English and French)

	Primary Reading (English)			Primary Reading (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2008	2009		2008	2009
Number of Students		109 634	104 704		6147	5967
Level 4		8.9%	7.9%		13.0%	19.8%
Level 3	1.23	58.5%	58.9%	0.81	50.6%	49.7%
Level 2	-0.34	25.8%	26.0%	-0.43	30.6%	24.9%
Level 1	-1.41	5.8%	6.2%	-1.62	5.2%	5.3%
NE1	-2.39	1.0%	1.0%	-2.82	0.6%	0.3%
% of Students at or Above the Provincial Standard		67.4%	66.8%		63.6%	69.5%

Table 5.4 Equating Results for Reading: Junior Division (English and French)

	Junior Reading (English)			Junior Reading (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2008	2009		2008	2009
Number of Students		123 549	117 264		5919	5794
Level 4		8.3%	12.2%		12.6%	16.2%
Level 3	1.07	63.8%	64.4%	0.92	66.1%	65.2%
Level 2	-0.65	23.2%	20.0%	-0.81	19.9%	17.7%
Level 1	-1.76	4.3%	3.3%	-2.23	1.3%	0.9%
NE1	-2.97	0.3%	0.2%	-3.09	0.2%	0.1%
% of Students at or Above the Provincial Standard		72.2%	76.5%		78.7%	81.4%

Table 5.5 Equating Results for Mathematics: Primary Division (English and French)

	Primary Mathematics (English)			Primary Mathematics (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2008	2009		2008	2009
Number of Students		107 372	101 992		6217	6046
Level 4		12.6%	13.8%		9.0%	13.5%
Level 3	1.04	60.1%	61.8%	1.02	55.9%	55.6%
Level 2	-0.66	25.0%	22.7%	-0.46	31.1%	27.5%
Level 1	-2.03	2.2%	1.6%	-1.79	3.7%	3.1%
NE1	-2.74	0.2%	0.1%	-2.61	0.4%	0.3%
% of Students at or Above the Provincial Standard		72.6%	75.6%		64.9%	69.1%

Table 5.6 Equating Results for Mathematics: Junior Division (English and French)

	Junior Mathematics (English)			Junior Mathematics (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2008	2009		2008	2009
Number of Students		125 452	119 484		5975	5900
Level 4		12.7%	14.7%		32.0%	37.5%
Level 3	1.01	53.4%	53.9%	0.33	49.6%	45.5%
Level 2	-0.45	28.6%	26.5%	-0.93	17.9%	16.8%
Level 1	-1.61	5.2%	4.8%	-2.37	0.3%	0.2%
NE1	-3.08	0.1%	0.1%	-2.59	0.2%	0.1%
% of Students at or Above the Provincial Standard		66.1%	68.6%		81.6%	83.0%

Alternative equating procedures were used for the writing component of the primary- and junior-division assessments due to anomalous field-test results. Inspection of the item statistics for the writing component of the 2009 primary- and junior-division assessments revealed that the 2007–2008 field-test results for the short-writing prompts that were placed on the 2008–2009 operational forms were very low for the English-language junior-division assessment and for the French-language primary- and junior-division assessments. These results were considerably lower than those for all short-writing prompts on recent operational assessments and field-tested short-writing prompts in previous years.

Consequently, concerns were raised that differences between operational and field-test results were due to factors other than changes in students’ writing skills.

Since the anomaly of lower-than-expected scores on the short-writing prompts was not observed for the writing component of the English-language primary-division assessment, the cut scores obtained from the forward-fixed common-item parameter equating process were used to determine student achievement levels for the 2008–2009 assessment. These equating results are presented in the left panel of Table 5.7. The IRT equating results for the French-language primary-division assessment and the English- and French-language junior-division assessments are also presented in Tables 5.7 and 5.8, but the cut scores in these tables were not the final cut scores used to assign achievement levels for these three 2008–2009 assessments. These cut scores were considered in combination with cut scores from the alternative equating process, which is described on the next page below Tables 5.7 and 5.8.

The reasons for these anomalous results are not obvious. One possibility is the location of the field-test short-writing prompts in the student booklets relative to the location of the operational prompts. However, this same booklet structure was used in previous years and did not show anomalous results. The structure of the student booklets has been modified for the 2009–2010 assessments so that the operational and field-test short-writing prompts are in similar locations within a session.

Table 5.7 Equating Results for Writing: Primary Division (English and French)

	Primary Writing (English)			Primary Writing (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2008	2009		2008	2009
Number of Students		109 829	105 112		6173	6036
Level 4		5.6%	4.9%		17.4%	18.7%
Level 3	1.46	65.4%	68.0%	0.85	60.3%	63.4%
Level 2	-0.54	28.9%	26.9%	-0.84	22.0%	17.7%
Level 1	-2.90	0.1%	0.1%	-2.47	0.3%	0.2%
NE1	-3.25	0.0%	0.1%	-3.40	0.0%	0.0%
% of Students at or Above the Provincial Standard		71.0%	72.9%		77.7%	82.1%

Table 5.8 Equating Results for Writing: Junior Division (English and French)

	Junior Writing (English)			Junior Writing (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2008	2009		2008	2009
Number of Students		122 515	117 136		5898	5823
Level 4		10.1%	12.7%		17.3%	27.0%
Level 3	1.03	63.6%	67.4%	0.53	67.1%	63.3%
Level 2	-0.75	25.8%	19.7%	-1.21	15.5%	9.7%
Level 1	-2.55	0.5%	0.2%	-2.96	0.0%	0.0%
NE1	-3.48	0.0%	0.0%	-3.35	0.0%	0.0%
% of Students at or Above the Provincial Standard		73.7%	80.1%		84.4%	90.3%

Since the field-test results for short-writing equating items could not be explained for the writing components of the French-language primary-division and the English- and French-language junior assessments, a second equating procedure, the summed-score equating process, was also applied to these assessments. The following steps were followed to determine the 3/4 cut point for each assessment:

1. The number of students below Level 4 in the 2007–2008 equating sample was determined.
2. The number of students at or below each summed score was tallied for the 2007–2008 equating score sample.
3. Adjacent summed scores that had numbers of students at or below them that were just below and just above the number from step 1 (above) were identified. These two summed scores were the lower and upper bounds for the cut score between Levels 3 and 4.
4. The number of students at or below the lower summed score in step 3 was subtracted from the value in step 1 to determine the number of students needed from the score interval for the upper summed score in step 3.

5. The number of students needed from the score interval in step 4 was divided by the total number in the score interval to determine the proportion of the interval that was needed to reach the number of students identified in step 1.
6. On the 2008–2009 test form, the same score values determined in step 3 were identified, based on an assumption that the test forms from the two years were constructed to be parallel and that the summed scores have the same meaning.
7. The number of students in the upper score category from step 6 was determined and then multiplied by the proportion from step 5. This is the number of students from that score interval needed to reach the number below the cut score.
8. The number in step 7 was added to the number of students at or below the lower score from step 6 to get the total number of students below the cut score for the 2008–2009 equating sample.
9. The theta value that had this number below it for the 2008–2009 equating calibration was computed for use as the cut score between Levels 3 and 4.

This procedure is based on an assumption that the test forms created in the two years are equivalent forms and that no equating is necessary. The standard IRT equating is based on an assumption that the forms are not equivalent and that adjustments need to be made. Because of the questionable data from some of the short-writing items, the standard equating process is suspect. Neither of these approaches is absolutely correct. The correct cut score is likely to be between the values obtained by the two methods.

These steps were repeated to determine the 2/3 cut point. Because the number of students at Level 1 and in the NE1 category is very small, the cut points from the IRT equating process were carried over for the 1/2 and NE1/1 cut scores.

The results of the two equating processes were reviewed in light of the item statistics and the test characteristics to reach decisions on the cut scores to use for each assessment. Results from the IRT equating, results from the raw score equating or an average of the two were considered. As indicated above, the results of the IRT equating process were used for the writing component of the English-language primary-division assessment. The cut scores from the IRT and raw score equating and the average of the two are presented in Table 5.9 for the other three writing components of the assessments.

Table 5.9 Cut Points from the Two Equating Procedures for the Writing Components of the Primary- and Junior-Division Assessments

Assessment	Levels	Theta Cut Score Based on IRT Equating	Theta Cut Score based on Summed-Score Equating	Average IRT and Summed-Score Cut Score
Primary French	3/4	0.8476	0.8742	0.8609
	2/3	-0.8363	-0.6491	-0.7427
	1/2	-2.4716	-2.4716*	-2.4716*
	NE1/1	-3.4022	-3.4022*	-3.4022*
Junior English	3/4	1.0313	1.2948	1.1631
	2/3	-0.7506	-0.4023	-0.5765
	1/2	-2.5495	-2.5495*	-2.5495*
	NE1/1	-3.3727	-3.3727*	-3.3727*
Junior French	3/4	0.5258	0.9287	0.7273
	2/3	-1.2140	-0.8377	-1.0259
	1/2	-3.1139	-3.1139*	-3.1139*
	NE1/1	-3.4766	-3.4766*	-3.4766*

* Values carried forward from the IRT equating results.

The cut scores used for assigning student achievement levels for the 2008–2009 assessments were as follows:

- Primary French (average cut scores)
- Junior English (average cut scores)
- Junior French (summed-score equating cut scores)

Since the assumptions were not fully satisfied for either equating procedure, taking the average cut score was a reasonable compromise. For the French-language junior-division assessment, the item information functions for the short-writing equating prompts were very different for the operational- and field-test administrations, so the summed-score approach was chosen.

Grade 9 Assessment of Mathematics

Description of the IRT Model

The 3PL model (equation 1) and GPCM model (equation 2) were used to estimate item and ability parameters. For the Grade 9 mathematics assessment (academic and applied), the 3PL model was modified by fixing the pseudo-guessing parameter at $0.20 \left(\frac{1}{k+1} \right)$, where k is the number of options) for the multiple-choice items to reflect the possibility of students with a very low ability to correctly answer an item.

The academic and applied versions of the mathematics assessment are administered twice (once in winter and once in spring) during the school year. The winter and spring versions of the assessment for each course have a set of common items and a set of unique items. The common items are used for equating across the two administrations (winter and spring).

Equating Sample

Equating samples for 2007–2008 and 2008–2009 were identified using a common set of selection rules. The equating samples for the academic and the applied courses were selected separately. However, the selection and exclusion rules for both courses and across the two administration years were the same. Students were excluded if they

- a) did not attend a publicly funded school;
- b) were home-schooled;
- c) produced no work by booklet and question type;
- d) received accommodations, except for English language learners who received accommodation for setting;
- e) received special provisions or
- f) did not attempt at least one question in each significant part of the test.

Table 5.10 presents the number of students in the Grade 9 equating samples for the 2007–2008 and the 2008–2009 assessments.

Table 5.10 Number of Grade 9 Students in the Equating Samples

Grade 9	2007–2008				2008–2009	
	English		French		English	French
	Calibration Sample	Equating Sample	Calibration Sample	Equating Sample	Calibration/Equating Sample	Calibration/Equating Sample
Academic	86 312	98 090	3 546	3 970	97 950	3 860
Applied	30 008	39 837	1 223	1 308	40 979	1 288

Equating Steps

The calibration and equating of the 2007–2008 and 2008–2009 assessments for the English-language and the French-language student populations in both academic and applied courses were conducted using the following steps:

1. Calibration and equating were conducted simultaneously for the 2008–2009 winter and spring samples.
2. A concurrent calibration was conducted for the 2007–2008 winter and spring samples (including the field-test items that were brought forward to the 2008–2009 operational test). In this calibration, the item parameters of the field-test items that were brought forward to the 2008–2009 operational tests were fixed at the values obtained from the 2008–2009 calibration runs (step 1). The parameters of the 2007–2008 operational items that were repeated on the 2008–2009 operational tests were also fixed at the values obtained from the 2008–2009 calibration runs (step 1). The 2007–2008 equating samples (operational items only) were scored using the scaled 2007–2008 operational item parameters.
3. The percentage of students in each achievement level was determined for the 2007–2008 equating sample from the levels assigned in 2007–2008. The theta-value cut points that replicated this distribution were identified for each boundary (0/1, 1/2, 2/3 and 3/4).
4. These theta values were then used as the cut points for the 2008–2009 assessments.
5. The item parameters for the 2008–2009 operational items obtained in step 1 were used to score the full student population.

6. The cut-score points identified in step 4 were applied to the 2008–2009 student theta values, students were assigned to levels and the percentage of students at each performance level was determined.

Eliminating Items and Collapsing of Score Categories

For the Grade 9 mathematics assessment, one multiple-choice item was eliminated from the equating process due to changes from field-test items to operational-test items for four assessments, as indicated in Table 5.11. None of the polytomously scored item categories were collapsed for any of the assessments.

Table 5.11 Number of Items Dropped from the Primary and Junior Assessments and from the Equating Process (2008–2009)

Assessment Component	No. of Items Not Used in Equating		No. of Items Dropped from the Assessment
	Multiple-Choice	Open-Response	
English, Applied, Winter	0	0	0
English, Applied, Spring	0	0	0
English, Academic, Winter	1	0	0
English, Academic, Spring	1	0	0
French, Applied, Winter	1	0	0
French, Applied, Spring	1	0	1
French, Academic, Winter	0	0	0
French, Academic, Spring	0	0	0

Equating Results

The equating results for the applied version of the Grade 9 mathematics assessment are summarized in Table 5.12. The results for the academic version of the Grade 9 mathematics assessment are summarized in Table 5.13. The theta cut scores and number of students in 2007 and 2008 at each achievement level is reported for both the English-language and French-language students. For example, the equated cut scores for the English-language applied version of the assessment were 1.44 for Levels 3 and 4, 0.20 for Levels 2 and 3, –0.83 for Levels 1 and 2 and –1.57 for Below Level 1 and Level 1.

Table 5.12 Equating Results for the Grade 9 Applied Mathematics

	English Applied			French Applied		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2008	2009		2008	2009
Number of Students		39 837	40 979		1308	1288
Level 4		5.4%	5.8%		5.3%	5.4%
Level 3	1.44	33.2%	36.5%	1.46	30.9%	37.1%
Level 2	0.20	40.5%	38.5%	0.21	42.5%	40.2%
Level 1	-0.83	16.2%	14.6%	-0.91	16.2%	13.7%
Below Level 1	-1.57	4.7%	4.6%	-1.63	5.1%	3.6%
% of Students at or Above the Provincial Standard		38.6%	42.3%		36.2%	42.5%

Table 5.13 Equating Results for the Grade 9 Academic Mathematics

	English Academic			French Academic		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2008	2009		2008	2009
Number of Students		98 090	97 950		3970	3860
Level 4		7.8%	8.2%		3.1%	3.5%
Level 3	1.37	69.1%	70.6%	1.76	65.2%	65.9%
Level 2	-0.77	16.4%	15.1%	-0.53	21.6%	20.9%
Level 1	-1.40	6.5%	5.9%	-1.19	10.1%	9.7%
Below Level 1	-2.49	0.2%	0.1%	-2.36	0.1%	0.1%
% of Students at or Above the Provincial Standard		76.9%	78.8%		68.3%	69.4%

Ontario Secondary School Literacy Test (OSSLT)

Description of the IRT Model

In contrast to the primary-division, junior-division and Grade 9 assessments, both the a -parameter and the c -parameter (see equation 1) were fixed, yielding a modified Rasch model. The a -parameter for all multiple-choice and open-response items was set at 0.588. The pseudo-guessing parameter for the multiple-choice items was set at $0.20 \left(\frac{1}{k+1} \right)$, where k is the number of options that reflect the possibility of students with a very low ability to correctly answer an item). The partial credit model (PCM) was used to estimate the item and ability parameters for the open-response items. The PCM is given by the following equation:

$$P_{ix}(\theta) = \frac{\exp \left[\sum_{j=0}^x (\theta - \delta_{ij}) \right]}{\sum_{r=0}^{m_i} \exp \left[\sum_{j=0}^r (\theta - \delta_{ij}) \right]}, \quad (\text{Equation \#3})$$

where $P_{ix}(\theta)$ is the probability of a student (with ability θ) receiving a particular score, x , on item i and

δ_{ij} is the item i step-difficulty parameter associated with a category score j .

Equating Sample

First-time eligible students from both publicly funded and private schools were selected for the 2008 and 2009 equating samples. The following categories of students were excluded from the equating sample:

- a) students with no work or incomplete work on a major section of the test;
- b) students receiving accommodations; however, students receiving an accommodation for extra time were not excluded;
- c) previously eligible students;
- d) students who were exempted, deferred or taking the Ontario Secondary School Literacy Course (OSSLC) and
- e) students who were home-schooled.

Table 5.14 presents the number of first-time eligible students in the OSSLT equating samples for the March 2008 and March 2009 tests.

Table 5.14 Number of First-Time Eligible OSSLT Students in the Equating Samples

OSSLT	2008		2009	
	English	French	English	French
First-Time Eligible	144 274	5240	140 328	5163

Equating Steps

The following steps were implemented to calibrate and equate the 2008 and 2009 OSSLT:

1. The parameters of the operational items administered in 2009 were calibrated using the 2009 equating sample.
2. The operational items that formed the 2008 test and the field-test items brought forward to the 2009 test were recalibrated using the 2008 equating sample. In this calibration, the common-item parameters were fixed at the values obtained in step 1.
3. The operational-item parameters of the 2008 test, obtained in step 2, were used to score the 2008 equating sample data.
4. The percentage of successful students was determined for the 2008 equating sample from the student results reported in 2008. The theta-value cut point that replicated this percentage was identified for the distribution of scores in the 2008 equating sample.
5. This theta value was applied to student scores for the 2009 assessment to determine which students would be successful. The results are presented in Table 5.13.

Scale Score

The reporting scale for the 2009 OSSLT, which ranges from 200 to 400, was generated using a linear transformation. The slope and intercept were obtained by fixing two points: the theta value -4.0 was fixed at the lowest value of the scale score (200), and the theta cut score obtained from the equating steps (step 5 above) was fixed at the scale score of 300.

Eliminating Items and Collapsing of Score Categories

Two multiple-choice items in the English-language OSSLT and three multiple-choice items in the French-language OSSLT were eliminated from the equating process due to modifications made to the items between the 2008 field test and the 2009 operational test.

The score 1.0 in the scoring rubric for one long-writing item was collapsed with the score 1.5 for the use of conventions section in the English-language test. The score 1.0 in the scoring rubric for the two long-writing items was collapsed with the score 1.5 for the topic development section in the French-language test. However, the score 1.0 was collapsed with the score 1.5 for only one long-writing item for the use of conventions section in the French-language test.

Equating Results

The equating results based on the equating samples for the OSSLT are summarized in Table 5.15. The theta cut score (see step 5 above) and the percentage of successful and unsuccessful students in 2008 and 2009 are reported for English-language and French-language students. For example, the equated cut score for the English-language test was -0.63. The percentage of successful students in the equating samples was 86.3% in 2008 and 87.3% in 2009.

Table 5.15 Equating Results for the OSSLT

	English-Language			French-Language		
	Theta Cut Point 2008–2009	Equating Sample 2008	Equating Sample 2009	Theta Cut Point 2008–2009	Equating Sample 2008	Equating Sample 2009
No. of Students		144 274	140 328		5240	5163
% Successful	-0.63	86.3%	87.3%	-0.99	84.8%	86.0%
% Unsuccessful		13.7%	12.7%		15.2%	14.0%

References

Kolen, M. J. & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (2nd ed.). New York: Springer-Verlag.

Muraki, E. & Bock, R.D. (2003). PARSCALE: IRT item analysis and test scoring for rating-scale data. (Version 4.1) [Computer Software]. Chicago, IL: Scientific Software International.

CHAPTER 6: REPORTING RESULTS

EQAO assessment results are reported at the student, school, school board and provincial levels. The results describe student achievement at critical stages in a student's education and provide useful information to support improvement planning at the school, school board and provincial levels.

EQAO publishes an annual provincial report for education stakeholders and the general public. The following reports for the 2009 assessments are available on the EQAO Web site (www.eqao.com):

- *EQAO's Provincial Report on the Results of the 2008–2009 Assessments of Reading, Writing and Mathematics, Primary Division (Grades 1–3) and Junior Division (Grades 4–6), and the Grade 9 Assessment of Mathematics (September 2009)* and
- *EQAO's Provincial Report on the Results of the 2008–2009 Ontario Secondary School Literacy Test (June 2009).*

Corresponding reports for the French-language assessments are also provided.

EQAO posts school and board results on its Web site for public access (www.eqao.com). However, EQAO does not publicly release school or board results when the number of students being reported on for a school or board is small enough that individual students could be identified (i.e., fewer than 15 students for achievement results and fewer than six students for questionnaire results).

The following two types of aggregate results are reported for schools, boards and the province:

1. the percentage of all students enrolled in Grade 3 and Grade 6, students enrolled in Grade 9 academic and applied mathematics courses and students eligible to write the OSSLT.
2. the percentage of students who participated in each assessment.

More detailed school and board results are posted on the secure section of the EQAO Web site and are available only to school board and school personnel through user identification numbers and passwords. These reports include results for small *n*-counts that are suppressed in the public reports and additional achievement results for sub-groups of the student population: English language learners, students with special education needs and French Immersion students in Grade 3. Results for male and female students are included in both the public and secure reports. In addition, schools and school boards receive data files with individual student achievement results for all students in the school and the board, and data files with aggregated results for each school, the board and the province.

Reporting the Results of the Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions

The four achievement levels defined in *The Ontario Curriculum* are used by EQAO to report student achievement in reading, writing and mathematics. Level 3 has been established as the provincial standard. Levels 1 and 2 indicate achievement below the provincial standard, and

Level 4 indicates achievement above the provincial standard. In addition to the four performance levels, the following three reporting categories are included:

- not enough evidence for Level 1 (NE1),
- no data and
- exempt.

Students with no data and students who were exempted are not included in the calculation of results for students participating in the assessments.

The Individual Student Report (ISR) for students in Grades 3 and 6 shows the overall achievement level for each subject (reading, writing and mathematics) at one of the five possible positions within the achievement level. School, school board and provincial results are included on the ISR to provide a context for interpreting student results. The five possible positions within the achievement levels are created from the distribution of student theta scores. Through the equating process, described in chapter 5, four theta values are identified that define the cut points between adjacent achievement levels (i.e., NE1/Level 1, Level 1/Level 2, Level 2/Level 3 and Level 3/Level 4). The width of each sub-category in a given level is determined by the range of theta values represented in the level divided by five. These results are designated in student data files as 1.1, 1.3, 1.5, 1.7, 1.9, 2.1, 2.3, and so on to 4.9. For students in Grade 6, the assessment results achieved in Grade 3, if available, are printed on their ISR.

The provincial, board and school reports provide the following:

- overall jurisdictional results for each subject (reading, writing and mathematics);
- longitudinal data showing jurisdictional results over time;
- overall jurisdictional results for each subject (reading, writing and mathematics) by gender and characteristics (such as English language learners and students with special education needs);
- areas of strength within the curriculum and areas for improvement;
- cohort-tracking results from Grade 3 to Grade 6;
- contextual data and
- student questionnaire results.

Results for the teacher and principal questionnaires are reported at the board and provincial levels. Some provincial results for the questionnaires are included in the provincial reports. Full results for all questionnaires are posted on EQAO's public Web site.

In addition, schools receive student rosters, which provide item results for each student who completed each assessment and summary item statistics for the school, school board and province. The data for individual students are also provided in data files. Cohort-tracking reports for schools and boards show a comparison of Grade 3 and Grade 6 results for the same students.

Directors of education are provided with graphs that show the number and percentage of schools with achievement levels in specified categories (e.g., 75% of their students achieving the provincial standard). The directors also receive school lists with achievement results over time, for convenient reference.

Reporting the Results of the Grade 9 Assessment of Mathematics

Reporting for the Grade 9 mathematics assessment is very similar to that for the primary- and junior-division assessments. The same four achievement levels with five sub-categories are used to report student achievement. The set of reports described above for the primary- and junior-division assessments is also produced for the Grade 9 assessment, with the exception of the cohort-tracking report. However, there are some differences between the reporting for the primary- and junior-division assessments and the Grade 9 assessment. For instance, the option to exempt students from the Grade 9 mathematics assessment was removed in 2007. Moreover, the reporting category “Not Enough Evidence for Level 1” is called “Below Level 1” in the Grade 9 assessment. In addition to the disaggregations identified for the primary- and junior-division assessments, results for the Grade 9 mathematics assessment are reported for semester 1, semester 2 and the full year. Furthermore, the ISR for Grade 9 does not include school and board results, and there is no principal questionnaire for the Grade 9 assessment. Summary cohort-tracking results from Grade 6 to Grade 9 are provided at the school, board and provincial levels, but the Grade 6 results for individual students are not recorded on the Grade 9 ISR.

Reporting the Results of the OSSLT

For the OSSLT, EQAO reports only two levels of achievement: successful and unsuccessful. A successful result on the OSSLT or the successful completion of the Ontario Secondary School Literacy Course (OSSLC) is required to meet the literacy requirement for graduation. Students must achieve a specified minimum theta score to receive a successful result. The process for establishing this minimum score is described in chapter 5. EQAO provides feedback to unsuccessful students to assist them in working to achieve the expected standard.

As with the other assessments, EQAO reports results for all students and for participating students. Students are considered to be not participating if they have been deferred to write in a later year, have opted to take the OSSLC or have no data for the current administration. Students who are not working toward the Ontario Secondary School Diploma are exempt from the OSSLT and are not included in either population used for reporting (i.e., all students or participating students). Aggregated results are reported separately for first-time eligible students and previously eligible students. Previously eligible students are those who were unsuccessful on a previous administration, were deferred from a previous administration or arrived in an Ontario school during their Grade 11 or Grade 12 year.

The OSSLT ISR provides the following:

- specification of successful or unsuccessful results and
- feedback for unsuccessful students to help them improve their level of achievement.

The student’s scale score is reported to each unsuccessful student, and these students are informed that a successful result requires a scale score of 300.

The OSSLT provincial, board and school reports provide the following:

- overall successful and unsuccessful jurisdictional results;
- overall successful and unsuccessful jurisdictional results by gender and characteristics such as English language learners and students with special education needs;

- results by exceptionality category and results for students receiving accommodations;
- results by type of English- or French-language course: academic, applied, locally developed or English as a Second Language (ESL)/English Literacy Development (ELD) or actualisation linguistique en français (ALF)/programme de perfectionnement du français (PDF);
- longitudinal data showing jurisdictional results over time;
- areas of strength within the curriculum and areas for improvement;
- cohort-tracking results from Grade 6 to OSSLT and
- results for the student questionnaire and contextual data.

In addition, schools receive student rosters that provide item results for each unsuccessful student who completed the test and summary item statistics for the school, board and province.

Interpretation Guides

Guidelines for interpreting results are included in the school and board reports. The following resource documents are also provided:

- *Using Data to Promote Student Success: A Brief Guide to Assist School Administrators in Interpreting Their Data*
 - *Assessments of Reading, Writing and Mathematics, Primary Division (Grades 1–3) and Junior Division (Grades 4–6) (August 2008)*
 - *Grade 9 Assessment of Mathematics (August 2008)*
 - *Ontario Secondary School Literacy Test (June 2008)*
- *Guide to Using the Item Information Report: Student Roster*
- *EQAO Questionnaire Data: Using Contextual Data to Inform Improvement Planning*
- *Profile Report Technical Paper*

Released test items and scoring guides are also posted on the EQAO Web site to assist with the interpretation of results.

CHAPTER 7: STATISTICAL AND PSYCHOMETRIC SUMMARIES

A variety of statistical and psychometric analyses were conducted for the 2009 Assessments of Reading, Writing and Mathematics, Primary Division and Junior Division; the Grade 9 Assessment of Mathematics and the Ontario Secondary School Literacy Test (OSSLT). The results from these analyses are summarized in this chapter, including results for Classical Test Theory (CTT), Item Response Theory (IRT), Differential Item Functioning (DIF) and decision accuracy and consistency. All IRT item parameters were obtained from the calibration process used for the equating samples (described in chapter 5). Detailed data for individual items are provided in appendix 7.1.

Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions

Classical Test Theory (CTT) Analyses

Table 7.1 presents descriptive statistics, Cronbach's alpha estimates of test score reliability and the standard error of measurement (SEM) for the English-language version and French-language version of the primary and junior assessments. The test means (converted to percentages) ranged from 59.5% (for junior English-language math) to 67.7% (for junior English- and French-language writing).

Reliability and the corresponding standard error of measurement refer to the precision of test scores, with higher reliability coefficients and lower standard errors of measurement indicating higher levels of precision. For the primary and junior assessments, Cronbach's alpha estimates range from 0.83 to 0.87 for reading, 0.77 to 0.80 for writing and 0.87 to 0.89 for mathematics. The corresponding standard errors of measurement range from 4.8% to 5.0% of the possible maximum score for reading, 6.7% to 6.8% of the possible maximum score for writing and 6.0% and 6.5% of the possible maximum score for mathematics. The reliabilities for writing are a little lower than those for reading and mathematics, and the corresponding standard errors are a little higher. This is attributable, in part, to the smaller number of writing items and the subjectivity in scoring writing performance¹. Taking these two factors into account, the obtained reliabilities and standard errors of measurement are acceptable and indicate that the test scores from these assessments provide a satisfactory level of precision.

¹ For example, if the number of writing items was increased to match the number of reading items, Cronbach's alpha estimate for the French-language version of the primary writing assessment would be about 0.85.

Table 7.1 Test Descriptive Statistics, Reliability and Standard Error of Measurement: Primary and Junior Divisions

Grade and Subject	No. of Items	Item Type		Possible Max. Score	No. of Students	Min.	Max.	Mean	SD	Cronbach's Coefficient Alpha	SEM
		MC	OR								
Primary Reading (English)	36	26	10	66	116 256	0	65	40.30	8.55	0.86	3.20
Junior Reading (English)	36	26	10	66	131 173	0	66	42.56	9.09	0.87	3.28
Primary Reading (French)	36	26	10	66	6 371	4	63	39.71	9.22	0.87	3.32
Junior Reading (French)	36	26	10	66	6 192	0	62	41.24	7.79	0.83	3.21
Primary Writing (English)	14	8	6*	29	116 812	0	29	19.03	4.41	0.80	1.97
Junior Writing (English)	14	8	6*	29	131 296	0	29	19.63	4.42	0.78	2.07
Primary Writing (French)	14	8	6*	29	6 379	0	29	18.93	4.45	0.80	1.99
Junior Writing (French)	14	8	6*	29	6 196	0	29	19.62	4.17	0.77	1.99
Primary Mathematics (English)	36	28	8	60	120 405	0	60	38.70	10.88	0.88	3.77
Junior Mathematics (English)	36	28	8	60	130 902	0	60	35.70	11.25	0.89	3.73
Primary Mathematics (French)	36	28	8	60	6 393	3	60	40.00	9.91	0.87	3.57
Junior Mathematics (French)	36	28	8	60	6 194	0	60	37.80	10.76	0.87	3.88

Note. MC = multiple-choice; OR = open-response; SD = standard deviation; SEM = standard error of measurement; * = short writing and long writing.

Item Response Theory (IRT) Analyses

In the IRT analysis, item parameters are estimated from the student responses used in the equating calibration sample (see chapter 5). The estimated item parameters are then used to score all student responses. The descriptive statistics for the IRT scores reported in Table 7.2 include the total population. The mean student ability scores are less than zero, and the standard

deviations are less than one (due to the inclusion of all students). The item parameters for individual items are presented in appendix 7.1.

Table 7.2 Descriptive Statistics of IRT Scores: Primary and Junior Divisions

Grade and Subject	No. of Students	Min.	Max.	Mean	SD
Primary Reading (English)	116 256	-3.86	3.44	-0.08	0.98
Junior Reading (English)	131 173	-3.89	3.23	-0.12	1.00
Primary Reading (French)	6 371	-3.84	2.84	-0.06	0.97
Junior Reading (French)	6 192	-3.92	3.02	-0.07	0.95
Primary Writing (English)	116 812	-3.60	2.51	-0.05	0.92
Junior Writing (English)	131 296	-3.69	2.30	-0.11	0.95
Primary Writing (French)	6 379	-3.66	2.45	-0.03	0.94
Junior Writing (French)	6 196	-3.78	2.36	-0.06	0.95
Primary Mathematics (English)	120 405	-3.79	2.48	-0.07	0.97
Junior Mathematics (English)	130 902	-3.71	2.62	-0.09	0.99
Primary Mathematics (French)	6 393	-3.75	2.59	-0.04	0.96
Junior Mathematics (French)	6 194	-3.70	2.59	-0.04	0.96

Note. SD = standard deviation.

The Test Characteristic Curves (TCCs) and the distribution of student thetas are provided in Figures 7.1 to 7.12. The TCCs slope upward from the lower left to the upper right. These curves can be used to translate a student ability score in IRT to a student ability score in CTT, as indicated by the left vertical axis. For example, a primary student with a theta score of -1.0 in English-language reading is expected to have an observed score of about 50%. The right vertical scale indicates the percentage of students at each theta value, and the theta cut points used to assign students to performance levels are marked on the graphs. The Test Information Functions (TIFs), which indicate where the components of each assessment contain the most information, are provided in Figures 7.13 to 7.24. For example, the maximum information provided by the reading component of the English-language primary-division assessment is at approximately theta -0.70 . The precision of the scores is greatest at this point. The theta cut points used to assign students to achievement levels are also marked to show the amount of information at each cut point.

Figure 7.1 Test Characteristic Curve and Distribution of Thetas for Reading: Primary Division (English)

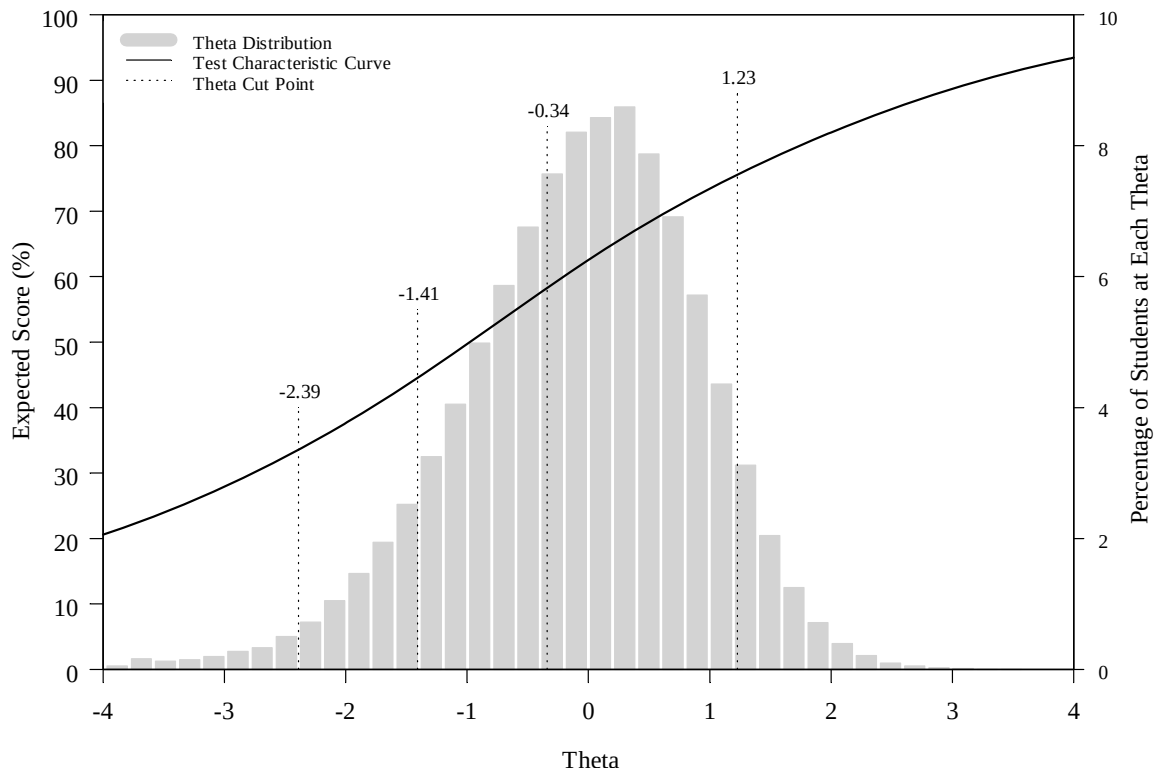


Figure 7.2 Test Characteristic Curve and Distribution of Thetas for Reading: Primary Division (French)

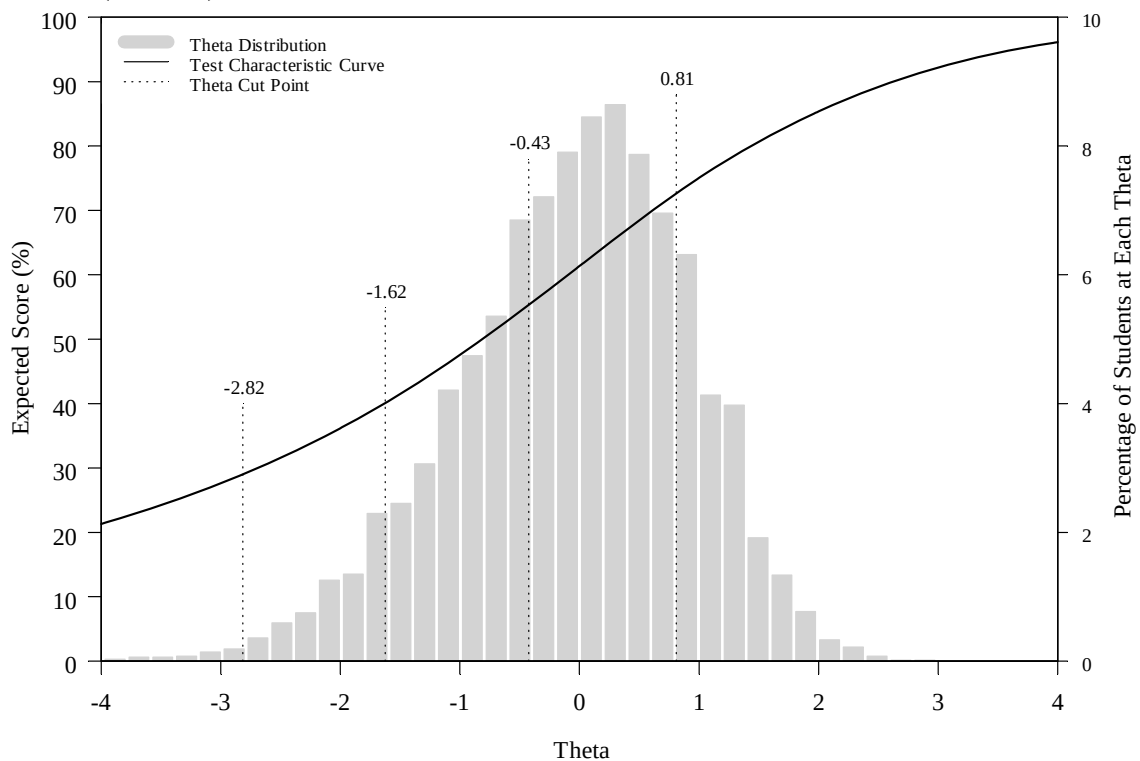


Figure 7.3 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (English)

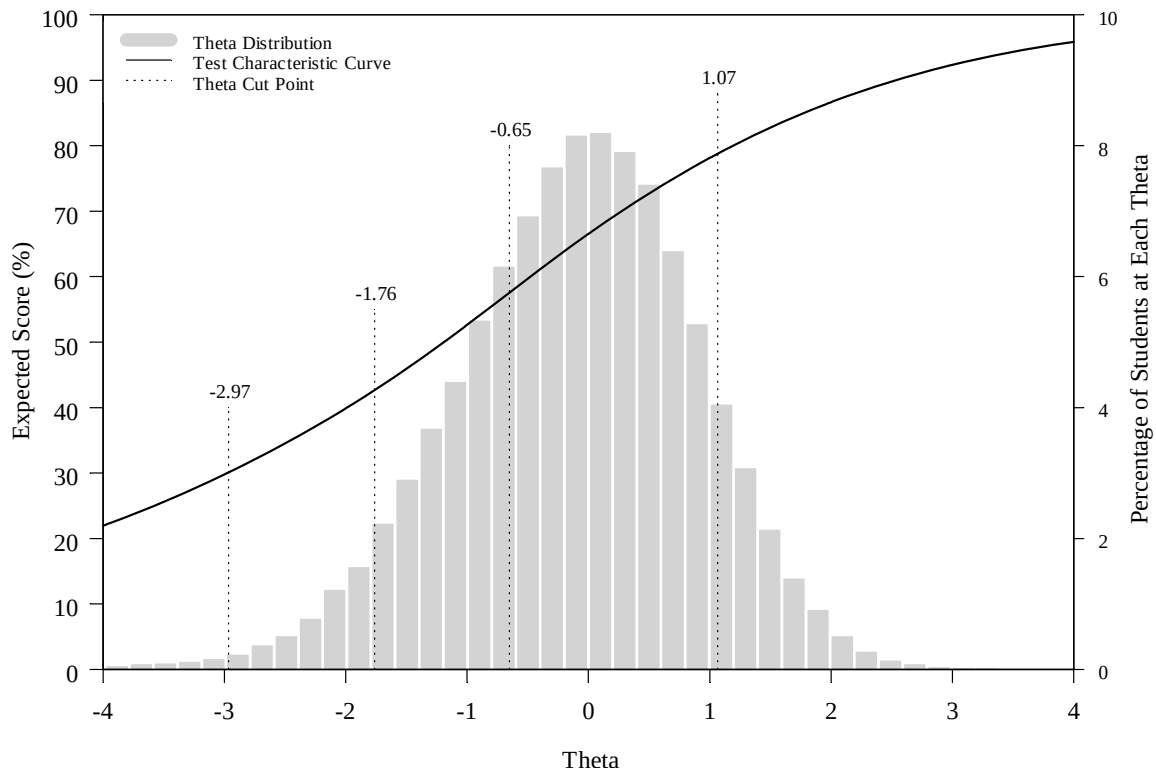


Figure 7.4 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (French)

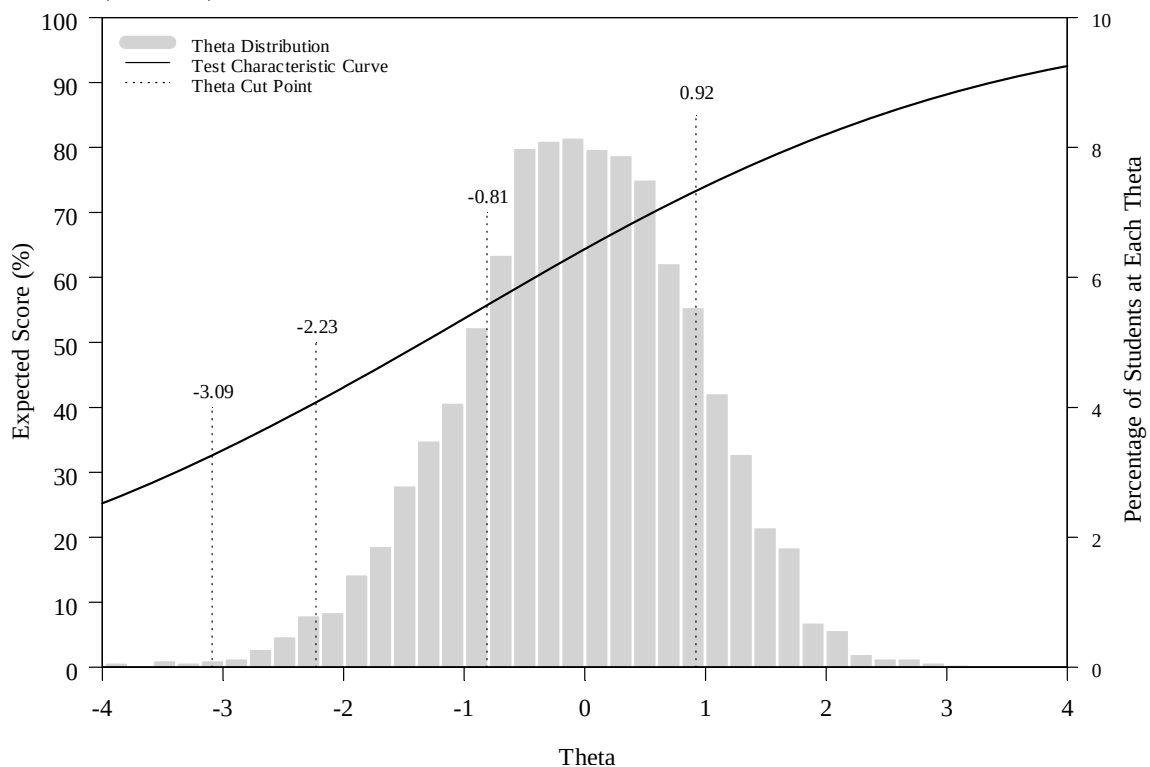


Figure 7.5 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (English)

