

EQAO's Technical Report

for the 2010–2011 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*



About the Education Quality and Accountability Office

The Education Quality and Accountability Office (EQAO) is an independent provincial agency funded by the Government of Ontario. EQAO's mandate is to conduct province-wide tests at key points in every student's primary, junior and secondary education and report the results to educators, parents and the public.

EQAO acts as a catalyst for increasing the success of Ontario students by measuring their achievement in reading, writing and mathematics in relation to *Ontario Curriculum* expectations. The resulting data provide a gauge of quality and accountability in the Ontario education system.

The objective and reliable assessment results are evidence that adds to current knowledge about student learning and serves as an important tool for improvement at all levels: for individual students, schools, boards and the province.

EQAO's Technical Report for the 2010–2011 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

**Education Quality and
Accountability Office**



2 Carlton Street, Suite 1200, Toronto ON M5B 2M9

Telephone: 1-888-327-7377 Web site: www.eqao.com

© 2012 Queen's Printer for Ontario

TABLE OF CONTENTS

CHAPTER 1: OVERVIEW OF THE ASSESSMENT PROGRAMS	1
THE EQAO ASSESSMENT PROGRAM: PRIMARY (GRADES 1–3), JUNIOR (GRADES 4–6), GRADE 9 AND THE ONTARIO SECONDARY SCHOOL LITERACY TEST	1
CHAPTER 2: ASSESSMENT DESIGN AND DEVELOPMENT	3
ASSESSMENT FRAMEWORKS	3
ASSESSMENT BLUEPRINTS.....	3
TEST CONSTRUCTION: SELECTING ITEMS FOR THE OPERATIONAL FORM	3
ITEM DEVELOPMENT	3
Item Developers	4
Training for Item Developers	4
EQAO Education Officer Review	5
Item Tryouts	5
THE ASSESSMENT DEVELOPMENT AND SENSITIVITY REVIEW COMMITTEES	5
The EQAO Assessment Development Committees	6
The EQAO Sensitivity Committee	6
FIELD TESTING	6
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 and Special Provisions.....	9
EQAO Policies and Procedures	10
QUALITY ASSURANCE	11
CHAPTER 4: SCORING	12
THE RANGE-FINDING PROCESS	12
Pre-Range Finding.....	12
Range Finding	13
Overview of the Range-Finding Process	14
PREPARING TRAINING MATERIALS FOR SCORING	14
FIELD-TEST SCORING	14
Training Field-Test Scoring Leaders and Scorers	14
Scoring Open-Response Field-Test Items	15
Developing Additional Scorer-Training Materials Before Scoring Operational Items	15
SCORING OPEN-RESPONSE OPERATIONAL ITEMS	15
Scoring Rooms for Scoring Open-Response Operational Items.....	16
Training for Scoring Open-Response Operational Items.....	16
Training of Scoring Leaders and Scoring Supervisors for Scoring Open-Response Operational Items	16
Training of Scorers for Scoring Open-Response Operational Items	17
PROCEDURES AT THE SCORING SITE.....	18
Students at Risk	18
Inappropriate Content, Cheating and Other Issues	19
Ongoing Daily Training	19
Daily Scoring-Centre Reports for Monitoring the Quality of Open-Response Item Scoring	20
Required Actions: Consequences of the Review and Analysis of Daily Scoring-Centre Data Reports	21
Auditing.....	22
SCORER VALIDITY AND RELIABILITY	22
Scoring Validity	22
Scorer Reliability.....	24

CHAPTER 5: EQUATING	25
IRT MODELS	25
EQUATING DESIGN	26
EQUATING SAMPLE	26
CALIBRATION SAMPLE	27
THE ASSESSMENTS OF READING, WRITING AND MATHEMATICS: PRIMARY AND JUNIOR DIVISIONS.....	27
Description of the IRT Model	27
Equating Sample: Exclusion Rules.....	27
Equating Steps.....	28
Eliminating Items and Collapsing of Score Categories	29
Equating Results.....	29
THE GRADE 9 ASSESSMENT OF MATHEMATICS	32
Description of the IRT Model	32
Equating Sample.....	32
Equating Steps.....	33
Eliminating Items and Collapsing of Score Categories	33
Equating Results.....	33
THE ONTARIO SECONDARY SCHOOL LITERACY TEST (OSSLT).....	34
Description of the IRT Model	34
Equating Sample.....	35
Equating Steps.....	35
Scale Score	35
Eliminating Items and Collapsing of Score Categories	35
Equating Results.....	36
REFERENCES	36
CHAPTER 6: REPORTING RESULTS	37
REPORTING THE RESULTS OF THE ASSESSMENTS OF READING, WRITING AND MATHEMATICS, PRIMARY AND JUNIOR DIVISIONS	37
REPORTING THE RESULTS OF THE GRADE 9 ASSESSMENT OF MATHEMATICS	38
REPORTING THE RESULTS OF THE OSSLT	39
INTERPRETATION GUIDES	39
CHAPTER 7: STATISTICAL AND PSYCHOMETRIC SUMMARIES	41
THE ASSESSMENTS OF READING, WRITING AND MATHEMATICS, PRIMARY AND JUNIOR DIVISIONS.....	41
Classical Test Theory (CTT) Analysis	41
Item Response Theory (IRT) Analysis	43
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)	56
THE GRADE 9 ASSESSMENT OF MATHEMATICS	58
Classical Test Theory (CTT) Analysis	58
Item Response Theory (IRT) Analysis	59
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)	64
THE ONTARIO SECONDARY SCHOOL LITERACY TEST (OSSLT).....	65
Classical Test Theory (CTT) Analysis	65
Item Response Theory (IRT) Analysis	65
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)	68
DIFFERENTIAL ITEM FUNCTIONING (DIF)	68
The Primary- and Junior-Division Assessments.....	70
The Grade 9 Mathematics Assessment.....	72
The OSSLT.....	74
DECISION ACCURACY AND CONSISTENCY	75
Accuracy.....	75
Consistency	75
Estimation from One Test Form.....	76
The Primary and Junior Assessments.....	77
The Grade 9 Assessment of Mathematics	78
The OSSLT.....	78
REFERENCES	78

CHAPTER 8: VALIDITY EVIDENCE	79
INTRODUCTION.....	79
The Purposes of EQAO Assessments.....	79
Conceptual Framework for the Validity Argument.....	79
VALIDITY EVIDENCE BASED ON THE CONTENT OF THE ASSESSMENTS AND THE ASSESSMENT PROCESSES...	80
Test Specifications for EQAO Assessments.....	80
Appropriateness of Test Questions.....	80
Quality Assurance in Administration	81
Scoring of Open-response Items.....	81
Equating	82
VALIDITY EVIDENCE BASED ON THE TEST CONSTRUCTS AND INTERNAL STRUCTURE.....	82
Test Dimensionality	82
Technical Quality of the Assessments	82
VALIDITY EVIDENCE BASED ON EXTERNAL ASSESSMENT DATA	83
Linkages to International Assessment Programs	83
VALIDITY EVIDENCE SUPPORTING APPROPRIATE INTERPRETATIONS OF RESULTS.....	84
Setting Standards	84
Reporting	84
REFERENCES	85
APPENDIX 4.1: SCORING VALIDITY AND INTERRATER RELIABILITY	86
APPENDIX 7.1: SCORE DISTRIBUTIONS AND ITEM STATISTICS FOR THE 2011 EQAO ASSESSMENTS	100

CHAPTER 1: OVERVIEW OF THE ASSESSMENT PROGRAMS

The EQAO Assessment Program: Primary (Grades 1–3), Junior (Grades 4–6), Grade 9 and the 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*. For example, students enrolled in inspected private schools are among those who write the OSSLT, as it is a graduation requirement for all students who wish to receive the Ontario Secondary School Diploma (OSSD).