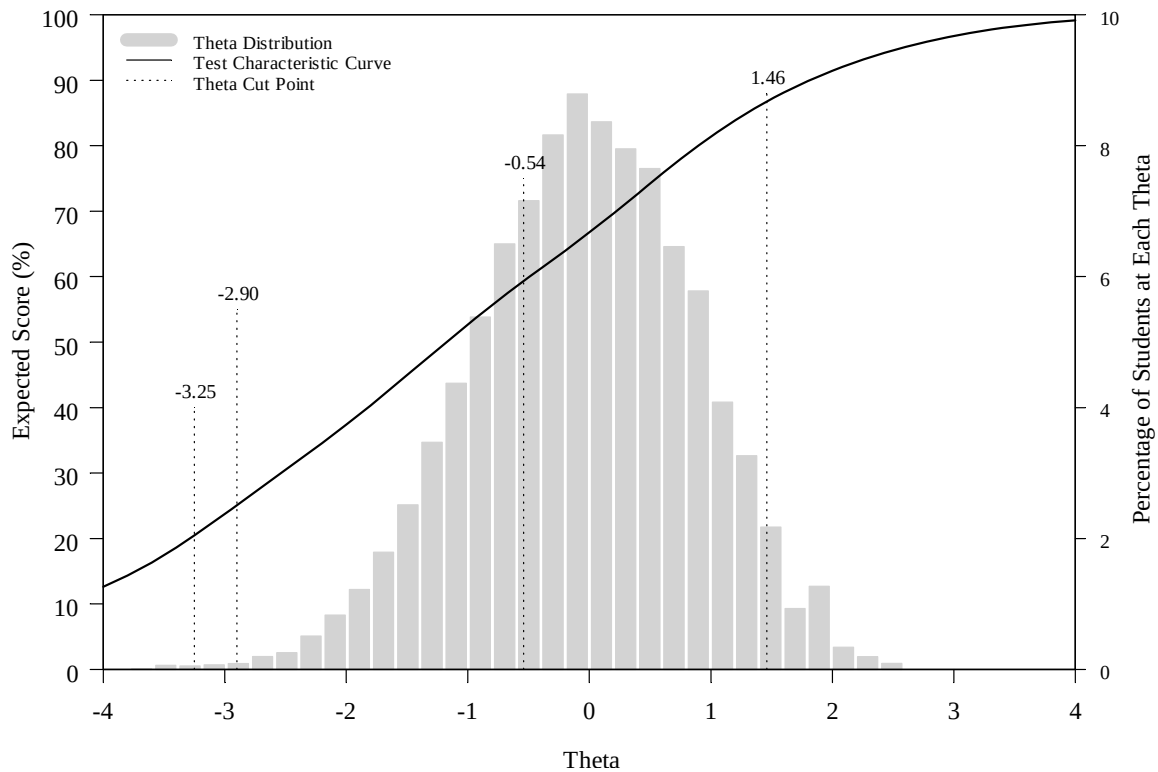


Figure 7.6 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (French)

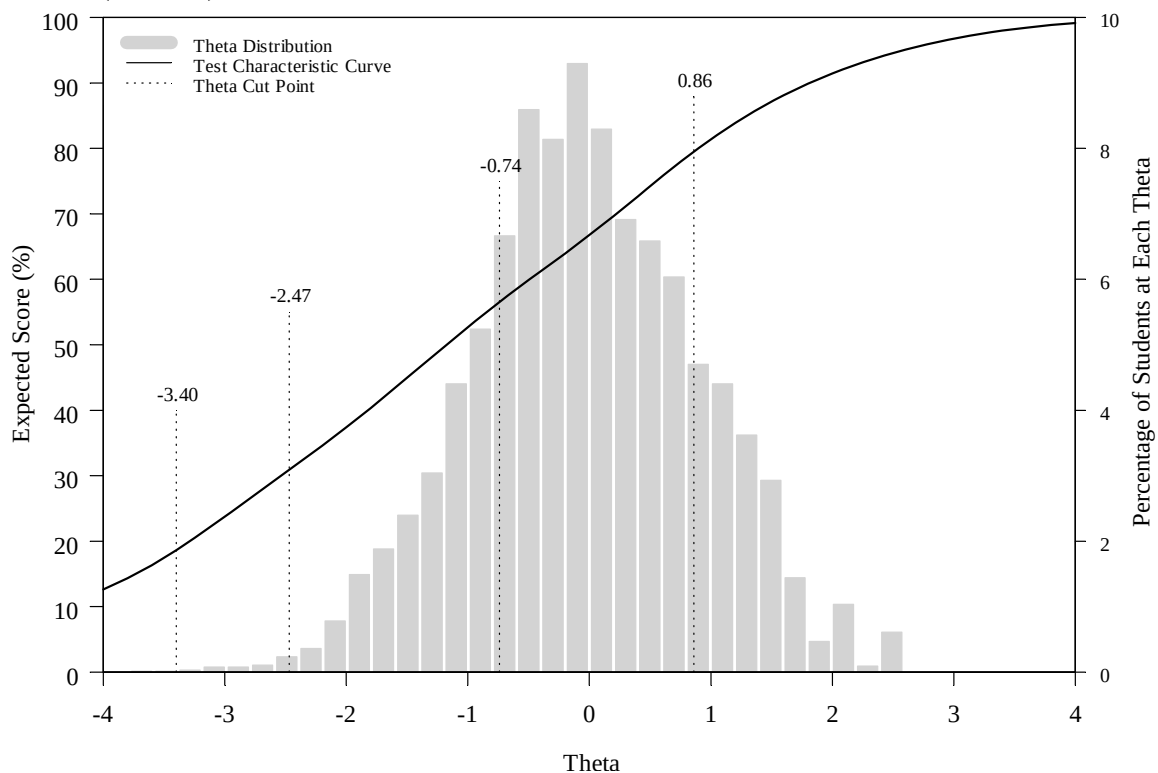


Figure 7.7 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (English)

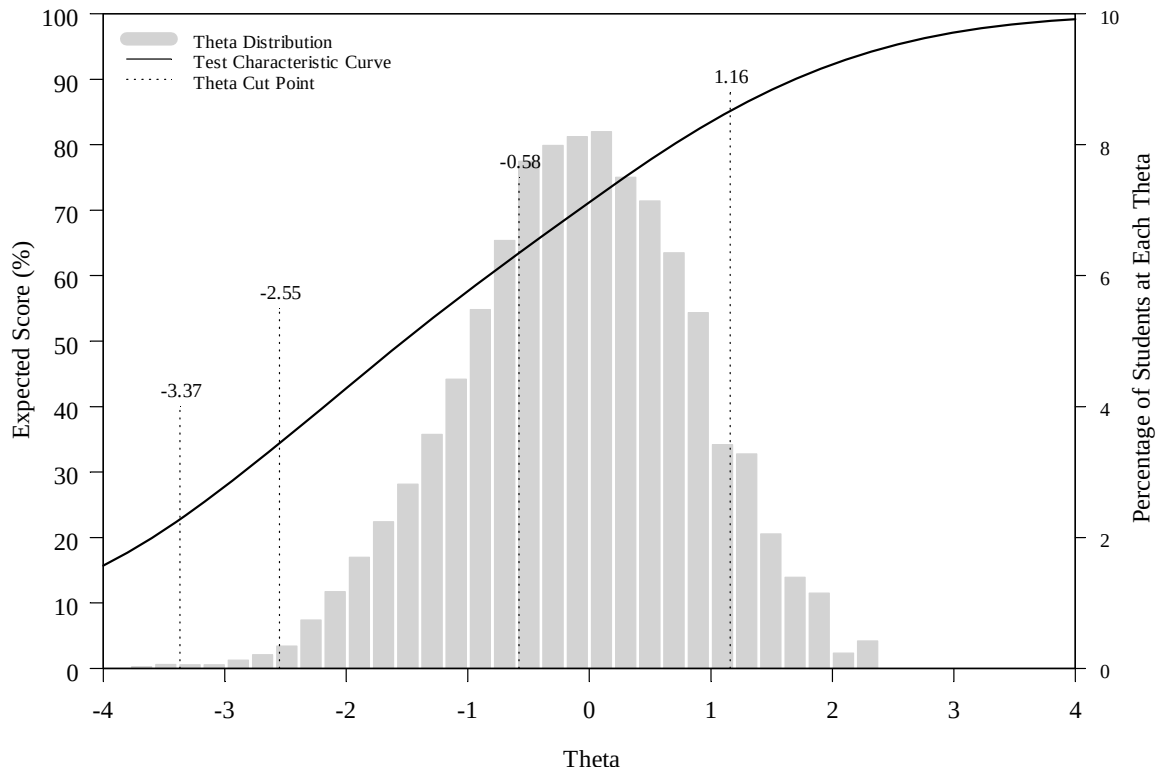


Figure 7.8 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (French)

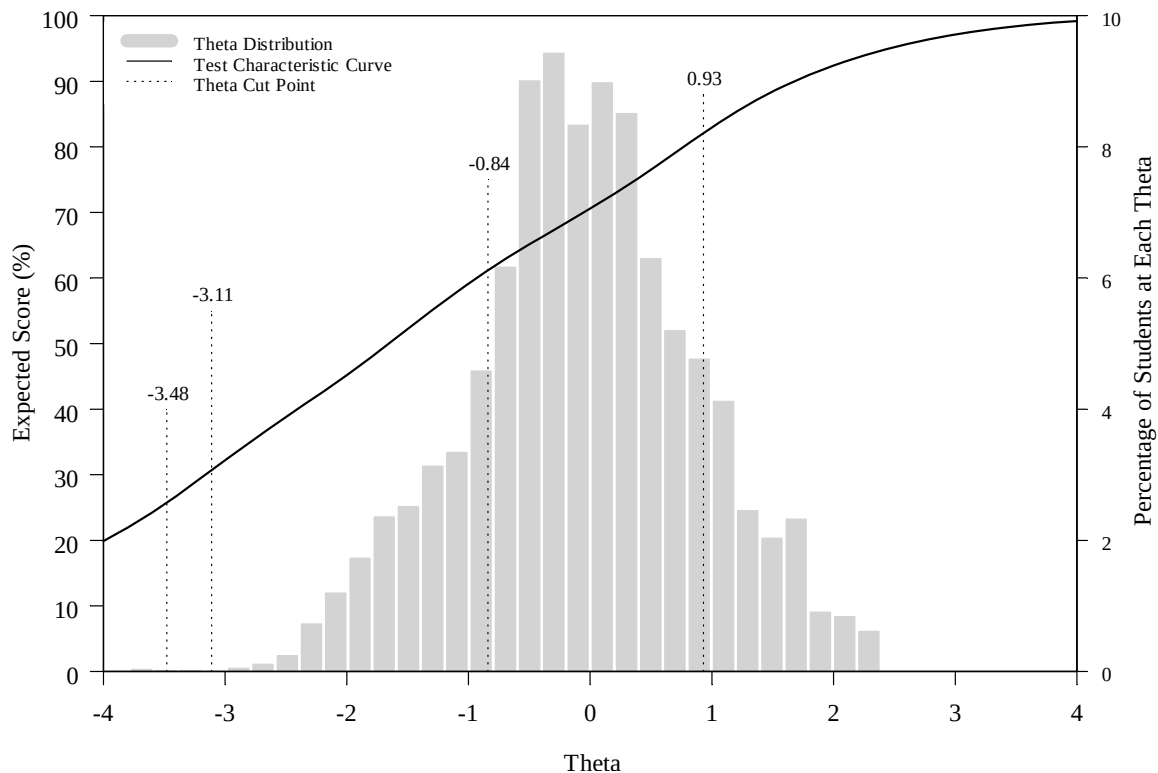


Figure 7.9 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (English)

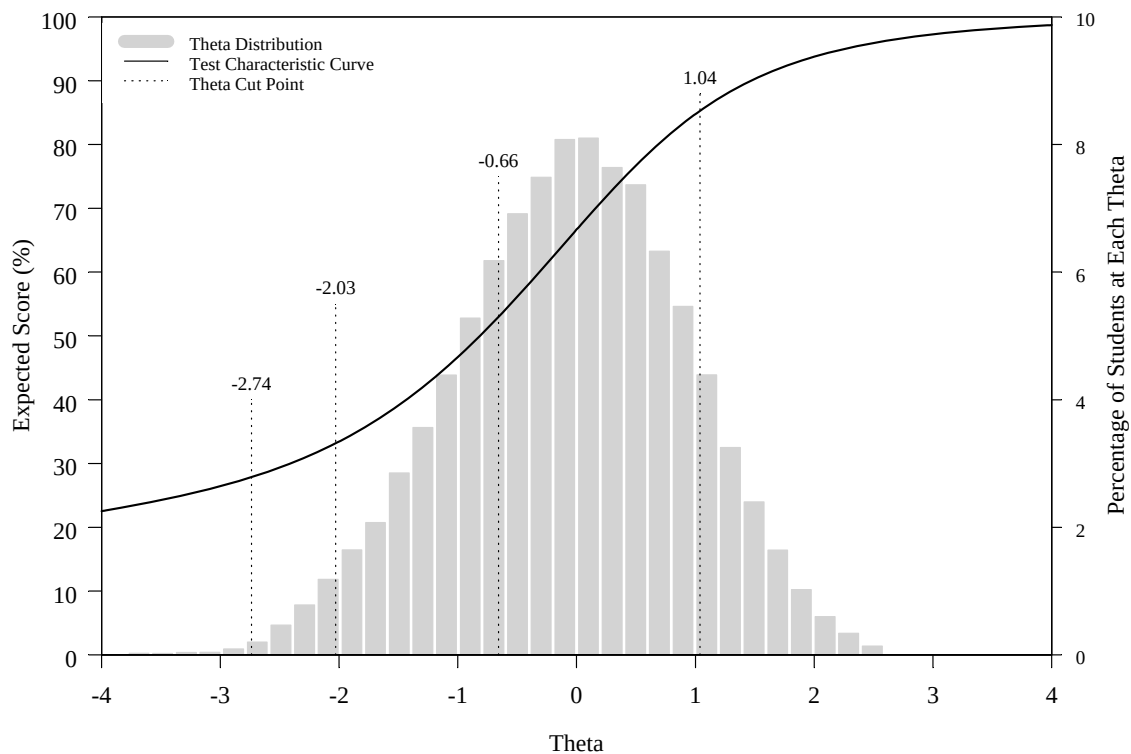


Figure 7.10 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (French)

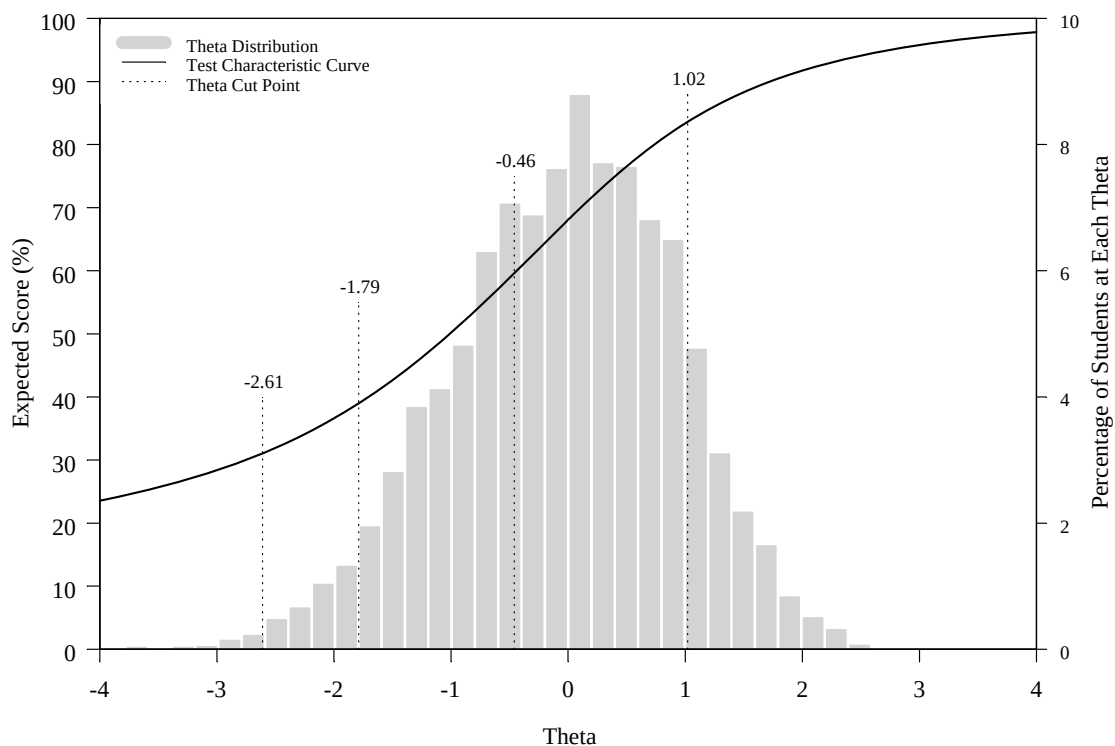


Figure 7.11 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (English)

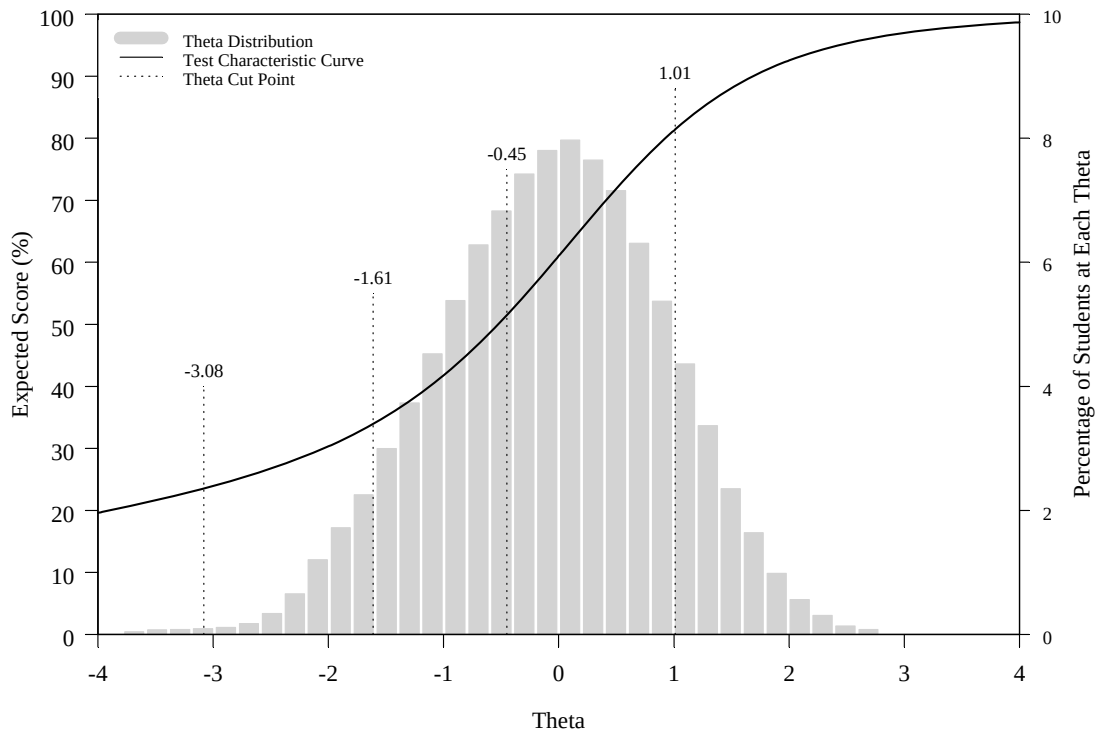


Figure 7.12 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (French)

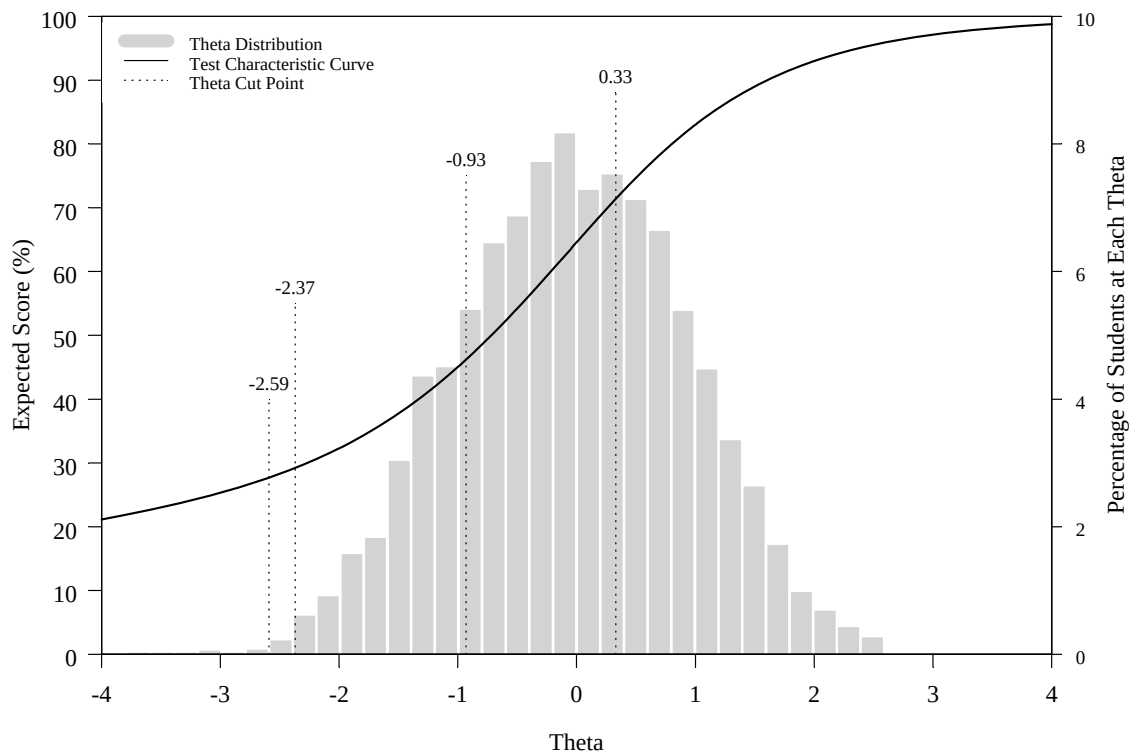


Figure 7.13 Test Information Function for Reading: Primary Division (English)

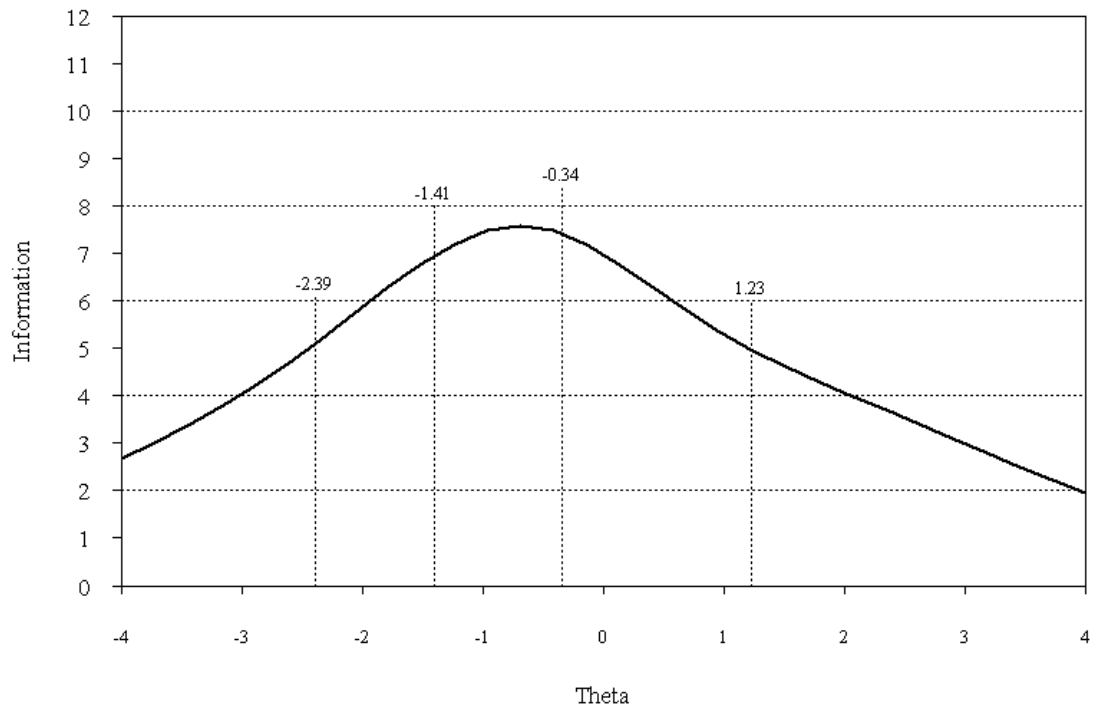


Figure 7.14 Test Information Function for Reading: Primary Division (French)

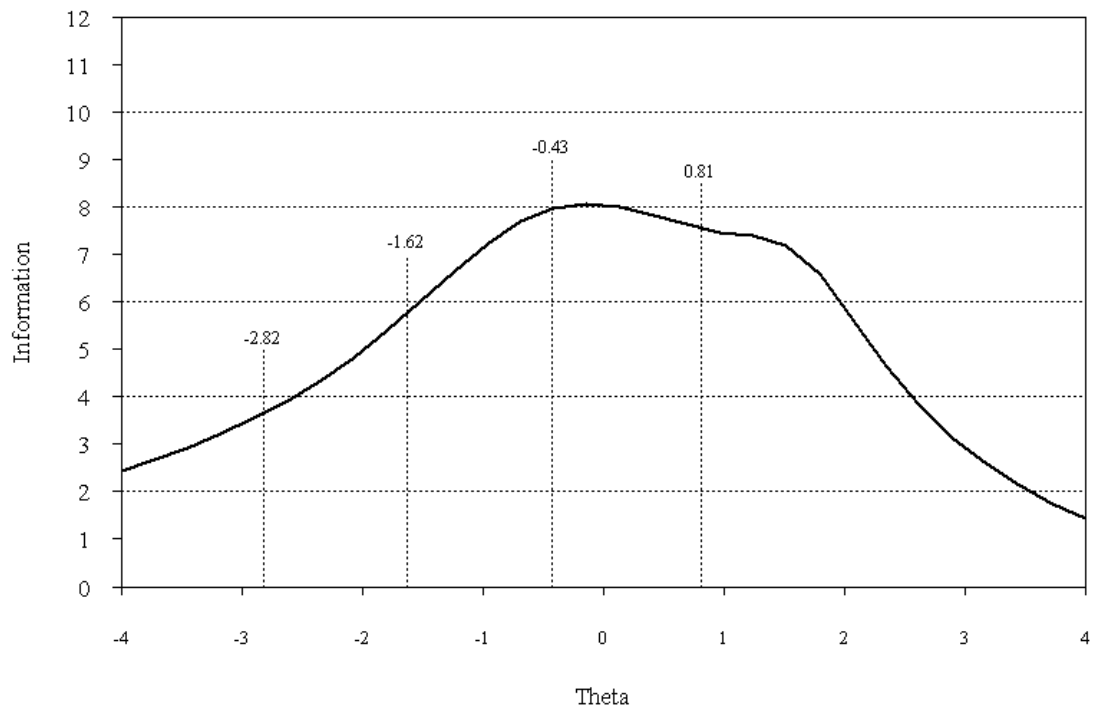


Figure 7.15 Test Information Function for Reading: Junior Division (English)

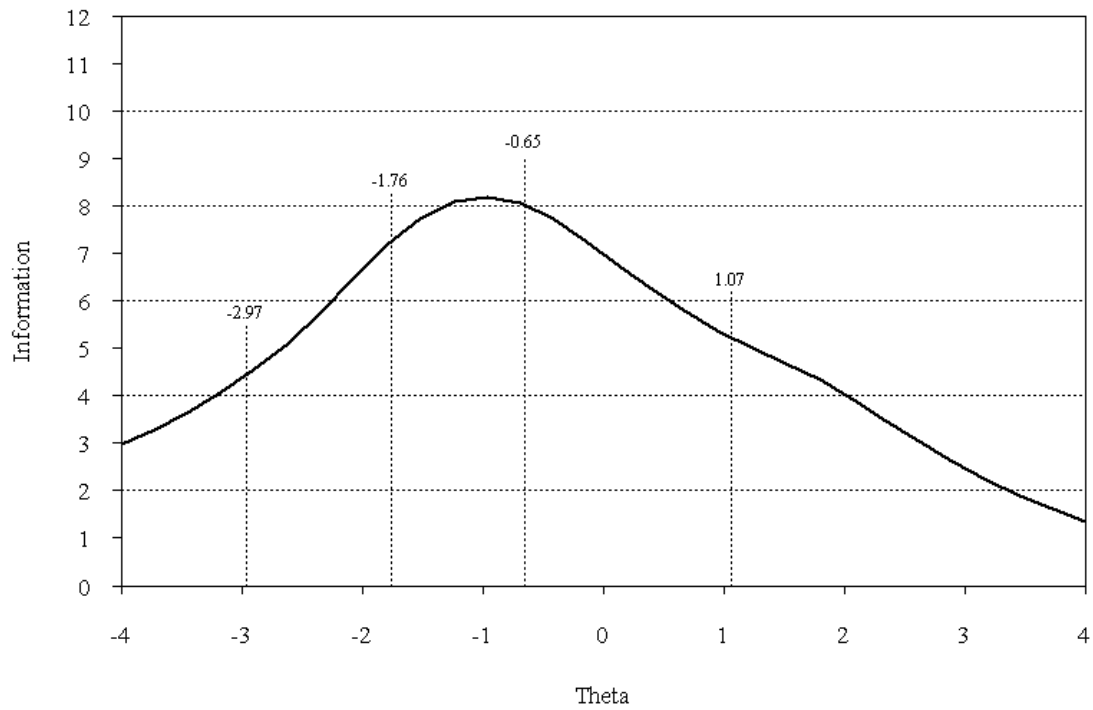


Figure 7.16 Test Information Function for Reading: Junior Division (French)

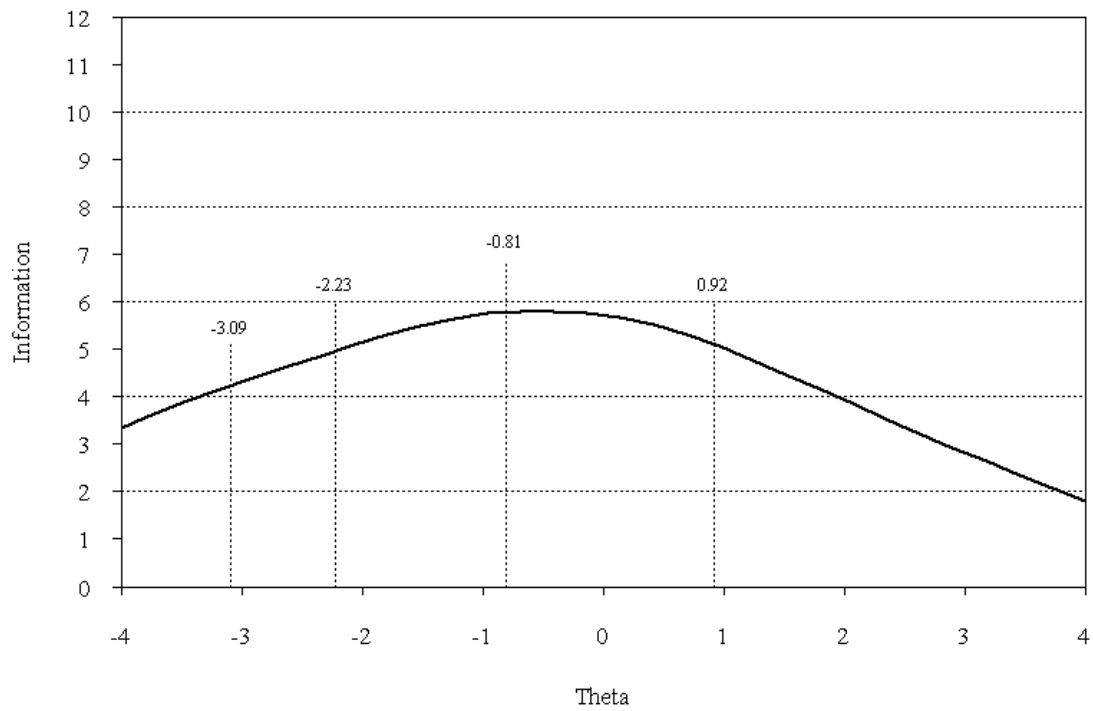


Figure 7.17 Test Information Function for Writing: Primary Division (English)

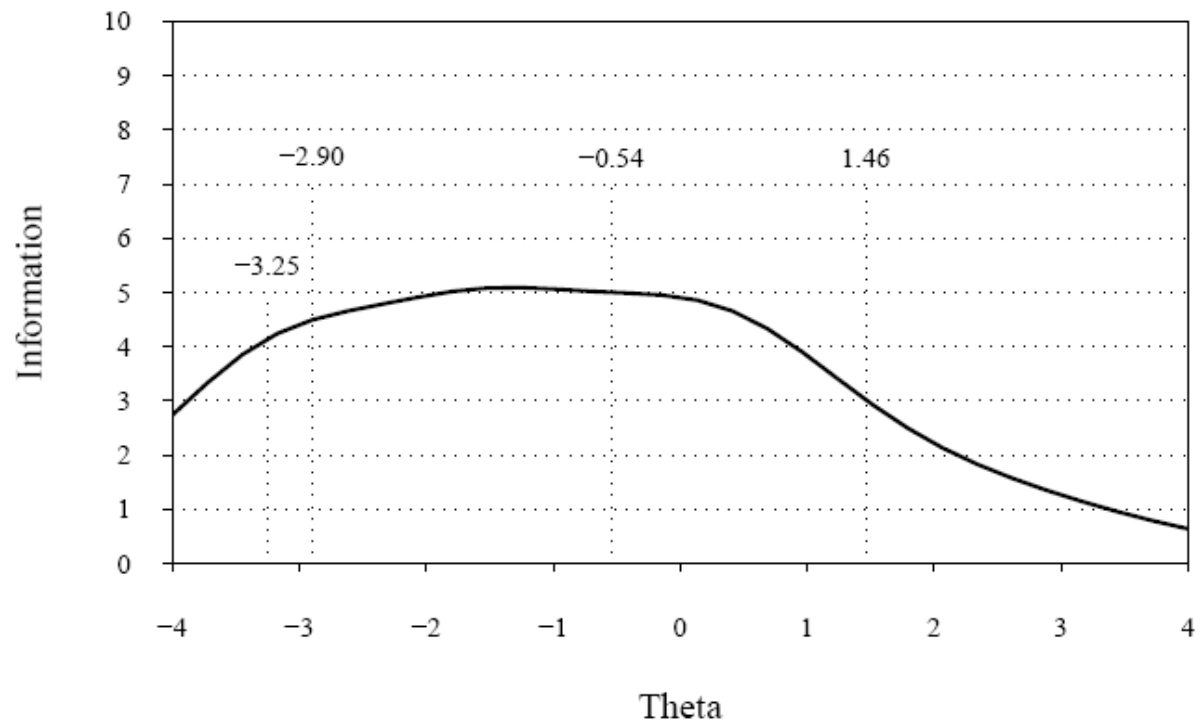


Figure 7.18 Test Information Function for Writing: Primary Division (French)

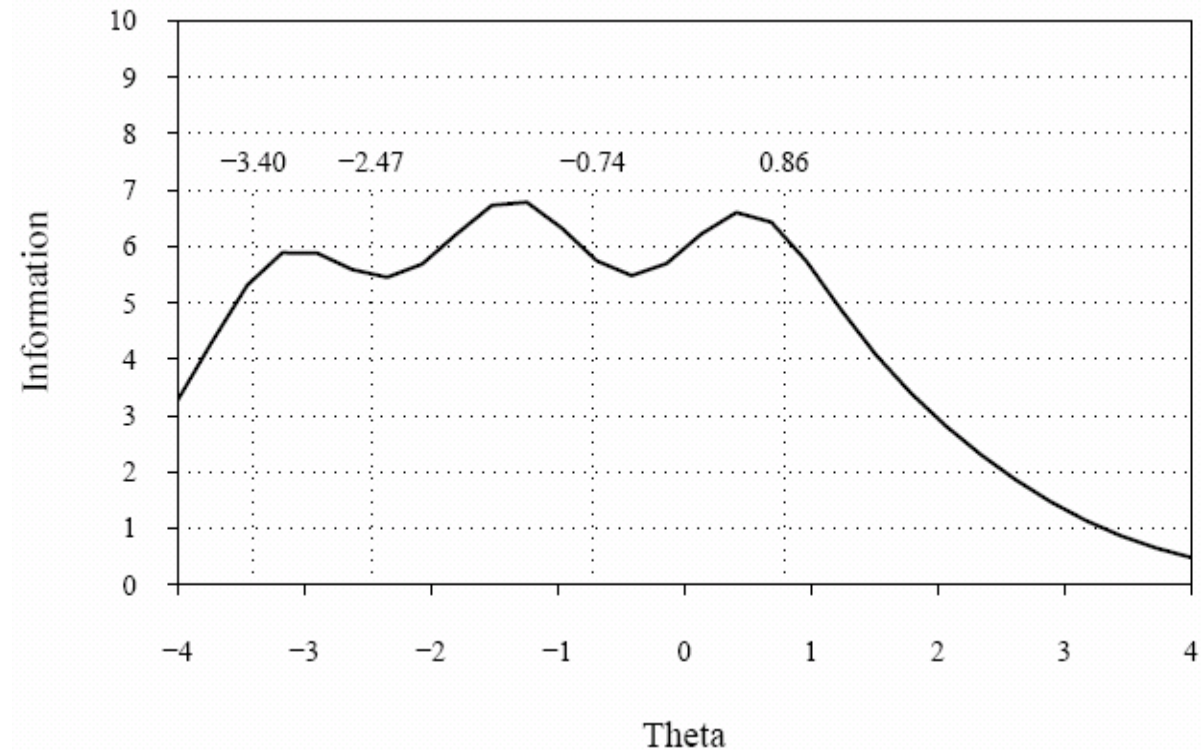


Figure 7.19 Test Information Function for Writing: Junior Division (English)

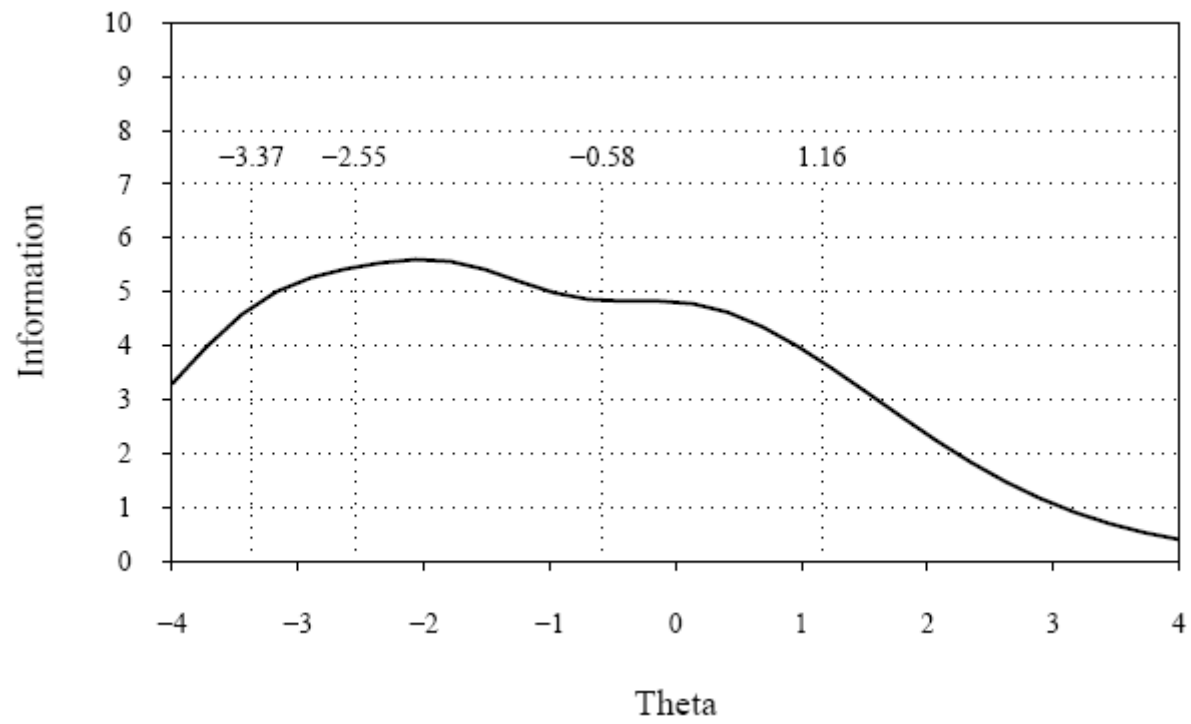


Figure 7.20 Test Information Function for Writing: Junior Division (French)

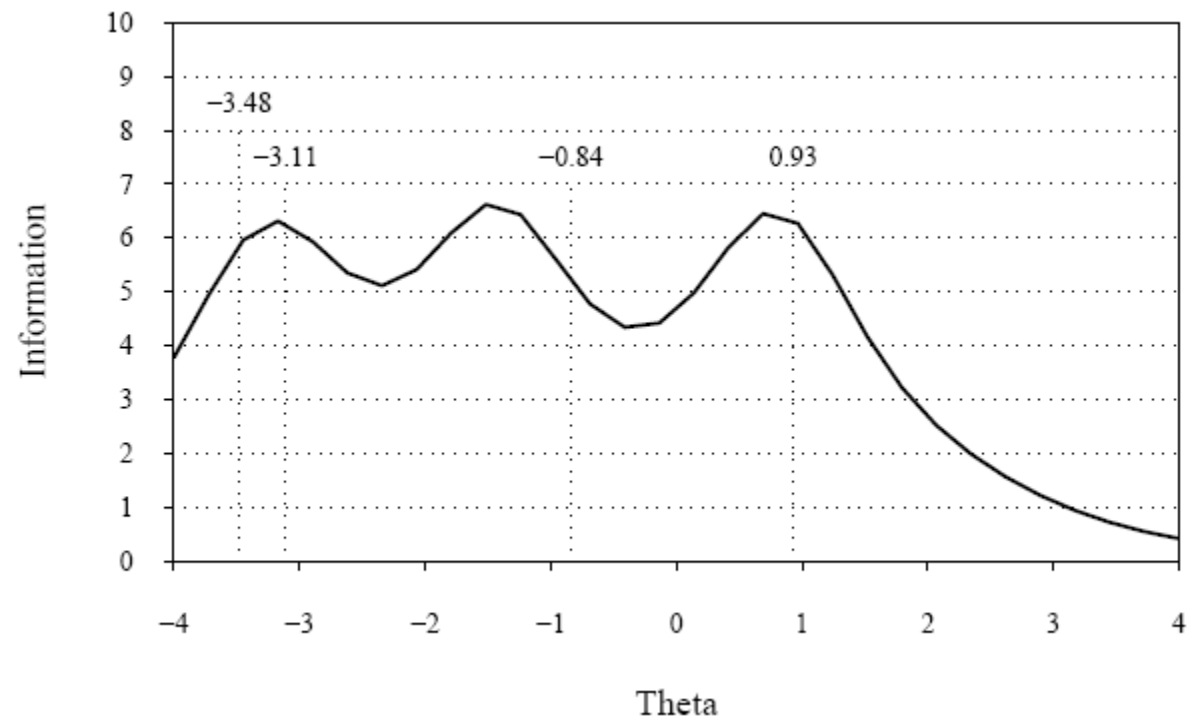


Figure 7.21 Test Information Function for Mathematics: Primary Division (English)

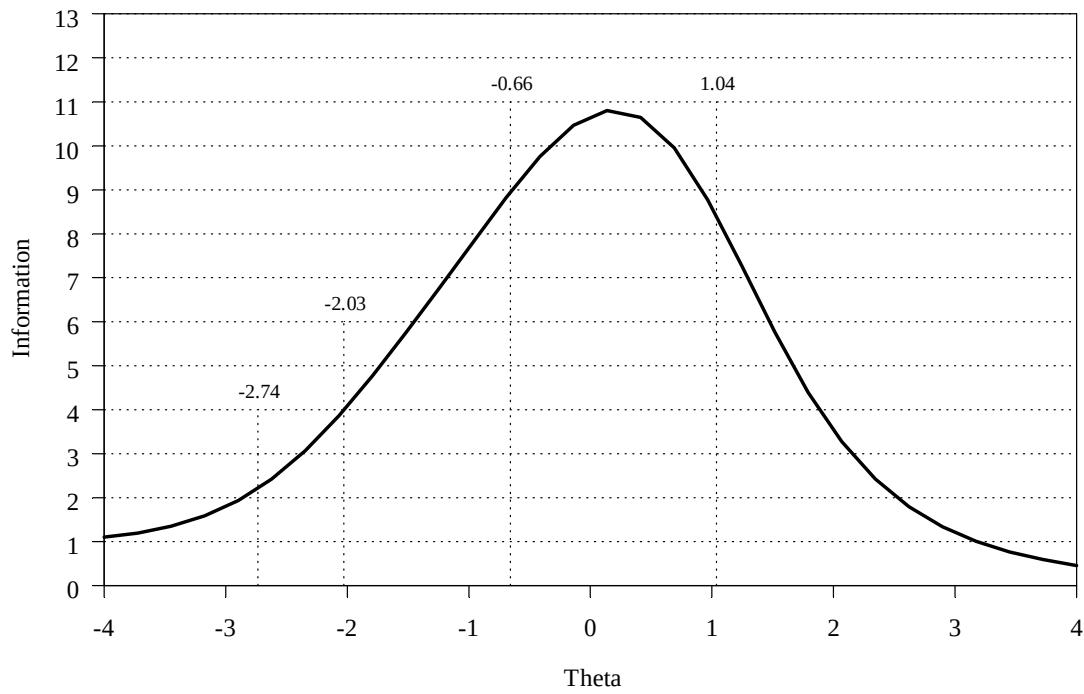


Figure 7.22 Test Information Function for Mathematics: Primary Division (French)

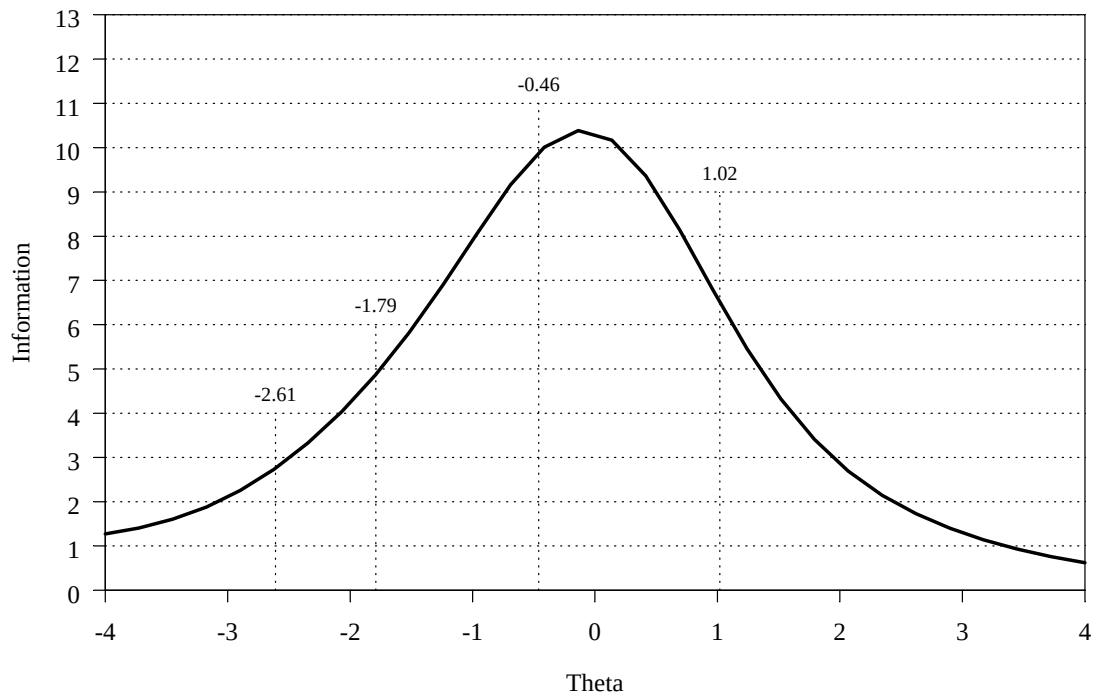


Figure 7.23 Test Information Function for Mathematics: Junior Division (English)

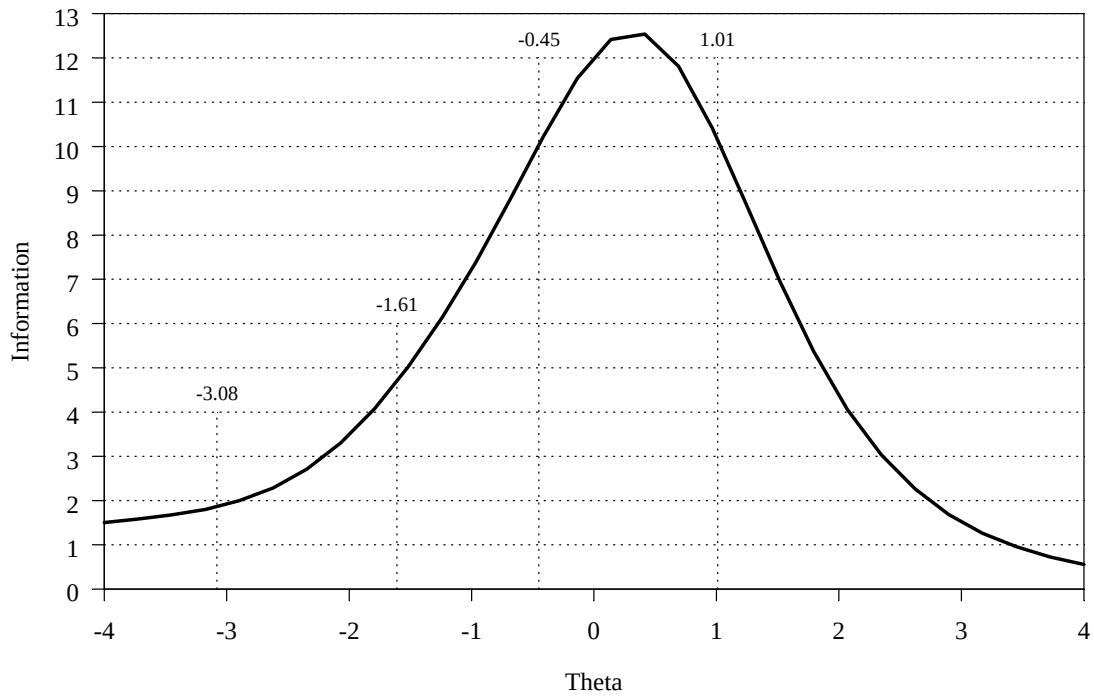
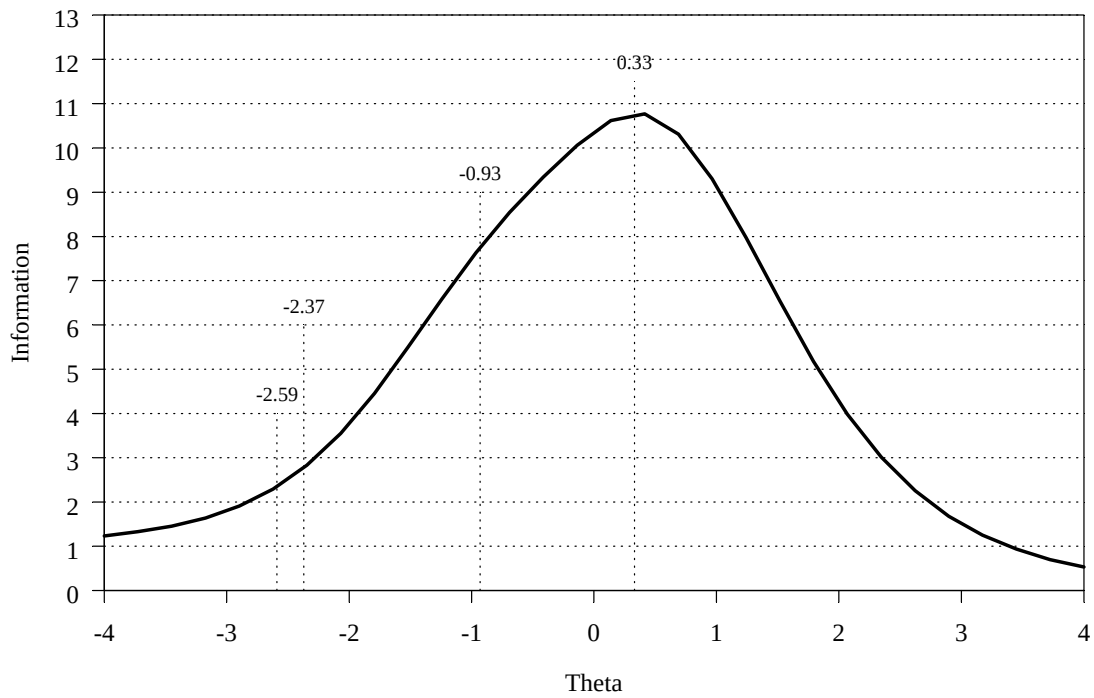


Figure 7.24 Test Information Function for Mathematics: Junior Division (French)



Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)

Table 7.3 contains a summary of both the CTT and IRT descriptive item statistics for the items included in the English-language and French-language versions of the primary- and junior-division assessments. These statistics were computed using the equating sample (see chapter 5). As expected, there is an inverse relationship between the CTT item difficulty estimates and the IRT location parameter estimates, due to the difference between the difficulty definitions in the two approaches. In contrast, there is a positive relationship between the CTT item-total correlation estimates and the IRT slope parameter estimates. Statistics for individual items are presented in appendix 7.1.

Table 7.3 Descriptive Statistics of CTT Item Statistics and IRT Item Parameter Estimates: Primary and Junior Divisions

Grade and Subject	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*	
			Item Difficulty	Item-total Correlation	Slope	Location
Primary Reading (English)	36	Min.	0.38	0.18	0.24	-2.22
		Max.	0.91	0.52	1.02	1.39
		Mean	0.67	0.36	0.59	-0.81
		SD	0.14	0.09	0.20	0.81
Junior Reading (English)	36	Min.	0.25	0.19	0.32	-2.67
		Max.	0.94	0.53	1.31	1.70
		Mean	0.69	0.37	0.64	-0.97
		SD	0.16	0.09	0.24	1.03
Primary Reading (French)	36	Min.	0.23	0.20	0.31	-2.21
		Max.	0.92	0.53	1.62	1.57
		Mean	0.65	0.37	0.66	-0.75
		SD	0.15	0.09	0.27	0.81
Junior Reading (French)	36	Min.	0.36	0.08	0.21	-4.69
		Max.	0.88	0.50	0.80	1.61
		Mean	0.66	0.31	0.50	-1.06
		SD	0.14	0.11	0.16	1.08
Primary Writing (English)	14	Min.	0.46	0.29	0.51	-1.68
		Max.	0.78	0.56	0.97	0.11
		Mean	0.66	0.43	0.70	-0.88
		SD	0.10	0.09	0.16	0.58
Junior Writing (English)	14	Min.	0.55	0.17	0.25	-2.20
		Max.	0.87	0.57	1.03	-0.11
		Mean	0.68	0.37	0.65	-1.11
		SD	0.10	0.16	0.23	0.66
Primary Writing (French)	14	Min.	0.32	0.15	0.46	-2.41
		Max.	0.89	0.81	1.44	1.36
		Mean	0.65	0.39	0.78	-0.85
		SD	0.15	0.26	0.31	0.95
Junior Writing (French)	14	Min.	0.50	0.16	0.28	-2.45
		Max.	0.81	0.77	1.59	0.53
		Mean	0.69	0.37	0.67	-1.25
		SD	0.08	0.26	0.59	0.25

Note. * = The guessing parameter was set at a constant of 0.2 for multiple-choice items.

SD = standard deviation.

Table 7.3 Descriptive Statistics of CTT Item Statistics and IRT Item Parameter Estimates: Primary and Junior Divisions (continued)

Grade and Subject	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*	
			Item Difficulty	Item-total Correlation	Slope	Location
Primary Mathematics (English)	36	Min.	0.39	0.26	0.30	-2.33
		Max.	0.86	0.60	1.17	0.91
		Mean	0.64	0.39	0.70	-0.68
		SD	0.13	0.09	0.23	0.98
Junior Mathematics (English)	36	Min.	0.38	0.23	0.36	-1.57
		Max.	0.87	0.63	1.30	1.03
		Mean	0.60	0.38	0.72	-0.39
		SD	0.12	0.11	0.24	0.79
Primary Mathematics (French)	36	Min.	0.42	0.19	0.22	-3.60
		Max.	0.91	0.57	1.27	1.36
		Mean	0.68	0.37	0.67	-0.89
		SD	0.12	0.09	0.27	1.01
Junior Mathematics (French)	36	Min.	0.34	0.22	0.28	-2.60
		Max.	0.84	0.61	1.53	1.26
		Mean	0.61	0.37	0.73	-0.40
		SD	0.16	0.10	0.29	1.14

Note. * = The guessing parameter was set at a constant of 0.2 for multiple-choice items.

SD = standard deviation.

Grade 9 Assessment of Mathematics

Classical Test Theory (CTT) Analyses

Table 7.4 presents descriptive statistics, Cronbach's alpha estimates of test score reliability and the standard error of measurement (SEM) for the English-language and French-language applied and academic versions of the Grade 9 mathematics assessment. The mean percentages ranged from 55.0% (English, winter) to 57.8% (French, spring) for applied mathematics. The mean percentages ranged from 55.7% (French, spring) to 67.2% (English, spring) for academic mathematics.

Cronbach's alpha estimates ranged from 0.81 to 0.84 for applied mathematics and from 0.83 to 0.85 for academic mathematics. The corresponding standard errors of measurement were likewise similar, ranging from 7.0% to 7.4% of the possible maximum score for applied mathematics and 7.2% to 7.6% of the possible maximum score for academic mathematics. The obtained reliabilities and standard errors of measurement were acceptable. This indicates that the test scores from these assessments provide a satisfactory level of precision.

Table 7.4 Test Descriptive Statistics, Reliability and Standard Error of Measurement: Grade 9 Mathematics

Program	No. of Items	Item Type		Possible Max. Score	No. of Students	Min.	Max.	Mean	SD	Cronbach's Coefficient Alpha	SEM
		MC	OR								
Applied, Winter Administration (English)	31	24	7	52	20 930	0	52	28.59	9.29	0.83	3.87
Applied, Spring Administration (English)	31	24	7	52	24 667	0	52	29.97	9.37	0.84	3.74
Academic, Winter Administration (English)	31	24	7	52	44 377	0	52	32.32	9.68	0.85	3.80
Academic, Spring Administration (English)	31	24	7	52	55 644	1	52	34.29	9.40	0.85	3.65
Applied, Winter Administration (French)	31	24	7	52	638	7	51	29.21	8.62	0.82	3.65
Applied, Spring Administration (French)	30	23	7	51	868	6	49	30.05	8.82	0.81	3.83
Academic, Winter Administration (French)	31	24	7	52	1 518	4	52	29.44	9.77	0.83	3.98
Academic, Spring Administration (French)	31	24	7	52	2 439	7	52	28.94	9.48	0.84	3.85

Note. MC = multiple-choice; OR = open-response; SD = standard deviation; SEM = standard error of measurement.

Item Response Theory (IRT) Analyses

In the IRT analysis, item parameters were estimated from the student responses used in the equating calibration sample (see chapter 5). The estimated item parameters were then used to score all the student responses. The descriptive statistics reported in Table 7.5 are for the total population of students. The mean student ability scores range from –0.17 to 0 for the applied version of the mathematics assessment and from –0.09 to 0.04 for the academic version of the assessment. The means that differ from zero and the standard deviations that differ from one are due to the inclusion of all students. The item parameters for individual items are presented in appendix 7.1.

Table 7.5 Descriptive Statistics of IRT Scores: Grade 9 Mathematics

Program	No. of Students	Min.	Max.	Mean	SD
Applied, Winter Administration (English)	20 930	-3.39	2.81	-0.12	0.99
Applied, Spring Administration (English)	24 667	-3.47	2.78	-0.04	0.98
Academic, Winter Administration (English)	44 377	-3.39	2.38	-0.09	0.93
Academic, Spring Administration (English)	55 644	-3.47	2.38	0.04	0.96
Applied, Winter Administration (French)	638	-3.25	2.47	-0.17	1.00
Applied, Spring Administration (French)	868	-2.94	2.31	0.00	0.93
Academic, Winter Administration (French)	1 518	-2.88	2.69	-0.02	0.95
Academic, Spring Administration (French)	2 439	-2.83	2.71	-0.02	0.92

Note. SD = standard deviation.

The TCCs and the distribution of student thetas are displayed in Figures 7.25 and 7.28. The TCCs follow the expected S-shaped distribution. The right vertical scale indicates the percentage of students at each theta value, and the theta cut points for assigning students to performance levels are marked on all graphs. The TCCs for all of the winter and spring assessments either overlap or are very close to each other. This indicates that the winter and spring applied and academic assessments possessed the same difficulty level. The TIFs, which are displayed in Figures 7.29 to 7.32, indicate that the academic version of the assessment provided more information than the applied version. Further, the applied version of the assessment provided most of its information at the 2/3 cut point, while the academic version yielded most of its information between the 2/3 and 3/4 cut points.

Figure 7.25 Test Characteristic Curves and Distribution of Thetas for Grade 9 Applied Mathematics (English)

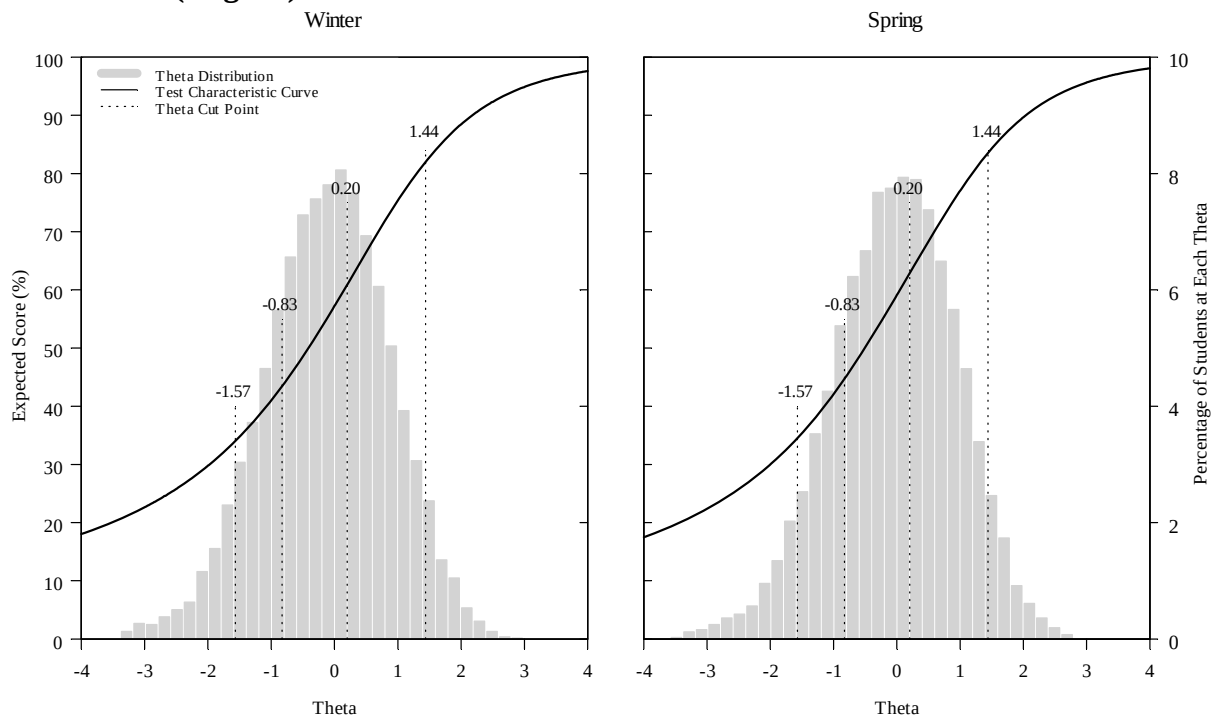


Figure 7.26 Test Characteristic Curves and Distribution of Thetas for Grade 9 Academic Mathematics (English)

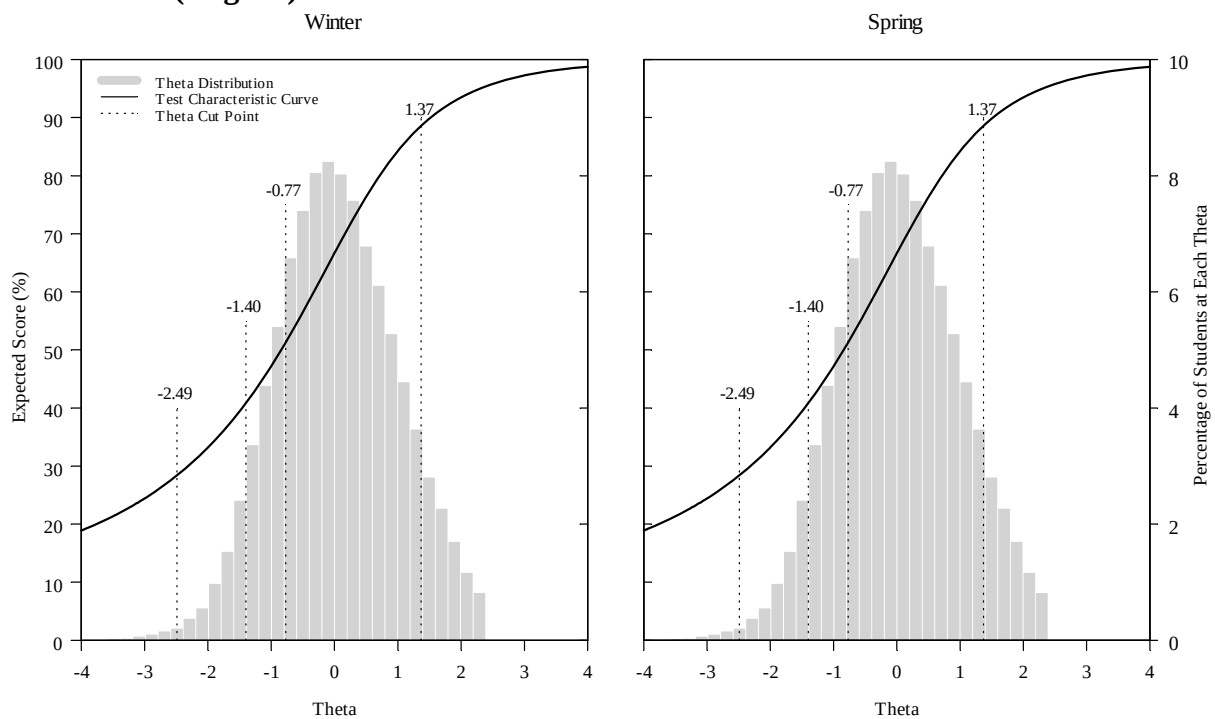


Figure 7.27 Test Characteristic Curves and Distribution of Thetas for Grade 9 Applied Mathematics (French)

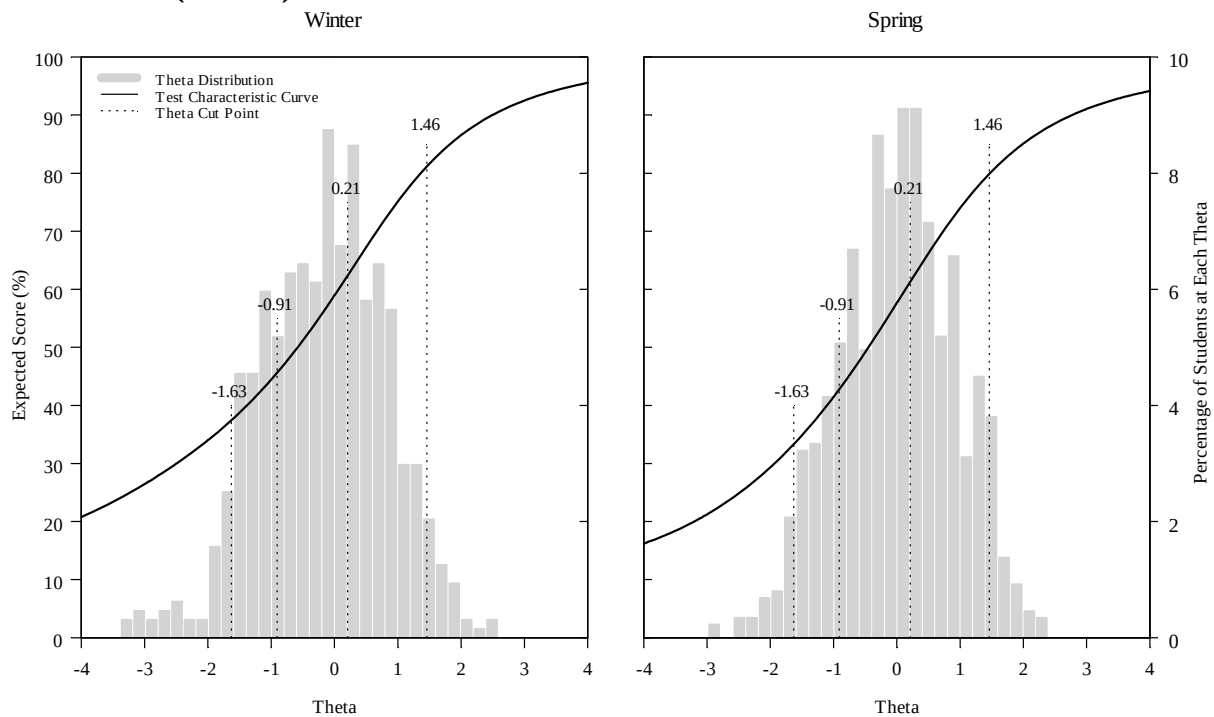


Figure 7.28 Test Characteristic Curves and Distribution of Thetas for Grade 9 Academic Mathematics (French)

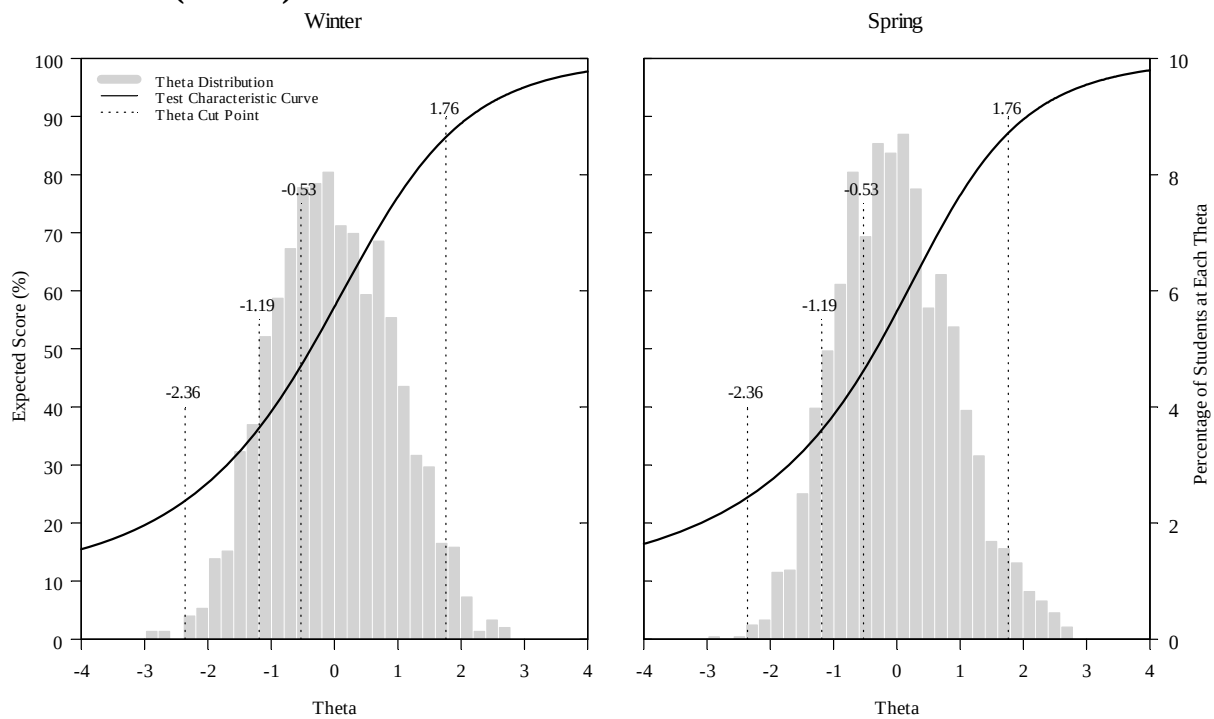


Figure 7.29 Test Information Function for Grade 9 Applied Mathematics (English)

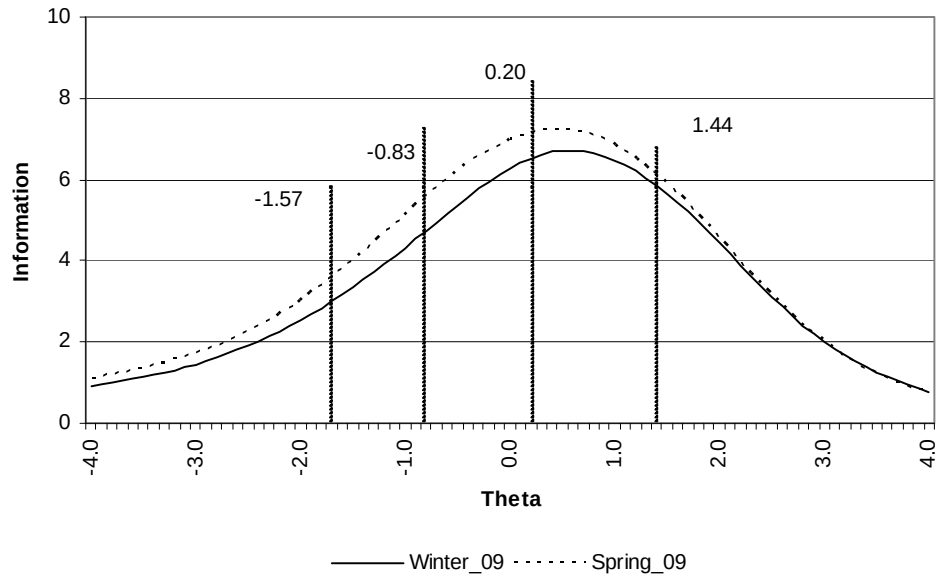


Figure 7.30 Test Information Function for Grade 9 Academic Mathematics (English)

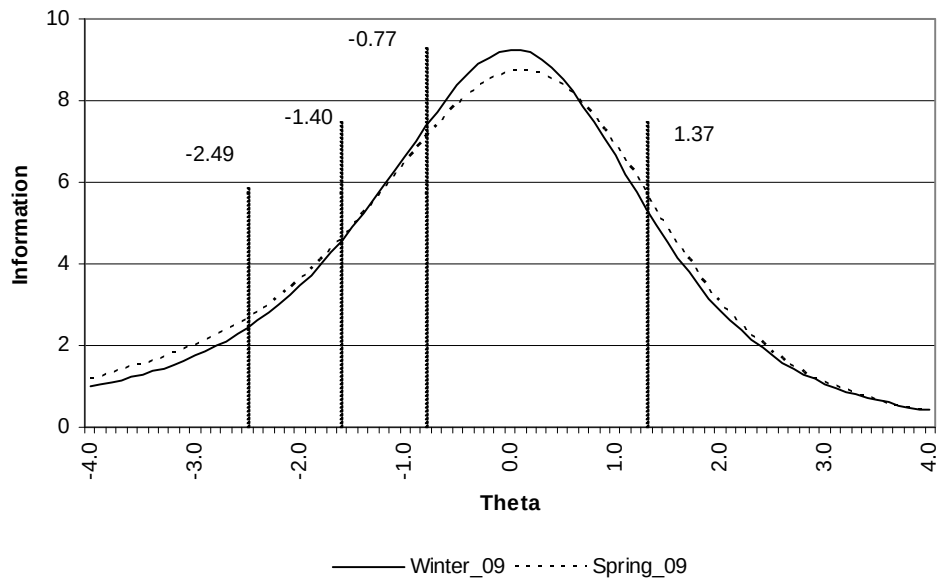


Figure 7.31 Test Information Function for Grade 9 Applied Mathematics (French)

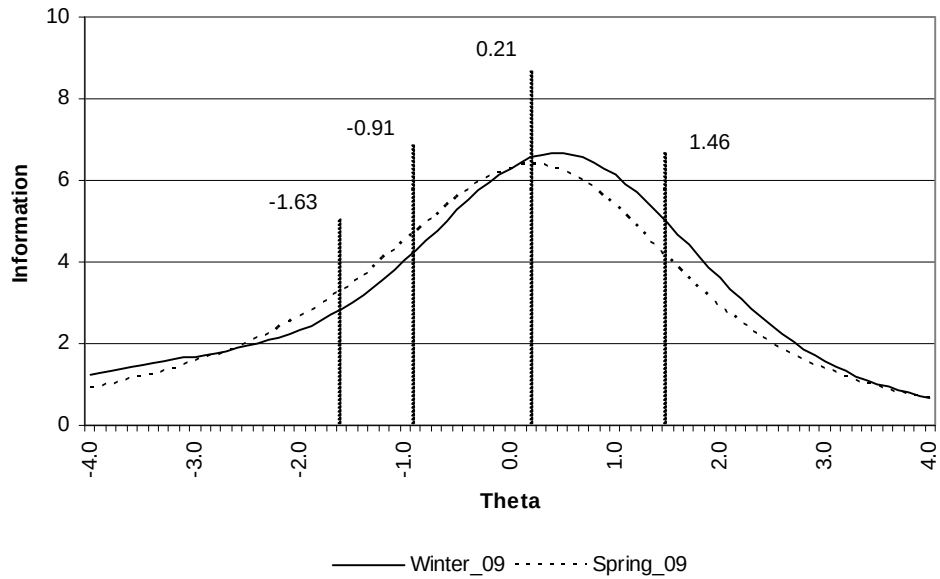
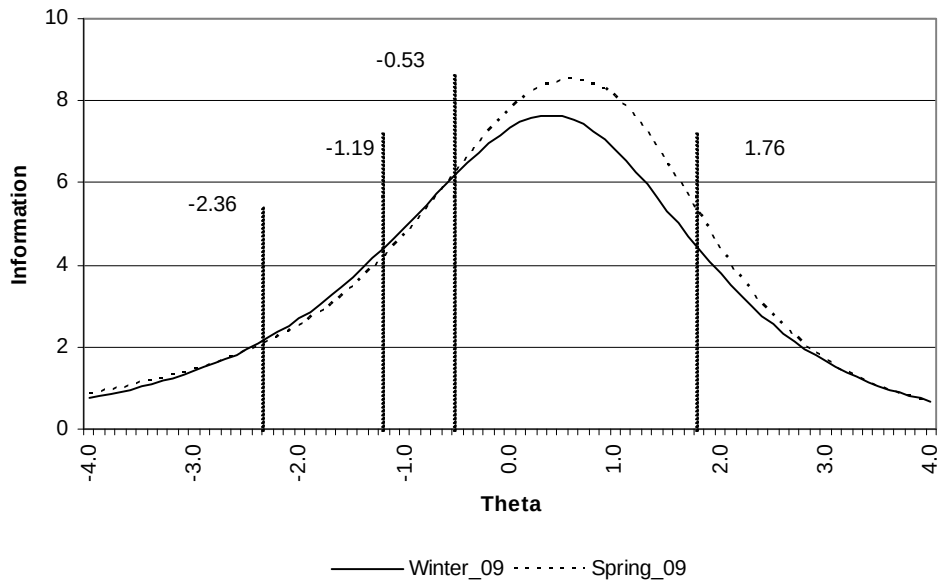


Figure 7.32 Test Information Function for Grade 9 Academic Mathematics (French)



Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)

Table 7.6 contains a summary of both the CTT and IRT item statistics for the items included in the Grade 9 mathematics assessment. The classical item statistics were computed using the provincial population. IRT item statistics were computed using the equating sample. As with the primary- and junior-division assessments, care must be taken to avoid matching the minimum CTT p -value with the minimum IRT location value, and the item-total correlation with the slope. As expected, there is an inverse relationship between the CTT item difficulty and the IRT location parameter, and there is a positive relationship between the CTT item-total correlation and the IRT slope parameter. The item difficulties and location parameters were in an acceptable range. Likewise, the point-biserial correlation coefficients were, for the most part, within an acceptable range, but values less than 0.20 are not recommended. Statistics for individual items are presented in appendix 7.1.

Table 7.6 Descriptive Statistics of CTT Item Statistics and IRT Item Parameter Estimates: Grade 9 Mathematics

Program	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters	
			Item Difficulty	Item-Total Correlation	Location	Slope
Applied, Winter Administration (English)	31	Min.	27.91	0.15	-2.22	0.24
		Max.	85.40	0.53	1.88	1.07
		Mean	56.96	0.32	-0.14	0.55
		SD	15.06	0.11	1.04	0.21
Applied, Spring Administration (English)	31	Min.	26.19	0.18	-1.63	0.27
		Max.	85.70	0.55	1.94	1.07
		Mean	58.83	0.33	-0.17	0.57
		SD	15.84	0.11	1.06	0.20
Academic, Winter Administration (English)	31	Min.	42.10	0.18	-2.78	0.33
		Max.	88.99	0.60	1.12	1.13
		Mean	62.78	0.37	-0.23	0.60
		SD	11.17	0.10	0.99	0.19
Academic, Spring Administration (English)	31	Min.	36.93	0.20	-4.06	0.17
		Max.	90.47	0.57	1.12	1.23
		Mean	66.42	0.37	-0.66	0.64
		SD	12.49	0.10	1.06	0.24
Applied, Winter Administration (French)	31	Min.	28.79	0.14	-1.63	0.11
		Max.	78.32	0.54	1.59	1.00
		Mean	57.74	0.32	-0.28	0.55
		SD	10.55	0.10	0.76	0.23
Applied, Spring Administration (French)	30	Min.	35.99	0.09	-1.30	0.19
		Max.	75.96	0.51	2.15	1.02
		Mean	61.05	0.32	-0.15	0.51
		SD	9.83	0.11	0.71	0.22
Academic, Winter Administration (French)	31	Min.	38.81	0.22	-1.83	0.19
		Max.	80.73	0.58	1.30	1.12
		Mean	58.26	0.35	-0.13	0.54
		SD	10.87	0.11	0.75	0.19
Academic, Spring Administration (French)	31	Min.	32.68	0.08	-3.20	0.13
		Max.	82.19	0.55	1.50	1.32
		Mean	56.37	0.34	-0.10	0.59
		SD	11.90	0.12	0.97	0.25

Note. IRT item parameters are estimated based on the equating sample; the guessing parameter was set at a constant of 0.2 for multiple-choice items.

SD = standard deviation.

Ontario Secondary School Literacy Test (OSSLT)

Classical Test Theory (CTT) Analyses

Table 7.7 presents descriptive statistics, Cronbach's alpha estimates of test-score reliability and the standard error of measurement (SEM) for the first-time eligible students of the English-language and French-language OSSLT. The mean percentages for first-time eligible students are 80.1% for English-language students and 78.3% for French-language students.

Cronbach's alpha estimates were 0.88 and 0.86, respectively. The corresponding standard errors of measurement were the same (4.0%). The obtained reliabilities and standard errors of measurement were acceptable and indicate that test scores from these assessments were at a satisfactory level of precision.

Table 7.7 Descriptive Test Statistics, Reliability and Standard Error of Measurement: OSSLT

Language	No. of Items	Item Type				Possible Max. Score	No. of Students	Min.	Max.	Mean	SD	R	SEM
		MC	OR	SW	LW								
English	47	39	4	2	2	81	142 394	3.0	81.0	64.9	9.45	0.88	3.21
French	47	39	4	2	2	81	5 480	19.0	81.0	63.4	8.66	0.86	3.21

Note. MC=multiple-choice; OR=open-response (reading); SW=short writing; LW=long writing; N=number of students; R=Cronbach's coefficient alpha; SD=standard deviation; SEM=standard error of measurement.

Item Response Theory (IRT) Analysis

In the IRT analysis, item parameters were estimated from the student responses used in the equating calibration sample (see chapter 5). The estimated item parameters were then used to score all student responses. The descriptive statistics reported in Table 7.8 are for the first-time eligible students in the provincial population. Due to the inclusion of all students for both the English- and French-language tests, the mean of student ability scores is not exactly zero and the standard deviation is less than one. The item parameters for individual items are presented in appendix 7.1.

Table 7.8 Descriptive Statistics for IRT Scores: OSSLT

Language	No. of Students	Min.	Max.	Mean	SD
English	142 394	-3.78	2.83	0.30	0.90
French	5 480	-3.63	2.70	-0.14	0.89

The TCCs and the distribution of student thetas are displayed in Figures 7.33 and 7.34 for the English-language and French-language students, respectively. The TCCs follow the expected S-shaped distribution. The distribution of student thetas is plotted on the TCC graphs, with the right vertical scale indicating the percentage of students at each theta value. The TIF plots for the English-language and French-language test administrations are shown in Figures 7.35 and 7.36. The theta cut point for assigning students to successful and unsuccessful levels of performance is marked on each TCC and TIF.