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, which are available on its Web site. The assessment results provide valuable information that supports 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. A parallel form of the Grade 9 Assessment of Mathematics is developed for each of the two courses (i.e., academic and applied), with one form administered toward the end of the first semester and the second 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 the 32 requirements for the OSSD.

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

- design and 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:

- diversity (cultural) and geographic location (to represent the northern, southern, eastern and western parts of the province);
- the representation of rural and urban regions;
- current elementary and secondary experience (teachers, administrators, subject experts and consultants);
- expertise in assessment, evaluation and large-scale assessment and
- previous experience working with EQAO.

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 it relates to *Ontario Curriculum* expectations. The English-language and French-language frameworks for the four EQAO assessments can be found on 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 relationship to *Ontario Curriculum* expectations (or “curriculum coverage”) 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.

Test Construction: Selecting Items for the Operational Form

Operational items are selected from the items that have been field tested in previous assessments. The collected operational items in an assessment constitute the operational form (or “operational assessment” or “operational test”). The operational form contains the items that are scored for inclusion in the reporting of student results. Field-test items do not count toward a student’s result. Several important factors are taken into consideration when 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:** It is important to note that while items are mapped to clusters of curriculum expectations, 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 www.eqao.com.

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 applicable grade levels;

- developing multiple-choice and open-response reading and writing or 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 for or 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 English and French language (reading and writing) and mathematics to participate in its item-writing committees. The item-writing committee for each assessment comprises 10–20 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 and writing or mathematics; open-response items in reading or 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 English and French language (reading and writing) or mathematics education, or both;
- familiarity with and knowledge of the elementary or secondary school curricula in Ontario (especially in language or mathematics);
- familiarity with the cross-curricular literacy requirements for elementary and secondary education in Ontario (especially for the OSSLT);
- expertise and experience in the application of elementary and secondary literacy and mathematics rubrics based on the achievement charts in *The Ontario Curriculum* (to identify varying levels of student performance);
- excellent written communication skills;
- comfort using computer software (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 materials for 2010–2011 were developed by EQAO in partnership with educators from across Ontario. EQAO led a two-day workshop for item developers and spent half a day introducing item developers to the criteria for item writing. EQAO provided an overview of the assessments, including a description of the frameworks, and provided details on the elements of effective item writing. The remaining time involved a guided item-writing session structured by EQAO education officers.

Each item developer was assigned to write items based on the blueprint for the specific assessment. After the session, the draft materials were reviewed by education officers and returned to the developers, who conducted cognitive labs of the items (i.e., trials with their own students). The results of the cognitive labs were discussed, and further changes to the items were made in a second two-day workshop. EQAO further refined and edited the items after the workshop and prepared the items for committee reviews in preparation for selection of items for the matrix field test.

EQAO Education Officer Review

When the first draft of the items and item-specific scoring rubrics is developed by the item developers, the items and rubrics are reviewed by EQAO education officers. The education officers ensure that each item is referenced correctly in terms of curriculum expectations and difficulty levels. For the multiple-choice items, the education officers consider the clarity and completeness of the stem, the integrity 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 and determine the scorability of the items.

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, with students, the items they have developed. 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.

The Assessment Development and Sensitivity Review Committees

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

The EQAO Assessment Development Committees

The Assessment Development Committee for each subject in each assessment comprises 10–12 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:

- language or mathematics in the primary division (Grades 1–3) for the primary assessment, administered in Grade 3;
- language or mathematics in the junior division (Grades 4–6) for the junior assessment, administered in Grade 6;
- mathematics in the intermediate division (Grades 7–10) for the Grade 9 assessment, administered in Grade 9 or
- literacy across the curriculum to the end of Grade 9 for the OSSLT, administered in Grade 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 revision of items.

The EQAO Sensitivity Committee

The Sensitivity Committee, which considers all four EQAO assessments, comprises 20–24 Ontario educators who serve for terms of one to five years. About 4–8 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 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 assessment. Scores on the field-test items are not used in determining student, school, school board or provincial results. The field-test items are arranged in the student 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 each assessment, among the operational items, to ensure 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 similarity is meant to counter the low motivation that students may feel when they know that items are field-test items and therefore do not count toward their score. No more than 20% of the items in an assessment are field-test items.

All items, except for the long-writing tasks on the primary- and junior-division assessments and the OSSLT, are field tested this way. Because of the length of time required to complete long-writing tasks, they are not embedded as field-test items with operational items. Long-writing prompts go through a rigorous process of committee reviews, and, for the OSSLT, field trials are conducted as part of the item development process to ensure their appropriateness. Long-writing tasks are not used for equating.

Questionnaires

EQAO develops Student, Teacher and Principal Questionnaires to collect information on factors inside and outside the classroom that affect student achievement, so that EQAO results can be used to make recommendations to improve student learning.

The Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions, include Student, Teacher and Principal Questionnaires. The Student Questionnaires include questions about the following: student engagement in reading, writing and mathematics (attitudes, perceptions of performance/confidence, learning strategies, reading and writing outside school); use of instructional resources in the classroom (e.g., use of calculator, computer, Internet, dictionaries); home environment (e.g., time spent doing extra-curricular activities; “screen time,” language(s) spoken at home by student and by others); parental engagement (home discussion, participation in child’s education) and the number of schools attended.

The Teacher Questionnaires include questions about the following: school learning environment (e.g., staff collaboration, school improvement planning); use of EQAO resources and data; use of resources in the classroom (e.g., use of calculator, computer and Internet by students and use of diverse materials by teacher); parental engagement in student learning (e.g., frequency and purposes of communication with parents); teacher’s information (e.g., background, experience, professional development) and classroom demographics (e.g., size and grade levels in class). The Principal Questionnaire includes questions about the following: principal’s information (e.g., gender, experience and teaching assignment); the school learning environment (e.g., staff collaboration, school improvement planning); use of EQAO data; parental engagement in student learning (e.g., communication with parents and parental participation) and school demographics (grades taught, enrolment, average percentage of students absent per day).

The Grade 9 Assessment of Mathematics also includes Student and Teacher Questionnaires. The Student Questionnaires include questions on student engagement in mathematics (attitudes, perceptions of performance/confidence, learning goals, learning strategies, what they attribute

success in mathematics to); use of instructional resources in the classroom (e.g., use of calculator, computer, Internet in the classroom); time spent on mathematics homework; attendance record in mathematics class; home environment (e.g., time spent doing extra-curricular activities, “screen time,” language(s) spoken at home by student); parental engagement (home discussion, participation in child’s education); future expectations (levels of schooling their parents expect them to complete, levels they expect to complete) and the number of elementary schools attended.

The Teacher Questionnaire includes questions about the school learning environment (e.g., staff collaboration, school improvement planning); use of EQAO resources and data; use and availability of resources in the classroom (e.g., use of calculator, computer and Internet by students); use of instructional practices in the classroom; parental engagement in student learning (e.g., frequency and purposes of communication with parents) and teacher’s information (e.g., background, experience, professional development).

In 2010 and in 2011, questions about the use of EQAO Grade 9 mathematics results as part of students’ course marks were added to the Student and Teacher Questionnaires.

The OSSLT includes a Student Questionnaire that asks students about their access to a computer at home; the amount of time spent reading in English or French outside school and the different types of materials read outside school; their access to reading materials and the language spoken at home; and the time spent writing in English or French outside school and on the different forms of writing they do 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 publishes an administration guide and a guide for accommodations and special provisions for English language learners annually for each assessment. The guides can be found on 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 and Special Provisions

The guide 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 list of accommodations and special provisions does not address a student's needs, exemption from participation in an assessment is allowed (for the primary and junior assessments only). Each year, EQAO reviews and updates these accommodations and provisions to ensure that they reflect Ministry of Education guidelines and new developments in the support 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 guides are based on three Ontario Ministry of Education policy documents: *Individual Education Plans: Standards for Development, Program Planning, and Implementation*; *English Language Learners / ESL and ELD Programs and Services: Policies and Procedures for Ontario Elementary and Secondary Schools, Kindergarten to Grade 12* (2007); and *Policy/Program Memorandum 127* (for Grade 9 and OSSLT only), 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 its 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 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 version 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 booklets
- one single-sided hard copy to be scanned for use with assistive devices and technology such as text-to-speech software, plus the required sets of regular-print booklets

EQAO Policies and Procedures

This document outlines EQAO’s policies and procedures related to the assessments (e.g., Consistency and Fairness, Student Participation, Absences and Lateness, School Emergency, Teacher Absences, Marking of Student Work by Classroom Teachers [Grade 9 only] and Request for a Student to Write at an Alternative Location).

Special Provisions for English Language Learners

“Special provisions” are adjustments for English language learners to the setting or timing of an assessment. 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 an assessment, even given accommodations or special provisions, the student may be exempted at the discretion of his or her school principal. A Grade 3 or 6 student *must* be exempted, however, if for reading, a teacher or another adult must read to him or her and, for mathematics, if mathematics terms have to be defined for him or her.

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

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

Deferrals (OSSLT Only)

All Ontario secondary school students are expected to write the OSSLT in their Grade 10 year. However, this requirement can be deferred for one year (every year until graduation) when a student is working toward the 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 the 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 and requires accommodations that cannot yet be provided.

All deferred students who wish to graduate with the OSSD must eventually complete the OSSLT requirement.

If a student has attempted and has been unsuccessful at least once in the OSSLT, the principal has the discretion to allow the student to take the Ontario Secondary Literacy Course (OSSLC).

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 analysis of student response data. The first analysis identifies student response patterns to multiple-choice items 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 a school's rate of success.
- Examination of test materials: Following each assessment, EQAO looks for evidence of possible irregularities in its administration. 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 responses to multiple-choice items, except on the primary assessment, are captured by a scanner. For multiple-choice items on the primary assessment, students fill in the circle corresponding to their response, and their choices are double-keyed in manually.

Item-specific, generic 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 under the supervision of EQAO staff. 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 for each assessment are included in each framework document on 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 that 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 OSSLT are pilot tested with a limited number of students. As a result, range finding for long-writing tasks uses student responses to operational items and occurs just prior to operational scoring.

Pre-Range Finding

During pre-range finding, practising educators work with EQAO staff to select 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, though a few minor variations of this process occur across assessments and between field-test and operational range finding:

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 task and operational long-writing task 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 tasks 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 of mixed range. The booklets chosen for the piles represent the full range of student responses, including off-topic, incorrect, typical and unusual responses. The mixed pile is determined after the other three piles.
5. Items and tasks that have been left unanswered (“blanks”) or that are difficult to read due to poor handwriting or light ink are not selected for pre-range finding.
6. A cover sheet for each range, showing item, task and booklet numbers, is printed and labelled “high,” “medium,” “low” or “mixed.”

Range Finding

During the range-finding process, subject experts from the Ontario education system, under the supervision of EQAO staff, meet to make recommendations about 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 performance within the scoring rubric codes and to 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., as training papers, qualifying test responses and possible responses for calibration activities).

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.

Members of the range-finding committees

- use their scoring expertise to assign the appropriate generic rubric or item-specific rubric codes to a set of student responses for each group of assessment items;
- share the codes they have assigned with the other members of the committees;
- work collaboratively with the other members of the committees, 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 committees are often divided into groups of three or four members.
3. Each group discusses a set of items, prompts and associated item-specific and generic scoring rubrics and recommends appropriate responses to be used as anchors, training papers and qualifying test items to train scorers for each task. The discussions focus on
 - the content and requirements of each item or task;
 - group agreement on the scores/codes for student responses and
 - scoring rules, as required, to ensure consistent scoring of each item or task.

Preparing Training Materials for Scoring

EQAO education officers prepare materials to train scorers for scoring 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.

Training materials include

- generic and/or item-specific rubrics;
- anchors that are a good (or “solid”) representation of the codes in the scoring rubrics;
- training papers that represent both solid score-point responses and unusual responses (e.g., shorter than average, atypical approaches, a 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;
- responses to be used for ongoing training during the daily calibration activity (operational scoring only) and
- solid responses used for monitoring validity (operational scoring only).

Field-Test Scoring

Field-test scoring occurs following operational scoring. Since field-test items are to be used in future assessments, they are scored according to the same high standards applied to the scoring of operational items. To ensure the consistency of year-to-year scoring and to reduce the time required for training, the most reliable and productive scoring leaders and scorers of operational items are selected to score field-test items similar to the operational items they have already scored. Education officers arrange for sufficient copies of materials to train the scorers of field-test items. All training materials are kept secure.

Training Field-Test Scoring Leaders and Scorers

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

Scoring leaders for each scoring room (designated according to open-response reading and mathematics items and short-writing tasks) 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 task and its scoring rubric, anchors and training papers);
- a qualifying test on the first item or task (when field-test scoring does not immediately follow operational scoring) and
- an introduction to subsequent items and tasks and their scoring rubrics, anchors and training papers prior to scoring them.

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

Scoring Open-Response Field-Test Items

A sample of approximately 1200 demographically representative English- and 500 French-language student responses for each field-test item or prompt is scored. One exception is the Grade 9 French-language Mathematics Assessment whereby an average of 50 to 350 French-language student responses for each field-test item or prompt is scored. The French Grade 9 number of mathematics field-test items scored varies according to the number of students enrolled in the applied and academic courses.

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

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.

Developing Additional Scorer-Training Materials Before Scoring Operational Items

When the full range of training materials has not been used for field-test scoring of open-response reading and mathematics items or writing tasks, EQAO develops additional scoring materials using the original range-finding data or field-test scoring data. In the latter case, education officers collect student responses in bundles of high, medium, low and mixed range, so that range finders can select additional scorer-training materials (e.g., 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 are to score operational items.

Scoring Open-Response Operational Items

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

The primary, junior and Grade 9 assessments are scored by qualified Ontario educators. The primary and junior assessments are scored by educators representing all the 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 before the end of the school year. EQAO recruits as many teacher-scorers (i.e., members of the Ontario College of Teachers) as possible and fills the complement of required scorers with retired educators and qualified non-educators (or “other-degree scorers”).

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 to score the test effectively.

Scoring Rooms for Scoring Open-Response Operational Items

A set of the operational assessment items is scored in a scoring room under the leadership of a scoring leader. He or she trains all the 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 score. 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 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 assessments and operational mathematics items for the Grade 9 assessment are single scored. To measure overall scorer reliability (consistency), the scores from sets of validity booklets are used in scoring those items that are not part of the validity process based on validity papers or validity booklets (Grade 9). More information about scoring validity is provided later in this chapter.

In scoring, "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.

Each open-response reading item and writing task on the OSSLT is scored by two trained scorers independently, 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 tasks) 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.

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 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 applicable 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 room.

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. 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 assigned to one room at the same time.