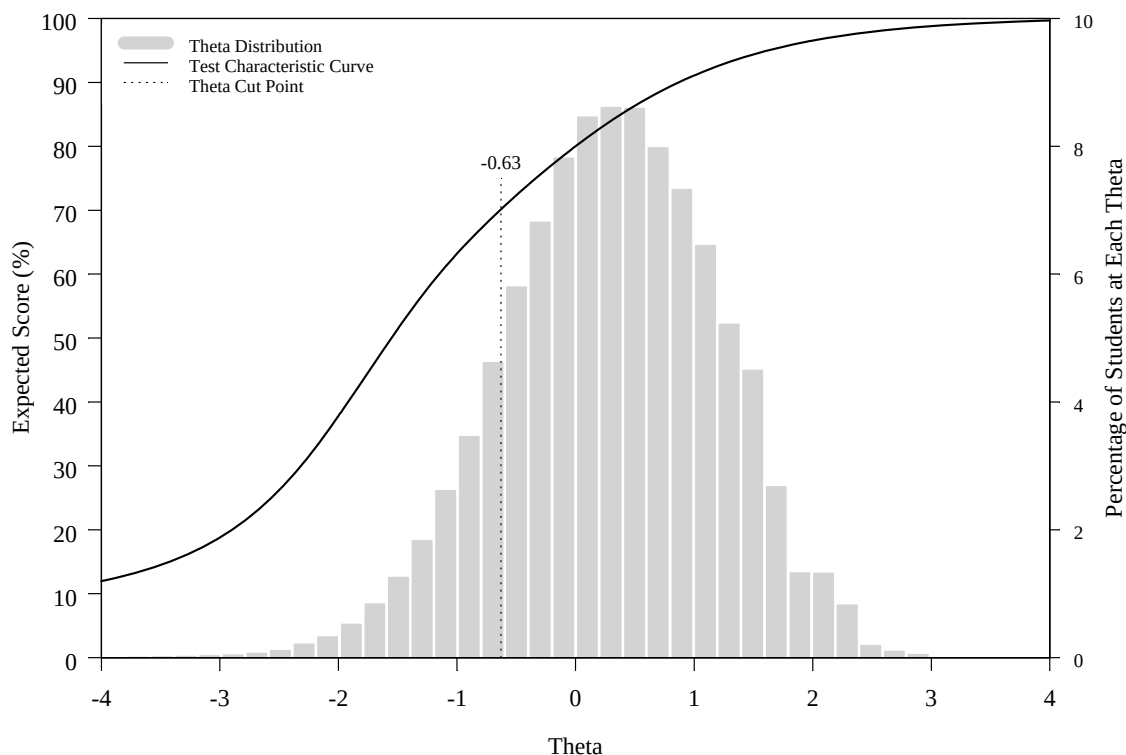
Figure 7.33 Test Characteristic Curve and Distribution of Theta: OSSLT (English)

Figure 7.34 Test Characteristic Curve and Distribution of Theta: OSSLT (French)

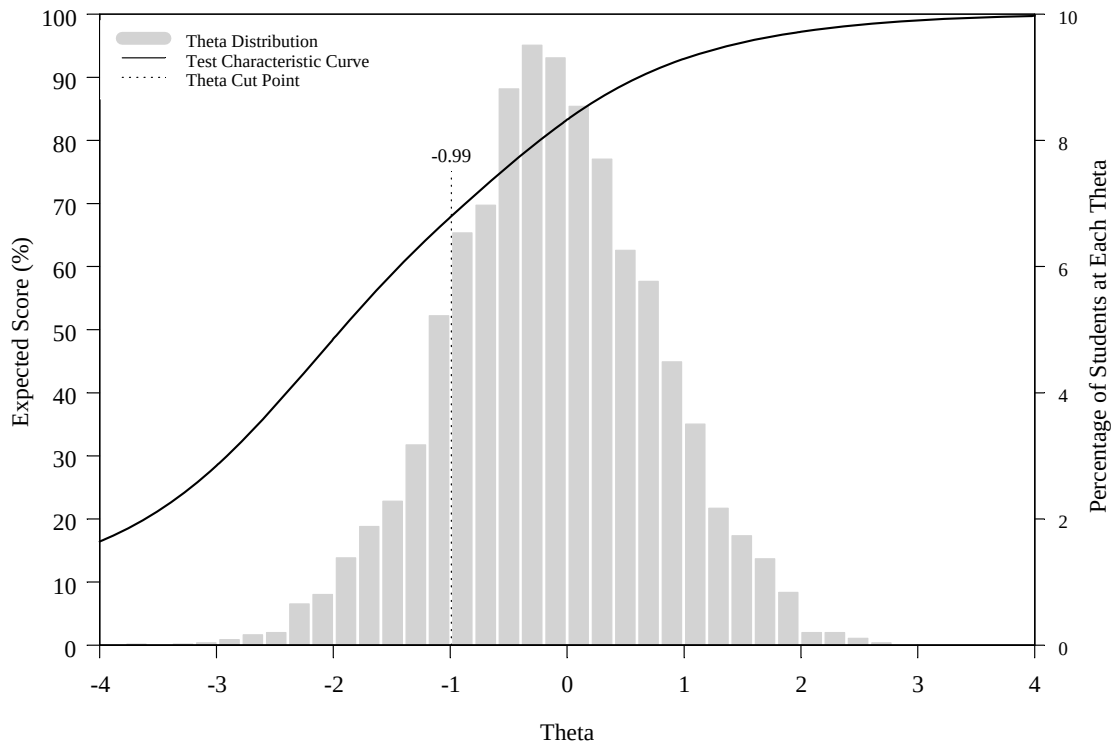


Figure 7.35 Test Information Function: OSSLT (English)

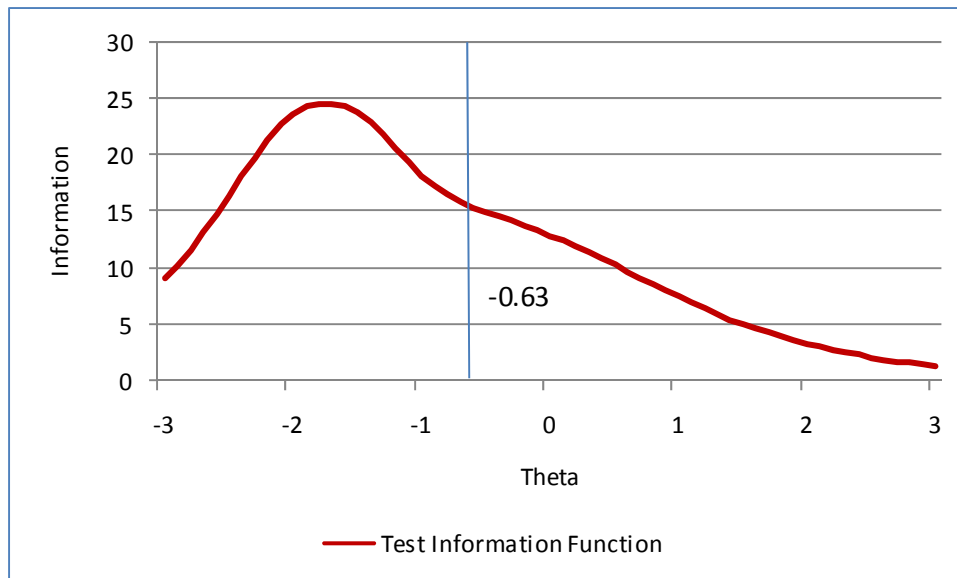
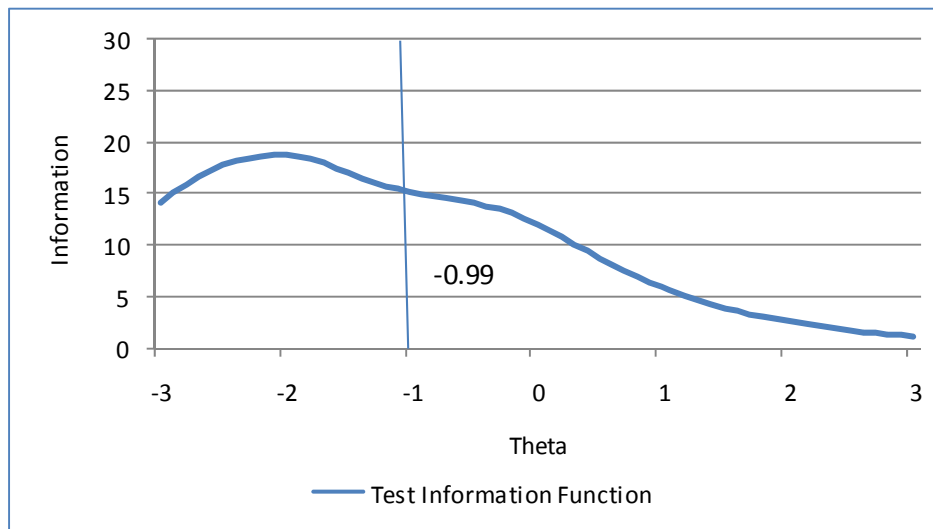


Figure 7.36 Test Information Function: OSSLT (French)



Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)

Table 7.9 contains a summary of both the CTT and IRT item statistics for the multiple-choice items on the OSSLT. As with the primary- and junior-division assessments, care must be taken to avoid matching the minimum CTT p -value with the minimum IRT location value, and the item-total correlation with the slope for the OSSLT. As expected, there is an inverse relationship between the CTT item-difficulty estimates and the IRT location parameter estimates due to the difference between the definitions of difficulty in the two approaches. Unlike the primary, junior and Grade 9 assessments, the a -parameter was set at a constant value for all items. Hence, it is not possible to determine the nature of the relationship between the CTT item-total correlation and the IRT slope parameter. However, the low values for the minimum point-biserial correlations suggest that some of the items did not reach the expected level (0.20), but the item difficulty values were within an acceptable range. Presented in appendix 7.1 are the statistics for individual items, the distribution of score points and threshold parameters for the open-response items and the analysis results for differential item functioning for all items.

Table 7.9 Descriptive Statistics of CTT Item Statistics and IRT Item Parameter Estimates: OSSLT

Language	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*
			Item Difficulty	Item-Total Correlation	
English	47	Min.	0.51	0.15	-3.16
		Max.	0.99	0.57	0.77
		Mean	0.79	0.34	-1.10
		SD	0.10	0.09	0.86
French	47	Min.	0.38	0.11	-3.91
		Max.	0.99	0.58	1.18
		Mean	0.79	0.30	-1.69
		SD	0.14	0.11	1.12

Note. * The slope was set at 0.588 for all items, and the guessing parameter was set at a constant of 0.20 for all multiple-choice items.

SD = standard deviation.

Differential Item Functioning (DIF)

One goal of test development is to assemble a set of items that provides an estimate of a student's ability that is as fair and accurate as possible for all groups within the student population. Differential Item Functioning (DIF) statistics are used to identify items on which students with the same level of ability, but from different identifiable groups, have different probabilities of answering correctly (e.g., females and males, English language learners [ELL] or students in the ALF/PDF programs, students with special education needs). If an item is more difficult for one subgroup than for another, the item may be measuring something different from the intended construct. However, it is important to recognize that DIF-flagged items might be related to actual differences in relevant knowledge or skill (item impact) or statistical Type I error. Therefore, items identified through DIF statistics must be reviewed by content experts and bias and sensitivity committees to determine possible sources and interpretations of the differences.

EQAO examined the 2009 assessments for gender- and ELL-based DIF, using SIBTEST for multiple-choice items and Poly-SIBTEST for open-response items. In all analyses, males and non-English language learners (non-ELLs) were the reference group, and females and ELLs were the focal, or studied, group. For the French-language version of the assessments, the corresponding groups were students in or not in the ALF/PDF programs. The valid subtest for each SIBTEST and Poly-SIBTEST run consisted of all the items that were not part of the suspect item. To estimate the expected score (ES) for a suspect item that is conditional on the primary dimension for the reference group (R) and the focal group (F), the SIBTEST and Poly-SIBTEST tests used the total test score for the matching subtest of items as defined below:

$$ES_R(t) = \sum_{k=1}^m kP_{Rk}(t) \text{ and } ES_F(t) = \sum_{k=1}^m kP_{Fk}(t),$$

where $P_{Gk}(t)$ is the empirical proportion of examinees in group G ($G = R$ or F) that has obtained score k on the studied item and that has valid subtest score t .

SIBTEST and Poly-SIBTEST use a regression-correction formula to correct for bias due to measurement error in the matching subtest that would indicate DIF favouring the group with the higher ability, when, in fact, no DIF exists (Shealy & Stout, 1993). The regression-correction formula adjusts the mean item scores to their unbiased estimates by computing a regression equation for each group that estimates a true score on the matching subtest as a function of the matching-subtest-observed score. This function is computed for each group using the estimated means, variances and reliabilities of the scores on the matching subtest. The resulting adjusted expected item scores, $ES_R^*(t)$ and $ES_F^*(t)$, are used for comparing group performance on the item to estimate the DIF index, as follows:

$$\hat{\beta} = \sum_{t=1}^T [ES_R^*(t) - ES_F^*(t)] \frac{N_R(t) - N_F(t)}{N},$$

where

T is the maximum score on the matching subtest,

$N_R(t)$ is the number of examinees in the reference group (R) with matching subtest score t ,

$N_F(t)$ is the number of examinees in the focal group (F) with matching subtest score t ,

N is the total number of examinees and

$\hat{\beta}$ is interpreted as the difference in the expected total score between the focal and reference groups.

Two statistics were produced by SIBTEST and Poly-SIBTEST: the $\hat{\beta}$, which is an indicator of effect size describing the size of the difference, and a p -value, which estimates the probability that an observed difference occurred by chance. For the 2009 EQAO assessments, the guidelines proposed by Roussos and Stout (1996) for SIBTEST were used to classify the multiple-choice items for which $\hat{\beta}$ was significantly different from zero: items with $|\hat{\beta}|$ less than 0.059 were considered to have negligible or A-level DIF; items with $|\hat{\beta}|$ equal to or greater than 0.059, but less than 0.088, were considered to have moderate or B-level DIF; items with $|\hat{\beta}|$ equal to or greater than 0.088 were considered to have large or C-Level DIF. In the case of open-response items, the same guidelines as SIBTEST were applied to poly-SIBTEST results to classify the items.

For each assessment, except for the French-language version of the Grade 9 assessment, two random samples stratified according to gender or designation as a second-language learner (ELL or ALF/PDF) were selected from the population of students who took the test under standard

administration conditions (i.e., no accommodations)². The use of two samples provided an estimate of the stability of the results in a cross-validation process. Items that were classified as having B- or C-level DIF in both samples were considered as DIF items.

The item-level results are provided in appendix 7.1. The results in each table are from the two random samples, which include the value of $\hat{\beta}$, the significance level, and the level of DIF severity. Negative estimates of $\hat{\beta}$ indicate that the girls outperformed the boys; positive $\hat{\beta}$ estimates indicate that the boys outperformed the girls. Items identified with DIF were determined to be statistically significant and had at least a moderate effect size in both samples.

Primary- and Junior-Division Assessments

For the 2009 primary- and junior-division reading, writing and mathematics components of the assessments for both languages, the summaries of the number of items that showed statistically significant gender-based DIF with at least a B-level or C-level effect size in both samples are reported in Tables 7.10 and 7.11. The numbers in the boys (or girls) column indicate the number of DIF items favouring boys (or girls).

Table 7.10 Number of B-Level Gender-Based DIF Items: Primary and Junior Assessments

Subject Area	Primary English		Junior English		Primary French		Junior French	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Reading <i>k</i> = 36	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 1 (OR)	0 (MC) 1 (OR)	0 (MC) 1 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)
Writing <i>k</i> = 14	1 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 1 (OR)	0 (MC) 0 (OR)	1 (MC) 1 (OR)	0 (MC) 1 (OR)	1 (MC) 3 (OR)	0 (MC) 1 (OR)
Math <i>k</i> = 36	2 (MC) 1 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	2 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)

Note. MC= multiple-choice; OR=open-response.

Table 7.11 Number of C-Level Gender-Based DIF Items: Primary and Junior Assessments

Subject Area	Primary English		Junior English		Primary French		Junior French	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Reading <i>k</i> = 36	1 (MC) 0 (OR)	0 (MC) 1 (OR)	2 (MC) 2 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 1 (OR)	0 (MC) 1 (OR)
Writing <i>k</i> = 14	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Math <i>k</i> = 36	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	2 (MC) 0 (OR)	0 (MC) 1 (OR)

Note. MC= multiple-choice; OR=open-response.

² During test administration, some students are accommodated for various reasons and do not write the test under standard conditions. These students were not included in the analyses.

Of the 344 items on the primary- and junior-division assessments, 32 items were found to have gender-based DIF. The majority of the DIF items had B-level DIF (21) and 11 items had C-level DIF. Of the 15 multiple-choice DIF items, 14 favoured the males, while one favoured the females. For open-response items, the number of DIF items favouring the males was almost the same as the number favouring the females: nine and seven items, respectively. Overall, more items were found to favour boys than girls. The reading component of the English-language junior-division assessment had the largest number of gender-based DIF items (six).

The summaries for second-language-learner (SLL)-based DIF are reported in Tables 7.12 and 7.13. The term second-language learner is used to represent ELLs for the English-language assessments and students in the ALF/PDF programs for the French-language assessments. Twenty out of the 344 items across all 12 assessments showed SLL-based DIF, with nine items having B-level DIF and 11 items having C-level DIF. Of the eight multiple-choice DIF items, five favoured non-SLL students and three favoured SLL students. For the 12 open-response DIF items, seven favoured non-SLL students and five favoured SLL students. More items were found to favour non-SLL students than SLL students. Again, the reading component of the English-language junior-division assessment was found to have the largest number of SLL-based DIF items.

Table 7.12 Number of B-Level SLL-Based DIF Items: Primary and Junior Assessments

Subject Area	Primary English		Junior English		Primary French		Junior French	
	Non-SLLs	SLLs	Non-SLLs	SLLs	Non-SLLs	SLLs	Non-SLLs	SLLs
Reading <i>k</i> = 36	1 (MC) 1 (OR)	0 (MC) 1 (OR)	0 (MC) 0 (OR)	0 (MC) 1 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Writing <i>k</i> = 14	1 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Math <i>k</i> = 36	1 (MC) 0 (OR)	0 (MC) 1 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)

Note. SLL= Second-language learner (English language learner or students in ALF/PDF programs); MC= multiple-choice; OR=open-response.

Table 7.13 Number of C-Level SLL-Based DIF Items: Primary and Junior Assessments

Subject Area	Primary English		Junior English		Primary French		Junior French	
	Non-SLLs	SLLs	Non-SLLs	SLLs	Non-SLLs	SLLs	Non-SLLs	SLLs
Reading <i>k</i> = 36	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 1 (OR)	0 (MC) 1 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 1 (OR)	1 (MC) 0 (OR)
Writing <i>k</i> = 14	0 (MC) 1 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)
Math <i>k</i> = 36	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 2 (OR)	0 (MC) 0 (OR)	0 (MC) 1 (OR)	0 (MC) 1 (OR)

Note. SLL= Second-language learner (English language learner or students in ALF/PDF programs); MC= multiple-choice; OR=open-response.

All items identified as having B-level or C-level DIF on the primary- and junior-division assessments were reviewed by the assessment team. Since the reviewers did not identify any apparent biases in the content of these items, none of the items were removed from the scoring.

Grade 9 Mathematics

The gender- and ELL-based (English only) DIF results for the academic and applied versions of the Grade 9 assessment of mathematics are provided in Tables 7.14 and 7.17. For gender-based DIF, it was not possible to have two random samples for the French-language academic and applied versions of the Grade 9 assessment of mathematics due to the small number of participating students. DIF analysis was not conducted for ALF/PDF students for the French-language version of the assessment due to the small number of students in this group.

Table 7.14 Number of B-Level Gender-Based DIF Items: Grade 9 Applied and Academic Mathematics

Assessment	English		French	
	Boys	Girls	Boys	Girls
Applied, Winter Administration k=31	0 (MC)	1 (MC)	0 (MC)	1 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Applied, Spring Administration k=31/30*	1(MC)	1 (MC)	2 (MC)	1 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Winter Administration k=31	0 (MC)	1 (MC)	1 (MC)	3 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Spring Administration k=30/31*	0 (MC)	1 (MC)	2 (MC)	2 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)

Note. MC= multiple-choice; OR=open-response; * = One item was dropped from the assessment.

Table 7.15 Number of C-Level Gender-Based DIF Items: Grade 9 Applied and Academic Mathematics

Assessment	English		French	
	Boys	Girls	Boys	Girls
Applied, Winter Administration k=31	0 (MC)	0 (MC)	3 (MC)	4 (MC)
	3 (OR)	2 (OR)	1 (OR)	1 (OR)
Applied, Spring Administration k=31/30*	0 (MC)	0 (MC)	1 (MC)	3 (MC)
	3 (OR)	3 (OR)	2 (OR)	3 (OR)
Academic, Winter Administration k=31	0 (MC)	0 (MC)	1 (MC)	0 (MC)
	2 (OR)	1 (OR)	2 (OR)	2 (OR)
Academic, Spring Administration k=31	0 (MC)	0 (MC)	1 (MC)	0 (MC)
	1 (OR)	2 (OR)	2 (OR)	3 (OR)

Note. MC= multiple-choice; OR=open-response; * = One item was dropped from the assessment.

Of the 247 items on the Grade 9 assessments, 17 multiple-choice items showed B-level gender-based DIF. Of the 17 B-level gender-based DIF items, six favoured boys and 11 favoured girls. Forty-six items showed C-level gender-based DIF, of which 13 were multiple-choice items (all French-language items) and 33 were open-response items. Of the 13 C-level multiple-choice items, six favoured boys and seven favoured girls. Of the 33 C-level open-response items, 16 favoured boys and 17 favoured girls.

Twelve to 16 items were used in both the winter and spring administrations across courses. None of the common items showed gender-based DIF in both administrations.

Table 7.16 Number of B-Level ELL-Based DIF Items: Grade 9 Applied and Academic Mathematics

Assessment	Applied		Academic	
	Non-ELLs	ELLs	Non-ELLs	ELLs
Winter Administration k=31	3 (MC)	0 (MC)	2 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Spring Administration k=31	0 (MC)	0 (MC)	0 (MC)	1 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)

Note. MC= multiple-choice; OR=open-response.

Table 7.17 Number of C-Level ELL-Based DIF Items: Grade 9 Applied and Academic Mathematics

Assessment	Applied		Academic	
	Non-ELLs	ELLs	Non-ELLs	ELLs
Winter Administration k=31	2 (MC)	2 (MC)	3 (MC)	4 (MC)
	3 (OR)	2 (OR)	3 (OR)	3 (OR)
Spring Administration k=31	3 (MC)	3 (MC)	3 (MC)	3 (MC)
	2 (OR)	3 (OR)	2 (OR)	2 (OR)

Note. MC= multiple-choice; OR=open-response.

Again, the six B-level ELL-based DIF items are all multiple-choice items, with five favouring non-ELL students and one favouring ELL students. Forty-three items (23 multiple-choice items and 20 open-response items) showed ELL-based DIF. Of the 23 multiple-choice items, 11 favoured non-ELL students and 12 favoured ELL students. Of the 20 open-response items, 10 favoured non-ELL students and 10 favoured ELL students. For the English-language applied version of the assessment, two common items showed ELL-based DIF in both the winter and spring administrations.

All items identified as B-level or C-level DIF items on the Grade 9 assessment were reviewed by the assessment team. Since the reviewers did not identify any apparent biases in the content of these items, none of the items were removed from the scoring process.

OSSLT

DIF results for the OSSLT are presented in Tables 7.18, 7.19, and 7.20. DIF analysis was not conducted for ALF/PDF students for the French-language version of test due to the small number of students in this group.

Table 7.18 Number of B-Level Gender-Based DIF Items: OSSLT

English (k=51)		French (k=51)	
Males	Females	Males	Females
0	5 (MC)	2 (MC) 2 (OR)	2 (MC) 2 (OR)

Note. MC= multiple-choice; OR=open-response.

Table 7.19 Number of C-Level Gender-Based DIF Items: OSSLT

English (k=51)		French (k=51)	
Males	Females	Males	Females
2 (MC)	0	1 (OR)	0

Note. MC= multiple-choice; OR=open-response.

There were five B-level DIF items and two C-level DIF items on the English-language version of the OSSLT. Five of the DIF items favoured the females, and two of the DIF items favoured the males.

There were eight B-level DIF items and one C-level DIF item on the French-language version of the OSSLT. Five of these DIF items favoured the males and four favoured the females. One open-response reading item exhibited B-level DIF favouring the females, and one open-response reading item exhibited C-level DIF favouring the males. Two short-writing prompts exhibited B-level DIF favouring the males in topic development. One long-writing prompt exhibited B-level DIF favouring the females in the use of conventions.

Table 7.20 Number of ELL-Based DIF Items: OSSLT (English)

English (k=51)			
B-Level DIF		C-Level DIF	
Non-ELLs	ELLs	Non-ELLs	ELLs
3 (OR)	3 (MC) 1 (OR)	3 (MC) 1 (OR)	2 (MC) 2 (OR)

Note. MC= multiple-choice; OR=open-response.

One short-writing prompt and one long-writing prompt exhibited B-level DIF favouring non-ELLs for topic development. Also, one short-writing item prompt exhibited B-level DIF favouring non-ELLs for the use of conventions. Three multiple-choice items and one open-response reading item exhibited B-level DIF favouring ELLs.

Three multiple-choice items and one long-writing item prompt exhibited C-level DIF favouring non-ELLs for the use of conventions. Two multiple-choice items, one open-response reading item and one long-writing item exhibited C-level DIF favouring non-ELLs for topic development.

All OSSLT items that were identified as exhibiting B-level or C-level DIF were reviewed by the assessment team. Since the reviewers did not identify any apparent biases in the content of these items, none of the items were removed from the scoring.

Decision Accuracy and Consistency

The four achievement levels defined in *The Ontario Curriculum* are used by EQAO to report student achievement in reading, writing and mathematics for the primary- and junior-division assessments and in the academic and applied versions of the Grade 9 mathematics assessment. Level 3 has been established as the provincial standard. Levels 1 and 2 indicate achievement below the provincial standard, and Level 4 indicates achievement above the provincial standard. In addition to these four performance levels, students with not enough evidence for Level 1 are placed in a level below Level 1³. Through the equating process described in chapter 5, four theta values are identified that define the cut points between adjacent levels (NE1/Level 1, Level 1/Level 2, Level 2/Level 3 and Level 3/Level 4). In the case of the OSSLT, EQAO reports only two levels of achievement: successful and unsuccessful. Thus, there is one cut score.

Two issues that arise when students are placed into categories, such as those used by EQAO, are the *accuracy* and *consistency* of the student classifications based on students' assessment scores. The term *accuracy* refers to the extent to which classifications based on observed student scores agree with classifications based on the true scores of these students. The term *consistency* refers to the extent to which classifications based on observed students' scores on one form of the assessment agree with the classifications of the observed scores of the same students on an interchangeable assessment form.

Accuracy: While observed scores include measurement error, true scores do not. Thus, classification decisions based on true scores are true classifications, or, put in another way, the right classifications. In contrast, classification decisions based on observed scores or scores derived from observed scores are not errorless. Since the errors may be positive, zero or negative, the observed score may be too low, just right or too high. This is illustrated in Table 7.21 for classifications in two adjacent categories (zero and one).

³ Students with no data and exempt students are not included in the calculation of results for students participating in the assessments.

Table 7.21 Demonstration of Classification Accuracy

		Classification Based on True Scores		Row Margins
		0	1	
Classification Based On Observed Scores	0	p_{00}	p_{01}	$p_{1.}$
	1	p_{10}	p_{11}	$p_{2.}$
Column Margins		$p_{.1}$	$p_{.2}$	1.00

The misclassifications, p_{01} and p_{10} , are attributable to the presence of measurement error. The sum of p_{00} and p_{11} equals the rate of classification accuracy, which should be high or close to one.

Consistency: In contrast to accuracy, both sets of observed scores on the two interchangeable tests are not errorless. For some students, their scores on one test will be higher than their scores on the second test. For other students, their scores will be equal, and still for other students, their scores on the first test will be lower than their scores on the second test. The differences, when they occur, may be so large that they may lead to different or inconsistent classifications. The classification on the observed score could be lower than, the same as or higher than the classification based on the second score. This is illustrated in Table 7.22 for classifications in two adjacent categories (zero and one).

Table 7.22 Demonstration of Classification Consistency

		Classification Based On Observed Scores 2		Row Margins
		0	1	
Classification Based On Observed Scores 1	0	p_{00}	p_{01}	$p_{1.}$
	1	p_{10}	p_{11}	$p_{2.}$
Column Margins		$p_{.1}$	$p_{.2}$	1.00

The different classifications, p_{01} and p_{10} , are attributable to the presence of measurement error. The sum of p_{00} and p_{11} equals the rate of classification consistency, which should be high or close to 1.00.

Estimation from One Test Form

There are several procedures for estimating decision accuracy and decision consistency. The procedure developed by Livingston and Lewis (1995) was used by EQAO because it yields estimates of both accuracy and consistency and because the procedure allows both multiple-

choice and open-response items. Further, this procedure is commonly used in large-scale assessment programs.

The Livingston-Lewis procedure uses the classical true score, τ , to determine classification accuracy. The true score, corresponding to an observed score X , is expressed as a proportion on a scale of 0 to 1:

$$\tau_p = \frac{\xi_f(X) - X_{\min}}{X_{\max} - X_{\min}},$$

where

τ_p is the proportional true score,

$\xi_f(X)$ is the expected value of a student's observed scores across $f \rightarrow \infty$ interchangeable forms and

$X_{\min}$ and $X_{\max}$ are, respectively, the minimum and maximum observed scores (Livingston & Lewis, 1995, p. 182).

Decision consistency is estimated using the joint distribution of reported performance-level classifications on the current form of the test and performance-level classifications on the alternate (parallel) test form. In each case, the proportion of performance-level classifications with exact agreement is the sum of the entries shown in the diagonal of the contingency table representing the joint distribution.

The Livingston-Lewis procedure requires the creation of an effective test length to model the complex data. The effective test length is determined by the “number of discrete, dichotomously scored, locally independent, equally difficult test items necessary to produce total scores having the same precision as the scores being used to classify the test takers” (Livingston & Lewis, 1995, p. 180). The formula for determining the effective test length is

$$\tilde{n} = \frac{(\hat{\mu}_X - X_{\min})(X_{\max} - \hat{\mu}_X) - r_{XX} \hat{\sigma}_X^2}{\hat{\sigma}_X^2 (1 - r_{XX})},$$

where

$\tilde{n}$ is the effective test length rounded to the nearest integer,

$\hat{\mu}_X$ is the mean of the observed scores,

$\hat{\sigma}_X^2$ is the unbiased estimator of the variance of the observed scores and

r_{XX} is the reliability of the observed scores.

The third step of the method requires that the observed scores in the original scale score for test X be transformed into a new scale score X' :

$$X' = \tilde{n} \frac{X - X_{\min}}{X_{\max} - X_{\min}}.$$

The distribution of true scores is estimated by fitting a four-parameter beta distribution, the parameters of which are estimated from the observed distribution of X' . In addition, the distribution of conditional errors is estimated by fitting a binomial model with regard to X' and $\hat{n}$. Both classification accuracy and classification consistency can then be determined by using these two distributions. The results for each are then adjusted so that the predicted marginal category proportions match those for the observed test. The computer program BB-CLASS (Brennan, 2004) was used to determine these estimates.

Primary and Junior Assessments

The classification indices for the primary- and junior-division assessments are presented in Table 7.23. The table includes the overall classification indices (NE1/Level 1, Level 1/Level 2, Level 2/Level 3 and Level 3/Level 4) and the indices for the cut point at the provincial standard (i.e., the Level 2/Level 3 cut, which divides students into those who met the provincial standard and those who did not). As expected, the indices for overall classification are lower than those for the provincial standard.

Table 7.23: Classification Accuracy and Consistency Indices: Primary- and Junior-Division Assessments

Grade and Subject	Overall Accuracy	Overall Consistency	Accuracy at the Provincial Standard Cut	Consistency at the Provincial Standard Cut
Primary Reading (English)	0.77	0.70	0.89	0.85
Junior Reading (English)	0.80	0.72	0.90	0.86
Primary Reading (French)	0.77	0.67	0.90	0.86
Junior Reading (French)	0.81	0.73	0.90	0.86
Primary Writing (English)	0.83	0.77	0.87	0.81
Junior Writing (English)	0.80	0.72	0.86	0.81
Primary Writing (French)	0.79	0.71	0.88	0.83
Junior Writing (French)	0.79	0.71	0.88	0.83
Primary Mathematics (English)	0.83	0.76	0.91	0.87
Junior Mathematics (English)	0.80	0.72	0.90	0.86
Primary Mathematics (French)	0.80	0.72	0.90	0.85
Junior Mathematics (French)	0.81	0.74	0.91	0.88

Grade 9 Assessment of Mathematics

The classification indices for the Grade 9 assessment are presented in Table 7.24. As is the case for the primary and junior assessments, the overall classification indices are lower than those for the provincial standard.

Table 7.24 Classification Accuracy and Consistency Indices: Grade 9 Mathematics

Program	Overall Accuracy	Overall Consistency	Accuracy at the Provincial Standard Cut	Consistency at the Provincial Standard Cut
Applied, Winter Administration (English)	0.68	0.57	0.87	0.82
Applied, Spring Administration (English)	0.70	0.59	0.86	0.83
Academic, Winter Administration (English)	0.80	0.74	0.90	0.85
Academic, Spring Administration (English)	0.81	0.74	0.90	0.86
Applied, Winter Administration (French)	0.67	0.57	0.87	0.82
Applied, Spring Administration (French)	0.70	0.59	0.86	0.81
Academic, Winter Administration (French)	0.78	0.71	0.88	0.83
Academic, Spring Administration (French)	0.78	0.71	0.88	0.83

OSSLT

For the OSSLT, students are classified into two categories: successful and unsuccessful. The classification indices for the English-language and French-language versions of the test are presented in Table 7.25, and they indicate high accuracy and consistency levels for both assessments.

Table 7.25 Classification Accuracy and Consistency Indices: OSSLT

	Accuracy (Successful or Unsuccessful)	Consistency (Successful or Unsuccessful)
English	0.93	0.90
French	0.92	0.88

References

- Brennan, R. L. (2004). BB-CLASS: A computer program that uses the beta-binomial model for classification consistency and accuracy [Computer software]. Iowa City, IA: The University of Iowa.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Roussos, L. A., & Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. *Journal of Educational Measurement*, 33, 215–230.
- Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58, 159–194.

Appendix 7.1: Score Distributions and Item Statistics for the 2009 EQAO Assessments

This appendix presents the classical item statistics and IRT item parameter estimates for the operational items and the DIF statistics for individual items with respect to gender, English-language learners (ELLs) and French-language learners (ALF/PDF). DIF analysis was not conducted for ALF/PDF students for the French-language version of the Grade 9 Assessment of Mathematics and the OSSLT due to small populations.

Classical item statistics and IRT item parameter estimates are combined into tables for each assessment: Tables 7.1.1 to 7.1.24 for the primary- and junior-division assessments, Tables 7.1.49 to 7.1.64 for the Grade 9 Assessment of Mathematics and Tables 7.1.77 and 7.1.82 for the OSSLT. The distribution of score points and the step parameters are also provided for open-response items.

Differential Item Functioning (DIF) statistics for individual items are shown in Tables 7.1.25 to 7.1.48 for the primary- and junior-division assessments, Tables 7.1.65 to 7.1.76 for the Grade 9 Assessment of Mathematics and Tables 7.1.83 and 7.1.85 for the OSSLT.

Primary and Junior Assessments

Classical Item Statistics and IRT Item Parameters

Table 7.1.1 Item Statistics: Primary Reading (English)

Item Code	Booklet	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQRE21002.083	1	1	R3.0	I	3	0.65	0.29	-0.53	0.44
EQRE21003.083	1	2	R3.0	C	1	0.72	0.37	-0.85	0.63
EQRE20997.083	1	3	R2.0	I	2	0.63	0.23	-0.40	0.35
EQRE20996.083	1	4	R1.0	C	4	0.69	0.35	-0.65	0.62
EQRE20994.083	1	5	R1.0	C	4*	0.59 (2.37)	0.40	-1.56	0.35
EQRE20999.083	1	6	R1.0	C	4*	0.55 (2.20)	0.41	-0.86	0.34
EQRE20986.082	1	7	R1.0	I	2	0.72	0.34	-0.74	0.67
EQRE20992.082	1	8	R3.0	E	1	0.38	0.18	1.39	0.65
EQRE20993.082	1	9	R3.0	C	3	0.72	0.44	-0.58	1.02
EQRE20990.082	1	10	R2.0	C	4	0.90	0.35	-2.00	0.79
EQRE20988.082	1	11	R1.0	C	4*	0.61 (2.45)	0.41	-0.97	0.41
EQRE20983.082	1	12	R1.0	E	4*	0.62 (2.48)	0.48	-0.98	0.45
EQRE20817.070	2	1	R1.0	I	1	0.91	0.32	-2.20	0.74
EQRE20833.070	2	2	R3.0	I	3	0.72	0.29	-0.87	0.53
EQRE20836.070	2	3	R3.0	I	4	0.51	0.30	0.49	0.58
EQRE20816.070	2	4	R1.0	E	3	0.87	0.29	-2.22	0.54
EQRE20832.070	2	5	R3.0	C	2	0.89	0.38	-1.91	0.83
EQRE20834.070	2	6	R2.0	I	3	0.84	0.42	-1.40	0.89
EQRE20830.070	2	7	R2.0	C	1	0.74	0.18	-2.05	0.24
EQRE20818.070	2	8	R1.0	I	1	0.80	0.41	-1.06	0.91
EQRE20826.070	2	9	R2.0	C	2	0.54	0.22	0.38	0.39
EQRE20819.070	2	10	R1.0	C	2	0.82	0.30	-1.71	0.51
EQRE20827.070	2	11	R2.0	I	4*	0.56 (2.22)	0.46	-0.65	0.52
EQRE20821.070	2	12	R1.0	C	4*	0.51 (2.02)	0.40	-0.16	0.44
EQRE20944.078	2	17	R2.0	C	4	0.65	0.30	-0.48	0.48
EQRE20948.078	2	18	R3.0	E	4	0.86	0.37	-1.66	0.80
EQRE20940.078	2	19	R1.0	I	3	0.58	0.40	0.02	0.97
EQRE20947.078	2	20	R3.0	I	2	0.71	0.43	-0.60	0.91
EQRE20939.078	2	21	R1.0	C	4*	0.47 (1.89)	0.44	-0.06	0.40
EQRE20945.078	2	22	R2.0	I	4*	0.46 (1.83)	0.52	0.12	0.54
EQRE21028.086	2	24	R1.0	I	3	0.79	0.30	-1.45	0.55
EQRE21031.086	2	25	R2.0	I	4	0.65	0.22	-0.64	0.32
EQRE21035.086	2	26	R3.0	I	1	0.73	0.40	-0.76	0.80
EQRE21027.086	2	27	R1.0	I	3	0.79	0.39	-1.15	0.76
EQRE21034.086	2	28	R2.0	C	4*	0.46 (1.85)	0.50	0.08	0.50
EQRE21030.086	2	29	R1.0	C	4*	0.57 (2.27)	0.51	-0.46	0.53

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

C = connections; E = explicit; I = implicit; R = reading.

**Table 7.1.2 Distribution of Score Points and Step Parameters for Open-Response Items:
Primary Reading (English)**

Item Code	Booklet	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
EQRE20994.083	1	5	% of Students	0.31	0.16	12.32	44.96	34.44	7.80
			Parameters			-4.63	-2.05	1.21	5.46
EQRE20999.083	1	6	% of Students	0.80	0.42	21.18	39.49	32.74	5.36
			Parameters			-4.61	-1.30	0.62	5.30
EQRE20988.082	1	11	% of Students	0.65	0.44	5.22	46.30	42.26	5.13
			Parameters			-4.53	-1.12	2.86	2.80
EQRE20983.082	1	12	% of Students	1.06	0.68	9.19	38.38	40.92	9.77
			Parameters			-3.38	-0.88	1.58	2.68
EQRE20827.070	2	11	% of Students	0.58	0.86	9.53	59.98	23.69	5.36
			Parameters			-3.13	-1.90	2.04	2.99
EQRE20821.070	2	12	% of Students	0.75	0.98	15.00	64.48	16.89	1.90
			Parameters			-3.92	-2.21	2.15	3.98
EQRE20939.078	2	21	% of Students	1.08	2.95	31.86	39.65	20.03	4.43
			Parameters			-3.11	-1.38	0.48	4.00
EQRE20945.078	2	22	% of Students	1.28	2.85	33.72	39.79	19.65	2.71
			Parameters			-3.02	-1.05	0.56	3.51
EQRE21034.086	2	28	% of Students	1.18	2.82	30.34	44.63	18.25	2.76
			Parameters			-3.05	-1.32	0.80	3.58
EQRE21030.086	2	29	% of Students	1.42	1.43	12.52	44.60	34.54	5.49
			Parameters			-3.25	-0.86	1.51	2.60

Note. The total number of students is 116 256.

Table 7.1.3 Item Statistics: Primary Reading (French)

Item Code	Booklet	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQLF20482.063	1	1	B	I	4	0.82	0.30	-1.59	0.56
OQLF20484.063	1	2	C	L	4	0.82	0.27	-1.79	0.48
OQLF20480.063	1	3	A	L	2	0.78	0.36	-1.15	0.69
OQLF20485.063	1	4	C	I	1	0.68	0.32	-0.56	0.55
OQLF20486.063	1	5	A	L	4*	0.54 (2.17)	0.49	-0.68	0.53
OQLF20489.063	1	6	A	L	4*	0.56 (2.22)	0.53	-0.74	0.61
OQLF20513.066	1	7	B	L	2	0.70	0.24	-1.02	0.36
OQLF20516.066	1	8	C	E	3	0.52	0.36	0.40	0.92
OQLF20514.066	1	9	C	L	4	0.86	0.35	-1.69	0.72
OQLF20510.066	1	10	A	I	4	0.70	0.43	-0.55	0.86
OQLF20518.066	1	11	A	I	4*	0.57 (2.28)	0.50	-1.20	0.40
OQLF20519.066	1	12	A	L	4*	0.50 (2.00)	0.51	-0.21	0.56
OQLF20452.061	2	1	B	L	2	0.79	0.26	-1.64	0.43
OQLF20459.061	2	2	C	L	3	0.81	0.39	-1.24	0.78
OQLF20449.061	2	3	A	I	1	0.92	0.32	-2.21	0.80
OQLF20463.061	2	4	C	I	4	0.57	0.27	0.18	0.49
OQLF20450.061	2	5	A	E	1	0.86	0.34	-1.64	0.71
OQLF20464.061	2	6	C	I	2	0.71	0.42	-0.54	0.90
OQLF20461.061	2	7	C	L	3	0.78	0.44	-0.93	0.95
OQLF20462.061	2	8	C	L	4	0.67	0.25	-0.53	0.45
OQLF20454.061	2	9	B	I	1	0.55	0.38	0.18	0.74
OQLF20456.061	2	10	C	I	1	0.23	0.25	1.57	1.62
OQLF20466.061	2	11	A	I	4*	0.52 (2.06)	0.41	-0.74	0.35
OQLF20467.061	2	12	A	L	4*	0.58 (2.32)	0.51	-0.84	0.55
OQLF20502.065	2	17	B	L	4	0.75	0.20	-1.64	0.31
OQLF20504.065	2	18	C	E	4	0.69	0.38	-0.53	0.71
OQLF20501.065	2	19	A	I	2	0.38	0.34	0.98	1.06
OQLF20503.065	2	20	C	I	1	0.68	0.35	-0.50	0.64
OQLF20509.065	2	21	A	I	4*	0.66 (2.65)	0.47	-1.93	0.39
OQLF20506.065	2	22	A	L	4*	0.43 (1.72)	0.36	-0.57	0.32
OQLF20532.068	2	24	A	I	3	0.71	0.40	-0.63	0.80
OQLF20536.068	2	25	C	I	3	0.69	0.43	-0.45	0.99
OQLF20531.068	2	26	A	I	3	0.51	0.34	0.42	0.75
OQLF20534.068	2	27	B	I	4	0.62	0.38	-0.16	0.76
OQLF20538.068	2	28	A	I	4*	0.63 (2.52)	0.45	-1.21	0.34
OQLF20539.068	2	29	B	L	4*	0.52 (2.06)	0.45	-0.62	0.43

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

C = connections; E = explicit; I = implicit; R = reading.

**Table 7.1.4 Distribution of Score Points and Step Parameters for Open-Response Items:
Primary Reading (French)**

Item Code	Booklet	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
OQLF20486.063	1	5	% of Students	0.52	0.50	16.54	51.98	25.25	5.21
			Parameters			-3.23	-1.67	1.16	3.74
OQLF20489.063	1	6	% of Students	0.54	0.41	13.21	54.86	24.98	6.01
			Parameters			-2.92	-1.67	1.29	3.30
OQLF20518.066	1	11	% of Students	0.61	0.83	27.48	33.48	17.16	20.43
			Parameters			-1.51	-2.30	-0.48	4.28
OQLF20519.066	1	12	% of Students	1.23	0.96	20.35	56.16	17.99	3.31
			Parameters			-2.89	-1.74	1.26	3.37
OQLF20466.061	2	11	% of Students	0.79	0.98	25.91	44.44	20.42	7.48
			Parameters			-3.09	-2.25	0.48	4.85
OQLF20467.061	2	12	% of Students	0.76	0.42	9.87	53.49	27.14	8.33
			Parameters			-2.83	-1.68	1.60	2.91
OQLF20509.065	2	21	% of Students	0.22	0.17	11.24	33.21	32.93	22.23
			Parameters			-2.93	-1.86	0.38	4.40
OQLF20506.065	2	22	% of Students	0.94	0.47	46.45	36.35	9.92	5.87
			Parameters			-2.32	-3.31	-0.90	6.53
OQLF20538.068	2	28	% of Students	1.26	0.87	15.50	26.74	39.86	15.77
			Parameters			-3.22	-0.48	0.24	3.45
OQLF20539.068	2	29	% of Students	1.01	0.61	22.46	50.48	19.06	6.37
			Parameters			-2.86	-2.22	0.83	4.24

Note. The total number of students is 6371.