Following training and prior to scoring, scoring leaders and scoring supervisors must pass a qualifying test that involves scoring 14–20 student responses for the items to be scored in their room. The items included in the qualifying test are selected during the range-finding process. Scoring leaders and supervisors must attain at least an 80% exact and a 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: generic or item-specific rubrics, anchors (real student responses illustrating work at each code in the rubric) and their annotations, training papers, a qualifying test, validity papers (primary, junior, OSSLT) or validity booklets (Grade 9) 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, as outlined below, are used during the scoring process to identify scorers who need to be retrained or dismissed from scoring.

All the scorers in one 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 its assessments.

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

- general instructions about the security, confidentiality and suitability of the scoring materials;
- instructions on entering scores into the Personal Digital Assistant (PDA) used to collect scoring data. For instance,
 - o prior to entering 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
 - o 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 a server;
- a thorough review and discussion of the scoring materials for each item to be scored (the item, generic or item-specific rubrics, anchors and their annotations):
 - o 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 and
 - o the anchors consist of responses that are typical of each achievement level (rather than unusual or uncommon) and solid (rather than controversial or "borderline") and
- the scoring of a series of validity papers or validity booklets (Grade 9), consisting of selected expertly scored student responses:
 - o validity papers or validity booklets (Grade 9) typically contain responses that are solid examples of student work for a given response code. Scorers will first score the responses and then synchronize the PDA and
 - o scorers will then discuss the attributes and results of each correct response with their scoring leader and supervisor. They 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 rather than 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, to avoid adjusting their scoring to take into account a characteristic they assume about a student (e.g., special education needs, being an English language learner).

Following training and prior to scoring, scorers must pass a qualifying test consisting of 14–20 student responses to all the items to be scored in a 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., response contains content that states or implies threats of violence to

oneself 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 the scoring site manager, 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 any such issues are sent to the exceptions room for consideration.

Offensive Content

Obscene, racist or sexist content in student response booklets is reviewed by EQAO staff to determine whether the school should be contacted. If the offensive content warrants it, EQAO will notify the school.

Cheating

When there is any evidence in a booklet that may indicate some form of irregularity (e.g., many changed answers, teacher interference), the booklet is reviewed by EQAO staff to determine whether the school should be notified. In cases where cheating is confirmed, no scores are provided for the student.

Damaged or Misprinted Booklets

In very few cases, booklets given to students are torn, 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 in these booklets should be pro-rated based on the results in booklets 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 rubrics and anchors. The purpose of the review is to refocus scorers and highlight any section of the rubrics that require 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 items. 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 papers. Scorers score up to 10 validity papers a day. The validity papers ensure that scorers are giving correct and accurate scores that compare to 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 are expected to maintain a minimum accuracy rate on the validity papers. The target 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, 65% exact and 95% exact-plus-adjacent agreement for five-point rubrics and 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 data reports summarize daily and cumulative levels of agreement (exact, adjacent and high or low non-adjacent agreement) on validity papers with pre-set 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 the scorers 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, re-scored.

Daily and Cumulative Mean-Score and Score-Point Distribution

Daily and cumulative mean-score and score-point distribution data reports are used to monitor room and individual scorer drift. They confirm validity and guide ongoing training (based on calibration items) at both the individual and room levels.

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

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 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.

The reports identify and summarize daily and cumulative levels of interrater agreement, 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.

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.

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, 65% on five-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
- productivity issues

Scoring leaders share the data and discuss data-related issues with the appropriate scoring supervisors so that interventions can be planned. 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.
- 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 and productivity in order to continue. If a scoring supervisor or scorer does not meet one or more of these 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.

Auditing

EQAO audits individual student score sheets (i.e., student records showing the scores assigned to selected open-response items) for inconsistencies that may indicate incomplete scoring. Any booklet scored entirely blank is rerouted for a second scoring.

Scorer Validity and Reliability

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

Scoring Validity

As described earlier in this chapter, scoring validity is assessed by having scorers assign scores to validity papers and validity booklets, which are student responses that have been scored by an expert panel. For the primary and junior assessments and for the OSSLT, a set of five validity papers is prepared, copied and distributed to all scorers each morning and afternoon. In addition, the original student booklets that these validity papers were copied from are used as blind validity booklets and circulated to provide additional validity material for the scorers. For Grade 9, only blind validity booklets are used, and they are circulated as frequently as possible so that most scorers can score at least ten validity booklets per day. The sets of validity papers are not used for Grade 9 because high levels of scorer consistency have been achieved over the years through the use of the blind validity booklets only.

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.

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

There are 10, three and eight open-response items for the reading, writing and mathematics components of the assessments, respectively. Four-point scoring rubrics are used for reading and mathematics. For the writing components, there are two short-writing prompts and one long-writing prompt 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 validity estimates for reading, writing and mathematics for the primary and junior divisions are presented in Tables 4.1.1–4.1.12 of Appendix 4.1. 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 separately.

In 2010–2011, the EQAO target of 95% exact-plus-adjacent agreement for validity was met for all but one item in the reading components, for all items in the writing components and for all but one item in the mathematics components of the assessments. The aggregate validity estimates for exact-plus-adjacent agreement ranged from 98.8–99.8%.

The 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 in mathematics courses in the first semester and in June for students in second-semester and full-year courses. The scoring validity estimates for the Grade 9 Assessment of Mathematics are presented in Tables 4.1.13–4.1.16 of Appendix 4.1 for both administrations. 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 both were scored during the same scoring session in July 2011. Seven questions were scored for each administration using four-point rubrics for a total of 56 questions across both administrations.

The EQAO target of 95% exact-plus-adjacent agreement for validity was met for all items and the aggregate validity estimates ranged from 98.9–100%.

The Ontario Secondary School Literacy Test (OSSLT)

The scoring validity estimates for the OSSLT are reported in Tables 4.1.17–4.1.20 of Appendix 4.1. For each test, four reading items were scored with three-point rubrics and two long-writing prompts were scored with a six-point rubric for topic development and a four-point rubric for conventions. Two short-writing prompts were scored with a three-point rubric for topic development and a two-point rubric for conventions, which were combined into a five-point rubric for the purposes of validity statistics. Aggregate statistics are provided separately for reading items, short-writing prompts, long-writing prompts and all writing prompts. In 2010–2011, the EQAO target of 95% exact-plus-adjacent agreement for validity was met for all reading and writing items on both the English-language and French-language tests. The aggregate validity estimates ranged from 96.7–99.9%.

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 for the primary, junior and Grade 9 assessments, EQAO tracks reliability using scores assigned to items in the pre-selected blind validity booklets that are not part of the validity process.

The process for determining the reliability of open-response scoring for the OSSLT does not require that selected student booklets be brought back to a scoring room (or “reinserted”). All student responses to open-response items are scored automatically by at least two scorers. Scoring reliability is determined from the scores assigned by the two independent scorers for each student response.

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 the interrater reliability: percentage of exact agreement, percentage of exact-plus-adjacent agreement, percentage of adjacent agreement and percentage of non-adjacent agreement.

“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. Scoring reliability estimates for the primary, junior and Grade 9 assessments and the OSSLT are presented in Tables 4.1.21–4.1.40 of Appendix 4.1.

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

In 2010–2011, the EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for 36 out of 40 reading items, all rubrics for writing and 30 out of 32 mathematics items. The aggregate reliability estimates ranged from 95.8–99.1%.

The Grade 9 Assessment of Mathematics: Academic and Applied

In 2010–2011, the EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for 54 out of 56 items, and the aggregate reliability estimates ranged from 96.3–100%.