Table 7.1.5 Item Statistics: Junior Reading (English)

Item Code	Booklet	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQRE21162.094	1	1	R2.0	C	2	0.77	0.19	-1.95	0.30
EQRE21167.094	1	2	R3.0	I	3	0.74	0.45	-0.79	0.96
EQRE21157.094	1	3	R1.0	E	2	0.90	0.25	-2.67	0.53
EQRE21158.094	1	4	R1.0	I	1	0.81	0.35	-1.47	0.63
EQRE21160.094	1	5	R1.0	C	4*	0.64 (2.54)	0.44	-1.35	0.42
EQRE21165.094	1	6	R2.0	I	4*	0.69 (2.77)	0.40	-2.42	0.32
EQRE21272.104	1	7	R1.0	I	1	0.94	0.33	-2.26	1.00
EQRE21278.104	1	8	R2.0	I	3	0.79	0.44	-1.08	0.92
EQRE21281.104	1	9	R3.0	I	2	0.65	0.29	-0.45	0.48
EQRE21271.104	1	10	R1.0	C	2	0.62	0.32	-0.28	0.55
EQRE21274.104	1	11	R1.0	C	4*	0.61 (2.44)	0.53	-1.07	0.54
EQRE21275.104	1	12	R1.0	C	4*	0.59 (2.37)	0.48	-1.05	0.46
EQRE21109.091	2	1	R1.0	I	1	0.91	0.41	-1.69	1.31
EQRE21116.091	2	2	R3.0	E	2	0.73	0.28	-1.02	0.49
EQRE21100.091	2	3	R1.0	E	3	0.85	0.32	-1.82	0.63
EQRE21112.091	2	4	R2.0	I	1	0.87	0.34	-1.75	0.76
EQRE21104.091	2	5	R1.0	E	1	0.70	0.43	-0.58	0.90
EQRE21113.091	2	6	R2.0	I	3	0.77	0.49	-0.90	1.14
EQRE21115.091	2	7	R3.0	C	1	0.25	0.29	1.70	1.03
EQRE21105.091	2	8	R1.0	I	1	0.57	0.32	0.10	0.60
EQRE21106.091	2	9	R1.0	I	2	0.87	0.31	-1.90	0.68
EQRE21101.091	2	10	R1.0	I	2	0.91	0.30	-2.23	0.72
EQRE21102.091	2	11	R1.0	I	4*	0.70 (2.78)	0.50	-1.59	0.49
EQRE21114.091	2	12	R2.0	I	4*	0.61 (2.44)	0.51	-0.93	0.50
EQRE21244.101	2	17	R1.0	I	4	0.40	0.23	1.42	0.46
EQRE21245.101	2	18	R2.0	C	1	0.76	0.40	-0.99	0.78
EQRE21248.101	2	19	R3.0	C	2	0.58	0.26	0.11	0.47
EQRE21240.101	2	20	R1.0	E	2	0.56	0.34	0.12	0.65
EQRE21243.101	2	21	R1.0	C	4*	0.48 (1.92)	0.49	-0.16	0.45
EQRE21242.101	2	22	R1.0	C	4*	0.56 (2.22)	0.44	-0.60	0.43
EQRE31959.172	2	24	R1.0	I	2	0.81	0.35	-1.56	0.61
EQRE31958.172	2	25	R1.0	E	2	0.86	0.23	-2.51	0.43
EQRE31961.172	2	26	R2.0	I	1	0.54	0.33	0.20	0.61
EQRE31963.172	2	27	R3.0	C	3	0.56	0.32	0.12	0.58
EQRE31965.172	2	28	R2.0	I	4*	0.61 (2.42)	0.42	-1.03	0.35
EQRE31964.172	2	29	R1.0	C	4*	0.54 (2.17)	0.52	-0.59	0.43

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

C = connections; E = explicit; I = implicit; R = reading.

**Table 7.1.6 Distribution of Score Points and Step Parameters for Open-Response Items:
Junior Reading (English)**

Item Code	Booklet	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
EQRE21160.094	1	5	% of Students	0.49	0.22	6.01	41.33	42.56	9.39
			Parameters			-3.97	-1.21	2.18	3.00
EQRE21165.094	1	6	% of Students	0.29	0.13	5.98	31.42	40.99	21.20
			Parameters			-3.90	-1.68	1.48	4.09
EQRE21274.104	1	11	% of Students	0.61	0.21	13.66	35.77	40.27	9.48
			Parameters			-3.28	-0.90	0.71	3.48
EQRE21275.104	1	12	% of Students	0.81	0.24	14.35	40.27	34.76	9.57
			Parameters			-3.30	-1.23	0.92	3.62
EQRE21102.091	2	11	% of Students	0.47	0.30	5.61	29.33	43.89	20.40
			Parameters			-2.88	-0.85	1.37	2.36
EQRE21114.091	2	12	% of Students	1.01	0.60	8.92	42.98	37.02	9.47
			Parameters			-3.12	-1.05	1.68	2.49
EQRE21243.101	2	21	% of Students	2.18	1.61	34.36	33.92	22.35	5.57
			Parameters			-2.76	-0.96	0.10	3.61
EQRE21242.101	2	22	% of Students	1.43	0.49	12.63	52.15	27.80	5.51
			Parameters			-3.50	-1.56	1.89	3.17
EQRE31965.172	2	28	% of Students	1.37	0.62	14.95	33.24	39.08	10.74
			Parameters			-3.66	-0.76	0.84	3.58
EQRE31964.172	2	29	% of Students	2.38	1.08	25.11	32.28	28.92	10.23
			Parameters			-2.63	-0.80	0.24	3.20

Note. The total number of students is 131 173.

Table 7.1.7 Item Statistics: Junior Reading (French)

Item Code	Booklet	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQLF20592.072	1	1	C	I	3	0.86	0.29	-1.96	0.57
OQLF20591.072	1	2	B	L	1	0.36	0.08	3.99	0.20
OQLF20594.072	1	3	C	I	4	0.52	0.23	0.55	0.43
OQLF20590.072	1	4	A	I	2	0.65	0.36	-0.33	0.68
OQLF20595.072	1	5	A	I	4*	0.54 (2.14)	0.40	-0.81	0.34
OQLF20596.072	1	6	A	L	4*	0.63 (2.51)	0.50	-1.21	0.56
OQLF20629.076	1	7	A	I	4	0.78	0.23	-1.71	0.38
OQLF20632.076	1	8	B	I	3	0.75	0.25	-1.35	0.41
OQLF20630.076	1	9	A	E	3	0.82	0.35	-1.43	0.71
OQLF20634.076	1	10	C	I	2	0.70	0.19	-0.62	0.39
OQLF20637.076	1	11	A	I	4*	0.63 (2.52)	0.39	-1.47	0.42
OQLF20635.076	1	12	A	L	4*	0.62 (2.47)	0.48	-1.19	0.53
OQLF20559.070	2	1	B	I	1	0.62	0.22	-0.26	0.36
OQLF20565.070	2	2	C	I	2	0.56	0.38	0.16	0.80
OQLF20555.070	2	3	A	E	3	0.88	0.33	-1.74	0.78
OQLF20563.070	2	4	C	I	1	0.86	0.13	-4.69	0.21
OQLF20554.070	2	5	A	E	2	0.76	0.27	-1.13	0.52
OQLF20558.070	2	6	A	I	3	0.78	0.35	-1.15	0.66
OQLF20562.070	2	7	C	L	2	0.44	0.16	1.61	0.30
OQLF20556.070	2	8	A	I	4	0.81	0.31	-1.51	0.58
OQLF20553.070	2	9	A	I	2	0.42	0.23	1.25	0.50
OQLF20552.070	2	10	A	I	2	0.82	0.20	-2.37	0.35
OQLF20566.070	2	11	A	I	4*	0.62 (2.47)	0.44	-1.24	0.44
OQLF20568.070	2	12	A	L	4*	0.64 (2.55)	0.48	-1.16	0.51
OQLF20611.074	2	17	A	I	2	0.76	0.22	-1.58	0.36
OQLF20609.074	2	18	A	E	3	0.79	0.37	-1.13	0.75
OQLF20612.074	2	19	B	L	4	0.74	0.33	-0.92	0.61
OQLF20614.074	2	20	C	L	3	0.66	0.32	-0.48	0.55
OQLF20617.074	2	21	A	L	4*	0.56 (2.23)	0.44	-0.47	0.57
OQLF20615.074	2	22	A	L	4*	0.53 (2.10)	0.44	-0.36	0.53
OQLF20663.079	2	24	B	I	3	0.67	0.31	-0.57	0.51
OQLF20665.079	2	25	C	L	1	0.77	0.15	-2.52	0.23
OQLF20661.079	2	26	A	I	2	0.49	0.36	0.51	0.77
OQLF20660.079	2	27	A	L	1	0.76	0.27	-1.32	0.45
OQLF20667.079	2	28	A	I	4*	0.50 (2.00)	0.31	-0.68	0.27
OQLF20669.079	2	29	B	L	4*	0.59 (2.37)	0.50	-0.84	0.55

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

L = connections; E = explicit; I = implicit; R = reading.

**Table 7.1.8 Distribution of Score Points and Step Parameters for Open-Response Items:
Junior Reading (French)**

Item Code	Booklet	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
OQLF20595.072	1	5	% of Students	0.71	1.05	23.64	40.94	25.83	7.83
			Parameters			-3.48	-1.75	0.54	4.69
OQLF20596.072	1	6	% of Students	0.24	0.37	6.83	43.45	39.58	9.53
			Parameters			-3.34	-1.24	1.57	3.01
OQLF20637.076	1	11	% of Students	0.21	0.29	3.98	46.36	40.92	8.24
			Parameters			-4.27	-1.58	2.80	3.04
OQLF20635.076	1	12	% of Students	0.31	0.28	7.75	43.90	39.14	8.62
			Parameters			-3.48	-1.25	1.45	3.28
OQLF20566.070	2	11	% of Students	0.26	0.26	10.89	37.82	42.54	8.24
			Parameters			-3.97	-1.04	1.08	3.94
OQLF20568.070	2	12	% of Students	0.39	0.23	9.29	32.20	50.29	7.61
			Parameters			-3.93	-0.54	1.07	3.41
OQLF20617.074	2	21	% of Students	0.65	0.18	8.27	59.90	28.69	2.31
			Parameters			-3.94	-1.37	2.18	3.13
OQLF20615.074	2	22	% of Students	0.73	0.23	14.98	59.30	22.51	2.25
			Parameters			-3.82	-1.67	1.66	3.83
OQLF20667.079	2	28	% of Students	0.86	0.15	30.87	37.70	27.52	2.90
			Parameters			-6.14	-1.52	-0.03	7.68
OQLF20669.079	2	29	% of Students	1.10	0.15	9.55	48.69	31.94	8.58
			Parameters			-2.94	-1.32	1.57	2.69

Note. The total number of students is 6192.

Table 7.1.9 Item Statistics: Primary Writing (English)

Item Code	Booklet	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQWE20250T	1	13	W2.0	T	4*	0.57 (2.29)	0.49	-0.66	0.70
EQWE20250V	1	13	W3.0	V	3*	0.75 (2.25)	0.54	-1.68	0.97
EQWE20245T	1	14	W2.0	T	4*	0.55 (2.19)	0.52	-0.68	0.60
EQWE20245V	1	14	W3.0	V	3*	0.70 (2.10)	0.56	-1.46	0.91
EQWE20233	2	13	W3.0	V	2	0.79	0.40	-1.32	0.67
EQWE20205	2	14	W1.0	T	1	0.77	0.41	-1.03	0.82
EQWE20206	2	15	W1.0	T	3	0.56	0.29	0.12	0.55
EQWE20224	2	16	W3.0	V	2	0.68	0.36	-0.62	0.74
EQWE30434T	2	23	W2.0	T	4*	0.46 (1.82)	0.46	-0.25	0.51
EQWE30434V	2	23	W3.0	V	3*	0.67 (2.02)	0.51	-1.39	0.94
EQWE20215	2	30	W2.0	T	3	0.78	0.40	-1.52	0.58
EQWE20230	2	31	W3.0	V	4	0.65	0.36	-0.56	0.73
EQWE20218	2	32	W2.0	T	3	0.56	0.29	0.01	0.53
EQWE20222	2	33	W3.0	V	3	0.73	0.36	-1.28	0.51

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

T = content; V = conventions.

Table 7.1.10 Distribution of Score Points and Step Parameters for Open-Response Items: Primary Writing (English)

Item Code	Booklet	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
EQWE20250T	1	13	% of Students	0.51	0.65	5.97	49.16	39.53	4.13
			Parameters			-3.03	-2.59	0.22	2.76
EQWE20250V	1	13	% of Students	0.51	0.18	8.83	48.43	42.05	
			Parameters			-3.41	-1.87	0.24	
EQWE20245T	1	14	% of Students	0.72	0.96	13.21	43.08	33.95	8.08
			Parameters			-3.42	-1.77	0.31	2.18
EQWE20245V	1	14	% of Students	0.72	0.28	15.26	47.02	36.63	
			Parameters			-3.42	-1.39	0.44	
EQWE30434T	2	23	% of Students	0.69	1.87	21.83	48.04	23.62	3.93
			Parameters			-3.69	-1.31	1.06	2.94
EQWE30434V	2	23	% of Students	0.69	0.50	12.94	57.80	28.07	
			Parameters			-3.39	-1.60	0.82	

Note. The total number of students is 116 809.

Table 7.1.11: Item Statistics: Primary Writing (French)

Item Code	Booklet	Sequence	Expectation	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQEF20220T	1	13	A	T	4*	0.50 (2.01)	0.71	-0.42	0.66
OQEF20220V	1	13	C	V	3*	0.70 (2.09)	0.79	-1.32	1.43
OQEF20226T	1	14	A	T	4*	0.56 (2.30)	0.79	-0.79	0.84
OQEF20226V	1	14	C	V	3*	0.70 (2.10)	0.81	-1.36	1.44
OQEF20195	2	13	A	T	4	0.53	0.16	0.47	0.47
OQEF20202	2	14	C	V	3	0.83	0.26	-1.52	0.71
OQEF20192	2	15	B	T	3	0.81	0.24	-1.65	0.53
OQEF20217	2	16	C	V	4	0.32	0.15	1.36	1.01
OQEF20230T	2	23	A	T	4*	0.55 (2.21)	0.34	-0.75	0.66
OQEF20230V	2	23	C	V	3*	0.67 (2.01)	0.40	-1.48	0.96
OQEF10026	2	30	A	T	2	0.89	0.26	-2.41	0.56
OQEF20199	2	31	B	T	3	0.69	0.22	-0.63	0.58
OQEF20212	2	32	C	V	3	0.76	0.24	-1.14	0.60
OQEF20207	2	33	C	V	4	0.62	0.22	-0.24	0.61

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

T = content; V = conventions.

Table 7.1.12 Distribution of Score Points and Step Parameters for Open-Response Items: Primary Writing (French)

Item Code	Booklet	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
OQEF20220T	1	13	% of Students	0.49	1.85	12.34	50.57	28.80	5.96
			Parameters			-2.82	-1.84	0.66	2.30
OQEF20220V	1	13	% of Students	0.49	0.11	14.56	51.23	33.61	
			Parameters			-3.15	-1.28	0.49	
OQEF20226T	1	14	% of Students	0.44	0.25	8.78	46.97	37.34	6.22
			Parameters			-3.55	-1.97	0.21	2.17
OQEF20226V	1	14	% of Students	0.44	0.11	13.23	53.25	32.97	
			Parameters			-3.23	-1.36	0.51	
OQEF20230T	2	23	% of Students	0.75	1.10	11.29	44.25	32.62	9.99
			Parameters			-3.22	-1.90	0.30	1.83
OQEF20230V	2	23	% of Students	0.75	0.09	15.50	57.09	26.56	
			Parameters			-3.86	-1.42	0.83	

Note. The total number of students is 6378.

Table 7.1.13 Item Statistics: Junior Writing (English)

Item Code	Booklet	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQWE20315T	1	13	W2.0	T	4*	0.60 (2.41)	0.51	-1.08	0.65
EQWE20315V	1	13	W3.0	V	3*	0.67 (2.00)	0.52	-1.43	0.97
EQWE20304T	1	14	W2.0	T	4*	0.60 (2.39)	0.56	-1.00	0.70
EQWE20304V	1	14	W3.0	V	3*	0.76 (2.28)	0.57	-1.74	0.98
EQWE20272	2	13	W2.0	T	4	0.87	0.29	-2.21	0.60
EQWE20291	2	14	W3.0	V	1	0.59	0.23	-0.11	0.49
EQWE20271	2	15	W1.0	T	3	0.64	0.17	-0.78	0.25
EQWE20278	2	16	W3.0	V	4	0.70	0.27	-0.82	0.62
EQWE20322T	2	23	W2.0	T	4*	0.55 (2.21)	0.53	-0.75	0.74
EQWE20322V	2	23	W3.0	V	3*	0.71 (2.13)	0.55	-1.56	1.03
EQWE20280	2	30	W3.0	V	1	0.81	0.30	-1.47	0.73
EQWE20276	2	31	W2.0	T	2	0.61	0.18	-0.35	0.32
EQWE20285	2	32	W3.0	V	2	0.81	0.24	-2.05	0.44
EQWE20290	2	33	W3.0	V	2	0.60	0.24	-0.18	0.60

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

T = content; V = conventions.

Table 7.1.14 Distribution of Score Points and Step Parameters for Open-Response Items: Junior Writing (English)

Item Code	Booklet	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
EQWE20315T	1	13	% of Students	0.48	0.26	7.13	45.28	37.01	9.84
			Parameters			-3.67	-2.63	0.09	1.89
EQWE20315V	1	13	% of Students	0.48	0.15	13.13	65.92	20.31	
			Parameters			-3.63	-1.74	1.09	
EQWE20304T	1	14	% of Students	0.72	0.22	11.13	38.26	38.26	11.41
			Parameters			-3.67	-1.94	-0.09	1.69
EQWE20304V	1	14	% of Students	0.72	0.08	9.27	43.54	46.39	
			Parameters			-3.32	-1.92	0.01	
EQWE20322T	2	23	% of Students	1.02	0.64	11.57	43.88	34.72	8.17
			Parameters			-3.23	-1.91	0.19	1.94
EQWE20322V	2	23	% of Students	1.02	0.17	12.62	46.21	39.97	
			Parameters			-3.28	-1.63	0.22	

Note. The total number of students is 131 296.

Table 7.1.15 Item Statistics: Junior Writing (French)

Item Code	Booklet	Sequence	Expectation	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQEF20268T	1	13	A	T	4*	0.60 (2.42)	0.74	-0.93	0.85
OQEF20268V	1	13	C	V	3*	0.69 (2.06)	0.76	-1.29	1.56
OQEF20278T	1	14	A	T	4*	0.61 (2.43)	0.76	-1.05	0.84
OQEF20278V	1	14	C	V	3*	0.69 (2.06)	0.77	-1.29	1.59
OQEF20241	2	13	A	T	2	0.71	0.19	-0.95	0.45
OQEF20248	2	14	C	V	1	0.76	0.21	-1.30	0.51
OQEF20243	2	15	A	T	2	0.74	0.20	-1.15	0.46
OQEF20266	2	16	C	V	3	0.50	0.18	0.53	0.57
OQEF30611T	2	23	A	T	4*	0.63 (2.54)	0.34	-1.39	0.54
OQEF30611V	2	23	C	V	3*	0.69 (2.06)	0.37	-1.81	0.81
OQEF20264	2	30	C	V	4	0.74	0.20	-1.33	0.42
OQEF20246	2	31	B	T	3	0.81	0.20	-2.45	0.34
OQEF20245	2	32	B	T	1	0.78	0.16	-2.36	0.28
OQEF20260	2	33	C	V	2	0.69	0.19	-0.81	0.43

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

T = content; V = conventions.

Table 7.1.16 Distribution of Score Points and Step Parameters for Open-Response Items: Junior Writing (French)

Item Code	Booklet	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
OQEF20268T	1	13	% of Students	0.41	0.23	7.33	43.25	40.24	8.55
			Parameters			-3.38	-2.21	0.00	1.88
OQEF20268V	1	13	% of Students	0.41	0.03	12.33	62.49	24.74	
			Parameters			-3.16	-1.47	0.77	
OQEF20278T	1	14	% of Students	0.39	0.11	8.62	41.22	40.19	9.47
			Parameters			-3.91	-2.05	-0.03	1.79
OQEF20278V	1	14	% of Students	0.39	0.02	14.49	62.40	22.71	
			Parameters			-3.38	-1.34	0.85	
OQEF30611T	2	23	% of Students	0.40	0.15	6.62	35.49	47.63	9.72
			Parameters			-4.64	-2.76	-0.46	2.31
OQEF30611V	2	23	% of Students	0.40	0.02	14.72	58.78	26.08	
			Parameters			-4.60	-1.71	0.90	

Note. The total number of students is 6225.

Table 7.1.17 Item Statistics: Primary Mathematics (English)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQME10073	1	NV2	KU	N	3	86.00	0.38	-1.51	0.96
EQME10167	2	NV1	PS	N	2	71.24	0.37	-0.74	0.69
EQME10038	3	PV2	AP	P	3	56.04	0.38	0.15	0.95
EQME10062	4	NV3	AP	N	4	47.08	0.35	0.56	0.88
EQME10042	5	PV1	AP	P	1	70.18	0.26	-0.76	0.50
EQME20936	6	NV3	PS	N	4	64.53	0.29	-0.30	0.62
EQME20944	7	PV1	AP	P	4*	71.50 (2.86)	0.56	-2.04	0.46
EQME20880	8	NV1	AP	N	4*	59.75 (2.39)	0.57	-1.46	0.44
EQME20811	9	GV3	AP	G	4*	61.50 (2.46)	0.48	-1.39	0.41
EQME20822	10	MV1	AP	M	4*	67.75 (2.71)	0.54	-2.19	0.35
EQME10126	11	MV1	KU	M	4	70.44	0.30	-0.70	0.58
EQME20894	12	PV2	KU	P	3	85.90	0.32	-1.85	0.65
EQME10115	13	MV1	KU	M	2	83.54	0.36	-1.51	0.77
EQME20883	14	DV3	AP	D	2	51.30	0.31	0.41	0.66
EQME20859	15	DV2	KU	D	2	61.22	0.32	-0.18	0.59
EQME10019	16	MV2	KU	M	3	76.45	0.37	-1.02	0.73
EQME20874	17	MV2	KU	M	1	53.19	0.26	0.40	0.56
EQME10079	18	NV1	AP	N	3	39.17	0.36	0.85	1.14
EQME20863	19	PV1	KU	P	1	86.43	0.32	-1.79	0.70
EQME10166	20	PV2	AP	P	3	73.70	0.37	-0.82	0.73
EQME20915	21	NV3	AP	N	2	52.65	0.33	0.34	0.77
EQME10058	22	GV2	KU	G	2	78.11	0.26	-1.57	0.46
EQME10046	23	GV3	AP	G	3	56.70	0.40	0.08	0.85
EQME20832	24	MV1	AP	M	4	52.42	0.41	0.22	0.99
EQME20833	25	MV2	AP	M	2	39.36	0.32	0.91	1.01
EQME10051	26	DV3	AP	D	3	72.29	0.39	-0.69	0.79
EQME10170	27	DV1	AP	D	4*	73.25 (2.93)	0.50	-2.24	0.36
EQME20825	28	NV3	PS	N	4*	60.25 (2.41)	0.53	-1.51	0.38
EQME20878	29	GV3	AP	G	4*	68.00 (2.72)	0.46	-2.33	0.30
EQME20941	30	DV2	PS	D	4*	62.25 (2.49)	0.60	-1.31	0.48
EQME10084	31	MV1	AP	M	3	56.49	0.41	0.07	0.85
EQME10108	32	NV3	KU	N	3	83.11	0.38	-1.38	0.81
EQME20893	33	PV2	PS	P	4	41.02	0.39	0.74	1.17
EQME20828	34	GV2	KU	G	1	56.74	0.34	0.07	0.71
EQME10029	35	MV2	PS	M	3	60.50	0.43	-0.07	1.00
EQME20871	36	GV1	AP	G	2	53.49	0.38	0.26	0.92

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

AP = algorithms and prior knowledge; KU = knowledge and understanding; PS = problem solving;

N = number sense and numeration; P = patterning and algebra; D = data management and probability;

M = measurement; G = geometry and spatial sense.

Table 7.1.18 Distribution of Score Points and Step Parameters for Open-Response Items: Primary Mathematics (English)

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
EQME20944	7	% of Students	0.29	0.09	15.52	24.47	17.35	42.28
		Parameters			4.10	-0.65	-2.26	-1.20
EQME20880	8	% of Students	0.72	0.2	32.68	23.86	11.94	30.6
		Parameters			4.53	-1.41	-2.45	-0.67
EQME20811	9	% of Students	0.49	0.18	15.21	34.31	37.18	12.62
		Parameters			4.29	0.40	-1.24	-3.45
EQME20822	10	% of Students	0.58	0.08	29.88	12.34	11.86	45.25
		Parameters			5.36	-3.04	-2.09	-0.24
EQME10170	27	% of Students	0.64	0.2	16.88	11.11	30.51	40.66
		Parameters			4.39	-2.12	-0.28	-1.99
EQME20825	28	% of Students	0.88	0.19	31.97	15.12	29.01	22.83
		Parameters			5.01	-2.21	-0.48	-2.33
EQME20878	29	% of Students	0.7	0.05	17.59	27.13	18.33	36.19
		Parameters			5.27	-1.07	-2.87	-1.33
EQME20941	30	% of Students	1.22	0.19	29.77	20.69	14.33	33.79
		Parameters			3.69	-1.21	-1.74	-0.74

Note. The total number of students is 120 258.

Table 7.1.19 Item Statistics: Primary Mathematics (French)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQMF10064	1	2	MA	N	2	83.95	0.33	-1.43	0.78
OQMF20541	2	1	MA	N	2	64.96	0.37	-0.32	0.77
OQMF20605	3	2	CC	A	3	70.40	0.33	-0.60	0.71
OQMF20560	4	4	MA	N	4	63.70	0.44	-0.19	1.16
OQMF20616	5	3	CC	A	1	63.08	0.38	-0.18	0.86
OQMF20540	6	1	CC	N	4	76.58	0.43	-0.78	1.11
OQMF20510	7	2	MA	G	4*	67.50 (2.70)	0.38	-2.63	0.22
OQMF20640	8	1	MA	N	4*	69.25 (2.77)	0.51	-2.09	0.39
OQMF20513	9	2	HP	G	4*	81.25 (3.25)	0.37	-3.60	0.23
OQMF10044	10	4	HP	M	4*	49.25 (1.97)	0.52	-1.05	0.46
OQMF20576	11	3	CC	M	3	66.25	0.33	-0.41	0.63
OQMF20590	12	1	MA	G	1	67.08	0.32	-0.46	0.65
OQMF20546	13	2	HP	N	4	53.57	0.43	0.19	1.03
OQMF20571	14	2	CC	M	3	88.09	0.19	-2.96	0.38
OQMF20628	15	1	MA	T	4	68.24	0.40	-0.43	0.86
OQMF20583	16	4	CC	M	3	62.30	0.29	-0.15	0.59
OQMF20630	17	1	HP	T	3	52.53	0.45	0.23	1.27
OQMF20635	18	2	MA	T	4	74.61	0.36	-0.83	0.73
OQMF10084	19	4	CC	N	2	78.02	0.33	-1.14	0.66
OQMF20572	20	2	MA	M	3	56.07	0.34	0.15	0.82
OQMF20587	21	1	CC	G	2	78.21	0.29	-1.41	0.51
OQMF20566	22	1	HP	M	4	50.09	0.38	0.41	0.87
OQMF20550	23	3	MA	N	3	66.93	0.41	-0.35	0.93
OQMF20611	24	3	MA	A	1	62.30	0.23	-0.19	0.43
OQMF10067	25	3	CC	N	4	91.04	0.31	-2.21	0.72
OQMF20593	26	1	HP	G	4	76.27	0.28	-1.25	0.49
OQMF20639	27	2	MA	T	4*	77.75 (3.11)	0.53	-1.93	0.46
OQMF20504	28	4	HP	N	4*	53.75 (2.15)	0.44	-0.94	0.28
OQMF20526	29	1	HP	T	4*	73.00 (2.92)	0.57	-1.94	0.48
OQMF20636	30	2	HP	A	4*	47.50 (1.90)	0.47	-0.35	0.50
OQMF20597	31	2	MA	G	2	67.54	0.28	-0.59	0.47
OQMF20556	32	3	HP	N	3	41.66	0.21	1.36	0.47
OQMF20622	33	1	CC	T	2	85.06	0.35	-1.59	0.73
OQMF20600	34	1	CC	A	3	75.64	0.29	-1.15	0.51
OQMF20564	35	1	MA	M	3	67.33	0.48	-0.36	1.23
OQMF20606	36	2	HP	A	1	75.58	0.35	-0.88	0.74

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

MA = algorithms and prior knowledge; CC = knowledge and understanding; HP = problem solving;

N = number sense and numeration; A = patterning and algebra; T = data management and probability;

M = measurement; G = geometry and spatial sense.

Table 7.1.20 Distribution of Score Points and Step Parameters for Open-Response Items: Primary Mathematics (French)

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
OQMF20510	7	% of Students	0.53	0.39	23.63	14.33	26.65	34.47
		Parameters			6.64	-3.56	-0.94	-2.14
OQMF20640	8	% of Students	0.23	0.11	18.60	16.38	32.92	31.75
		Parameters			4.96	-1.66	-0.83	-2.48
OQMF20513	9	% of Students	0.27	0.20	12.91	12.77	9.03	64.83
		Parameters			5.87	-3.07	-4.22	1.42
OQMF10044	10	% of Students	0.67	0.16	51.68	17.29	10.48	19.72
		Parameters			5.12	-2.21	-1.96	-0.95
OQMF20639	27	% of Students	0.55	0.41	8.68	10.89	37.82	41.65
		Parameters			2.63	-0.68	0.01	-1.96
OQMF20504	28	% of Students	2.13	1.77	37.79	16.12	24.08	18.12
		Parameters			4.65	-2.46	-0.20	-1.99
OQMF20526	29	% of Students	0.52	0.08	13.64	19.92	25.00	40.84
		Parameters			3.67	-0.64	-1.39	-1.63
OQMF20636	30	% of Students	0.84	0.28	30.04	49.54	16.34	2.96
		Parameters			4.75	0.54	-2.01	-3.27

Note. The total number of students is 6393.

Table 7.1.21 Item Statistics: Junior Mathematics (English)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
3459716	1	NV1	KU	N	4	80.51	0.26	-1.57	0.52
EQME21115	2	NV2	AP	N	4	69.36	0.30	-0.67	0.59
EQME10180	3	MV2	AP	M	3	50.77	0.43	0.28	1.30
EQME10294	4	MV2	AP	M	1	62.19	0.35	-0.20	0.73
EQME21102	5	PV2	AP	P	1	86.90	0.36	-1.55	0.95
EQME10193	6	PV1	KU	P	3	65.20	0.32	-0.39	0.63
EQME21089	7	GV1	AP	G	4*	63.00 (2.52)	0.52	-1.23	0.36
EQME21119	8	DV3	PS	D	4*	46.50 (1.86)	0.51	-1.22	0.40
EQME21138	9	PV1	PS	P	4*	56.25 (2.25)	0.49	-1.35	0.37
EQME21043	10	MV2	AP	M	4*	61.50 (2.46)	0.63	-1.09	0.55
EQME21140	11	DV2	AP	D	4	64.15	0.26	-0.39	0.48
EQME10211	12	DV3	AP	D	3	53.71	0.31	0.29	0.70
EQME21031	13	DV2	AP	D	2	62.61	0.33	-0.22	0.69
EQME10282	14	PV1	KU	P	4	82.12	0.38	-1.21	0.97
EQME21070	15	PV2	PS	P	4	55.61	0.40	0.10	0.90
EQME21021	16	GV3	AP	G	2	52.34	0.28	0.40	0.61
EQME21063	17	GV1	KU	G	4	65.76	0.42	-0.31	1.03
EQME21080	18	GV1	PS	G	1	39.92	0.31	0.89	0.98
EQME21130	19	MV1	KU	M	1	53.41	0.27	0.34	0.69
EQME21083	20	MV2	PS	M	1	51.06	0.29	0.44	0.59
EQME10185	21	MV2	KU	M	3	74.60	0.23	-1.39	0.38
EQME21142	22	NV3	PS	N	2	37.83	0.29	1.03	0.96
EQME10237	23	NV1	PS	N	2	61.68	0.34	-0.17	0.70
EQME21008	24	PV1	AP	P	4	78.46	0.28	-1.50	0.50
EQME21133	25	PV1	AP	P	1	69.81	0.37	-0.58	0.80
EQME21060	26	GV3	PS	G	4	49.34	0.37	0.41	1.01
EQME21104	27	DV1	PS	D	4*	59.25 (2.37)	0.53	-0.93	0.45
EQME10229	28	NV2	PS	N	4*	64.00 (2.56)	0.55	-1.25	0.42
EQME21137	29	NV1	PS	N	4*	55.75 (2.23)	0.63	-0.87	0.57
EQME21135	30	GV3	AP	G	4*	64.50 (2.58)	0.56	-1.39	0.43
EQME21048	31	NV2	AP	N	2	45.83	0.24	0.77	0.75
EQME21099	32	NV2	KU	N	1	70.16	0.4	-0.56	0.90
EQME10256	33	MV1	AP	M	2	45.56	0.34	0.60	0.95
EQME10259	34	MV2	PS	M	1	49.90	0.36	0.41	0.99
EQME21062	35	DV3	PS	D	3	64.64	0.36	-0.31	0.80
EQME21139	36	DV2	KU	D	1	52.88	0.43	0.20	1.09

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

AP = algorithms and prior knowledge; KU = knowledge and understanding; PS = problem solving;

N = number sense and numeration; P = patterning and algebra; D = data management and probability;

M = measurement; G = geometry and spatial sense.

Table 7.1.22 Distribution of Score Points and Step Parameters for Open-Response Items: Junior Mathematics (English)

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
EQME21089	7	% of Students	2.25	0.96	25.07	17.94	23.82	29.97
		Parameters			3.15	-1.22	-0.71	-1.22
EQME21119	8	% of Students	0.63	0.10	55.59	19.25	5.98	18.45
		Parameters			6.00	-2.53	-3.22	-0.25
EQME21138	9	% of Students	0.91	0.21	25.6	40.75	12.17	20.35
		Parameters			4.58	-0.14	-3.39	-1.05
EQME21043	10	% of Students	1.50	0.24	28.43	20.89	19.8	29.13
		Parameters			3.12	-0.83	-1.13	-1.16
EQME21104	27	% of Students	1.30	0.23	22.74	22.84	43.28	9.61
		Parameters			3.85	-0.30	-0.11	-3.44
EQME10229	28	% of Students	2.24	0.45	17.51	32.44	15.36	31.99
		Parameters			2.57	0.31	-2.23	-0.65
EQME21137	29	% of Students	1.76	0.24	31.78	31.98	10.18	24.06
		Parameters			3.17	-0.38	-2.17	-0.62
EQME21135	30	% of Students	1.52	0.39	20.03	28.67	16.53	32.87
		Parameters			3.15	-0.27	-2.03	-0.85

Note. The total number of students in Primary Reading (English) is 130 484.

Table 7.1.23 Item Statistics: Junior Mathematics (French)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQMF20716	1	2	MA	N	3	80.98	0.26	-1.58	0.53
OQMF20745	2	2	HP	M	2	57.14	0.43	0.08	1.16
OQMF20816	3	1	HP	T	4	33.94	0.32	1.19	1.03
OQMF10189	4	2	CC	N	3	74.48	0.31	-0.87	0.67
OQMF20703	5	1	MA	N	4	59.38	0.38	-0.02	0.90
OQMF20723	6	3	MA	N	2	40.54	0.29	0.99	0.82
OQMF10131	7	1	MA	G	4*	75.00 (3.00)	0.49	-2.33	0.37
OQMF20796	8	1	HP	A	4*	65.75 (2.63)	0.44	-2.40	0.28
OQMF20718	9	3	MA	N	4*	82.75 (3.31)	0.42	-2.60	0.32
OQMF20844	10	2	HP	G	4*	48.00 (1.92)	0.48	-0.75	0.38
OQMF20832	11	2	CC	T	1	47.50	0.22	0.82	0.51
OQMF10139	12	1	HP	G	3	80.25	0.42	-0.98	1.15
OQMF20778	13	2	HP	G	4	38.03	0.37	0.89	1.17
OQMF10162	14	2	CC	M	2	51.66	0.33	0.41	0.72
OQMF20827	15	2	MA	T	2	78.94	0.35	-1.07	0.79
OQMF20802	16	2	HP	A	3	39.53	0.36	0.85	1.17
OQMF20751	17	3	MA	M	2	58.67	0.41	-0.01	0.92
OQMF20729	18	1	CC	M	2	83.56	0.36	-1.36	0.83
OQMF20701	19	1	CC	N	3	72.98	0.32	-0.78	0.68
OQMF20753	20	3	CC	M	1	59.11	0.38	0.00	0.85
OQMF20814	21	1	MA	T	2	54.02	0.25	0.38	0.49
OQMF10137	22	1	MA	G	3	82.30	0.37	-1.19	0.95
OQMF20829	23	2	HP	T	3	36.25	0.26	1.26	0.82
OQMF10125	24	2	MA	A	2	59.56	0.38	-0.04	0.84
OQMF10115	25	1	CC	A	4	47.49	0.49	0.39	1.53
OQMF10149	26	2	CC	G	3	50.28	0.29	0.53	0.66
OQMF20710	27	1	HP	N	4*	68.00 (2.72)	0.59	-1.42	0.46
OQMF20747	28	2	MA	M	4*	57.25 (2.29)	0.61	-0.84	0.55
OQMF10203	29	1	HP	T	4*	74.25 (2.97)	0.48	-2.13	0.34
OQMF10127	30	2	MA	A	4*	63.75 (2.55)	0.55	-1.61	0.34
OQMF20790	31	1	MA	A	4	82.96	0.36	-1.33	0.82
OQMF10156	32	1	HP	M	2	62.73	0.34	-0.19	0.71
OQMF20760	33	1	CC	G	3	46.29	0.24	0.85	0.55
OQMF20734	34	1	MA	M	1	39.55	0.29	1.06	0.76
OQMF20725	35	3	HP	N	3	47.49	0.22	0.85	0.51
OQMF20773	36	2	CC	G	1	82.98	0.31	-1.55	0.63

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

MA = algorithms and prior knowledge; CC = knowledge and understanding; HP = problem solving;

N = number sense and numeration; A = patterning and algebra; T = data management and probability;

M = measurement; G = geometry and spatial sense.

Table 7.1.24 Distribution of Score Points and Step Parameters for Open-Response Items: Junior Mathematics (French)

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
OQMF10131	7	% of Students	0.44	0.13	10.95	28.37	8.09	52.03
		Parameters			3.67	-0.06	-4.11	0.50
OQMF20796	8	% of Students	0.47	0.19	23.78	21.96	18.71	34.88
		Parameters			6.06	-2.06	-2.67	-1.34
OQMF20718	9	% of Students	0.74	0.10	8.65	13.11	13.55	63.85
		Parameters			2.83	-0.93	-2.14	0.25
OQMF20844	10	% of Students	1.55	0.55	44.21	27.66	11.79	14.24
		Parameters			4.75	-1.25	-2.35	-1.15
OQMF20710	27	% of Students	1.40	0.31	22.67	17.99	17.18	40.45
		Parameters			3.05	-1.04	-1.35	-0.65
OQMF20747	28	% of Students	1.39	0.60	29.52	26.53	21.93	20.04
		Parameters			3.18	-0.46	-1.15	-1.57
OQMF10203	29	% of Students	0.92	0.11	14.18	16.02	24.69	44.08
		Parameters			3.64	-1.21	-1.14	-1.30
OQMF10127	30	% of Students	2.00	0.32	40.50	4.10	6.36	46.71
		Parameters			4.42	-5.14	-0.81	1.53

Note. The total number of students is 6217.

Differential Item Functioning (DIF) Analysis Results

The gender-based, ELL-based and ALF/PDF-based DIF results are provided for the primary- and junior-division assessments in Tables 7.1.25 to 7.1.48. Results are presented for two random samples. Each table indicates the value of $\hat{\beta}$ and the significance level for each item. The level of DIF (B or C) is also provided for those items that had significant DIF with at least a B-level or C-level effect size. Results for items with B-Level or C-level DIF in both samples are presented in boldface. For gender-based DIF, negative estimates of $\hat{\beta}$ indicate that the girls outperformed the boys; positive $\hat{\beta}$ estimates indicate that the boys outperformed the girls. For ELL- and ALF/PDF-based DIF, negative estimates of $\hat{\beta}$ indicate that the ELLs and ALF/PDFs outperformed the non-ELLs and non-ALF/PDFs, respectively; positive $\hat{\beta}$ estimates indicate that the non-ELLs and non-ALF/PDFs outperformed the ELLs and ALF/PDFs, respectively.