The Ontario Secondary School Literacy Test (OSSLT)

The EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for all reading items, for all short-writing prompts, for all conventions rubrics for the long-writing prompts and for none of the topic-development rubrics for the long-writing prompts. The aggregate reliability estimates ranged from 90.3–97.9%.

CHAPTER 5: EQUATING

For security purposes, EQAO constructs different assessments every year while ensuring that content and statistical specifications are similar to those of the assessments from 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 between assessments from year to year (Kolen & Brennan, 2004). Equating ensures that students in one year are not given an unfair advantage over students in another and that reported changes in achievement levels are due to differences in student performance and not to differences in assessment difficulty.

From time to time, the Ministry of Education makes modifications to *The Ontario Curriculum*, and EQAO assessments are modified accordingly in content and length. The new assessments differ in content and statistical specifications from those constructed in previous years, prior to the curriculum revisions. In such cases, EQAO uses a process called scaling to link the previous years' assessments with the current year's modified ones.

The processes used in equating and scaling are similar, but their purposes are different. Equating is used to adjust for differences in difficulty among assessments that are similar in content and statistical specifications. Scaling is used to link two assessments that are different in content and statistical specifications (Kolen & Brennan, 2004). Since there were no significant changes to the test specifications from 2009–2010 to 2010–2011, only equating procedures were used in 2010–2011.

The following sections describe the Item Response Theory (IRT) models, equating design, equating samples and calibration procedures used during the 2010–2011 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 (GPC) model are the general models used by EQAO to estimate the parameters of multiple-choice and open-response items and the ability parameters. The 3PL model (see Yen & Fitzpatrick, 2006, for example) is given by Equation 1:

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

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 constant equal to 1.7.

The GPC model (Muraki, 1997) is given by Equation 2:

$$P_{ih}(\theta) = \frac{\exp\left[\sum_{v=0}^h Da_i(\theta - b_i + d_h)\right]}{\sum_{c=0}^{M_i} \exp\left[\sum_{v=0}^c Da_i(\theta - b_i + d_h)\right]}, \quad h = 0, 1, \dots, M_i, \quad (2)$$

where $P_{ih}(\theta)$ is the probability of a student with ability θ choosing the h th score category for item i ;

a_i is the slope parameter for item i ;

b_i is the difficulty parameter for item i ;

d_h is the category parameter for category h of item i ;

D is a scaling constant equal to 1.7 and

M_i is the maximum score on item i .

Equating Design

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

1. Operational item parameters in the current year's assessments are calibrated.
2. Operational items and field-test items from the previous year are brought forward to the current year's assessments and recalibrated:
 - This is done by fixing the parameters of the items common to the two years at the values obtained in Step 1.
 - 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 theta 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 Sample

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 assessment 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 Sample

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 is conducted separately. The calibrations are conducted using the program PARSCALE 4.1 (Muraki & Bock, 2003).

Before final equating is conducted, several steps are taken to determine which common items are outliers and which of the outlying items should be excluded from the final equating process. First, scatter-plots are produced to compare the common-item parameter estimates (both difficulty and discrimination estimates) from field testing (FT) to operational (OP) testing. Ninety-five percent confidence intervals are constructed for both FT- and OP-item parameter estimates, and the best-fit line is also estimated. An item is flagged as an outlier if neither its field-test confidence interval nor its operational confidence interval crosses the best-fit line. Second, further investigation is focused on an outlying item that meets most of the following conditions: if it is flagged by both the OP-FT difficulty and OP-FT discrimination plots, if it has a large difference between OP and FT classical item statistics, if it is a multiple-choice item (instead of an open-response item) and if its position in the booklets changes drastically from FT to OP administration. Third, once it is decided which outliers to focus on, these items are excluded from the common-item set, and sensitivity analyses are conducted to evaluate the impact on equating results. The resulting theta cut scores and percentages at each achievement level are compared with those from the initial round of equating when no item was excluded from equating, and the resulting achievement levels of students are compared with their initial levels. A lot of judgment is required during the second and the third steps. Finally, another measure is taken to inform the final decision-making: whether the slope of the best-fit line in the OP-FT item difficulty plot is the reciprocal of that in the OP-FT item discrimination plot, since theoretically both lines should be the same.

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

Description of the IRT Model

A modified 3PL model (Equation 1) is used for multiple-choice items on the primary- and junior-division assessments, with the pseudo-guessing parameter fixed at $0.20 [1/(k + 1)]$, where k is the number of options] to reflect the possibility of students with a very low ability to correctly answer an item. The GPC model (Equation 2) is used for open-response items.

Equating Sample: Exclusion Rules

The following categories of students were excluded from the equating samples for 2009–2010 and 2010–2011:

1. students who were not attending a publicly funded school;
2. students who were home-schooled;
3. French Immersion students;
4. students who completed no work on an item (by booklet and item type);
5. students receiving accommodations;
6. students who were exempted and
7. 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, as presented in Table 5.1:

1. students from the 2009–2010 population who responded to both the 2010–2011 operational-test items and the field-test items that had been brought forward to form the 2010–2011 operational assessments and who were not excluded by the rules stated above (calibration sample);
2. students from the 2009–2010 population who wrote the operational assessment and who were not excluded by the rules stated above (equating sample) and
3. students from the 2010–2011 population who wrote the operational assessment and who were not excluded by the rules stated below (calibration and equating samples).

Table 5.1 Number of Students in the Calibration and Equating Samples for the 2009–2010 and 2010–2011 Primary- and Junior-Division Assessments (English and French)

Assessment	Number of Students		
	2010–2011 Calibration and Equating Samples	2009–2010 Calibration Sample	2009–2010 Equating Sample
Primary Reading (English)	100 912	29 082	105 282
Junior Reading (English)	6 179	1 994	6 082
Primary Reading (French)	111 685	31 550	114 842
Junior Reading (French)	5 839	1 846	5 662
Primary Writing (English)	101 359	48 364	105 792
Junior Writing (English)	111 861	51 966	115 136
Primary Writing (French)	6 221	3 831	6 143
Junior Writing (French)	5 856	3 359	5 700
Primary Mathematics (English)	97 787	79 282	102 286
Junior Mathematics (English)	113 490	99 698	116 706
Primary Mathematics (French)	6 272	5 727	6 153
Junior Mathematics (French)	5 927	5 700	5 702

Equating Steps

In equating the 2009–2010 and 2010–2011 tests, the forward fixed common-item-parameter non-equivalent group design was implemented as follows:

1. The 2010–2011 operational items were calibrated independently to obtain item parameter estimates and student-ability scores for the 2010–2011 calibration and equating sample.
2. The 2009–2010 operational items were calibrated (using the calibration sample) together with the field-test items that were brought forward to the 2010–2011 operational assessments. In this calibration, the item parameter estimates of the field-test items were fixed at the values obtained from the 2010–2011 calibration runs (Step 1).
3. The 2009–2010 equating sample was scored using the operational item parameter estimates obtained in Step 2.
4. The percentage of students at each achievement level was determined for the 2009–2010 equating sample from the levels assigned in 2009–2010. The theta value of the cut points that replicated this distribution was 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 2010–2011.
6. The operational item parameter estimates of 2010–2011 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 2010–2011 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, eleven multiple-choice items and six open-response items across the 12 assessment components other than the primary French writing, (English and French reading, writing and mathematics) were excluded from equating. These items were either modified between field- and operational-test administrations or were not field tested in the 2009–2010 assessment year. For the primary French writing assessment, three multiple-choice items were excluded from equating. For these items, it was observed that there were large differences in the p -values between the field test and the operational test. The number of items not used in the equating process and the number of items dropped from each assessment component are presented in Table 5.2.