Table 7.1.25 Gender-Based DIF Statistics: Primary Reading (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQRE21002.083	1	1	0.117	0.000	C+	0.092	0.000	C+
EQRE21003.083	1	2	0.054	0.000		0.048	0.000	
EQRE20997.083	1	3	0.030	0.044		0.002	0.885	
EQRE20996.083	1	4	0.041	0.003		0.029	0.034	
EQRE20994.083	1	5	0.050	0.037		0.078	0.001	B+
EQRE20999.083	1	6	0.013	0.626		0.016	0.523	
EQRE20986.082	1	7	-0.022	0.114		0.002	0.892	
EQRE20992.082	1	8	0.007	0.638		0.018	0.249	
EQRE20993.082	1	9	0.054	0.000		0.046	0.000	
EQRE20990.082	1	10	0.014	0.122		0.016	0.069	
EQRE20988.082	1	11	-0.011	0.605		-0.008	0.685	
EQRE20983.082	1	12	0.053	0.035		0.097	0.000	C+
EQRE20817.070	2	1	-0.003	0.775		-0.005	0.587	
EQRE20833.070	2	2	-0.053	0.000		-0.045	0.001	
EQRE20836.070	2	3	0.009	0.561		-0.002	0.920	
EQRE20816.070	2	4	-0.005	0.596		-0.012	0.229	
EQRE20832.070	2	5	-0.003	0.742		0.007	0.418	
EQRE20834.070	2	6	-0.012	0.262		0.001	0.945	
EQRE20830.070	2	7	-0.006	0.670		-0.004	0.768	
EQRE20818.070	2	8	-0.029	0.015		-0.024	0.038	
EQRE20826.070	2	9	0.053	0.001		0.046	0.003	
EQRE20819.070	2	10	0.001	0.910		-0.013	0.269	
EQRE20827.070	2	11	-0.028	0.179		-0.008	0.697	
EQRE20821.070	2	12	0.025	0.216		0.027	0.179	
EQRE20944.078	2	17	-0.047	0.001		-0.058	0.000	
EQRE20948.078	2	18	-0.027	0.007		-0.018	0.084	
EQRE20940.078	2	19	-0.061	0.000	B-	-0.032	0.029	
EQRE20947.078	2	20	-0.007	0.602		-0.011	0.399	
EQRE20939.078	2	21	0.013	0.618		-0.010	0.708	
EQRE20945.078	2	22	-0.016	0.504		-0.042	0.084	
EQRE21028.086	2	24	-0.034	0.007		-0.019	0.125	
EQRE21031.086	2	25	-0.056	0.000		-0.040	0.008	
EQRE21035.086	2	26	-0.003	0.803		0.023	0.075	
EQRE21027.086	2	27	-0.034	0.006		-0.004	0.768	
EQRE21034.086	2	28	-0.130	0.000	C-	-0.118	0.000	C-
EQRE21030.086	2	29	0.029	0.223		-0.016	0.498	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.26 Gender-Based DIF Statistics: Primary Reading (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	p-value	Level of DIF	$\hat{\beta}$	p-value	Level of DIF
OQLF20482.063	1	1	-0.008	0.480		0.007	0.541	
OQLF20484.063	1	2	0.013	0.265		-0.020	0.098	
OQLF20480.063	1	3	0.019	0.134		0.017	0.180	
OQLF20485.063	1	4	-0.011	0.450		-0.014	0.344	
OQLF20486.063	1	5	-0.038	0.098		-0.060	0.008	B-
OQLF20489.063	1	6	-0.028	0.201		-0.062	0.004	B-
OQLF20513.066	1	7	-0.039	0.008		-0.028	0.056	
OQLF20516.066	1	8	0.009	0.535		0.022	0.143	
OQLF20514.066	1	9	-0.017	0.094		-0.010	0.347	
OQLF20510.066	1	10	0.023	0.086		0.015	0.260	
OQLF20518.066	1	11	0.003	0.931		0.027	0.387	
OQLF20519.066	1	12	-0.075	0.000	B-	-0.062	0.004	B-
OQLF20452.061	2	1	-0.013	0.312		-0.002	0.881	
OQLF20459.061	2	2	-0.029	0.014		-0.013	0.282	
OQLF20449.061	2	3	-0.011	0.165		-0.006	0.463	
OQLF20463.061	2	4	-0.036	0.018		-0.038	0.013	
OQLF20450.061	2	5	-0.004	0.707		0.004	0.692	
OQLF20464.061	2	6	0.004	0.743		-0.018	0.182	
OQLF20461.061	2	7	-0.068	0.000	B-	-0.058	0.000	
OQLF20462.061	2	8	0.008	0.566		0.010	0.507	
OQLF20454.061	2	9	-0.015	0.322		-0.022	0.150	
OQLF20456.061	2	10	0.015	0.233		0.014	0.281	
OQLF20466.061	2	11	0.002	0.925		0.000	0.992	
OQLF20467.061	2	12	-0.047	0.036		-0.033	0.144	
OQLF20502.065	2	17	0.013	0.322		0.002	0.867	
OQLF20504.065	2	18	0.003	0.858		0.008	0.589	
OQLF20501.065	2	19	0.027	0.059		0.029	0.045	
OQLF20503.065	2	20	0.010	0.469		0.024	0.093	
OQLF20509.065	2	21	0.008	0.783		0.006	0.838	
OQLF20506.065	2	22	0.057	0.032		0.079	0.003	B+
OQLF20532.068	2	24	0.032	0.015		0.035	0.010	
OQLF20536.068	2	25	0.039	0.003		0.022	0.105	
OQLF20531.068	2	26	0.020	0.180		0.019	0.222	
OQLF20534.068	2	27	-0.019	0.190		-0.032	0.026	
OQLF20538.068	2	28	0.065	0.022	B+	0.075	0.009	B+
OQLF20539.068	2	29	0.061	0.015	B+	0.036	0.141	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.27 Gender-Based DIF Statistics: Junior Reading (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQRE21162.094	1	1	-0.047	0.000		-0.034	0.010	
EQRE21167.094	1	2	0.034	0.008		0.001	0.935	
EQRE21157.094	1	3	0.007	0.450		-0.007	0.456	
EQRE21158.094	1	4	-0.025	0.028		0.002	0.856	
EQRE21160.094	1	5	-0.015	0.508		-0.065	0.003	B-
EQRE21165.094	1	6	-0.029	0.255		-0.008	0.753	
EQRE21272.104	1	7	0.012	0.085		0.003	0.637	
EQRE21278.104	1	8	0.017	0.156		0.013	0.259	
EQRE21281.104	1	9	-0.033	0.027		-0.038	0.011	
EQRE21271.104	1	10	-0.068	0.000	B-	-0.063	0.000	B-
EQRE21274.104	1	11	0.095	0.000	C+	0.111	0.000	C+
EQRE21275.104	1	12	-0.022	0.369		0.055	0.026	
EQRE21109.091	2	1	-0.021	0.008		-0.017	0.038	
EQRE21116.091	2	2	0.007	0.600		0.033	0.015	
EQRE21100.091	2	3	-0.018	0.097		-0.011	0.299	
EQRE21112.091	2	4	0.020	0.049		0.015	0.131	
EQRE21104.091	2	5	0.020	0.141		0.040	0.003	
EQRE21113.091	2	6	-0.048	0.000		-0.026	0.024	
EQRE21115.091	2	7	-0.035	0.007		-0.062	0.000	B-
EQRE21105.091	2	8	0.098	0.000	C+	0.092	0.000	C+
EQRE21106.091	2	9	0.005	0.654		0.032	0.002	
EQRE21101.091	2	10	-0.024	0.006		-0.021	0.013	
EQRE21102.091	2	11	-0.073	0.002	B-	-0.085	0.000	B-
EQRE21114.091	2	12	-0.035	0.134		-0.030	0.208	
EQRE21244.101	2	17	-0.027	0.077		-0.007	0.668	
EQRE21245.101	2	18	-0.020	0.110		-0.023	0.062	
EQRE21248.101	2	19	0.095	0.000	C+	0.092	0.000	C+
EQRE21240.101	2	20	0.007	0.659		0.020	0.185	
EQRE21243.101	2	21	0.034	0.220		-0.016	0.550	
EQRE21242.101	2	22	-0.046	0.049		-0.065	0.004	B-
EQRE31959.172	2	24	-0.027	0.023		-0.015	0.191	
EQRE31958.172	2	25	-0.013	0.239		-0.018	0.107	
EQRE31961.172	2	26	-0.021	0.168		-0.044	0.004	
EQRE31963.172	2	27	0.095	0.000	C+	0.034	0.024	
EQRE31965.172	2	28	0.140	0.000	C+	0.158	0.000	C+
EQRE31964.172	2	29	-0.014	0.633		-0.042	0.137	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.28 Gender-Based DIF Statistics: Junior Reading (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQLF20592.072	1	1	-0.030	0.006		-0.021	0.049	
OQLF20591.072	1	2	-0.015	0.329		-0.017	0.260	
OQLF20594.072	1	3	0.008	0.622		0.019	0.221	
OQLF20590.072	1	4	0.001	0.949		0.007	0.600	
OQLF20595.072	1	5	-0.001	0.968		0.012	0.661	
OQLF20596.072	1	6	0.028	0.204		0.013	0.557	
OQLF20629.076	1	7	0.037	0.004		0.035	0.007	
OQLF20632.076	1	8	-0.017	0.206		-0.016	0.220	
OQLF20630.076	1	9	-0.015	0.198		-0.007	0.540	
OQLF20634.076	1	10	0.050	0.001		0.076	0.000	B+
OQLF20637.076	1	11	-0.012	0.582		-0.030	0.159	
OQLF20635.076	1	12	0.033	0.128		0.057	0.011	
OQLF20559.070	2	1	-0.005	0.746		0.001	0.937	
OQLF20565.070	2	2	-0.038	0.011		-0.029	0.047	
OQLF20555.070	2	3	0.003	0.789		0.009	0.385	
OQLF20563.070	2	4	-0.025	0.020		-0.033	0.003	
OQLF20554.070	2	5	0.029	0.029		0.021	0.103	
OQLF20558.070	2	6	0.009	0.453		0.019	0.123	
OQLF20562.070	2	7	-0.015	0.338		-0.021	0.175	
OQLF20556.070	2	8	-0.025	0.045		-0.019	0.123	
OQLF20553.070	2	9	-0.009	0.571		0.003	0.868	
OQLF20552.070	2	10	-0.042	0.000		-0.048	0.000	
OQLF20566.070	2	11	-0.004	0.869		-0.005	0.823	
OQLF20568.070	2	12	0.033	0.152		0.021	0.340	
OQLF20611.074	2	17	0.021	0.113		-0.001	0.954	
OQLF20609.074	2	18	0.011	0.367		-0.007	0.543	
OQLF20612.074	2	19	0.004	0.750		0.007	0.615	
OQLF20614.074	2	20	-0.020	0.159		-0.030	0.039	
OQLF20617.074	2	21	-0.073	0.000	B-	-0.058	0.002	
OQLF20615.074	2	22	-0.088	0.000	C-	-0.106	0.000	C-
OQLF20663.079	2	24	-0.048	0.001		-0.050	0.001	
OQLF20665.079	2	25	0.021	0.118		0.001	0.921	
OQLF20661.079	2	26	0.061	0.000	B+	0.063	0.000	B+
OQLF20660.079	2	27	0.003	0.829		0.000	0.984	
OQLF20667.079	2	28	0.155	0.000	C+	0.148	0.000	C+
OQLF20669.079	2	29	-0.025	0.282		-0.026	0.285	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.29 ELL-Based DIF Statistics: Primary Reading (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	p-value	Level of DIF	$\hat{\beta}$	p-value	Level of DIF
EQRE21002.083	1	1	0.019	0.211		0.045	0.003	
EQRE21003.083	1	2	-0.004	0.784		-0.044	0.001	
EQRE20997.083	1	3	-0.020	0.196		-0.042	0.006	
EQRE20996.083	1	4	0.030	0.039		0.015	0.297	
EQRE20994.083	1	5	-0.073	0.002	B-	-0.086	0.000	B-
EQRE20999.083	1	6	-0.004	0.879		-0.032	0.203	
EQRE20986.082	1	7	-0.005	0.709		0.010	0.458	
EQRE20992.082	1	8	0.014	0.372		0.027	0.080	
EQRE20993.082	1	9	0.031	0.026		0.016	0.251	
EQRE20990.082	1	10	0.016	0.104		0.017	0.072	
EQRE20988.082	1	11	-0.072	0.001	B-	-0.041	0.051	
EQRE20983.082	1	12	-0.002	0.945		0.021	0.401	
EQRE20817.070	2	1	-0.015	0.125		-0.001	0.932	
EQRE20833.070	2	2	0.048	0.001		0.044	0.002	
EQRE20836.070	2	3	0.009	0.568		0.012	0.417	
EQRE20816.070	2	4	0.005	0.657		-0.019	0.084	
EQRE20832.070	2	5	-0.043	0.000		-0.012	0.241	
EQRE20834.070	2	6	-0.005	0.640		0.009	0.444	
EQRE20830.070	2	7	-0.010	0.471		-0.017	0.251	
EQRE20818.070	2	8	0.003	0.785		-0.020	0.094	
EQRE20826.070	2	9	-0.042	0.008		-0.034	0.033	
EQRE20819.070	2	10	0.038	0.004		0.029	0.025	
EQRE20827.070	2	11	0.013	0.548		-0.028	0.198	
EQRE20821.070	2	12	-0.029	0.139		-0.021	0.301	
EQRE20944.078	2	17	0.036	0.021		-0.001	0.942	
EQRE20948.078	2	18	-0.032	0.002		-0.057	0.000	
EQRE20940.078	2	19	-0.030	0.043		-0.004	0.803	
EQRE20947.078	2	20	0.063	0.000	B+	0.062	0.000	B+
EQRE20939.078	2	21	0.031	0.245		0.029	0.276	
EQRE20945.078	2	22	-0.012	0.622		0.019	0.439	
EQRE21028.086	2	24	-0.025	0.056		0.002	0.886	
EQRE21031.086	2	25	-0.048	0.001		-0.043	0.005	
EQRE21035.086	2	26	-0.018	0.184		-0.035	0.009	
EQRE21027.086	2	27	-0.013	0.317		-0.010	0.431	
EQRE21034.086	2	28	0.107	0.000	C+	0.076	0.002	B+
EQRE21030.086	2	29	-0.001	0.957		0.034	0.161	

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.30 ALF/PDF-Based DIF Statistics: Primary Reading (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	p-value	Level of DIF	$\hat{\beta}$	p-value	Level of DIF
OQLF20482.063	1	1	0.007	0.720		0.005	0.783	
OQLF20484.063	1	2	0.043	0.027		0.035	0.077	
OQLF20480.063	1	3	0.033	0.085		0.034	0.078	
OQLF20485.063	1	4	-0.017	0.432		-0.001	0.952	
OQLF20486.063	1	5	-0.001	0.966		0.001	0.965	
OQLF20489.063	1	6	0.044	0.179		0.022	0.499	
OQLF20513.066	1	7	-0.004	0.841		0.031	0.167	
OQLF20516.066	1	8	0.020	0.385		-0.012	0.609	
OQLF20514.066	1	9	-0.025	0.141		-0.028	0.086	
OQLF20510.066	1	10	0.031	0.140		0.024	0.253	
OQLF20518.066	1	11	-0.117	0.015	C-	-0.091	0.064	
OQLF20519.066	1	12	0.049	0.140		0.036	0.273	
OQLF20452.061	2	1	0.002	0.933		0.059	0.002	B+
OQLF20459.061	2	2	-0.007	0.702		-0.027	0.128	
OQLF20449.061	2	3	-0.011	0.400		-0.002	0.893	
OQLF20463.061	2	4	-0.032	0.168		-0.028	0.229	
OQLF20450.061	2	5	0.020	0.230		0.010	0.573	
OQLF20464.061	2	6	0.021	0.322		0.016	0.429	
OQLF20461.061	2	7	0.027	0.155		-0.004	0.820	
OQLF20462.061	2	8	-0.001	0.963		0.016	0.488	
OQLF20454.061	2	9	0.047	0.042		0.034	0.134	
OQLF20456.061	2	10	-0.017	0.388		-0.059	0.002	B-
OQLF20466.061	2	11	0.003	0.949		-0.006	0.871	
OQLF20467.061	2	12	0.053	0.133		0.056	0.105	
OQLF20502.065	2	17	-0.008	0.728		0.018	0.388	
OQLF20504.065	2	18	-0.054	0.013		-0.069	0.001	B-
OQLF20501.065	2	19	-0.003	0.878		0.006	0.791	
OQLF20503.065	2	20	-0.001	0.953		-0.009	0.665	
OQLF20509.065	2	21	-0.143	0.001	C-	-0.068	0.104	
OQLF20506.065	2	22	-0.069	0.081		-0.081	0.043	B-
OQLF20532.068	2	24	-0.026	0.223		-0.002	0.935	
OQLF20536.068	2	25	-0.002	0.939		0.016	0.454	
OQLF20531.068	2	26	-0.022	0.354		0.003	0.907	
OQLF20534.068	2	27	0.006	0.794		-0.011	0.636	
OQLF20538.068	2	28	0.072	0.104		0.082	0.063	
OQLF20539.068	2	29	0.065	0.088		-0.017	0.652	

Note. B = moderate DIF; C = large DIF; - = favouring ALF/PDF students; + = favouring non-ALF/PDF students.

Table 7.1.31 ELL-Based DIF Statistics: Junior Reading (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQRE21162.094	1	1	0.012	0.398		-0.012	0.415	
EQRE21167.094	1	2	0.019	0.177		0.025	0.076	
EQRE21157.094	1	3	-0.010	0.345		0.007	0.493	
EQRE21158.094	1	4	0.029	0.026		0.019	0.147	
EQRE21160.094	1	5	-0.007	0.768		-0.063	0.006	B-
EQRE21165.094	1	6	0.004	0.876		0.014	0.587	
EQRE21272.104	1	7	-0.002	0.789		-0.002	0.775	
EQRE21278.104	1	8	0.003	0.825		-0.007	0.574	
EQRE21281.104	1	9	-0.068	0.000	B-	-0.045	0.004	
EQRE21271.104	1	10	0.009	0.564		0.004	0.786	
EQRE21274.104	1	11	-0.047	0.055		-0.035	0.151	
EQRE21275.104	1	12	-0.115	0.000	C-	-0.115	0.000	C-
EQRE21109.091	2	1	0.011	0.255		0.001	0.921	
EQRE21116.091	2	2	-0.050	0.000		-0.028	0.053	
EQRE21100.091	2	3	0.024	0.042		0.018	0.128	
EQRE21112.091	2	4	0.005	0.651		0.004	0.757	
EQRE21104.091	2	5	0.044	0.002		0.046	0.001	
EQRE21113.091	2	6	-0.058	0.000		-0.072	0.000	B-
EQRE21115.091	2	7	0.049	0.000		0.007	0.558	
EQRE21105.091	2	8	-0.001	0.943		-0.023	0.140	
EQRE21106.091	2	9	0.025	0.029		0.022	0.056	
EQRE21101.091	2	10	-0.003	0.775		-0.011	0.253	
EQRE21102.091	2	11	-0.024	0.339		-0.041	0.093	
EQRE21114.091	2	12	0.001	0.953		-0.027	0.265	
EQRE21244.101	2	17	0.019	0.208		0.052	0.001	
EQRE21245.101	2	18	-0.003	0.808		-0.034	0.013	
EQRE21248.101	2	19	0.101	0.000	C+	0.101	0.000	C+
EQRE21240.101	2	20	-0.001	0.924		0.004	0.784	
EQRE21243.101	2	21	0.049	0.069		0.028	0.298	
EQRE21242.101	2	22	-0.079	0.001	B-	-0.093	0.000	C-
EQRE31959.172	2	24	0.051	0.000		0.075	0.000	B+
EQRE31958.172	2	25	-0.035	0.002		-0.038	0.001	
EQRE31961.172	2	26	-0.049	0.002		-0.045	0.004	
EQRE31963.172	2	27	-0.052	0.000		-0.032	0.041	
EQRE31965.172	2	28	-0.069	0.012	B-	-0.029	0.303	
EQRE31964.172	2	29	0.274	0.000	C+	0.323	0.000	C+

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.32 ALF/PDF-Based DIF Statistics: Junior Reading (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	p-value	Level of DIF	$\hat{\beta}$	p-value	Level of DIF
OQLF20592.072	1	1	-0.004	0.877		-0.018	0.423	
OQLF20591.072	1	2	-0.012	0.688		0.030	0.332	
OQLF20594.072	1	3	0.013	0.663		-0.014	0.649	
OQLF20590.072	1	4	0.053	0.064		0.034	0.240	
OQLF20595.072	1	5	-0.041	0.449		-0.056	0.305	
OQLF20596.072	1	6	0.120	0.005	C+	0.091	0.040	C+
OQLF20629.076	1	7	0.013	0.614		-0.007	0.789	
OQLF20632.076	1	8	0.001	0.976		0.033	0.222	
OQLF20630.076	1	9	-0.041	0.081		-0.025	0.283	
OQLF20634.076	1	10	-0.037	0.223		0.014	0.630	
OQLF20637.076	1	11	-0.026	0.523		0.020	0.615	
OQLF20635.076	1	12	-0.018	0.662		0.068	0.097	
OQLF20559.070	2	1	0.019	0.545		0.053	0.076	
OQLF20565.070	2	2	0.011	0.713		-0.009	0.762	
OQLF20555.070	2	3	-0.016	0.433		-0.011	0.594	
OQLF20563.070	2	4	0.046	0.052		0.050	0.033	
OQLF20554.070	2	5	-0.025	0.357		-0.018	0.484	
OQLF20558.070	2	6	0.018	0.499		0.025	0.329	
OQLF20562.070	2	7	-0.095	0.003	C-	-0.094	0.003	C-
OQLF20556.070	2	8	0.019	0.448		0.000	0.987	
OQLF20553.070	2	9	0.012	0.703		0.033	0.266	
OQLF20552.070	2	10	-0.020	0.379		-0.058	0.014	
OQLF20566.070	2	11	0.015	0.751		-0.073	0.128	
OQLF20568.070	2	12	-0.005	0.914		-0.031	0.493	
OQLF20611.074	2	17	0.010	0.724		0.010	0.706	
OQLF20609.074	2	18	0.014	0.571		-0.020	0.414	
OQLF20612.074	2	19	0.002	0.940		0.003	0.909	
OQLF20614.074	2	20	-0.041	0.151		-0.051	0.079	
OQLF20617.074	2	21	-0.009	0.811		0.015	0.690	
OQLF20615.074	2	22	0.008	0.846		-0.007	0.848	
OQLF20663.079	2	24	0.033	0.249		-0.032	0.271	
OQLF20665.079	2	25	-0.020	0.444		-0.014	0.576	
OQLF20661.079	2	26	-0.040	0.162		-0.026	0.384	
OQLF20660.079	2	27	0.033	0.217		0.026	0.337	
OQLF20667.079	2	28	-0.049	0.344		-0.041	0.433	
OQLF20669.079	2	29	0.066	0.139		0.018	0.697	

Note. B = moderate DIF; C = large DIF; - = favouring ALF/PDF students; + = favouring non-ALF/PDF students.

Table 7.1.33 Gender-Based DIF Statistics: Primary Writing (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQWE20250T	1	13	-0.030	0.298		-0.043	0.138	
EQWE20250V	1	13	0.005	0.854		0.026	0.287	
EQWE20245T	1	14	0.074	0.032	B+	0.056	0.094	
EQWE20245V	1	14	0.040	0.145		0.019	0.476	
EQWE20233	2	13	0.007	0.728		0.049	0.017	
EQWE20205	2	14	0.096	0.000	B+	0.071	0.001	B+
EQWE20206	2	16	0.036	0.159		-0.051	0.048	
EQWE20224	2	16	-0.019	0.425		-0.003	0.893	
EQWE30434T	2	23	-0.025	0.468		-0.024	0.492	
EQWE30434V	2	23	-0.021	0.400		0.007	0.775	
EQWE20215	2	30	0.027	0.169		0.010	0.604	
EQWE20230	2	31	-0.067	0.005		0.028	0.231	
EQWE20218	2	32	0.028	0.271		0.052	0.043	
EQWE20222	2	33	-0.016	0.463		0.052	0.018	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.34 Gender-Based DIF Statistics: Primary Writing (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQEF20220T	1	13	0.063	0.054	B+	0.072	0.026	B+
OQEF20220V	1	13	0.036	0.117		0.026	0.276	
OQEF20226T	1	14	0.012	0.669		0.064	0.022	B+
OQEF20226V	1	14	0.024	0.304		0.030	0.205	
OQEF20195	2	13	-0.003	0.907		-0.033	0.179	
OQEF20202	2	14	-0.003	0.85		-0.038	0.026	
OQEF20192	2	16	-0.008	0.668		-0.027	0.13	
OQEF20217	2	16	-0.027	0.248		0.025	0.275	
OQEF20230T	2	23	-0.069	0.034	B-	-0.058	0.092	B-
OQEF20230V	2	23	-0.028	0.246		-0.113	0.000	
OQEF10026	2	30	0.016	0.254		0.027	0.064	
OQEF20199	2	31	0.066	0.003	B+	0.053	0.019	
OQEF20212	2	32	-0.025	0.224		-0.010	0.634	
OQEF20207	2	33	0.029	0.211		0.068	0.003	B+

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.35 Gender-Based DIF Statistics: Junior Writing (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQWE20315T	1	13	-0.031	0.319		0.053	0.092	
EQWE20315V	1	13	0.029	0.213		0.005	0.815	
EQWE20304T	1	14	0.121	0.000	C+	0.062	0.071	B+
EQWE20304V	1	14	0.026	0.295		-0.022	0.376	
EQWE20272	2	13	0.001	0.952		0.059	0.000	B+
EQWE20291	2	14	0.031	0.221		0.077	0.002	B+
EQWE20271	2	16	0.088	0.000	B+	-0.012	0.607	
EQWE20278	2	16	0.006	0.767		0.024	0.272	
EQWE20322T	2	23	-0.050	0.122		-0.027	0.394	
EQWE20322V	2	23	-0.015	0.585		0.002	0.954	
EQWE20280	2	30	-0.038	0.044		-0.01	0.605	
EQWE20276	2	31	0.025	0.305		0.049	0.043	
EQWE20285	2	32	-0.019	0.319		0.022	0.257	
EQWE20290	2	33	-0.031	0.206		0.011	0.657	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.36 Gender-Based DIF Statistics: Junior Writing (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQEF20268T	1	13	0.054	0.055		0.004	0.882	
OQEF20268V	1	13	0.030	0.167		0.013	0.571	
OQEF20278T	1	14	0.100	0.001	C+	0.061	0.046	B+
OQEF20278V	1	14	0.064	0.004	B+	0.069	0.003	B+
OQEF20241	2	13	-0.060	0.010	B-	-0.050	0.028	
OQEF20248	2	14	-0.008	0.711		-0.022	0.298	
OQEF20243	2	16	0.022	0.326		0.103	0.000	C+
OQEF20266	2	16	0.004	0.870		0.040	0.120	
OQEF30611T	2	23	-0.093	0.002	B-	-0.023	0.463	
OQEF30611V	2	23	-0.082	0.001	B-	-0.069	0.007	B-
OQEF20264	2	30	0.046	0.043		0.043	0.055	
OQEF20246	2	31	0.035	0.070		0.026	0.191	
OQEF20245	2	32	0.044	0.035		0.033	0.110	
OQEF20260	2	33	0.080	0.001	B+	0.064	0.006	B+

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.37 ELL-Based DIF Statistics: Primary Writing (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQWE20250T	1	13	0.041	0.154		-0.004	0.897	
EQWE20250V	1	13	-0.079	0.001	B-	-0.024	0.342	
EQWE20245T	1	14	0.092	0.006	C+	0.104	0.002	C+
EQWE20245V	1	14	0.004	0.865		0.015	0.581	
EQWE20233	2	13	0.081	0.000	B+	0.115	0.000	C+
EQWE20205	2	14	0.024	0.258		-0.039	0.077	
EQWE20206	2	16	-0.025	0.338		-0.009	0.717	
EQWE20224	2	16	-0.044	0.061		-0.075	0.001	
EQWE30434T	2	23	-0.030	0.355		-0.045	0.175	
EQWE30434V	2	23	-0.044	0.065		-0.04	0.103	
EQWE20215	2	30	0.045	0.030		0.063	0.002	
EQWE20230	2	31	-0.077	0.001	B-	-0.022	0.359	
EQWE20218	2	32	0.069	0.007	B+	0.028	0.284	
EQWE20222	2	33	-0.104	0.000	C-	-0.071	0.001	B-

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.38 ALF/PDF-Based DIF Statistics: Primary Writing (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQEF20220T	1	13	0.034	0.296		0.052	0.034	
OQEF20220V	1	13	-0.005	0.828		-0.008	0.026	
OQEF20226T	1	14	0.001	0.967		-0.017	0.030	
OQEF20226V	1	14	-0.035	0.148		-0.038	0.025	
OQEF20195	2	13	-0.022	0.400		-0.038	0.026	
OQEF20202	2	14	-0.037	0.061		-0.015	0.020	
OQEF20192	2	16	0.014	0.525		0.034	0.021	
OQEF20217	2	16	-0.024	0.322		-0.007	0.024	
OQEF20230T	2	23	0.003	0.940		-0.003	0.036	
OQEF20230V	2	23	-0.007	0.779		-0.008	0.027	
OQEF10026	2	30	0.002	0.922		0.013	0.017	
OQEF20199	2	31	-0.035	0.140		-0.055	0.024	
OQEF20212	2	32	0.020	0.380		0.021	0.023	
OQEF20207	2	33	0.037	0.143		0.026	0.025	

Note. B = moderate DIF; C = large DIF; - = favouring ALF/PDF students; + = favouring non-ALF/PDF students.

Table 7.1.39 ELL-Based DIF Statistics: Junior Writing (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	p-value	Level of DIF	$\hat{\beta}$	p-value	Level of DIF
EQWE20315T	1	13	-0.042	0.167		-0.005	0.863	
EQWE20315V	1	13	-0.022	0.325		-0.033	0.130	
EQWE20304T	1	14	-0.074	0.019		-0.011	0.737	
EQWE20304V	1	14	0.020	0.428		0.016	0.531	
EQWE20272	2	13	0.040	0.019		0.021	0.226	
EQWE20291	2	14	0.048	0.044		0.007	0.766	
EQWE20271	2	16	-0.018	0.451		0.021	0.393	
EQWE20278	2	16	-0.038	0.097		-0.024	0.293	
EQWE20322T	2	23	0.048	0.128		-0.021	0.504	
EQWE20322V	2	23	0.046	0.082		0.044	0.087	
EQWE20280	2	30	-0.035	0.058		-0.077	0.000	
EQWE20276	2	31	0.024	0.330		0.07	0.004	
EQWE20285	2	32	-0.029	0.135		-0.006	0.752	
EQWE20290	2	33	-0.011	0.641		-0.048	0.050	

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.40 ALF/PDF-Based DIF Statistics: Junior Writing (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	p-value	Level of DIF	$\hat{\beta}$	p-value	Level of DIF
OQEF20268T	1	13	0.026	0.427		0.036	0.269	
OQEF20268V	1	13	-0.005	0.858		-0.005	0.847	
OQEF20278T	1	14	-0.015	0.664		-0.035	0.303	
OQEF20278V	1	14	-0.004	0.871		-0.007	0.783	
OQEF20241	2	13	0.052	0.055		0.033	0.223	
OQEF20248	2	14	-0.100	0.000	C-	-0.092	0.000	C-
OQEF20243	2	16	0.047	0.083		0.041	0.129	
OQEF20266	2	16	0.014	0.645		-0.046	0.128	
OQEF30611T	2	23	-0.016	0.660		0.007	0.836	
OQEF30611V	2	23	0.008	0.791		0.008	0.781	
OQEF20264	2	30	0.012	0.667		0.039	0.140	
OQEF20246	2	31	0.011	0.634		0.015	0.526	
OQEF20245	2	32	0.003	0.911		0.016	0.515	
OQEF20260	2	33	-0.063	0.023	B-	-0.044	0.106	

Note. B = moderate DIF; C = large DIF; - = favouring ALF/PDF students; + = favouring non-ALF/PDF students.

Table 7.1.41 Gender-Based DIF Statistics: Primary Mathematics (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME10073	1	0.029	0.003		0.033	0.001	
EQME10167	2	-0.048	0.000		-0.048	0.000	
EQME10038	3	0.004	0.759		-0.027	0.061	
EQME10062	4	-0.019	0.198		-0.008	0.589	
EQME10042	5	0.042	0.003		0.006	0.686	
EQME20936	6	0.010	0.504		0.003	0.830	
EQME20944	7	0.005	0.883		-0.010	0.759	
EQME20880	8	0.060	0.072		0.109	0.001	C+
EQME20811	9	-0.065	0.012		-0.017	0.510	
EQME20822	10	-0.080	0.026		-0.057	0.118	
EQME10126	11	0.042	0.003		0.050	0.000	
EQME20894	12	0.010	0.343		0.025	0.017	
EQME10115	13	0.031	0.004		0.052	0.000	
EQME20883	14	0.035	0.022		0.055	0.000	
EQME20859	15	-0.066	0.000		-0.051	0.001	
EQME10019	16	-0.014	0.273		-0.001	0.925	
EQME20874	17	0.022	0.145		0.052	0.001	
EQME10079	18	0.105	0.000	C+	0.085	0.000	B+
EQME20863	19	-0.053	0.000		-0.036	0.000	
EQME10166	20	-0.004	0.773		-0.007	0.597	
EQME20915	21	-0.013	0.404		-0.033	0.028	
EQME10058	22	-0.009	0.481		-0.017	0.176	
EQME10046	23	-0.037	0.013		-0.040	0.007	
EQME20832	24	0.021	0.140		0.027	0.065	
EQME20833	25	0.060	0.000	B+	0.066	0.000	B+
EQME10051	26	-0.008	0.548		-0.022	0.089	
EQME10170	27	-0.064	0.035		-0.116	0.000	C-
EQME20825	28	-0.018	0.575		-0.081	0.016	
EQME20878	29	0.048	0.138		0.069	0.036	B+
EQME20941	30	0.119	0.000	C+	0.081	0.015	B+
EQME10084	31	-0.033	0.022		-0.031	0.037	
EQME10108	32	-0.042	0.000		-0.042	0.000	
EQME20893	33	-0.025	0.082		-0.033	0.022	
EQME20828	34	-0.026	0.077		-0.017	0.241	
EQME10029	35	-0.003	0.852		-0.024	0.095	
EQME20871	36	-0.046	0.002		-0.036	0.017	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.42 Gender-Based DIF Statistics: Primary Mathematics (French)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQMF10064	1	0.021	0.052		0.012	0.253	
OQMF20541	2	-0.014	0.310		-0.003	0.809	
OQMF20605	3	-0.049	0.000		-0.027	0.053	
OQMF20560	4	-0.023	0.081		-0.018	0.180	
OQMF20616	5	0.006	0.685		0.015	0.295	
OQMF20540	6	0.064	0.000	B+	0.071	0.000	B+
OQMF20510	7	0.014	0.683		0.025	0.473	
OQMF20640	8	0.002	0.960		0.002	0.939	
OQMF20513	9	-0.040	0.229		-0.049	0.140	
OQMF10044	10	0.197	0.000	C+	0.232	0.000	C+
OQMF20576	11	-0.041	0.004		-0.045	0.002	
OQMF20590	12	-0.047	0.001		-0.032	0.025	
OQMF20546	13	0.054	0.000		0.054	0.000	
OQMF20571	14	-0.047	0.000		-0.051	0.000	
OQMF20628	15	-0.057	0.000		-0.068	0.000	
OQMF20583	16	0.060	0.000	B+	0.052	0.000	
OQMF20630	17	0.043	0.002		0.039	0.006	
OQMF20635	18	-0.044	0.001		-0.045	0.001	
OQMF10084	19	0.003	0.834		0.006	0.619	
OQMF20572	20	0.129	0.000	C+	0.117	0.000	C+
OQMF20587	21	-0.022	0.081		-0.043	0.001	
OQMF20566	22	-0.004	0.768		0.006	0.665	
OQMF20550	23	0.080	0.000	B+	0.060	0.000	B+
OQMF20611	24	-0.005	0.757		0.009	0.573	
OQMF10067	25	-0.016	0.064		-0.014	0.098	
OQMF20593	26	-0.022	0.090		-0.010	0.459	
OQMF20639	27	-0.052	0.051		-0.073	0.006	
OQMF20504	28	-0.174	0.000	C-	-0.150	0.000	C-
OQMF20526	29	-0.006	0.829		-0.006	0.848	
OQMF20636	30	0.038	0.088		0.019	0.401	
OQMF20597	31	-0.038	0.008		-0.035	0.015	
OQMF20556	32	-0.071	0.000		-0.086	0.000	
OQMF20622	33	-0.062	0.000		-0.057	0.000	
OQMF20600	34	-0.036	0.006		-0.056	0.000	
OQMF20564	35	0.102	0.000	C+	0.107	0.000	C+
OQMF20606	36	-0.017	0.194		-0.003	0.796	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.43 Gender-Based DIF Statistics: Junior Mathematics (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
3459716	1	0.022	0.068		0.026	0.033	
EQME21115	2	-0.096	0.000	C-	-0.087	0.000	
EQME10180	3	-0.026	0.066		-0.027	0.057	
EQME10294	4	-0.017	0.238		-0.014	0.328	
EQME21102	5	0.010	0.287		-0.008	0.416	
EQME10193	6	-0.026	0.071		0.007	0.608	
EQME21089	7	-0.029	0.394		-0.008	0.820	
EQME21119	8	0.025	0.444		0.035	0.272	
EQME21138	9	-0.023	0.456		0.006	0.834	
EQME21043	10	0.037	0.237		-0.001	0.981	
EQME21140	11	0.032	0.029		0.020	0.180	
EQME10211	12	0.020	0.189		0.044	0.003	
EQME21031	13	-0.020	0.177		-0.041	0.004	
EQME10282	14	0.037	0.000		0.017	0.128	
EQME21070	15	0.007	0.634		0.013	0.351	
EQME21021	16	0.048	0.002		0.063	0.000	B+
EQME21063	17	-0.009	0.484		-0.041	0.003	
EQME21080	18	0.012	0.417		0.008	0.581	
EQME21130	19	-0.043	0.006		-0.021	0.162	
EQME21083	20	0.039	0.010		0.026	0.094	
EQME10185	21	0.008	0.541		-0.027	0.042	
EQME21142	22	-0.041	0.006		-0.070	0.000	
EQME10237	23	0.099	0.000	C+	0.099	0.000	C+
EQME21008	24	-0.007	0.564		-0.034	0.008	
EQME21133	25	-0.028	0.036		-0.036	0.007	
EQME21060	26	-0.018	0.231		0.005	0.748	
EQME21104	27	-0.059	0.028		-0.067	0.012	
EQME10229	28	-0.015	0.643		-0.023	0.472	
EQME21137	29	0.051	0.080		0.020	0.515	
EQME21135	30	0.042	0.186		0.063	0.051	
EQME21048	31	0.087	0.000	B+	0.092	0.000	C+
EQME21099	32	-0.040	0.002		-0.029	0.032	
EQME10256	33	0.000	0.977		0.011	0.455	
EQME10259	34	-0.030	0.044		-0.030	0.040	
EQME21062	35	0.028	0.044		0.049	0.001	
EQME21139	36	-0.030	0.039		-0.058	0.000	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.44 Gender-Based DIF Statistics: Junior Mathematics (French)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQMF20716	1	-0.041	0.001		-0.043	0.000	
OQMF20745	2	-0.008	0.574		0.005	0.712	
OQMF20816	3	-0.044	0.002		-0.037	0.012	
OQMF10189	4	0.044	0.001		0.033	0.013	
OQMF20703	5	0.092	0.000	C+	0.097	0.000	C+
OQMF20723	6	-0.031	0.035		-0.027	0.067	
OQMF10131	7	0.087	0.007	B+	0.057	0.075	
OQMF20796	8	-0.005	0.875		-0.029	0.399	
OQMF20718	9	0.027	0.376		0.034	0.272	
OQMF20844	10	0.056	0.072		0.053	0.084	
OQMF20832	11	0.058	0.000		0.048	0.002	
OQMF10139	12	-0.024	0.034		-0.019	0.083	
OQMF20778	13	-0.008	0.585		0.006	0.683	
OQMF10162	14	0.006	0.688		-0.001	0.935	
OQMF20827	15	-0.022	0.070		-0.010	0.412	
OQMF20802	16	-0.005	0.729		-0.004	0.807	
OQMF20751	17	0.127	0.000	C+	0.110	0.000	C+
OQMF20729	18	-0.030	0.005		-0.031	0.005	
OQMF20701	19	-0.002	0.878		-0.006	0.634	
OQMF20753	20	0.013	0.379		0.018	0.208	
OQMF20814	21	-0.039	0.012		-0.042	0.006	
OQMF10137	22	-0.029	0.008		-0.033	0.004	
OQMF20829	23	0.065	0.000	B+	0.053	0.000	
OQMF10125	24	-0.025	0.087		-0.043	0.003	
OQMF10115	25	-0.040	0.004		-0.042	0.002	
OQMF10149	26	0.003	0.851		-0.010	0.509	
OQMF20710	27	-0.022	0.518		0.000	1.000	
OQMF20747	28	0.056	0.063		0.029	0.325	
OQMF10203	29	-0.129	0.000	C-	-0.140	0.000	C-
OQMF10127	30	-0.082	0.039		-0.027	0.497	
OQMF20790	31	0.004	0.705		-0.003	0.804	
OQMF10156	32	-0.011	0.440		0.015	0.301	
OQMF20760	33	-0.006	0.716		-0.013	0.397	
OQMF20734	34	-0.017	0.256		-0.014	0.355	
OQMF20725	35	-0.008	0.606		-0.002	0.904	
OQMF20773	36	-0.035	0.002		-0.044	0.000	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.45 ELL-Based DIF Statistics: Primary Mathematics (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME10073	1	0.001	0.920		0.009	0.387	
EQME10167	2	-0.032	0.018		-0.044	0.001	
EQME10038	3	-0.033	0.023		-0.033	0.025	
EQME10062	4	-0.052	0.001		-0.054	0.000	
EQME10042	5	0.000	0.974		-0.033	0.022	
EQME20936	6	0.025	0.098		0.027	0.070	
EQME20944	7	-0.037	0.232		-0.036	0.238	
EQME20880	8	-0.072	0.028		-0.008	0.803	
EQME20811	9	0.026	0.329		-0.009	0.732	
EQME20822	10	0.038	0.298		-0.012	0.752	
EQME10126	11	0.053	0.000		0.034	0.020	
EQME20894	12	-0.028	0.007		-0.051	0.000	
EQME10115	13	-0.039	0.000		-0.043	0.000	
EQME20883	14	0.027	0.079		0.076	0.000	B+
EQME20859	15	-0.003	0.862		0.010	0.511	
EQME10019	16	0.005	0.682		-0.014	0.276	
EQME20874	17	0.065	0.000	B+	0.058	0.000	
EQME10079	18	-0.015	0.275		-0.012	0.388	
EQME20863	19	-0.009	0.378		0.020	0.073	
EQME10166	20	-0.020	0.126		-0.027	0.042	
EQME20915	21	-0.039	0.008		-0.058	0.000	
EQME10058	22	-0.026	0.040		-0.043	0.001	
EQME10046	23	0.006	0.691		0.005	0.727	
EQME20832	24	0.120	0.000	C+	0.117	0.000	C+
EQME20833	25	0.015	0.310		0.027	0.063	
EQME10051	26	0.030	0.023		0.019	0.172	
EQME10170	27	-0.092	0.004	C-	-0.149	0.000	C-
EQME20825	28	0.030	0.356		0.115	0.001	C+
EQME20878	29	0.096	0.003	C+	0.029	0.387	
EQME20941	30	-0.005	0.878		0.049	0.146	
EQME10084	31	-0.022	0.131		-0.045	0.003	
EQME10108	32	-0.021	0.062		-0.033	0.004	
EQME20893	33	0.045	0.001		0.023	0.110	
EQME20828	34	-0.018	0.232		0.005	0.723	
EQME10029	35	-0.002	0.885		0.006	0.702	
EQME20871	36	0.014	0.333		-0.001	0.944	

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.46 ALF/PDF-Based DIF Statistics: Primary Mathematics (French)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQMF10064	1	-0.004	0.797		-0.009	0.597	
OQMF20541	2	0.025	0.245		0.021	0.318	
OQMF20605	3	-0.029	0.179		-0.007	0.750	
OQMF20560	4	-0.016	0.447		-0.003	0.883	
OQMF20616	5	0.032	0.143		0.034	0.125	
OQMF20540	6	-0.021	0.264		-0.014	0.457	
OQMF20510	7	0.132	0.014	C+	0.185	0.001	C+
OQMF20640	8	-0.095	0.047	C-	-0.083	0.085	
OQMF20513	9	-0.013	0.802		0.006	0.901	
OQMF10044	10	-0.079	0.116		-0.158	0.002	C-
OQMF20576	11	0.053	0.014		0.033	0.133	
OQMF20590	12	0.003	0.876		0.001	0.954	
OQMF20546	13	-0.015	0.504		-0.020	0.354	
OQMF20571	14	0.001	0.936		0.006	0.690	
OQMF20628	15	-0.032	0.140		-0.016	0.441	
OQMF20583	16	-0.014	0.541		-0.010	0.647	
OQMF20630	17	-0.020	0.343		-0.012	0.586	
OQMF20635	18	0.005	0.824		0.024	0.239	
OQMF10084	19	0.015	0.434		-0.010	0.600	
OQMF20572	20	-0.028	0.213		-0.049	0.030	
OQMF20587	21	0.008	0.674		0.014	0.493	
OQMF20566	22	-0.012	0.606		0.013	0.576	
OQMF20550	23	-0.029	0.166		-0.027	0.199	
OQMF20611	24	0.034	0.148		0.024	0.305	
OQMF10067	25	0.023	0.091		0.023	0.081	
OQMF20593	26	0.023	0.271		0.036	0.077	
OQMF20639	27	-0.070	0.101		-0.059	0.158	
OQMF20504	28	0.140	0.007	C+	0.114	0.027	C+
OQMF20526	29	-0.031	0.508		-0.002	0.957	
OQMF20636	30	-0.024	0.480		-0.003	0.933	
OQMF20597	31	0.048	0.029		0.045	0.046	
OQMF20556	32	-0.032	0.176		-0.043	0.072	
OQMF20622	33	0.044	0.009		0.034	0.051	
OQMF20600	34	0.004	0.849		-0.007	0.740	
OQMF20564	35	0.026	0.179		0.013	0.523	
OQMF20606	36	-0.046	0.023		-0.021	0.276	

Note. B = moderate DIF; C = large DIF; - = favouring ALF/PDF students; + = favouring non-ALF/PDF students.

Table 7.1.47 ELL-Based DIF Statistics: Junior Mathematics (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
3459716	1	-0.015	0.230		-0.015	0.218	
EQME21115	2	-0.037	0.008		-0.019	0.170	
EQME10180	3	-0.038	0.007		-0.017	0.228	
EQME10294	4	-0.007	0.606		-0.015	0.303	
EQME21102	5	-0.003	0.736		-0.007	0.503	
EQME10193	6	-0.009	0.516		-0.010	0.489	
EQME21089	7	0.020	0.562		0.005	0.883	
EQME21119	8	-0.004	0.897		0.052	0.093	
EQME21138	9	-0.007	0.806		0.022	0.457	
EQME21043	10	0.055	0.082		0.011	0.731	
EQME21140	11	-0.007	0.634		0.015	0.297	
EQME10211	12	0.013	0.408		0.038	0.012	
EQME21031	13	0.039	0.007		0.055	0.000	
EQME10282	14	0.003	0.784		0.014	0.219	
EQME21070	15	-0.015	0.300		0.007	0.620	
EQME21021	16	0.013	0.393		0.012	0.437	
EQME21063	17	0.016	0.254		0.038	0.006	
EQME21080	18	-0.011	0.455		-0.036	0.015	
EQME21130	19	-0.041	0.007		-0.034	0.027	
EQME21083	20	0.001	0.932		-0.044	0.004	
EQME10185	21	-0.013	0.331		0.007	0.590	
EQME21142	22	0.035	0.018		0.015	0.296	
EQME10237	23	0.043	0.003		0.029	0.051	
EQME21008	24	0.020	0.127		-0.008	0.504	
EQME21133	25	0.004	0.761		-0.004	0.767	
EQME21060	26	0.025	0.087		0.011	0.440	
EQME21104	27	0.049	0.068		0.026	0.334	
EQME10229	28	-0.151	0.000	C-	-0.067	0.037	
EQME21137	29	0.014	0.643		-0.028	0.346	
EQME21135	30	0.001	0.979		-0.016	0.612	
EQME21048	31	0.035	0.022		0.037	0.016	
EQME21099	32	-0.057	0.000		-0.029	0.026	
EQME10256	33	0.028	0.051		0.047	0.001	
EQME10259	34	-0.029	0.046		-0.024	0.098	
EQME21062	35	0.014	0.346		0.008	0.572	
EQME21139	36	-0.046	0.001		-0.073	0.000	

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.48 ALF/PDF-Based DIF Statistics: Junior Mathematics (French)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQMF20716	1	0.040	0.092		0.052	0.028	
OQMF20745	2	0.005	0.841		-0.041	0.132	
OQMF20816	3	-0.074	0.008		-0.088	0.002	
OQMF10189	4	-0.013	0.601		-0.015	0.557	
OQMF20703	5	0.002	0.950		0.038	0.156	
OQMF20723	6	0.001	0.963		0.028	0.330	
OQMF10131	7	-0.048	0.419		-0.069	0.252	
OQMF20796	8	-0.001	0.988		-0.061	0.358	
OQMF20718	9	-0.077	0.183		-0.070	0.223	
OQMF20844	10	0.066	0.257		0.089	0.144	
OQMF20832	11	-0.047	0.112		-0.033	0.273	
OQMF10139	12	0.008	0.703		0.006	0.801	
OQMF20778	13	-0.062	0.018		-0.057	0.034	
OQMF10162	14	0.014	0.636		0.005	0.858	
OQMF20827	15	0.004	0.860		0.024	0.270	
OQMF20802	16	-0.060	0.027		-0.046	0.102	
OQMF20751	17	-0.030	0.263		-0.034	0.196	
OQMF20729	18	0.025	0.237		0.011	0.629	
OQMF20701	19	-0.098	0.000	C-	-0.038	0.124	
OQMF20753	20	-0.021	0.449		0.011	0.692	
OQMF20814	21	0.023	0.442		0.093	0.002	C+
OQMF10137	22	0.021	0.341		-0.005	0.809	
OQMF20829	23	0.000	0.986		0.005	0.859	
OQMF10125	24	0.002	0.934		-0.011	0.707	
OQMF10115	25	0.048	0.071		0.022	0.422	
OQMF10149	26	0.065	0.024	B+	0.034	0.252	
OQMF20710	27	-0.254	0.000	C-	-0.253	0.000	C-
OQMF20747	28	-0.049	0.399		-0.073	0.210	
OQMF10203	29	0.268	0.000	C+	0.283	0.000	C+
OQMF10127	30	0.124	0.107		0.140	0.072	
OQMF20790	31	0.045	0.041		0.004	0.851	
OQMF10156	32	0.002	0.935		0.018	0.519	
OQMF20760	33	0.039	0.188		-0.002	0.939	
OQMF20734	34	-0.046	0.102		-0.066	0.021	
OQMF20725	35	0.061	0.039	B+	0.060	0.045	B+
OQMF20773	36	-0.038	0.084		-0.010	0.631	

Note. B = moderate DIF; C = large DIF; - = favouring ALF/PDF students; + = favouring non-ALF/PDF students.

Grade 9 Assessment of Mathematics

Classical Item Statistics and IRT Item Parameters

Table 7.1.49 Item Statistics: Grade 9 Applied Mathematics, Winter Administration (English)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQME10578	1	NA1.03	PS	N	1	32.71	0.19	1.88	0.62
EQME10621	2	NA1.04	AP	N	3	57.21	0.22	0.20	0.39
EQME10667	3	NA2.02	AP	N	4	38.25	0.27	1.19	0.74
EQME10607	4	NA2.05	KU	N	3	50.58	0.26	0.56	0.52
EQME10425	5	NA2.07	KU	N	4	46.37	0.28	0.76	0.63
EQME21429	6	NA2.08	AP	N	4*	58.00 (2.32)	0.52	-0.63	0.32
EQME10616	7	LR1.01	AP	R	2	65.30	0.26	-0.48	0.43
EQME10533	8	LR2.01	KU	R	3	69.72	0.32	-0.70	0.59
EQME10534	9	LR2.03	AP	R	4	70.87	0.29	-0.74	0.53
EQME10536	10	LR3.01	KU	R	2	84.11	0.31	-1.55	0.69
EQME21407	11	LR3.02	PS	R	4	53.10	0.38	0.30	0.84
EQME10666	12	LR4.01	KU	R	4	73.07	0.32	-0.80	0.63
EQME10635	13	LR4.02	PS	R	2	66.94	0.22	-0.68	0.37
EQME10537	14	LR4.03	PS	R	3	65.64	0.19	-0.63	0.35
EQME21498	15	LR3.01	AP	R	4*	45.50 (1.82)	0.51	-0.27	0.39
EQME21472	16	LR4.06	PS	R	4*	60.25 (2.41)	0.53	-0.85	0.43
EQME21402	17	MG2.02	PS	M	3	63.16	0.15	-0.44	0.24
EQME10582	18	MG2.05	PS	M	4	27.91	0.25	1.72	1.07
EQME10629	19	MG2.05	AP	M	1	29.26	0.18	1.88	0.83
EQME10544	20	MG3.01	AP	M	3	61.75	0.33	-0.13	0.66
EQME21492	21	MG3.02	AP	M	4*	39.75 (1.59)	0.44	0.07	0.29
EQME10549	NR	LR3.04	AP	R	1	44.14	0.26	0.97	0.61
EQME10522	NR	MG3.01	KU	M	1	85.40	0.25	-1.94	0.54
EQME21452	NR	NA1.06	PS	N	2	48.76	0.38	0.49	0.95
EQME21453	NR	LR1.04	AP	R	4	69.29	0.31	-0.61	0.58
EQME10505	NR	MG1.03	KU	M	1	64.68	0.32	-0.31	0.59
EQME10599	NR	LR4.03	AP	R	4	42.08	0.31	0.93	0.78
EQME10515	NR	NA1.03	KU	N	2	65.10	0.28	-0.38	0.50
EQME10643	NR	NA1.05	PS	N	4*	70.75 (2.83)	0.45	-1.67	0.33
EQME21413	NR	LR2.03	AP	R	4*	69.25 (2.77)	0.36	-2.22	0.24
EQME21454	NR	MG2.03	PS	M	4*	46.75 (1.87)	0.53	-0.18	0.45

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

AP = application; KU = knowledge and understanding; PS = problem solving; M = measurement and geometry;

N = number sense and algebra; R = linear relations; NR = not released.

**Table 7.1.50 Distribution of Score Points and Step Parameters for Open-Response Items:
Grade 9 Applied Mathematics, Winter Administration (English)**

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
EQME10643	NR	% of Students	1.81	0.36	17.46	14.41	27.13	38.83
		Parameters			-4.74	-0.25	-1.30	-0.39
EQME21413	NR	% of Students	1.38	0.14	18.52	15.34	30.32	34.30
		Parameters			-7.00	0.00	-1.81	-0.08
EQME21429	6	% of Students	12.96	2.15	28.53	8.76	4.43	43.17
		Parameters			-1.87	1.88	1.31	-3.85
EQME21454	NR	% of Students	5.68	1.18	40.18	25.66	14.10	13.20
		Parameters			-3.06	0.39	1.09	0.86
EQME21472	16	% of Students	2.64	1.27	26.82	17.87	27.15	24.24
		Parameters			-3.59	0.09	-0.56	0.64
EQME21492	21	% of Students	13.61	1.42	51.60	9.77	6.76	16.84
		Parameters			-2.94	3.32	1.07	-1.17
EQME21498	15	% of Students	6.30	1.70	47.96	17.08	8.08	18.88
		Parameters			-3.32	1.39	1.41	-0.56

Note. The total number of students is 18 497.

NR = not released.

Table 7.1.51 Item Statistics: Grade 9 Applied Mathematics, Spring Administration (English)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQME10526	1	NA1.05	PS	N	2	51.95	0.30	0.52	0.60
EQME10667	2	NA2.02	AP	N	4	39.46	0.30	1.19	0.74
EQME10607	3	NA2.05	KU	N	3	52.84	0.26	0.56	0.52
EQME10425	4	NA2.07	KU	N	4	48.09	0.29	0.76	0.63
EQME21426	5	NA1.05	PS	N	4*	50.50 (2.02)	0.53	-0.27	0.39
EQME10616	6	LR1.01	AP	R	2	67.13	0.24	-0.48	0.43
13641	7	LR1.04	AP	R	3	74.66	0.21	-1.37	0.35
EQME10533	8	LR2.01	KU	R	3	72.87	0.31	-0.70	0.59
EQME10534	9	LR2.03	AP	R	4	71.19	0.30	-0.74	0.53
EQME10536	10	LR3.01	KU	R	2	85.70	0.31	-1.55	0.69
EQME10604	11	LR3.04	AP	R	1	29.27	0.22	1.88	0.80
EQME10666	12	LR4.01	KU	R	4	73.85	0.32	-0.80	0.63
EQME10635	13	LR4.02	PS	R	2	68.65	0.22	-0.68	0.37
EQME10537	14	LR4.03	PS	R	3	67.78	0.22	-0.63	0.35
EQME10631	15	LR2.02	AP	R	4*	71.50 (2.86)	0.51	-1.16	0.41
EQME21494	16	LR4.06	PS	R	4*	55.75 (2.23)	0.53	-0.46	0.41
EQME10601	17	MG1.03	KU	M	2	77.79	0.25	-1.36	0.46
EQME10582	18	MG2.05	PS	M	4	26.19	0.26	1.72	1.07
EQME10544	19	MG3.01	AP	M	3	63.16	0.33	-0.13	0.66
EQME10521	20	MG3.02	KU	M	4	77.50	0.25	-1.34	0.45
EQME21474	21	MG2.05	PS	M	4*	50.25 (2.01)	0.54	-0.62	0.49
EQME10520	NR	LR3.05	PS	R	4	50.25	0.27	0.66	0.55
EQME10617	NR	NA1.06	PS	N	1	35.91	0.18	1.94	0.48
EQME10625	NR	NA1.01	KU	N	1	70.88	0.33	-0.64	0.60
EQME10662	NR	NA1.04	AP	N	2	33.28	0.26	1.50	0.84
EQME10538	NR	MG2.03	PS	M	3	51.03	0.40	0.43	1.02
EQME10579	NR	MG2.05	AP	M	4	52.94	0.42	0.33	0.91
3434098	NR	LR4.04	AP	R	4	70.05	0.28	-0.68	0.49
EQME21441	NR	LR3.01	AP	R	4*	67.75 (2.71)	0.45	-1.63	0.27
EQME21427	NR	NA2.04	AP	N	4*	73.25 (2.93)	0.55	-1.60	0.48
EQME21425	NR	MG3.01	AP	M	4*	42.25 (1.69)	0.50	0.03	0.42

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

AP = application; KU = knowledge and understanding; PS = problem solving; M = measurement and geometry; N = number sense and algebra; R = linear relations; NR = not released.

**Table 7.1.52 Distribution of Score Points and Step Parameters for Open-Response Items:
Grade 9 Applied Mathematics, Spring Administration (English)**

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
EQME10631	15	% of Students	4.02	0.81	9.59	24.94	16.25	44.38
		Parameters			-2.04	-1.97	0.46	-1.11
EQME21425	NR	% of Students	8.2	0.47	47.66	22.53	8.12	13.02
		Parameters			-2.95	1.00	1.86	0.21
EQME21426	5	% of Students	7.78	1.74	34.53	20.31	15.41	20.23
		Parameters			-2.61	0.62	0.67	0.25
EQME21427	NR	% of Students	1.01	0.08	20.70	11.47	17.75	48.99
		Parameters			-4.78	0.05	-0.73	-0.95
EQME21441	NR	% of Students	3.36	0.30	24.05	16.50	8.82	46.96
		Parameters			-4.90	0.43	1.34	-3.39
EQME21474	21	% of Students	1.86	0.27	43.05	24.17	12.66	17.99
		Parameters			-4.41	0.48	1.08	0.36
EQME21494	16	% of Students	6.49	0.38	24.02	28.66	20.61	19.84
		Parameters			-2.59	-0.56	0.64	0.67

Note. The total number of students is 22 482.

NR = not released.

Table 7.1.53 Item Statistics: Grade 9 Academic Mathematics, Winter Administration (English)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQME21269	1	NA1.01	KU	N	4	59.79	0.35	-0.07	0.71
EQME10332	2	NA1.03	AP	N	2	78.12	0.29	-1.11	0.59
EQME10409	3	NA2.03	AP	N	1	62.64	0.27	-0.26	0.48
EQME10390	4	NA2.09	PS	N	2	88.99	0.18	-2.83	0.44
EQME21282	5	NA2.09	AP	N	4*	68.00 (2.72)	0.39	-1.72	0.24
EQME10456	6	LR2.03	AP	R	1	49.10	0.41	0.39	1.03
EQME10352	7	LR2.05	KU	R	3	49.67	0.33	0.45	0.73
EQME10388	8	LR3.02	AP	R	2	58.94	0.33	-0.07	0.65
EQME10311	9	LR3.03	KU	R	3	84.39	0.34	-1.38	0.86
EQME21225	10	LR2.04	PS	R	4*	72.00 (2.88)	0.43	-2.04	0.30
EQME10319	11	AG1.03	AP	G	4	59.12	0.44	-0.05	1.12
EQME10548	12	AG2.01	AP	G	4	55.55	0.33	0.10	0.71
EQME10341	13	AG2.04	PS	G	2	54.68	0.37	0.07	0.78
EQME10472	14	AG3.01	KU	G	3	69.88	0.40	-0.57	0.89
EQME10395	15	AG3.02	AP	G	2	45.04	0.35	0.63	0.88
EQME10374	16	AG3.05	PS	G	4*	65.75 (2.63)	0.53	-1.06	0.44
EQME10338	17	MG2.02	AP	M	3	62.53	0.34	-0.20	0.67
EQME10447	18	MG2.03	KU	M	3	63.96	0.34	-0.30	0.66
EQME10312	19	MG3.01	KU	M	3	57.55	0.41	-0.06	0.89
EQME10313	20	MG3.01	AP	M	3	70.09	0.30	-0.75	0.59
EQME21232	21	MG2.03	PS	M	4*	69.25 (2.77)	0.50	-1.15	0.34
EQME10320	NR	AG1.02	KU	G	4	66.74	0.29	-0.51	0.55
EQME21251	NR	LR1.01	PS	R	3	42.10	0.30	0.96	0.66
EQME10438	NR	MG2.06	PS	M	2	57.15	0.36	0.07	0.73
EQME10433	NR	AG3.05	PS	G	2	75.87	0.30	-1.10	0.57
EQME10346	NR	NA2.05	KU	N	4	76.31	0.30	-1.07	0.62
EQME10361	NR	MG1.05	PS	M	1	64.41	0.26	-0.41	0.45
EQME10384	NR	LR2.05	PS	R	2	58.33	0.24	0.06	0.45
EQME21256	NR	LR3.01	PS	R	4*	44.75 (1.79)	0.51	-0.37	0.38
EQME21275	NR	AG1.03	AP	G	4*	57.50 (2.30)	0.57	-0.49	0.43
EQME21284	NR	MG3.01	PS	M	4*	58.00 (2.32)	0.60	-0.66	0.52

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

AP = application; KU = knowledge and understanding; PS = problem solving; G = analytic geometry;

M = measurement and geometry; N = number sense and algebra; R = linear relations; NR = not released.

**Table 7.1.54 Distribution of Score Points and Step Parameters for Open-Response Items:
Grade 9 Academic Mathematics, Winter Administration (English)**

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
EQME10374	16	% of Students	3.09	0.36	16.57	20.08	33.13	26.77
		Parameters			-3.15	-0.87	-0.83	0.63
EQME21225	10	% of Students	1.18	0.20	14.42	19.29	25.01	39.91
		Parameters			-5.51	-1.17	-0.74	-0.74
EQME21232	21	% of Students	6.79	0.64	17.73	10.53	19.09	45.22
		Parameters			-2.41	0.33	-1.23	-1.28
EQME21256	NR	% of Students	5.66	0.89	51.95	16.06	7.29	18.15
		Parameters			-3.83	1.62	1.46	-0.73
EQME21275	NR	% of Students	13.28	1.32	23.44	12.93	15.80	33.23
		Parameters			-1.45	0.41	-0.29	-0.61
EQME21282	5	% of Students	3.01	0.75	22.41	9.63	26.97	37.23
		Parameters			-5.20	1.64	-2.72	-0.63
EQME21284	NR	% of Students	5.58	0.80	24.83	27.01	13.50	28.28
		Parameters			-2.49	-0.57	0.78	-0.35

Note. The total number of students is 43 579.

NR = not released.

Table 7.1.55 Item Statistics: Grade 9 Academic Mathematics, Spring Administration (English)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQME10332	1	NA1.03	AP	N	2	76.72	0.30	-1.11	0.59
EQME10422	2	NA2.04	KU	N	4	79.27	0.30	-1.17	0.61
EQME10390	3	NA2.09	PS	N	2	90.47	0.20	-2.83	0.44
EQME21213	4	NA2.02	PS	N	4*	65.00 (2.60)	0.45	-1.27	0.27
EQME21222	5	LR1.01	PS	R	1	71.39	0.40	-0.55	0.78
EQME10352	6	LR2.05	KU	R	3	52.67	0.34	0.45	0.73
EQME10388	7	LR3.02	AP	R	2	62.81	0.36	-0.07	0.65
EQME10311	8	LR3.03	KU	R	3	85.82	0.36	-1.38	0.86
EQME21219	9	LR3.02	PS	R	4*	68.75 (2.75)	0.56	-1.35	0.49
EQME10360	10	AG1.01	KU	G	3	71.67	0.31	-0.64	0.62
EQME10548	11	AG2.01	AP	G	4	59.92	0.36	0.10	0.71
EQME10341	12	AG2.04	PS	G	2	61.35	0.38	0.07	0.78
EQME10472	13	AG3.01	KU	G	3	73.20	0.40	-0.57	0.89
EQME10395	14	AG3.02	AP	G	2	48.52	0.35	0.63	0.88
EQME21276	15	AG2.04	AP	G	4*	53.25 (2.13)	0.57	-0.19	0.50
EQME10338	16	MG2.02	AP	M	3	64.44	0.34	-0.20	0.67
EQME10469	17	MG2.03	PS	M	2	71.03	0.32	-0.62	0.60
EQME10314	18	MG2.06	PS	M	1	36.93	0.36	1.12	1.03
EQME10312	19	MG3.01	KU	M	3	63.32	0.42	-0.06	0.89
EQME10313	20	MG3.01	AP	M	3	73.32	0.32	-0.75	0.59
EQME21241	21	MG3.01	PS	M	4*	57.50 (2.30)	0.54	-0.51	0.34
EQME10413	NR	LR2.01	AP	R	4	86.09	0.25	-1.96	0.53
EQME10421	NR	NA1.01	KU	N	2	61.66	0.29	-0.04	0.59
EQME21264	NR	AG1.03	AP	G	1	60.93	0.44	0.00	1.10
EQME21274	NR	NA2.07	AP	N	2	80.67	0.24	-1.65	0.44
EQME10367	NR	AG3.03	PS	G	1	65.67	0.28	-0.35	0.48
EQME21230	NR	MG1.01	KU	M	3	56.66	0.33	0.23	0.62
EQME21254	NR	LR2.05	PS	R	3	42.85	0.42	0.74	1.23
EQME21259	NR	LR1.04	AP	R	4*	79.25 (3.17)	0.28	-4.06	0.17
EQME21239	NR	AG3.05	PS	G	4*	75.25 (3.01)	0.52	-1.55	0.38
EQME10476	NR	MG2.06	PS	M	4*	62.50 (2.50)	0.56	-0.82	0.48

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

AP = application; KU = knowledge and understanding; PS = problem solving; G = analytic geometry;

M = measurement and geometry; N = number sense and algebra; R = linear relations; NR = not released.

**Table 7.1.56 Distribution of Score Points and Step Parameters for Open-Response Items:
Grade 9 Academic Mathematics, Spring Administration (English)**

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
EQME10476	NR	% of Students	2.63	0.56	15.92	32.08	25.39	23.42
		Parameters			-2.96	-1.34	0.32	0.69
EQME21213	4	% of Students	3.60	1.34	19.36	23.90	13.90	37.91
		Parameters			-3.65	-0.81	1.18	-1.79
EQME21219	9	% of Students	1.01	0.20	13.18	29.43	21.75	34.43
		Parameters			-3.98	-1.58	0.26	-0.10
EQME21239	NR	% of Students	2.46	0.13	12.35	19.49	13.10	52.48
		Parameters			-3.42	-1.32	0.41	-1.87
EQME21241	21	% of Students	11.97	1.38	26.92	12.96	9.54	37.23
		Parameters			-1.83	0.99	0.64	-1.82
EQME21259	NR	% of Students	0.70	0.08	11.94	15.80	12.16	59.33
		Parameters			-10.22	-1.36	0.80	-5.43
EQME21276	15	% of Students	7.80	1.19	26.77	21.96	26.37	15.92
		Parameters			-2.06	-0.04	0.01	1.36

Note. The total number of students is 54 371.

NR = not released.

Table 7.1.57 Item Statistics: Grade 9 Applied Mathematics, Winter Administration (French)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQMF10264	1	PN02	CC	N	3	56.07	0.34	-0.02	0.52
OQMF21041	2	PN05	CC	N	3	71.03	0.35	-0.72	0.71
OQMF10272	3	PN06	CC	N	2	69.35	0.38	-0.66	0.65
OQMF21323	4	PN06	MA	N	1	43.55	0.39	0.78	0.89
OQMF10279	5	PN09	MA	N	3	65.05	0.32	-0.46	0.47
OQMF10269	6	PN02	MA	N	4*	50.25 (2.01)	0.32	-0.62	0.19
OQMF21335	7	PR01	CC	R	4	68.97	0.16	-1.14	0.27
13466	8	PR04	HP	R	2	57.57	0.33	-0.02	0.49
OQMF10300	9	PR06	CC	R	2	71.40	0.20	-0.83	0.38
OQMF10302	10	PR09	MA	R	3	56.26	0.28	0.01	0.54
OQMF10311	11	PR14	MA	R	3	56.07	0.46	-0.09	0.89
OQMF21377	12	PR14	HP	R	2	45.79	0.29	0.73	0.64
OQMF10296	13	PR03	MA	R	4*	64.75 (2.59)	0.49	-1.36	0.49
13471	14	PM03	HP	M	2	61.31	0.35	-0.26	0.77
OQMF21438	15	PM05	MA	M	2	54.77	0.32	0.19	0.62
OQMF21407	16	PM05	HP	M	1	28.79	0.21	1.59	1.00
18936	17	PM14	HP	M	3	68.22	0.37	-0.55	0.71
13456	18	PM17	MA	M	1	70.47	0.23	-0.88	0.48
13454	19	PM17	HP	M	3	60.75	0.14	-0.23	0.25
OQMF21390	20	PM04	MA	M	4*	50.75 (2.03)	0.54	-0.66	0.50
OQMF21370	NR	PR12	CC	R	3	54.02	0.39	0.18	0.77
OQMF21409	NR	PM15	MA	M	4	59.07	0.40	-0.05	0.89
OQMF21306	NR	PN02	MA	N	1	42.99	0.32	0.77	0.84
OQMF21403	NR	PM05	CC	M	2	58.13	0.33	0.00	0.62
OQMF21340	NR	PR04	MA	R	3	78.32	0.29	-1.40	0.52
OQMF10281	NR	PN09	HP	N	2	47.29	0.21	0.68	0.50
OQMF21428	NR	PM18	CC	M	4	50.47	0.14	0.82	0.22
OQMF21333	NR	PN09	HP	N	4*	67.00 (2.68)	0.43	-1.30	0.37
OQMF21364	NR	PR10	MA	R	4*	50.75 (2.03)	0.37	-0.50	0.28
OQMF21420	NR	PM15	HP	M	4*	50.25 (2.01)	0.45	-0.95	0.32
OQMF10313	NR	PR14	HP	R	4*	60.50 (2.42)	0.23	-1.63	0.11

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

MA = application; CC = knowledge and understanding; HP = problem solving;

M = measurement and geometry; N = number sense and algebra; R = linear relations; NR = not released.

**Table 7.1.58 Distribution of Score Points and Step Parameters for Open-Response Items:
Grade 9 Applied Mathematics, Winter Administration (French)**

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
OQMF10269	6	% of Students	7.29	0.75	28.04	37.20	8.41	18.32
		Parameters			-3.89	-1.27	3.77	-1.09
OQMF10296	13	% of Students	0.93	0.19	6.36	45.05	26.92	20.56
		Parameters			-3.53	-3.15	0.48	0.76
OQMF10313	NR	% of Students	7.85	1.87	19.25	19.63	21.87	29.53
		Parameters			-4.02	-0.38	-0.65	-1.47
OQMF21333	NR	% of Students	4.30	0.56	7.10	36.07	19.44	32.52
		Parameters			-2.08	-3.26	0.61	-0.48
OQMF21364	NR	% of Students	4.86	0.75	24.49	42.43	16.07	11.40
		Parameters			-3.85	-1.47	2.14	1.20
OQMF21390	20	% of Students	2.99	0.00	38.69	29.72	9.91	18.69
		Parameters			-3.98	-0.05	1.47	-0.08
OQMF21420	NR	% of Students	2.62	1.12	46.17	18.32	8.41	23.36
		Parameters			-5.33	1.38	1.53	-1.36

Note. The total number of students is 535.

NR = not released.

Table 7.1.59 Item Statistics: Grade 9 Applied Mathematics, Spring Administration (French)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQMF10264	1	PN02	CC	N	3	63.48	0.24	-0.02	0.52
OQMF21307	2	PN02	MA	N	4	61.22	0.34	0.01	0.67
OQMF10272	3	PN06	CC	N	2	72.78	0.30	-0.66	0.65
OQMF21323	4	PN06	MA	N	1	43.29	0.34	0.78	0.89
OQMF10279	5	PN09	MA	N	3	67.46	0.23	-0.46	0.47
OQMF10269	6	PN02	MA	N	4*	52.25 (2.09)	0.32	-0.62	0.19
13470	7	PR02	MA	R	1	56.18	0.42	0.25	1.02
13466	8	PR04	HP	R	2	62.42	0.22	-0.02	0.49
OQMF10300	9	PR06	CC	R	2	68.39	0.26	-0.83	0.38
OQMF10302	10	PR09	MA	R	3	62.42	0.30	0.01	0.54
OQMF21367	11	PR12	CC	R	2	61.22	0.18	-0.05	0.30
OQMF10311	11	PR14	MA	R	3	65.87	0.39	-0.09	0.89
13536	13	PR01	MA	R	4*	64.50 (2.58)	0.43	-1.03	0.34
13471	14	PM03	HP	M	2	68.26	0.37	-0.26	0.77
OQMF21402	15	PM05	CC	M	3	75.96	0.36	-0.75	0.78
OQMF21410	16	PM14	MA	M	1	53.92	0.24	0.52	0.48
18936	17	PM14	HP	M	3	71.31	0.34	-0.55	0.71
13456	18	PM17	MA	M	1	73.44	0.29	-0.88	0.48
13454	19	PM17	HP	M	3	62.68	0.14	-0.23	0.25
OQMF21365	20	PM15	HP	M	4*	74.25 (2.97)	0.45	-1.22	0.34
OQMF10295	NR	PR02	CC	R	3	71.05	0.26	-0.70	0.48
OQMF21406	NR	PM05	HP	M	3	35.99	0.16	2.15	0.43
OQMF21312	NR	PN07	CC	N	4	57.10	0.35	0.23	0.70
OQMF21374	NR	PR14	HP	R	4	53.65	0.29	0.47	0.51
OQMF10281	NR	PN09	HP	N	2	52.32	0.28	0.68	0.50
OQMF21428	NR	PM18	CC	M	4	56.97	0.09	0.82	0.22
OQMF21333	NR	PN09	HP	N	4*	71.75 (2.87)	0.48	-1.30	0.37
OQMF21392	NR	PM02	MA	M	4*	48.25 (1.93)	0.42	-0.27	0.23
OQMF21363	NR	PR10	MA	R	4*	51.75 (2.07)	0.51	-0.20	0.35
OQMF21038	NR	PR15	HP	R	4*	51.50 (2.06)	0.49	-0.30	0.34

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

MA = application; CC = knowledge and understanding; HP = problem solving;

M = measurement and geometry; N = number sense and algebra; R = linear relations; NR = not released.

**Table 7.1.60 Distribution of Score Points and Step Parameters for Open-Response Items:
Grade 9 Applied Mathematics, Spring Administration (French)**

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
13536	13	% of Students	3.45	0.40	7.70	42.50	18.33	27.62
		Parameters			-2.04	-3.36	1.51	-0.24
OQMF10269	6	% of Students	6.11	2.26	23.77	36.65	13.01	18.19
		Parameters			-3.89	-1.27	3.77	-1.09
OQMF21038	NR	% of Students	10.36	1.33	30.01	26.43	4.78	27.09
		Parameters			-2.15	0.09	3.19	-2.33
OQMF21333	NR	% of Students	2.12	0.13	5.71	32.27	22.31	37.45
		Parameters			-2.08	-3.26	0.61	-0.48
OQMF21363	NR	% of Students	10.36	2.26	29.88	17.93	16.6	22.97
		Parameters			-1.99	0.71	0.39	0.10
OQMF21365	20	% of Students	5.31	0.66	7.97	14.61	25.5	45.95
		Parameters			-1.43	-1.59	-1.10	-0.74
OQMF21392	NR	% of Students	12.35	3.19	42.23	4.65	9.30	28.29
		Parameters			-2.87	5.51	-1.48	-2.25

Note. The total number of students is 753.

NR = not released.

Table 7.1.61 Item Statistics: Grade 9 Academic Mathematics, Winter Administration (French)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQMF10383	1	DN05	CC	N	4	55.44	0.22	0.38	0.40
OQMF10388	2	DN09	CC	N	2	66.19	0.40	-0.06	0.72
OQMF10392	3	DN11	MA	N	2	55.58	0.27	0.31	0.50
OQMF20943	4	DN16	HP	N	1	54.16	0.25	0.43	0.46
13809	5	DN18	MA	N	2	65.58	0.32	-0.44	0.58
OQMF10397	6	DN13	MA	N	4*	49.50 (1.98)	0.58	-0.19	0.51
OQMF20988	7	DR01	HP	R	3	42.46	0.27	1.10	0.61
OQMF10408	8	DR03	MA	R	4	51.25	0.45	0.44	1.12
OQMF10414	9	DR07	MA	R	2	64.84	0.30	-0.49	0.51
13437	10	DR15	MA	R	3	80.73	0.23	-1.83	0.43
OQMF20919	11	DR14	HP	R	4*	65.50 (2.62)	0.38	-1.54	0.19
OQMF20981	12	DG07	CC	G	2	73.56	0.27	-0.94	0.51
13425	13	DG09	CC	G	3	77.35	0.28	-1.32	0.55
12038	14	DG09	HP	G	4	38.81	0.32	1.30	0.76
OQMF20924	15	DG07	MA	G	4*	47.50 (1.90)	0.57	-0.02	0.44
OQMF20999	16	DM01	HP	M	2	55.44	0.22	0.38	0.41
OQMF21014	17	DM05	MA	M	2	48.61	0.32	0.53	0.65
OQMF10357	18	DM15	CC	M	1	62.61	0.29	-0.12	0.48
OQMF10363	19	DM18	HP	M	1	41.11	0.36	0.77	0.88
18537	20	DM20	MA	M	3	59.70	0.22	0.10	0.35
OQMF10370	21	DM20	HP	M	4*	51.50 (2.06)	0.49	-0.19	0.27
OQMF20967	NR	DR14	HP	R	3	56.66	0.22	0.29	0.38
OQMF20976	NR	DN03	MA	N	4	70.45	0.35	-0.58	0.69
OQMF10338	NR	DG07	HP	G	1	42.19	0.28	1.03	0.71
OQMF10332	NR	DG04	MA	G	3	57.07	0.37	0.16	0.77
OQMF10373	NR	DM20	CC	M	2	71.26	0.31	-0.71	0.57
13429	NR	DR09	CC	R	2	61.80	0.35	-0.28	0.64
OQMF10351	NR	DM06	CC	M	2	64.84	0.37	-0.11	0.72
OQMF20905	NR	DN05	MA	N	4*	46.50 (1.86)	0.43	-0.67	0.29
OQMF20926	NR	DM15	HP	M	4*	67.75 (2.71)	0.53	-0.98	0.39
OQMF20913	NR	DR02	MA	R	4*	60.00 (2.40)	0.56	-0.65	0.42

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

MA = application; CC = knowledge and understanding; HP = problem solving; G = analytic geometry;

M = measurement and geometry; N = number sense and algebra; R = linear relations; NR=not released.

**Table 7.1.62 Distribution of Score Points and Step Parameters for Open-Response Items:
Grade 9 Academic Mathematics, Winter Administration (French)**

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
OQMF10370	21	% of Students	16.16	7.03	19.54	17.17	8.25	31.85
		Parameters			-0.15	0.09	1.75	-2.46
OQMF10397	6	% of Students	5.14	1.08	35.29	25.08	20.69	12.71
		Parameters			-2.82	0.14	0.51	1.39
OQMF20905	NR	% of Students	2.64	1.69	47.33	24.41	5.48	18.46
		Parameters			-5.38	1.20	3.28	-1.80
OQMF20913	NR	% of Students	6.63	1.35	20.96	23.66	17.92	29.48
		Parameters			-2.22	-0.59	0.43	-0.19
OQMF20919	11	% of Students	6.02	2.30	26.17	10.89	4.26	50.37
		Parameters			-4.06	2.35	2.81	-7.25
OQMF20924	15	% of Students	11.16	4.33	27.45	26.98	11.36	18.73
		Parameters			-1.43	-0.16	1.45	0.08
OQMF20926	NR	% of Students	5.41	1.89	13.52	23.26	13.12	42.80
		Parameters			-1.88	-1.36	0.75	-1.44

Note. The total number of students is 1479.

NR = not released.

Table 7.1.63 Item Statistics: Grade 9 Academic Mathematics, Spring Administration (French)

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
13432	1	DN01	MA	N	2	38.43	0.21	1.50	0.56
OQMF10383	2	DN05	CC	N	4	55.35	0.22	0.38	0.40
OQMF10388	3	DN09	CC	N	2	57.75	0.34	-0.06	0.72
13809	4	DN18	MA	N	2	67.95	0.30	-0.44	0.58
OQMF10380	5	DN03	MA	N	4*	52.00 (2.08)	0.51	-0.43	0.39
OQMF10408	6	DR03	MA	R	4	47.21	0.46	0.44	1.12
OQMF10414	7	DR07	MA	R	2	68.79	0.27	-0.49	0.51
13437	8	DR15	MA	R	3	82.19	0.21	-1.83	0.43
OQMF20978	9	DR14	HP	R	1	43.01	0.24	1.23	0.49
OQMF20918	10	DR04	MA	R	4*	44.25 (1.77)	0.55	-0.01	0.45
OQMF20985	11	DG05	MA	G	2	54.64	0.34	0.29	0.70
OQMF20960	12	DG07	HP	G	1	32.68	0.35	1.20	1.32
13425	13	DG09	CC	G	3	80.22	0.27	-1.32	0.55
12038	14	DG09	HP	G	4	35.53	0.28	1.30	0.76
OQMF20925	15	DG07	MA	G	4*	50.75 (2.03)	0.55	-0.19	0.42
OQMF21014	16	DM05	MA	M	2	51.78	0.32	0.53	0.65
OQMF10357	17	DM15	CC	M	1	61.11	0.25	-0.12	0.48
OQMF10363	18	DM18	HP	M	1	45.23	0.40	0.77	0.88
18537	19	DM20	MA	M	3	58.50	0.19	0.10	0.35
13415	20	DM20	CC	M	3	65.35	0.31	-0.33	0.58
OQMF10353	21	DM09	HP	M	4*	59.25 (2.37)	0.51	-0.77	0.39
OQMF10394	NR	DN12	MA	N	1	57.29	0.37	0.12	0.76
OQMF20951	NR	DN15	HP	N	2	46.66	0.26	0.87	0.55
OQMF10329	NR	DG02	CC	G	4	73.12	0.08	-3.20	0.13
OQMF21001	NR	DM04	HP	M	2	49.35	0.34	0.55	0.77
13433	NR	DR01	HP	R	1	61.74	0.40	-0.10	0.89
13429	NR	DR09	CC	R	2	66.44	0.32	-0.28	0.64
OQMF10351	NR	DM06	CC	M	2	60.39	0.34	-0.11	0.72
OQMF10367	NR	DM20	HP	M	4*	58.75 (2.35)	0.53	-0.83	0.37
OQMF20920	NR	DR15	HP	R	4*	55.25 (2.21)	0.50	-0.47	0.36
OQMF20912	NR	DN13	MA	N	4*	66.50 (2.66)	0.47	-1.22	0.30

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items.

* = Maximum score code for open-response (OR) items; () = Mean score for OR items.

MA = application; CC = knowledge and understanding; HP = problem solving; G = analytic geometry; M = measurement and geometry; N = number sense and algebra; R = linear relations; NR=not released.

**Table 7.1.64 Distribution of Score Points and Step Parameters for Open-Response Items:
Grade 9 Academic Mathematics, Spring Administration (French)**

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
OQMF10353	21	% of Students	3.53	1.76	20.54	32.09	16.00	26.08
		Parameters			-2.87	-1.07	1.09	-0.24
OQMF10367	NR	% of Students	3.32	3.57	34.36	13.19	8.23	37.34
		Parameters			-3.32	1.20	0.80	-1.99
OQMF10380	5	% of Students	5.17	1.43	26.12	38.47	10.25	18.56
		Parameters			-2.79	-0.86	2.17	-0.24
OQMF20912	NR	% of Students	3.23	2.65	26.75	4.96	19.87	42.55
		Parameters			-3.73	2.86	-2.80	-1.20
OQMF20918	10	% of Students	8.90	3.36	39.02	25.24	6.30	17.18
		Parameters			-2.12	0.44	2.16	-0.51
OQMF20920	NR	% of Students	6.13	1.81	24.53	25.75	21.84	19.95
		Parameters			-2.57	-0.39	0.37	0.69
OQMF20925	15	% of Students	10.29	2.10	25.83	27.72	14.24	19.82
		Parameters			-1.71	-0.35	1.12	0.19

Note. The total number of students is 2381.

NR = not released.

Differential Item Functioning (DIF) Analysis Results

The gender- and ELL-based DIF results for the Grade 9 Assessment of Mathematics are provided in Tables 7.1.65 to 7.1.76. The results in each table are from the two random samples, except for the French-language populations. Each table indicates the value of $\hat{\beta}$, the significance level and the category of the effect size for each item for B- and C-level DIF items. For gender-based DIF, negative estimates of $\hat{\beta}$ indicate that the girls outperformed the boys; positive $\hat{\beta}$ estimates indicate that the boys outperformed the girls. For ELL-based DIF, negative estimates of $\hat{\beta}$ indicate that the ELLs outperformed the non-ELLs; positive $\hat{\beta}$ estimates indicate that the non-ELLs outperformed the ELLs. DIF analysis was not conducted for ALF/PDF students for the French-language version of the Grade 9 mathematics assessment due to small populations.