Table 5.2 Number of Items Excluded from the Equating Process and Dropped from the Primary- and Junior-Division Assessments (2010–2011)

Assessment	No. of Items Excluded from Equating*		No. of Items Dropped from the Assessment
	Multiple-Choice	Open-Response	
Primary Reading (English)	0	0	0
Junior Reading (English)	2	0	0
Primary Reading (French)	0	0	0
Junior Reading (French)	1	0	0
Primary Writing (English)	0	1	0
Junior Writing (English)	0	1	0
Primary Writing (French)	3	1	0
Junior Writing (French)	0	1	0
Primary Mathematics (English)	1	1	0
Junior Mathematics (English)	0	0	0
Primary Mathematics (French)	1	1	0
Junior Mathematics (French)	3	0	0

*Long-writing prompts for the current year are not field tested in 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 writing components of the assessments are provided in Tables 5.3–5.6, and the results for the mathematics assessments are in Tables 5.7 and 5.8. The theta cut scores and the number of students at each achievement level in 2009–2010 and 2010–2011 are reported for both English-language and French-language students. For example, the theta cut scores for the reading component of the English-language primary-division assessment were 1.15 for Levels 3 and 4, –0.48 for Levels 2 and 3, –1.62 for Levels 1 and 2 and –2.66 for “not enough evidence for Level 1” (NE1) and Level 1.

Since the 2009–2010 and 2010–2011 student thetas are on the same scale, the theta cut scores in the following tables apply to the assessments for 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	
		2010	2011		2010	2011
Number of Students		105 282	100 912		6082	6179
Level 4		9.1%	9.0%		21.9%	27.1%
Level 3	1.15	59.9%	63.3%	0.60	50.4%	48.8%
Level 2	-0.48	24.5%	22.5%	-0.63	23.6%	21.4%
Level 1	-1.62	5.6%	4.4%	-1.89	3.8%	2.8%
NE1	-2.66	0.9%	0.7%	-3.17	0.4%	0.1%
% of Students at or Above the Provincial Standard		69.0%	72.4%		72.3%	75.8%

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	
		2010	2011		2010	2011
Number of Students		114 842	111 685		5662	5839
Level 4		12.6%	13.3%		19.5%	23.9%
Level 3	1.02	66.3%	68.7%	0.69	65.1%	63.7%
Level 2	-0.83	18.1%	16.1%	-1.08	14.8%	11.9%
Level 1	-2.05	2.9%	1.8%	-2.55	0.7%	0.5%
NE1	-3.29	0.1%	0.1%	-3.51	0.0%	0.0%
% of Students at or Above the Provincial Standard		78.9%	82.0%		84.6%	87.6%

Since the percentages of students in the 2009–2010 equating sample are very low at NE1 for the writing component of the French-language primary and junior divisions, the cut points were determined by an alternative method for the 2010–2011 assessments. To determine the 0/1 cut point for the 2010–2011 assessments, regression equations that predicted the 1/2 and 2/3 2010–2011 cuts from the 1/2 and 2/3 2009–2010 cuts were determined. The constant and multiplier values obtained from these regression equations were used to predict the 2010–2011 0/1 cut from the 2009–2010 0/1 cut for these assessments.

Table 5.5 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	
		2010	2011		2010	2011
Number of Students		105 792	101 359		6143	6221
Level 4		5.1%	5.3%		23.5%	27.2%
Level 3	1.42	70.1%	72.9%	0.56	60.9%	59.5%
Level 2	-0.72	24.8%	21.7%	-1.04	15.3%	13.2%
Level 1	-3.21	0.0%	0.1%	-2.76	0.3%	0.1%
NE1	-3.44	0.0%	0.0%	-3.67	0.0%	0.0%
% of Students at or Above the Provincial Standard		75.2%	78.2%		84.4%	86.7%

Table 5.6 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	
		2010	2011		2010	2011
Number of Students		115 136	111 861		5700	5856
Level 4		10.3%	11.5%		17.0%	19.2%
Level 3	1.09	67.3%	69.3%	0.81	68.2%	68.3%
Level 2	-0.77	22.1%	19.1%	-1.02	14.8%	12.4%
Level 1	-2.93	0.2%	0.2%	-3.37	0.1%	0.1%
NE1	-3.61	0.0%	0.0%	-3.74	0.0%	0.0%
% of Students at or Above the Provincial Standard		77.7%	80.7%		85.1%	87.5%

Table 5.7 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	
		2010	2011		2010	2011
Number of Students		102 286	97 787		6153	6272
Level 4		13.5%	13.3%		17.8%	20.0%
Level 3	1.04	62.9%	61.4%	0.81	57.4%	57.6%
Level 2	-0.61	21.7%	23.2%	-0.70	22.4%	20.4%
Level 1	-2.03	1.7%	2.0%	-2.07	2.2%	1.9%
NE1	-2.75	0.2%	0.1%	-2.86	0.2%	0.1%
% of Students at or Above the Provincial Standard		76.5%	74.7%		75.2%	77.6%

Table 5.8 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	
		2010	2011		2010	2011
Number of Students		116 706	113 490		5702	5927
Level 4		15.6%	14.3%		40.4%	40.9%
Level 3	1.02	51.2%	50.7%	0.24	45.1%	43.1%
Level 2	-0.35	28.5%	28.8%	-0.98	14.1%	15.5%
Level 1	-1.49	4.6%	6.2%	-2.38	0.2%	0.3%
NE1	-3.04	0.1%	0.1%	-2.79	0.2%	0.1%
% of Students at or Above the Provincial Standard		66.8%	65.0%		85.5%	84.0%

The Grade 9 Assessment of Mathematics

Description of the IRT Model

The 3PL model (Equation 1) and GPC model (Equation 2) were used to estimate item and ability parameters. For the Grade 9 academic and applied mathematics assessments, the 3PL model was modified by fixing the pseudo-guessing parameter at $0.20 [1/(k + 1)]$, where k is the number of options] for 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 in one school year—in winter and in spring. The winter and spring assessments for each version have a set of common items and a set of unique items. The common items are used for equating across the winter and spring administrations.

Equating Sample

Equating samples for 2009–2010 and 2010–2011 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 years were the same. Students were excluded if they

- did not attend a publicly funded school;
- were home-schooled;
- completed no work on an item (by booklet and item type);
- received accommodations, except for English language learners who received the accommodation for setting, and
- did not attempt at least one question in each significant part of the test.

Table 5.9 presents the number of students in the Grade 9 equating and calibration samples for the 2009–2010 and the 2010–2011 assessments.

Table 5.9 Number of Grade 9 Students in the Equating Samples

Version	2009–2010				2010–2011	
	English		French		English	French
	Calibration Sample	Equating Sample	Calibration Sample	Equating Sample	Calibration and Equating Sample	Calibration and Equating Sample
Academic	85 549	98 729	3 720	4 026	96 675	3 982
Applied	29 426	40 313	1 267	1 268	37 521	1 166

Equating Steps

The calibration and equating of the 2009–2010 and 2010–2011 assessments for the English-language and the French-language student populations in both the academic and applied courses were conducted using the following steps:

1. A concurrent calibration was conducted simultaneously for the 2010–2011 winter and spring samples.
2. Calibration and equating were conducted for the 2009–2010 winter and spring samples (including the field-test items that were brought forward to the 2010–2011 operational test). In this calibration, the item parameter estimates of the field-test items that were brought forward to the 2010–2011 operational tests were fixed at the values obtained from the 2010–2011 calibration runs (Step 1). The parameter estimates of the 2009–2010 operational items that were repeated on the 2010–2011 operational tests were also fixed at the values obtained from the 2010–2011 calibration runs (Step 1). The 2009–2010 equating samples (operational items only) were scored using the scaled 2009–2010 operational item parameter estimates.
3. The percentage of students in each achievement level was determined for the 2009–2010 equating sample from the levels assigned in 2009–2010. 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 2010–2011 assessments.
5. The item parameter estimates for the 2010–2011 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 2010–2011 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 assessments, eight items were excluded from the equating process. Out of the eight items, six were from 2011 winter matrix multiple-choice items, one was from 2008 matrix multiple choice items, and one was from 2011 matrix open-response items. (see Table 5.10).

Table 5.10 Number of Items Excluded from the Equating Process and Dropped: Grade 9 (2010–2011)

Assessment Version	No. of Items Excluded from Equating		No. of Items Dropped from the Assessment
	Multiple-Choice	Open-Response	
Applied, Winter (English)	0	0	0
Applied, Spring (English)	6	1	0
Academic, Winter (English)	0	0	0
Academic, Spring (English)	0	0	0
Applied, Winter (French)	1	0	0
Applied, Spring (French)	0	0	0
Academic, Winter (French)	0	0	0
Academic, Spring (French)	0	0	0

Equating Results

The equating results for the applied version of the Grade 9 mathematics assessments are summarized in Table 5.11. The results for the academic version of the Grade 9 mathematics assessments are summarized in Table 5.12. The theta cut scores and number of students in 2009–2010 and 2010–2011 at each achievement level are 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 1.34 for Levels 3 and 4, 0.10 for Levels 2 and 3, -0.92 for Levels 1 and 2 and -1.68 for “below Level 1” and Level 1.

Table 5.11 Equating Results for the Grade 9 Applied Mathematics Assessment

	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	
		2010	2011		2010	2011
Number of Students		40 313	37 521		1268	1166
Level 4		7.4%	7.7%		6.1%	6.5%
Level 3	1.34	37.2%	37.9%	1.42	32.7%	33.7%
Level 2	0.10	37.8%	37.5%	0.22	45.3%	41.6%
Level 1	-0.92	14.2%	13.5%	-0.87	14.3%	14.6%
Below Level 1	-1.68	3.5%	3.4%	-1.58	1.6%	3.6%
% of Students at or Above the Provincial Standard		44.6%	45.6%		38.8%	40.2%

Table 5.12 Equating Results for the Grade 9 Academic Mathematics Assessment

	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	
		2010	2011		2010	2011
Number of Students		98 729	96 675		4026	3982
Level 4		10.3%	10.8%		3.3%	4.8%
Level 3	1.21	73.4%	73.6%	1.59	68.5%	66.9%
Level 2	-0.96	12.1%	11.6%	-0.57	18.8%	17.9%
Level 1	-1.59	4.1%	3.9%	-1.19	9.2%	9.9%
Below Level 1	-2.65	0.1%	0.1%	-2.25	0.2%	0.6%
% of Students at or Above the Provincial Standard		83.7%	84.4%		71.7%	71.6%

The 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 for the OSSLT, yielding a modified Rasch model for multiple-choice items. The a -parameter for all multiple-choice and open-response items was set at 0.588. The pseudo-guessing parameter for multiple-choice items was set at 0.20 $[1/(k + 1)]$, where k is the number of options that reflect the possibility of students with a very low ability to correctly answer an item]. The GPC model (see Equation 2), with a constant slope parameter of 0.588, was used to estimate the item and ability parameters for open-response items.

Equating Sample

First-time eligible students from both publicly funded and private schools were selected for the 2009–2010 and 2010–2011 equating samples. The following categories of students were excluded from the equating samples:

- a) students with no work or incomplete work on a major section of the test;
- b) students receiving the following accommodations on: assistive devices and technology, sign language, Braille, an audio recording or verbatim reading of the test, a computer, audio- or video-recorded responses and scribing;
- 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 5.13 presents the number of first-time eligible students in the OSSLT equating samples for the 2009–2010 and 2010–2011 tests.

Table 5.13 Number of First-Time Eligible OSSLT Students in the Equating Samples

OSSLT	2010		2011	
	English	French	English	French
First-Time Eligible	139 717	4930	137 736	5224

Equating Steps

The following steps were implemented to calibrate and equate the 2010 and 2011 OSSLT:

1. The parameter estimates of the operational items administered in 2011 were calibrated using the 2011 equating sample.
2. The operational items that formed the 2010 test and the field-test items brought forward to the 2011 test were recalibrated using the 2010 equating sample. In this calibration, the parameter estimates of the common items were fixed at the values obtained in Step 1.
3. The operational item parameter estimates of the 2010 test, obtained in Step 2, were used to score the 2010 equating sample data.
4. The percentage of successful students was determined for the 2010 equating sample from the student results reported in 2010. The theta-value cut point that replicated this percentage was identified for the distribution of scores in the 2010 equating sample.
5. This theta value was applied to student scores for the 2011 assessment to determine which students would be successful. The results are presented in Table 5.14.

Scale Score

The reporting scale scores for the 2011 OSSLT, which range from 200 to 400, were 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 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 four multiple-choice items in the French-language OSSLT were excluded from the equating process due to modifications made to the items between the 2010 field-test and the 2011 operational-test administrations.

The score 1.0 in the scoring rubric for long-writing prompts was collapsed with the score 1.5 for topic development and the use of conventions in both the English- and the French-language tests.

Equating Results

The equating results based on the equating samples for the OSSLT are summarized in Table 5.14. The theta cut score and the percentages of successful and unsuccessful students in 2010 and 2011 are reported for English-language and French-language students. For example, the equated cut score for the English-language test was -0.78 . The percentage of successful students in the equating samples was 86.2% in 2010 and 85.8% in 2011.

Table 5.14 Equating Results for the OSSLT

	English-Language			French-Language		
	Theta Cut Point 2010–2011	Equating Sample 2010	Equating Sample 2011	Theta Cut Point 2010–2011	Equating Sample 2010	Equating Sample 2011
No. of Students		139 717	137 736		4930	5224
% Successful	-0.78	86.2%	85.8%	-0.67	85.2%	85.5%
Unsuccessful		13.8%	14.2%		14.8%	14.5%

References

- Kolen, M. J. & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (2nd ed.). New York: Springer-Verlag.
- Muraki, E. (1997). A generalized partial credit model. In W. J. Van der Linden, & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 153-164). 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.
- Yen, W. M. & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp.111-153). Westport, CT: American Council on Education and Praeger Publishers.

CHAPTER 6: REPORTING RESULTS

EQAO assessment results are reported at the student, school, school board and provincial levels.

EQAO publishes an annual provincial report for education stakeholders and the general public. The reports for the 2010–2011 English-language assessments are available on www.eqao.com:

- *EQAO's Provincial Elementary School Report on the Results of the 2010–2011 Assessments of Reading, Writing and Mathematics, Primary Division (Grades 1–3) and Junior Division (Grades 4–6)* and
- *EQAO's Provincial Secondary School Report on the Results of the Grade 9 Assessment of Mathematics and the Ontario Secondary School Literacy Test, 2010–2011.*

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

EQAO posts school and board results on www.eqao.com for public access. However, EQAO does not publicly release school or board results when the number of students having written an assessment therein 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).

Two types of aggregate results are reported for schools, boards and the province:

1. the percentage of all students enrolled in Grades 3 and 6, students enrolled in Grade 9 academic and applied mathematics courses and students eligible to write the OSSLT and
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 and school board personnel through user identification numbers and passwords. These reports include the results for small *n*-counts that are suppressed in the public reports and additional achievement results for sub-groups of the student population (i.e., English language learners, students with special education needs, 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 their students and data files with aggregated results for each school, 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 on 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. There are three reporting categories in addition to the four performance levels, as follows:

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

Two sets of results are reported: those based on all students and those based on participating students. Students without data and exempted students are not included in the calculation of results for participating students.