Table 7.1.65 Gender-Based DIF Statistics: Grade 9 Applied Mathematics, Winter Administration (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME10578	1	0.029	0.051		-0.017	0.252	
EQME10621	2	-0.007	0.662		0.037	0.018	
EQME10667	3	0.001	0.949		0.000	0.975	
EQME10607	4	-0.076	0	B-	-0.034	0.031	
EQME10425	5	-0.049	0.001		-0.042	0.007	
EQME21429	6	-0.087	0.050		-0.110	0.012	
EQME10616	7	0.026	0.086		0.016	0.275	
EQME10533	8	-0.014	0.314		-0.033	0.021	
EQME10534	9	-0.038	0.008		-0.030	0.039	
EQME10536	10	0.050	0		0.061	0	B+
EQME21407	11	0.050	0.001		0.067	0	B+
EQME10666	12	-0.077	0	B-	-0.082	0	B-
EQME10635	13	0.011	0.444		-0.002	0.902	
EQME10537	14	0.022	0.150		0.000	0.975	
EQME21498	15	0.165	0	C+	0.164	0	C+
EQME21472	16	0.223	0	C+	0.202	0	C+
EQME21402	17	0.050	0.001		0.053	0.001	
EQME10582	18	-0.003	0.841		-0.007	0.610	
EQME10629	19	-0.011	0.433		-0.042	0.004	
EQME10544	20	0.010	0.527		0.005	0.754	
EQME21492	21	-0.013	0.718		-0.015	0.689	
EQME10549	NR	0.033	0.032		0.058	0	
EQME10522	NR	-0.022	0.048		-0.015	0.175	
EQME21452	NR	0.049	0.001		0.016	0.276	
EQME21453	NR	0.024	0.090		0.027	0.060	
EQME10505	NR	-0.047	0.002		-0.031	0.036	
EQME10599	NR	-0.018	0.238		-0.019	0.221	
EQME10515	NR	-0.051	0.001		-0.023	0.117	
EQME10643	NR	-0.206	0	C-	-0.154	0	C-
EQME21413	NR	-0.257	0	C-	-0.211	0	C-
EQME21454	NR	0.189	0	C+	0.106	0.001	C+

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.66 Gender-Based DIF Statistics: Grade 9 Applied Mathematics, Spring Administration (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME10526	1	0.028	0.074		0.039	0.012	
EQME10667	2	-0.016	0.288		0.008	0.608	
EQME10607	3	-0.066	0	B-	-0.048	0.002	
EQME10425	4	-0.082	0	B-	-0.066	0	B-
EQME21426	5	0.093	0.009	C+	0.125	0	C+
EQME10616	6	0.041	0.006		0.037	0.013	
13641	7	-0.005	0.727		0.022	0.107	
EQME10533	8	0.008	0.589		0.004	0.786	
EQME10534	9	-0.006	0.667		-0.018	0.213	
EQME10536	10	0.068	0	B+	0.072	0	B+
EQME10604	11	0.001	0.935		-0.008	0.582	
EQME10666	12	-0.083	0	B-	-0.095	0	C-
EQME10635	13	0.040	0.006		0.018	0.239	
EQME10537	14	0.049	0.001		0.014	0.343	
EQME10631	15	-0.109	0.002	C-	-0.157	0	C-
EQME21494	16	0.183	0	C+	0.197	0	C+
EQME10601	17	0.018	0.181		-0.014	0.303	
EQME10582	18	-0.006	0.653		0.008	0.557	
EQME10544	19	0.025	0.091		0.028	0.063	
EQME10521	20	0.004	0.750		0.008	0.561	
EQME21474	21	-0.283	0	C-	-0.239	0	C-
EQME10520	NR	0.003	0.850		0.049	0.002	
EQME10617	NR	-0.002	0.872		0.014	0.345	
EQME10625	NR	-0.007	0.626		-0.018	0.192	
EQME10662	NR	0.010	0.492		0.001	0.964	
EQME10538	NR	-0.022	0.141		-0.035	0.021	
EQME10579	NR	-0.067	0	B-	-0.051	0.001	
3434098	NR	0.024	0.087		0.011	0.460	
EQME21441	NR	0.228	0	C+	0.237	0	C+
EQME21427	NR	-0.152	0	C-	-0.251	0	C-
EQME21425	NR	0.067	0.038	B+	0.094	0.004	C+

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.67 Gender-Based DIF Statistics: Grade 9 Academic Mathematics, Winter Administration (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME21269	1	-0.020	0.180		0.003	0.822	
EQME10332	2	0.026	0.043		0.036	0.005	
EQME10409	3	-0.047	0.002		-0.018	0.222	
EQME10390	4	-0.007	0.481		-0.002	0.865	
EQME21282	5	0.235	0	C+	0.183	0	C+
EQME10456	6	0.054	0		0.029	0.056	
EQME10352	7	0.017	0.254		0	0.975	
EQME10388	8	0.064	0	B+	0.050	0.001	
EQME10311	9	0.004	0.747		-0.007	0.524	
EQME21225	10	-0.046	0.162		-0.017	0.606	
EQME10319	11	-0.050	0.001		-0.034	0.020	
EQME10548	12	0.006	0.674		0.013	0.408	
EQME10341	13	0.009	0.549		0.002	0.895	
EQME10472	14	-0.019	0.165		-0.048	0	
EQME10395	15	-0.009	0.537		0.017	0.258	
EQME10374	16	0.172	0	C+	0.149	0	C+
EQME10338	17	-0.020	0.180		-0.039	0.008	
EQME10447	18	0.003	0.837		0.015	0.307	
EQME10312	19	-0.014	0.337		-0.012	0.409	
EQME10313	20	0.052	0		0.028	0.054	
EQME21232	21	-0.129	0.001		-0.082	0.030	B-
EQME10320	NR	-0.077	0	B-	-0.067	0	B-
EQME21251	NR	0.031	0.046		0.015	0.319	
EQME10438	NR	-0.024	0.116		-0.018	0.241	
EQME10433	NR	0.003	0.830		0.023	0.088	
EQME10346	NR	-0.042	0.001		-0.039	0.003	
EQME10361	NR	0.025	0.092		0.040	0.007	
EQME10384	NR	0.028	0.073		0.037	0.017	
EQME21256	NR	-0.021	0.531		-0.018	0.596	
EQME21275	NR	-0.354	0	C-	-0.282	0	C-
EQME21284	NR	0.100	0.003	C+	0.034	0.319	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.68 Gender-Based DIF Statistics: Grade 9 Academic Mathematics, Spring Administration (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME10332	1	0.022	0.093		0.016	0.231	
EQME10422	2	-0.020	0.113		-0.027	0.032	
EQME10390	3	-0.005	0.562		0.004	0.672	
EQME21213	4	0.312	0	C+	0.293	0	C+
EQME21222	5	0.118	0	C+	0.080	0	B+
EQME10352	6	-0.010	0.515		0.051	0.001	
EQME10388	7	0.049	0.001		0.038	0.012	
EQME10311	8	0.024	0.026		0.022	0.035	
EQME21219	9	0.100	0.001	C+	0.086	0.004	B+
EQME10360	10	-0.014	0.325		-0.042	0.003	
EQME10548	11	0.028	0.065		0.004	0.776	
EQME10341	12	-0.003	0.826		0.014	0.363	
EQME10472	13	-0.024	0.075		-0.033	0.014	
EQME10395	14	0.007	0.649		0.020	0.187	
EQME21276	15	-0.368	0	C-	-0.354	0	C-
EQME10338	16	-0.017	0.256		-0.011	0.442	
EQME10469	17	0.014	0.335		-0.002	0.869	
EQME10314	18	0	0.985		-0.001	0.954	
EQME10312	19	0.007	0.634		-0.004	0.775	
EQME10313	20	0.037	0.009		0.025	0.070	
EQME21241	21	-0.066	0.112		-0.058	0.170	
EQME10421	NR	-0.031	0.040		-0.011	0.484	
EQME21264	NR	-0.082	0	B-	-0.092	0	C-
EQME21274	NR	0.028	0.028		0.038	0.002	
EQME10413	NR	0.003	0.760		0.005	0.674	
EQME10367	NR	-0.066	0	B-	-0.043	0.004	
EQME21230	NR	0.012	0.428		0.006	0.692	
EQME21254	NR	-0.063	0	B-	-0.062	0	B-
EQME21259	NR	0.007	0.845		0.044	0.191	
EQME21239	NR	0.113	0.001	C+	0.064	0.052	
EQME10476	NR	-0.127	0	C-	-0.099	0.001	C-

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.69 Gender-Based DIF Statistics: Grade 9 Applied Mathematics, Winter Administration (French)

Item Code	Sequence	Sample 1		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQMF10264	1	0.132	0	C+
OQMF21041	2	-0.128	0	C-
OQMF10272	3	0.012	0.729	
OQMF21323	4	-0.036	0.346	
OQMF10279	5	0.116	0.002	C+
OQMF10269	6	0.194	0.031	C+
OQMF21335	7	-0.123	0.001	C-
13466	8	0.009	0.815	
OQMF10300	9	-0.023	0.549	
OQMF10302	10	0.073	0.064	
OQMF10311	11	0.042	0.260	
OQMF21377	12	0.041	0.289	
OQMF10296	13	-0.191	0.004	C-
13471	14	-0.110	0.004	C-
OQMF21438	15	-0.017	0.661	
OQMF21407	16	0.015	0.687	
18936	17	-0.065	0.081	
13456	18	-0.072	0.047	B-
13454	19	0.031	0.442	
OQMF21390	20	-0.190	0.025	C-
OQMF10281	NR	0.057	0.155	
OQMF21428	NR	0.012	0.777	
OQMF21370	NR	-0.023	0.549	
OQMF21409	NR	0.055	0.152	
OQMF21306	NR	0.096	0.013	C+
OQMF21403	NR	-0.064	0.096	
OQMF21340	NR	0.061	0.065	
OQMF21364	NR	-0.070	0.383	
OQMF21420	NR	-0.015	0.871	
OQMF10313	NR	0.132	0.212	
OQMF21333	NR	0.151	0.082	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.70 Gender-Based DIF Statistics: Grade 9 Applied Mathematics, Spring Administration (French)

Item Code	Sequence	Sample 1		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQMF10264	1	0.061	0.065	
OQMF21307	2	0.057	0.072	
OQMF10272	3	-0.006	0.855	
OQMF21323	4	-0.034	0.303	
OQMF10279	5	-0.043	0.173	
OQMF10269	6	0.197	0.013	C+
13470	7	0.036	0.259	
13466	8	0.084	0.012	B+
OQMF10300	9	-0.111	0.001	C-
OQMF10302	10	0.056	0.090	
OQMF21367	11	-0.018	0.596	
OQMF10311	12	0.026	0.412	
13536	13	-0.218	0.001	C-
13471	14	-0.099	0.001	C-
OQMF21402	15	-0.069	0.016	B-
OQMF21410	16	-0.019	0.575	
18936	17	-0.031	0.324	
13456	18	-0.032	0.274	
13454	19	-0.340	0.325	
OQMF21365	20	0.122	0.109	
OQMF10295	NR	0.017	0.579	
OQMF10281	NR	0.106	0.001	C+
OQMF21428	NR	-0.008	0.818	
OQMF21406	NR	0.034	0.310	
OQMF21312	NR	-0.112	0.001	C-
OQMF21374	NR	0.079	0.016	B+
OQMF21392	NR	-0.208	0.027	C-
OQMF21363	NR	0.027	0.739	
OQMF21333	NR	0.285	0	C+
OQMF21038	NR	-0.191	0.018	C-

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.71 Gender-Based DIF Statistics: Grade 9 Academic Mathematics, Winter Administration (French)

Item Code	Sequence	Sample 1		
		$\hat{\beta}$	p-value	Level of DIF
OQMF10383	1	-0.036	0.161	
OQMF10388	2	-0.016	0.496	
OQMF10392	3	-0.059	0.022	B-
OQMF20943	4	0.068	0.008	B+
13809	5	-0.054	0.025	
OQMF10397	6	0.289	0	C+
OQMF20988	7	0.127	0	C+
OQMF10408	8	-0.028	0.239	
OQMF10414	9	-0.043	0.082	
13437	10	0.044	0.032	
OQMF20919	11	0.515	0	C+
OQMF20981	12	-0.038	0.107	
13425	13	-0.048	0.032	
12038	14	-0.022	0.375	
OQMF20924	15	-0.448	0	C-
OQMF20999	16	0.025	0.327	
OQMF21014	17	-0.011	0.671	
OQMF10357	18	0.021	0.403	
OQMF10363	19	0.016	0.531	
18537	20	-0.030	0.250	
OQMF10370	21	-0.139	0.053	
13429	NR	0.021	0.387	
OQMF20967	NR	0.042	0.104	
OQMF20976	NR	0.049	0.028	
OQMF10338	NR	-0.060	0.016	B-
OQMF10332	NR	-0.079	0.001	B-
OQMF10351	NR	-0.015	0.539	
OQMF10373	NR	0.028	0.221	
OQMF20905	NR	-0.073	0.199	
OQMF20926	NR	-0.084	0.171	
OQMF20913	NR	-0.150	0.010	C-

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.72 Gender-Based DIF Statistics: Grade 9 Academic Mathematics, Spring Administration (French)

Item Code	Sequence	Sample 1		
		$\hat{\beta}$	p-value	Level of DIF
13432	1	0.005	0.792	
OQMF10383	2	-0.025	0.215	
OQMF10388	3	0.028	0.151	
13809	4	-0.043	0.024	
OQMF10380	5	-0.055	0.190	
OQMF10408	6	-0.039	0.042	
OQMF10414	7	0.019	0.319	
13437	8	0.030	0.047	
OQMF20978	9	0.114	0	C+
OQMF20918	10	-0.121	0.005	C-
OQMF20985	11	-0.078	0	B-
OQMF20960	12	0.029	0.115	
13425	13	-0.031	0.053	
12038	14	-0.069	0	B-
OQMF20925	15	-0.260	0	C-
OQMF21014	16	-0.030	0.135	
OQMF10357	17	0.018	0.350	
OQMF10363	18	-0.006	0.757	
18537	19	-0.007	0.728	
13415	20	0.066	0	B+
OQMF10353	21	0.051	0.249	
13429	NR	0.046	0.015	
OQMF10394	NR	-0.042	0.032	
OQMF20951	NR	0.071	0	B+
OQMF10329	NR	0.031	0.096	
OQMF21001	NR	-0.020	0.308	
13433	NR	-0.021	0.268	
OQMF10351	NR	-0.011	0.583	
OQMF10367	NR	-0.145	0.004	C-
OQMF20920	NR	0.277	0	C+
OQMF20912	NR	0.248	0	C+

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.73 ELL-Based DIF Statistics: Grade 9 Applied Mathematics, Winter Administration (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME21413	1	-0.028	0.189		-0.044	0.038	
EQME21498	2	0.055	0.017		0.013	0.570	
EQME10667	3	-0.029	0.187		-0.033	0.134	
EQME10534	4	-0.116	0	C-	-0.153	0	C-
EQME10425	5	-0.114	0	C-	-0.108	0	C-
EQME21429	6	-0.171	0.009	C-	-0.228	0	C-
EQME10536	7	0.138	0	C+	0.150	0	C+
EQME10533	8	0.066	0.003	B+	0.086	0	B+
EQME10666	9	-0.032	0.133		-0.039	0.067	
EQME10643	10	0.048	0.007		0.023	0.206	
EQME21407	11	0.081	0	B+	0.060	0.006	B+
EQME21492	12	-0.024	0.232		-0.014	0.491	
EQME10616	13	0.022	0.322		0.036	0.107	
EQME10537	14	0.009	0.702		0.054	0.013	
EQME10629	15	-0.081	0.102		0.024	0.624	
EQME21472	16	0.142	0.006	C+	0.119	0.016	C+
EQME21402	17	0.037	0.109		0.025	0.280	
EQME10578	18	-0.004	0.851		-0.038	0.049	
EQME10582	19	-0.045	0.027		-0.008	0.698	
EQME10544	20	-0.031	0.164		-0.011	0.614	
EQME10515	21	-0.204	0	C-	-0.177	0.001	C-
EQME10549	NR	-0.025	0.267		-0.016	0.462	
EQME10522	NR	0.011	0.544		0.009	0.627	
EQME21452	NR	0.061	0.005	B+	0.104	0	C+
EQME21453	NR	-0.012	0.573		0.021	0.317	
EQME10621	NR	0.084	0	B+	0.070	0.001	B+
EQME10505	NR	-0.008	0.698		-0.047	0.032	
EQME10635	NR	-0.043	0.049		-0.023	0.280	
EQME10599	NR	0.184	0	C+	0.117	0.021	C+
EQME10607	NR	0.134	0.015	C+	0.154	0.004	C+
EQME21454	NR	-0.065	0.177		-0.080	0.079	

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.74 ELL-Based DIF Statistics: Grade 9 Applied Mathematics, Spring Administration (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME10526	1	0.100	0		0.076	0	
EQME10667	2	0.018	0.331		0.029	0.103	
EQME10607	3	-0.181	0	C-	-0.175	0	C-
EQME10425	4	-0.186	0	C-	-0.175	0	C-
EQME21426	5	-0.005	0.913		0.047	0.274	
EQME10616	6	0.142	0	C+	0.169	0	C+
13641	7	0.093	0	C+	0.115	0	C+
EQME10533	8	0.045	0.010		0.065	0	
EQME10534	9	-0.046	0.007		-0.042	0.014	
EQME10536	10	0.038	0.010		0.052	0	
EQME10604	11	-0.030	0.090		-0.060	0	
EQME10666	12	-0.037	0.027		-0.033	0.049	
EQME10635	13	0.042	0.024		0.049	0.008	
EQME10537	14	0.011	0.556		0.004	0.830	
EQME10631	15	-0.106	0.011		-0.085	0.041	
EQME21494	16	0.291	0	C+	0.274	0	C+
EQME10601	17	0.031	0.062		0.024	0.142	
EQME10582	18	-0.010	0.521		-0.012	0.474	
EQME10544	19	-0.016	0.369		-0.042	0.019	
EQME10521	20	0.033	0.043		0.058	0	
EQME21474	21	-0.168	0	C-	-0.106	0.005	C-
EQME10520	NR	0.021	0.262		0.023	0.231	
EQME10617	NR	-0.023	0.208		-0.025	0.170	
EQME10625	NR	-0.040	0.021		-0.032	0.064	
EQME10662	NR	-0.013	0.450		0.004	0.832	
EQME10538	NR	-0.005	0.761		0.015	0.403	
EQME10579	NR	-0.119	0	C-	-0.118	0	C-
3434098	NR	0.145	0	C+	0.134	0	C+
EQME21441	NR	0.249	0	C+	0.181	0	C+
EQME21427	NR	-0.191	0	C-	-0.196	0	C-
EQME21425	NR	-0.173	0	C-	-0.165	0	C-

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.75 ELL-Based DIF Statistics: Grade 9 Academic Mathematics, Winter Administration (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME21269	1	-0.130	0	C-	-0.129	0	C-
EQME10332	2	-0.075	0		-0.103	0	
EQME10409	3	-0.044	0.010		-0.061	0	
EQME10390	4	0.076	0	B+	0.078	0	B+
EQME21282	5	0.207	0	C+	0.153	0	C+
EQME10456	6	0.019	0.268		0.044	0.010	
EQME10352	7	-0.007	0.708		0.012	0.503	
EQME10388	8	0.104	0	C+	0.092	0	C+
EQME10311	9	0.030	0.019		0.027	0.035	
EQME21225	10	0.493	0	C+	0.457	0	C+
EQME10319	11	-0.117	0	C-	-0.113	0	C-
EQME10548	12	0.033	0.060		0.026	0.135	
EQME10341	13	0.062	0	B+	0.061	0	B+
EQME10472	14	-0.014	0.359		-0.010	0.505	
EQME10395	15	-0.128	0	C-	-0.099	0	C-
EQME10374	16	0.247	0	C+	0.206	0	C+
EQME10338	17	0.016	0.352		-0.005	0.761	
EQME10447	18	0.057	0.001		0.056	0.001	
EQME10312	19	-0.040	0.018		-0.055	0.001	
EQME10313	20	0.041	0.012		0.026	0.108	
EQME21232	21	-0.119	0.006	C-	-0.100	0.018	C-
EQME10320	NR	-0.035	0.033		-0.061	0	
EQME21251	NR	0.023	0.187		0.018	0.295	
EQME10438	NR	0.007	0.685		0.013	0.427	
EQME10433	NR	0.129	0	C+	0.143	0	C+
EQME10346	NR	-0.120	0	C-	-0.116	0	C-
EQME10361	NR	0.133	0	C+	0.135	0	C+
EQME10384	NR	0.020	0.251		0.024	0.177	
EQME21256	NR	-0.171	0		-0.084	0.031	
EQME21275	NR	-0.396	0	C-	-0.358	0	C-
EQME21284	NR	-0.271	0	C-	-0.286	0	C-

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.76 ELL-Based DIF Statistics: Grade 9 Academic Mathematics, Spring Administration (English)

Item Code	Sequence	Sample 1			Sample 2		
		$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQME10332	1	-0.058	0		-0.055	0	
EQME10422	2	-0.061	0	B-	-0.064	0	B-
EQME10390	3	0.054	0		0.060	0	
EQME21213	4	-0.042	0.258		-0.005	0.900	
EQME21222	5	0.107	0	C+	0.122	0	C+
EQME10352	6	-0.003	0.827		0.009	0.554	
EQME10388	7	0.097	0	C+	0.093	0	C+
EQME10311	8	0.050	0		0.040	0	
EQME21219	9	0.038	0.233		-0.001	0.971	
EQME10360	10	0.001	0.921		-0.003	0.826	
EQME10548	11	0.040	0.008		0.038	0.012	
EQME10341	12	0.059	0		0.056	0	
EQME10472	13	-0.003	0.819		-0.012	0.354	
EQME10395	14	-0.117	0		-0.087	0	
EQME21276	15	-0.266	0	C-	-0.260	0	C-
EQME10338	16	-0.037	0.011		-0.015	0.314	
EQME10469	17	0.030	0.036		0.036	0.011	
EQME10314	18	0.021	0.147		0.016	0.272	
EQME10312	19	-0.026	0.073		-0.051	0	
EQME10313	20	0.020	0.144		-0.005	0.742	
EQME21241	21	-0.256	0	C-	-0.272	0	C-
EQME10421	NR	-0.099	0	C-	-0.112	0	C-
EQME21264	NR	-0.113	0	C-	-0.110	0	C-
EQME21274	NR	0	0.990		0.004	0.737	
EQME10413	NR	0.018	0.109		0.019	0.105	
EQME10367	NR	0.120	0	C+	0.108	0	C+
EQME21230	NR	0.060	0		0.046	0.003	
EQME21254	NR	-0.097	0	C-	-0.088	0	C-
EQME21259	NR	0.281	0	C+	0.389	0	C+
EQME21239	NR	0.172	0	C+	0.158	0	C+
EQME10476	NR	0.003	0.918		-0.023	0.467	

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Ontario Secondary School Literacy Test (OSSLT)

Classical Item Statistics and IRT Item Parameters

Table 7.1.77 Item Statistics: OSSLT (English)

Item Code	Section	Sequence	Cognitive Skill	Answer Key/Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
EQRE21425.116	I	1	R1	4	0.93	0.34	-2.39
EQRE21420.116	I	2	R1	2	0.68	0.15	-0.08
EQRE21426.116	I	3	R2	1	0.69	0.36	-0.25
EQRE21421.116	I	4	R2	2	0.77	0.38	-0.79
EQRE21424.116	I	5	R3	2	0.79	0.35	-0.89
EQRE21427.116	I	6	R3	3*	0.85 (2.56)	0.35	-1.73
EQWE20346	II	1	W3	2	0.86	0.36	-1.64
EQWE20359	II	2	W3	2	0.79	0.38	-1.02
EQWE20339	II	3	W2	1	0.96	0.31	-2.99
EQWE20333	II	4	W1	1	0.82	0.32	-1.17
EQWE20375T	III	1	W4	3*	0.86 (2.57)	0.37	-1.83
EQWE20375V	III	1	W3	2*	0.99 (1.97)	0.21	-2.87
1W3_057T	IV	1	W4	6*	0.75 (4.49)	0.57	-0.84
1W3_057V	IV	1	W3	4*	0.85 (3.39)	0.54	-1.55
EQRE21388.113	V	1	R3	3	0.89	0.24	-1.75
EQRE21391.113	V	2	R2	1	0.82	0.39	-1.14
EQRE21394.113	V	3	R1	1	0.51	0.33	0.77
EQRE21385.113	V	4	R1	4	0.74	0.29	-0.51
EQRE21395.113	V	5	R2	3	0.79	0.40	-0.92
EQRE21396.113	V	6	R2	1	0.87	0.26	-1.62
EQRE21398.113	V	7	R2	3*	0.76 (2.29)	0.36	-0.90
1W2_054T	VI	1	W4	6*	0.74 (4.46)	0.57	-1.04
1W2_054V	VI	1	W3	4*	0.79 (3.14)	0.52	-1.49
EQWE20358	VII	1	W3	2	0.75	0.38	-0.62
EQWE20354	VII	2	W3	2	0.72	0.33	-0.42
EQWE20344	VII	3	W2	1	0.92	0.39	-2.31
EQWE20336	VII	4	W1	3	0.86	0.33	-1.47
ES6_0028_6101	VIII	1	R1	3	0.73	0.24	-0.41
ES6_0028_6202	VIII	2	R3	4	0.87	0.33	-1.61
ES6_0028_6203	VIII	3	R2	3	0.61	0.27	0.29
ES6_0028_6204	VIII	4	R2	3	0.62	0.31	0.19
ES6_0028_6105	VIII	5	R3	2	0.73	0.30	-0.47
ES6_0028_6210	VIII	6	R2	2	0.69	0.37	-0.27
ES6_0028_6108	VIII	7	R2	4	0.68	0.25	-0.09
ES6_0028_6209	VIII	8	R1	2	0.65	0.41	-0.03
ES6_0028_6111	VIII	9	R2	2	0.83	0.33	-1.27
EQRE21464.120	IX	1	R2	3	0.97	0.19	-3.16
EQRE21468.120	IX	2	R2	2	0.93	0.26	-2.41
EQRE21465.120	IX	3	R3	4	0.75	0.27	-0.57

Table 7.1.77 Item Statistics: OSSLT (English) (continued)

Item Code	Section	Sequence	Cognitive Skill	Answer Key/Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
EQRE21467.120	IX	4	R2	4	0.73	0.24	-0.39
EQRE21471.120	IX	5	R2	3	0.78	0.32	-0.80
EQRE21474.120	IX	6	R2	3*	0.86 (2.58)	0.49	-1.61
EQRE21475.120	IX	7	R3	3*	0.87 (2.62)	0.30	-1.52
EQWE20381T	X	1	W4	3*	0.86 (2.59)	0.48	-1.52
EQWE20381V	X	1	W3	2*	0.98 (1.95)	0.34	-2.33
EQRE21584.128	XI	1	R3	2	0.74	0.38	-0.59
EQRE21586.128	XI	2	R2	3	0.81	0.36	-1.09
EQRE21583.128	XI	3	R1	4	0.83	0.31	-1.30
EQRE21587.128	XI	4	R2	4	0.65	0.33	-0.03
EQRE21582.128	XI	5	R2	4	0.76	0.35	-0.72
EQRE21588.128	XI	6	R2	1	0.79	0.29	-0.89

Note. The slope parameter was set at 0.588 for all items, and the guessing parameter was set at 0.20 for all multiple-choice items.

¹ = Answer key for multiple-choice items and maximum score category for open-response items.

* = Open-response items (OR, SW and LW); () = Mean score for open-response items.

R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections;

W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development.

Table 7.1.78 Distribution of Score Points and Step Parameters for Open-Response Reading and Short-Writing Items: OSSLT (English)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1	Score 2	Score 3
EQRE21427.116	I	6	% of Students	0.36	0.73	0.04	4.23	31.69	62.95
			Parameters				-2.37	-2.44	-0.39
EQWE20375T	III	1	% of Students	0.28	0.57	0.01	4.86	30.84	63.45
			Parameters				-2.78	-2.30	-0.43
EQWE20375V	III	1	% of Students	0.28	0.57	0.01	1.20	97.94	
			Parameters				-1.32	-4.42	
EQRE21398.113	V	7	% of Students	0.90	3.82	0.12	17.53	21.76	55.88
			Parameters				-1.89	-0.38	-0.44
EQRE21474.120	IX	6	% of Students	0.65	0.84	0.03	8.85	19.99	69.64
			Parameters				-2.61	-1.25	-0.98
EQRE21475.120	IX	7	% of Students	0.37	1.73	0.06	3.73	24.41	69.70
			Parameters				-1.56	-2.23	-0.77
EQWE20381T	X	1	% of Students	0.71	1.22	0.04	6.11	23.13	68.79
			Parameters				-1.97	-1.76	-0.84
EQWE20381V	X	1	% of Students	0.71	1.22	0.04	1.50	96.53	
			Parameters				-0.43	-4.23	

Note. The total number of students is 142 329.

Table 7.1.79 Distribution of Score Points and Step Parameters for Long-Writing Items: OSSLT (English)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1.0	Score 1.5	Score 2.0	Score 2.5
1W3_057T	IV	1	% of Students	0.05	0.69	0.02	0.41	0.38	0.88	1.54
			Parameters				-1.21	-1.01	-1.78	-1.39
1W3_057V	IV	1	% of Students		0.69	0.02	0.16	0.13	1.85	6.23
			Parameters				-0.64*		-2.93	-2.19
1W2_054T	VI	1	% of Students	0.01	0.34	0.02	0.06	0.11	0.56	1.37
			Parameters				-0.61	-1.78	-2.79	-1.95
1W2_054V	VI	1	% of Students		0.34	0.02	0.12	0.36	3.84	12.59
			Parameters				-0.93	-2.49	-3.35	-2.09

Note. The total number of students is 142 329.

* = Scores 1.0 and 1.5 have only one step parameter, since the two categories were collapsed.

(continued)

Item Code	Section	Sequence		Score 3.0	Score 3.5	Score 4.0	Score 4.5	Score 5.0	Score 5.5	Score 6.0
1W3_057T	IV	1	% of Students	3.91	9.05	19.24	23.19	19.06	13.92	7.65
			Parameters	-1.68	-1.46	-1.24	-0.53	0.02	0.29	0.75
1W3_057V	IV	1	% of Students	27.85	33.72	29.34				
			Parameters	-2.34	-0.88	-0.34				
1W2_054T	VI	1	% of Students	3.91	9.54	22.94	24.34	19.93	12.36	4.51
			Parameters	-2.00	-1.68	-1.50	-0.51	-0.06	0.38	1.09
1W2_054V	VI	1	% of Students	42.67	27.94	12.13				
			Parameters	-1.96	-0.13	0.49				

Note. The total number of students is 142 329.

Table 7.1.80 Item Statistics: OSSLT (French)

Item Code	Section	Sequence	Cognitive Skill	Answer Key/Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
OQLF20717.083	I	1	R2	1	0.93	0.17	-2.79
OQLF20712.083	I	2	R1	2	0.86	0.25	-1.97
OQLF20711.083	I	3	R1	1	0.97	0.15	-3.91
OQLF20714.083	I	4	R2	3	0.38	0.28	1.18
OQLF20719.083	I	5	R3	1	0.83	0.22	-1.68
OQLF20722.083	I	6	R3	3*	0.81 (2.44)	0.33	-2.19
OQEF20306	II	1	W1	4	0.87	0.27	-2.05
OQEF20307	II	2	W2	4	0.81	0.31	-1.55
OQEF20323	II	3	W3	2	0.88	0.34	-2.21
OQEF20343	II	4	W3	4	0.88	0.34	-2.18
OQEF20359T	III	1	W4	3*	0.89 (2.68)	0.39	-2.41
OQEF20359V	III	1	W3	2*	0.99 (1.98)	0.19	-3.31
2É3_024T	IV	1	W4	6*	0.71 (4.25)	0.57	-1.17
2É3_024V	IV	1	W3	4*	0.81 (3.23)	0.58	-1.88
OQLF20823.080	V	1	R2	1	0.68	0.44	-0.69
OQLF20680.080	V	2	R3	1	0.81	0.21	-1.47
OQLF20674.080	V	3	R2	3	0.73	0.29	-0.92
OQLF20672.080	V	4	R1	1	0.73	0.20	-0.91
OQLF20670.080	V	5	R1	3	0.95	0.28	-3.17
OQLF20824.080	V	6	R2	2	0.72	0.11	-0.79
OQLF20681.080	V	7	R2	3*	0.77 (2.30)	0.35	-1.38
2É2_025T	VI	1	W4	6*	0.65 (3.87)	0.55	-1.09
2É2_025V	VI	1	W3	4*	0.73 (2.91)	0.58	-1.76
OQEF20340	VII	1	W3	3	0.53	0.20	0.28
OQEF20305	VII	2	W2	4	0.73	0.36	-0.99
OQEF20293	VII	3	W1	1	0.91	0.30	-2.52
OQEF20337	VII	4	W3	2	0.47	0.31	0.55
OQLF20835.086	VIII	1	R2	4	0.73	0.21	-0.94
OQLF20756.086	VIII	2	R1	1	0.95	0.21	-3.29
OQLF20771.086	VIII	3	R2	3	0.88	0.28	-2.14
OQLF20759.086	VIII	4	R1	4	0.96	0.28	-3.65
OQLF20768.086	VIII	5	R2	2	0.92	0.11	-2.67
OQLF20765.086	VIII	6	R2	2	0.53	0.28	0.26
OQLF20774.086	VIII	7	R3	2	0.92	0.18	-2.58
OQLF20775.086	VIII	8	R3	4	0.70	0.31	-0.80
OQLF20772.086	VIII	9	R2	4	0.78	0.34	-1.32
OQLF20745.085	IX	1	R2	4	0.88	0.31	-2.18
OQLF20746.085	IX	2	R2	3	0.85	0.28	-1.86
OQLF20749.085	IX	3	R3	3	0.87	0.32	-2.04
OQLF20750.085	IX	4	R2	1	0.86	0.33	-1.96
OQLF20741.085	IX	5	R2	4	0.71	0.29	-0.82
OQLF20751.085	IX	6	R2	3*	0.69 (2.06)	0.44	-1.14

Table 7.1.80 Item Statistics: OSSLT (French) (continued)

Item Code	Section	Sequence	Cognitive Skill	Answer Key/Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
OQLF20754.085	IX	7	R3	3*	0.75 (2.25)	0.50	-1.47
OQEF20379T	X	1	W4	3*	0.87 (2.61)	0.34	-1.92
OQEF20379V	X	1	W3	2*	0.98 (1.95)	0.21	-2.67
OQLF20821.089	XI	1	R3	4	0.66	0.32	-0.50
OQLF20810.089	XI	2	R1	4	0.96	0.20	-3.64
OQLF20820.089	XI	3	R2	1	0.51	0.24	0.39
OQLF20815.089	XI	4	R2	3	0.94	0.31	-3.05
OQLF20844.089	XI	5	R2	3	0.85	0.31	-1.87
OQLF20812.089	XI	6	R2	2	0.78	0.26	-1.31

Note. The slope parameter was set at 0.588 for all items, and the guessing parameter was set at 0.20 for all multiple-choice items.

¹ = Answer key for multiple-choice items and maximum score category for open-response items.

* = Open-response items (OR, SW and LW); () = Mean score for open-response items.

R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections;

W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development.

Table 7.1.81 Distribution of Score Points and Step Parameters for Open-Response Reading and Short-Writing Items: OSSLT (French)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1	Score 2	Score 3
OQLF20722.083	I	6	% of Students	0.51	0.16		5.37	43.46	50.49
			Parameters				-3.48	-2.85	-0.24
OQEF20359T	III	1	% of Students	0.20	0.49	0.02	1.22	27.51	70.55
			Parameters				-1.85	-4.13	-1.24
OQEF20359V	III	1	% of Students	0.20	0.49	0.02	0.99	98.30	N/A
			Parameters				-1.46	-5.17	N/A
OQLF20681.080	V	7	% of Students	0.88	3.58	0.04	13.18	29.96	52.37
			Parameters				-2.11	-1.42	-0.61
OQLF20751.085	IX	6	% of Students	3.74	0.95	0.02	16.87	46.33	32.09
			Parameters				-2.37	-1.49	0.45
OQLF20754.085	IX	7	% of Students	1.57	1.53	0.07	10.00	45.91	40.91
			Parameters				-2.32	-2.18	0.10
OQEF20379T	X	1	% of Students	0.13	1.97	0.20	0.82	30.03	66.85
			Parameters				-0.25	-4.47	-1.05
OQEF20379V	X	1	% of Students	0.13	1.97	0.20	0.68	97.02	N/A
			Parameters				0.25	-5.59	N/A

Note. The total number of students is 5478.

Table 7.1.82 Distribution of Score Points and Step Parameters for Long-Writing Items: OSSLT (French)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1.0	Score 1.5	Score 2.0	Score 2.5
2É3_024T	IV	1	% of Students	0.04	0.38	0.02	0.62	0.31	1.19	2.30
			Parameters				-2.96	-1.49	-1.82	N/A
2É3_024V	IV	1	% of Students		0.38	0.02	0.33	0.46	2.81	11.52
			Parameters				-2.01*		-2.89	-2.78
2É2_025T	VI	1	% of Students		0.27	0.02	0.16	0.51	2.57	5.39
			Parameters				-3.17	-2.62	-1.86	N/A
2É2_025V	VI	1	% of Students		0.27	0.02	1.06	3.34	12.41	22.18
			Parameters				-3.03	-2.77	-2.74	-1.77

Note. The total number of students is 5478.

* = Scores 1.0 and 1.5 have only one step parameter, since the two categories were collapsed.

(continued)

Item Code	Section	Sequence		Score 3.0	Score 3.5	Score 4.0	Score 4.5	Score 5.0	Score 5.5	Score 6.0
2É3_024T	IV	1	% of Students	7.74	15.02	21.74	22.51	10.86	12.36	4.91
			Parameters	-2.35	-1.61	-1.19	-0.70	0.17	-0.50	0.70
2É3_024V	IV	1	% of Students	36.04	28.66	19.79	N/A	N/A	N/A	N/A
			Parameters	-2.36	-0.81	-0.46	N/A	N/A	N/A	N/A
2É2_025T	VI	1	% of Students	19.97	20.92	15.37	15.99	9.16	6.06	3.60
			Parameters	-2.24	-0.89	-0.34	-0.58	0.19	0.17	0.45
2É2_025V	VI	1	% of Students	27.51	20.72	12.49	N/A	N/A	N/A	N/A
			Parameters	-1.24	-0.56	-0.17	N/A	N/A	N/A	N/A

Note. The total number of students is 5478.

Differential Item Functioning (DIF) Analysis Results

The gender- and ELL-based DIF results for the OSSLT are provided in Tables 7.1.83 and 7.1.85. The results in each table are from two random samples of 2000, except for the French-language version of the assessment. Each table indicates the value of $\hat{\beta}$, the significance level and the category of the effect size for each item for B- and C-level DIF items. For gender-based DIF, negative estimates of $\hat{\beta}$ indicate that the girls outperformed the boys; positive $\hat{\beta}$ estimates indicate that the boys outperformed the girls. For ELL-based DIF, negative estimates of $\hat{\beta}$ indicate that the ELLs outperformed the non-ELLs; positive $\hat{\beta}$ estimates indicate that the non-ELLs outperformed the ELLs. DIF analysis was not conducted for ALF/PDF students for the French-language version of the OSSLT due to small populations.

Table 7.1.83 Gender-Based DIF Statistics: OSSLT (English)

Item Code	Section	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	p-value	Level of DIF	$\hat{\beta}$	p-value	Level of DIF
EQRE21425.116	I	1	0.005	0.495		-0.003	0.679	
EQRE21420.116	I	2	0.041	0.005		0.006	0.667	
EQRE21426.116	I	3	-0.010	0.475		-0.005	0.742	
EQRE21421.116	I	4	0.016	0.187		0.030	0.016	
EQRE21424.116	I	5	0.004	0.728		0.035	0.004	
EQRE21427.116	I	6	0.090	0.000	B+	0.055	0.006	
EQWE20346	II	1	-0.067	0.000	B-	-0.076	0.000	B-
EQWE20359	II	2	-0.073	0.000	B-	-0.074	0.000	B-
EQWE20339	II	3	0.009	0.081		0.000	0.971	
EQWE20333	II	4	0.024	0.037		-0.004	0.720	
EQWE20375T	III	1	0.034	0.059		0.035	0.064	
EQWE20375V	III	1	-0.004	0.449		0.004	0.514	
1W3_057T	IV	1	0.007	0.783		0.002	0.948	
1W3_057V	IV	1	-0.006	0.732		-0.049	0.005	
EQRE21388.113	V	1	-0.013	0.177		-0.006	0.587	
EQRE21391.113	V	2	-0.004	0.717		-0.001	0.937	
EQRE21394.113	V	3	-0.083	0.000	B-	-0.071	0.000	B-
EQRE21385.113	V	4	0.027	0.043		0.008	0.571	
EQRE21395.113	V	5	-0.041	0.001		-0.051	0.000	
EQRE21396.113	V	6	0.002	0.857		0.017	0.089	
EQRE21398.113	V	7	0.009	0.742		0.030	0.286	
1W2_054T	VI	1	0.044	0.053		-0.030	0.185	
1W2_054V	VI	1	0.017	0.310		-0.013	0.430	
EQWE20358	VII	1	-0.003	0.799		0.008	0.541	
EQWE20354	VII	2	-0.073	0.000	B-	-0.080	0.000	B-
EQWE20344	VII	3	-0.025	0.002		-0.046	0.000	
EQWE20336	VII	4	-0.023	0.026		-0.034	0.002	
ES6_0028_6101	VIII	1	0.025	0.074		0.008	0.585	
ES6_0028_6202	VIII	2	0.046	0.000		0.038	0.000	
ES6_0028_6203	VIII	3	0.042	0.005		0.006	0.667	
ES6_0028_6204	VIII	4	0.005	0.744		0.003	0.864	
ES6_0028_6105	VIII	5	-0.045	0.001		-0.047	0.001	
ES6_0028_6210	VIII	6	0.027	0.049		0.030	0.030	
ES6_0028_6108	VIII	7	0.119	0.000	C+	0.097	0.000	C+
ES6_0028_6209	VIII	8	0.044	0.002		0.046	0.001	
ES6_0028_6111	VIII	9	-0.044	0.000		-0.035	0.003	
EQRE21464.120	IX	1	-0.011	0.041		-0.018	0.003	
EQRE21468.120	IX	2	-0.018	0.018		-0.004	0.652	
EQRE21465.120	IX	3	-0.105	0.000	C-	-0.078	0.000	B-
EQRE21467.120	IX	4	0.052	0.000		0.050	0.000	
EQRE21471.120	IX	5	-0.074	0.000	B-	-0.056	0.000	
EQRE21474.120	IX	6	0.013	0.510		-0.009	0.666	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.83 Gender-Based DIF Statistics: OSSLT (English) (continued)

Item Code	Section	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQRE21475.120	IX	7	-0.045	0.017		-0.070	0.001	B-
EQWE20381T	X	1	-0.001	0.970		0.045	0.029	
EQWE20381V	X	1	0.012	0.108		0.033	0.001	
EQRE21584.128	XI	1	0.096	0.000	C+	0.112	0.000	C+
EQRE21586.128	XI	2	0.041	0.000		0.048	0.000	
EQRE21583.128	XI	3	-0.060	0.000	B-	-0.046	0.000	
EQRE21587.128	XI	4	0.073	0.000	B+	0.036	0.013	
EQRE21582.128	XI	5	0.052	0.000		0.059	0.000	B+
EQRE21588.128	XI	6	0.035	0.005		0.059	0.000	B+

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.84 Gender-Based DIF Statistics: OSSLT (French)

Item Code	Section	Sequence	Sample 1		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQLF20717.083	I	1	0.004	0.665	
OQLF20712.083	I	2	0.015	0.156	
OQLF20711.083	I	3	-0.003	0.553	
OQLF20714.083	I	4	0.079	0.000	B+
OQLF20719.083	I	5	0.028	0.017	
OQLF20722.083	I	6	0.173	0.000	C+
OQEF20306	II	1	-0.028	0.011	
OQEF20307	II	2	0.020	0.096	
OQEF20323	II	3	-0.049	0.000	
OQEF20343	II	4	-0.062	0.000	B-
OQEF20359T	III	1	0.087	0.000	B+
OQEF20359V	III	1	0.005	0.411	
2É3_024T	IV	1	-0.013	0.646	
2É3_024V	IV	1	-0.055	0.001	
OQLF20823.080	V	1	0.080	0.000	B+
OQLF20680.080	V	2	0.018	0.146	
OQLF20674.080	V	3	-0.022	0.109	
OQLF20672.080	V	4	-0.031	0.031	
OQLF20670.080	V	5	0.005	0.500	
OQLF20824.080	V	6	0.013	0.353	
OQLF20681.080	V	7	-0.038	0.154	
2É2_025T	VI	1	0.008	0.778	
2É2_025V	VI	1	-0.083	0.000	B-
OQEF20340	VII	1	-0.011	0.474	
OQEF20305	VII	2	-0.041	0.003	
OQEF20293	VII	3	0.001	0.889	
OQEF20337	VII	4	0.008	0.593	
OQLF20835.086	VIII	1	0.005	0.748	
OQLF20756.086	VIII	2	-0.002	0.772	
OQLF20771.086	VIII	3	-0.008	0.421	
OQLF20759.086	VIII	4	0.005	0.350	
OQLF20768.086	VIII	5	0.010	0.264	
OQLF20765.086	VIII	6	0.021	0.164	
OQLF20774.086	VIII	7	-0.046	0.000	
OQLF20775.086	VIII	8	-0.016	0.255	
OQLF20772.086	VIII	9	-0.062	0.000	B-
OQLF20745.085	IX	1	-0.031	0.001	
OQLF20746.085	IX	2	0.031	0.006	
OQLF20749.085	IX	3	0.013	0.210	
OQLF20750.085	IX	4	0.009	0.412	
OQLF20741.085	IX	5	-0.015	0.296	
OQLF20751.085	IX	6	0.014	0.570	
OQLF20754.085	IX	7	-0.078	0.000	B-
OQEF20379T	X	1	0.099	0.000	B+

Table 7.1.84 Gender-Based DIF Statistics: OSSLT (French) (continued)

Item Code	Section	Sequence	Sample 1		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF
OQEF20379V	X	1	0.023	0.017	
OQLF20821.089	XI	1	0.024	0.094	
OQLF20810.089	XI	2	-0.003	0.576	
OQLF20820.089	XI	3	0.019	0.214	
OQLF20815.089	XI	4	-0.013	0.073	
OQLF20844.089	XI	5	0.010	0.371	
OQLF20812.089	XI	6	0.009	0.476	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students.

Table 7.1.85 ELL-Based DIF Statistics: OSSLT (English)

Item Code	Section	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQRE21425.116	I	1	-0.029	0.003		-0.017	0.094	
EQRE21420.116	I	2	-0.047	0.003		-0.015	0.370	
EQRE21426.116	I	3	0.042	0.005		0.022	0.159	
EQRE21421.116	I	4	0.027	0.053		-0.003	0.823	
EQRE21424.116	I	5	0.051	0.000		0.102	0.000	C+
EQRE21427.116	I	6	-0.071	0.000	B-	-0.098	0.000	B-
EQWE20346	II	1	0.005	0.654		-0.029	0.020	
EQWE20359	II	2	-0.073	0.000	B-	-0.084	0.000	B-
EQWE20339	II	3	0.133	0.000	C+	0.270	0.000	C+
EQWE20333	II	4	-0.029	0.025		-0.003	0.816	
EQWE20375T	III	1	-0.039	0.035		-0.107	0.000	C-
EQWE20375V	III	1	0.009	0.189		0.024	0.008	
1W3_057T	IV	1	0.048	0.078		0.080	0.008	B+
1W3_057V	IV	1	0.085	0.000	B+	0.178	0.000	C+
EQRE21388.113	V	1	0.014	0.211		0.062	0.000	B+
EQRE21391.113	V	2	0.000	0.979		0.103	0.000	C+
EQRE21394.113	V	3	-0.107	0.000	C-	-0.117	0.000	C-
EQRE21385.113	V	4	0.009	0.518		-0.059	0.000	B-
EQRE21395.113	V	5	-0.030	0.022		-0.029	0.037	
EQRE21396.113	V	6	0.012	0.313		0.030	0.016	
EQRE21398.113	V	7	-0.278	0.000	C-	-0.406	0.000	C-
1W2_054T	VI	1	-0.160	0.000	C-	-0.213	0.000	C-
1W2_054V	VI	1	0.104	0.000	C+	0.217	0.000	C+
EQWE20358	VII	1	0.090	0.000	C+	0.208	0.000	C+
EQWE20354	VII	2	-0.010	0.507		-0.053	0.000	
EQWE20344	VII	3	0.002	0.853		0.029	0.010	
EQWE20336	VII	4	0.068	0.000	B+	0.043	0.001	
ES6_0028_6101	VIII	1	-0.038	0.010		-0.080	0.000	B-
ES6_0028_6202	VIII	2	-0.013	0.263		-0.003	0.812	
ES6_0028_6203	VIII	3	0.007	0.639		-0.003	0.834	
ES6_0028_6204	VIII	4	-0.015	0.336		-0.028	0.085	
ES6_0028_6105	VIII	5	0.014	0.341		0.037	0.020	
ES6_0028_6210	VIII	6	-0.096	0.000	C-	-0.105	0.000	C-
ES6_0028_6108	VIII	7	-0.036	0.016		0.001	0.951	
ES6_0028_6209	VIII	8	-0.032	0.031		-0.039	0.007	
ES6_0028_6111	VIII	9	-0.027	0.028		-0.028	0.034	
EQRE21464.120	IX	1	0.011	0.103		0.014	0.067	
EQRE21468.120	IX	2	0.030	0.001		0.040	0.000	
EQRE21465.120	IX	3	0.123	0.000	C+	0.105	0.000	C+
EQRE21467.120	IX	4	-0.023	0.134		-0.051	0.001	
EQRE21471.120	IX	5	0.023	0.096		0.057	0.000	
EQRE21474.120	IX	6	0.053	0.012		-0.009	0.696	
EQRE21475.120	IX	7	0.030	0.152		0.026	0.220	

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students.

Table 7.1.85 ELL-Based DIF Statistics: OSSLT (English) (continued)

Item Code	Section	Sequence	Sample 1			Sample 2		
			$\hat{\beta}$	<i>p</i> -value	Level of DIF	$\hat{\beta}$	<i>p</i> -value	Level of DIF
EQWE20381T	X	1	0.075	0.000	B+	0.104	0.000	C+
EQWE20381V	X	1	0.067	0.000	B+	0.139	0.000	C+
EQRE21584.128	XI	1	0.071	0.000	B+	0.013	0.352	
EQRE21586.128	XI	2	-0.051	0.000		-0.066	0.000	B-
EQRE21583.128	XI	3	-0.078	0.000	B-	-0.085	0.000	B-
EQRE21587.128	XI	4	-0.087	0.000	B-	-0.138	0.000	C-
EQRE21582.128	XI	5	0.036	0.009		-0.024	0.101	
EQRE21588.128	XI	6	0.028	0.041		0.014	0.344	

Note. B = moderate DIF; C = large DIF; - = favouring ELL students; + = favouring non-ELL students..