The provincial, board and school reports provide the following:

- overall jurisdictional results for each component;
- longitudinal data showing jurisdictional results over time;
- overall jurisdictional results for each component by gender and other relevant characteristics (e.g., English language learners, special education needs);
- results by exceptionality category and results for students receiving accommodations;
- areas of strength with respect to 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 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 Item Information Report: Student Rosters, which provide item results for each student who has 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 having achieved the provincial standard). The directors also receive school lists with achievement results over time, for convenient reference.

The Individual Student Report (ISR) for students in Grades 3 and 6 shows the overall achievement level for each component (reading, writing and mathematics) at one of five positions. For example, Level 1 includes the sub-categories 1.1, 1.3, 1.5, 1.7 and 1.9. The five sub-categories 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 and Level 1, Level 1 and Level 2, Level 2 and Level 3 and Level 3 and 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 accordingly as 1.1, 1.3, 1.5, 1.7, and so on, to 4.9. School, school board and provincial results are included on the ISR to provide a context for interpreting student results. For students in Grade 6, the assessment results they achieved in Grade 3, if available, are printed on their ISR.

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 in the reports for 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" for 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, 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. In 2011, cohort tracking from Grade 3 to Grade 9 was also conducted at the provincial level. Mathematics assessment results achieved in Grades 3 and 6, if available, are printed 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 OSSLC) is required to meet the literacy requirement for graduation. Students must achieve a minimum theta score to receive a successful result on the OSSLT. 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 minimum score.

As with the other assessments, EQAO reports results for all and for participating students. Students are considered to be “not participating” if they were deferred, opted to take the OSSLC or have no data for the current administration. Students who are not working toward the OSSD are exempt from the OSSLT and are not included in either reported population. 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 12 year.

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 other characteristics (e.g., English language learners, 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, English as a second language (ESL) or English literacy development (ELD), “actualisation linguistique en français” (ALF) or “programme d’appui aux nouveaux arrivants” (PANA);
- longitudinal data showing jurisdictional results over time;
- areas of strength relative to the curriculum and areas for improvement;
- cohort-tracking results from Grade 6 to the OSSLT and
- results for the student questionnaire and contextual data.

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

The OSSLT ISR provides the following:

- the statement of a successful or unsuccessful result;
- the student’s scale score;
- the median scale score for the school and province and
- feedback for students on areas of strength and areas for improvement.

Each unsuccessful student is informed that a successful result requires a scale score of 300.

Interpretation Guides

Guides for interpreting results are included in the school and board reports, and released test items and scoring guides are posted on the EQAO Web site.

EQAO also produces the following resource documents:

- “Using Data to Promote Student Success: A Brief Guide to Assist School Administrators in Interpreting Their Data”;
- the “Guide to Using the Item Information Report: Student Roster”;
- “EQAO Questionnaire Data: Using Contextual Data to Inform Improvement Planning” and
- the “Technical Paper: School and School Board Profiles of Strengths and Areas for Improvement.”

CHAPTER 7: STATISTICAL AND PSYCHOMETRIC SUMMARIES

A variety of statistical and psychometric analyses were conducted for the 2009–2010 assessments. 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 parameter estimates were obtained from the calibration process used for the equating samples (described in Chapter 5). Detailed data for individual items appear in Appendix 7.1.

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

Classical Test Theory (CTT) Analysis

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 and French-language versions of the primary and junior assessments. The test means (converted to percentages) ranged from 57.4% (for primary English-language reading) to 66.2% (for junior French-language writing).

Reliability and the corresponding SEMs refer to the precision of test scores, with higher reliability coefficients and lower SEMs indicating higher levels of precision. For the primary and junior assessments, Cronbach's alpha estimates range from 0.86 to 0.88 for reading, 0.80 to 0.82 for writing and 0.87 to 0.89 for mathematics. The corresponding standard errors of measurement range from 4.6% to 4.9% of the possible maximum score for reading, 6.6% to 7.2% of the possible maximum score for writing and 5.7% to 6.9% of the possible maximum score for mathematics. The reliability coefficients for writing are a little lower than those for reading and mathematics. This is attributable, in part, to the smaller number of writing items and the subjectivity in scoring writing performance.¹ Taking these two factors into account, the obtained reliability coefficients 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 0.91.

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

Assessment	No. of Items	Item Type		Max. Score	No. of Students	Min.	Max.	Mean	SD	Alpha	SEM
		MC	OR								
Primary Reading (English)	36	26	10	66	115 950	0	65	37.88	8.90	0.87	3.21
Junior Reading (English)	36	26	10	66	128 687	0	66	39.79	8.84	0.88	3.06
Primary Reading (French)	36	26	10	66	6 766	6	63	39.92	8.67	0.86	3.24
Junior Reading (French)	36	26	10	66	6 442	4	66	41.31	8.96	0.87	3.23
Primary Writing (English)	14	8	6*	29	116 331	0	29	17.89	4.70	0.80	2.10
Junior Writing (English)	14	8	6*	29	128 827	0	29	18.29	4.52	0.82	1.92
Primary Writing (French)	14	8	6*	29	6 773	2	29	17.35	4.39	0.81	1.91
Junior Writing (French)	14	8	6*	29	6 440	2	29	19.21	4.29	0.80	1.92
Primary Mathematics (English)	36	28	8	60	120 441	0	60	38.66	11.51	0.87	4.15
Junior Mathematics (English)	36	28	8	60	128 474	0	60	37.92	11.69	0.89	3.88
Primary Mathematics (French)	36	28	8	60	6 786	7	60	39.17	10.31	0.89	3.42
Junior Mathematics (French)	36	28	8	60	6 428	0	60	36.99	11.55	0.89	3.83

Note. MC = multiple choice; OR = open response; SD = standard deviation; Alpha = Cronbach's alpha; SEM = standard error of measurement.

*Short writing and long writing.

Item Response Theory (IRT) Analysis

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 refer to 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 parameter estimates for individual items are presented in Appendix 7.1.

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

Assessment	No. of Students	Min.	Max.	Mean	SD
Primary Reading (English)	115 950	-3.85	3.50	-0.12	1.00
Junior Reading (English)	128 687	-3.93	3.49	-0.15	1.02
Primary Reading (French)	6 766	-3.82	2.95	-0.08	0.98
Junior Reading (French)	6 442	-3.93	3.20	-0.11	0.99
Primary Writing (English)	116 331	-3.56	2.64	-0.08	0.93
Junior Writing (English)	128 827	-3.66	2.58	-0.14	0.98
Primary Writing (French)	6 773	-3.61	2.82	-0.04	0.94
Junior Writing (French)	6 440	-3.74	2.40	-0.09	0.96
Primary Mathematics (English)	120 441	-3.71	2.40	-0.10	1.00
Junior Mathematics (English)	128 474	-3.68	2.44	-0.13	1.01
Primary Mathematics (French)	6 786	-3.47	2.54	-0.07	0.98
Junior Mathematics (French)	6 428	-3.72	2.56	-0.08	0.98

Note. SD = standard deviation.

The Test Characteristic Curves (TCCs) and the distribution of student thetas are provided in Figures 7.1–7.12. 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 48%. 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 7.13–7.24. For example, the maximum information provided by the reading component of the English-language primary-division assessment is at approximately theta -0.90 . 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)

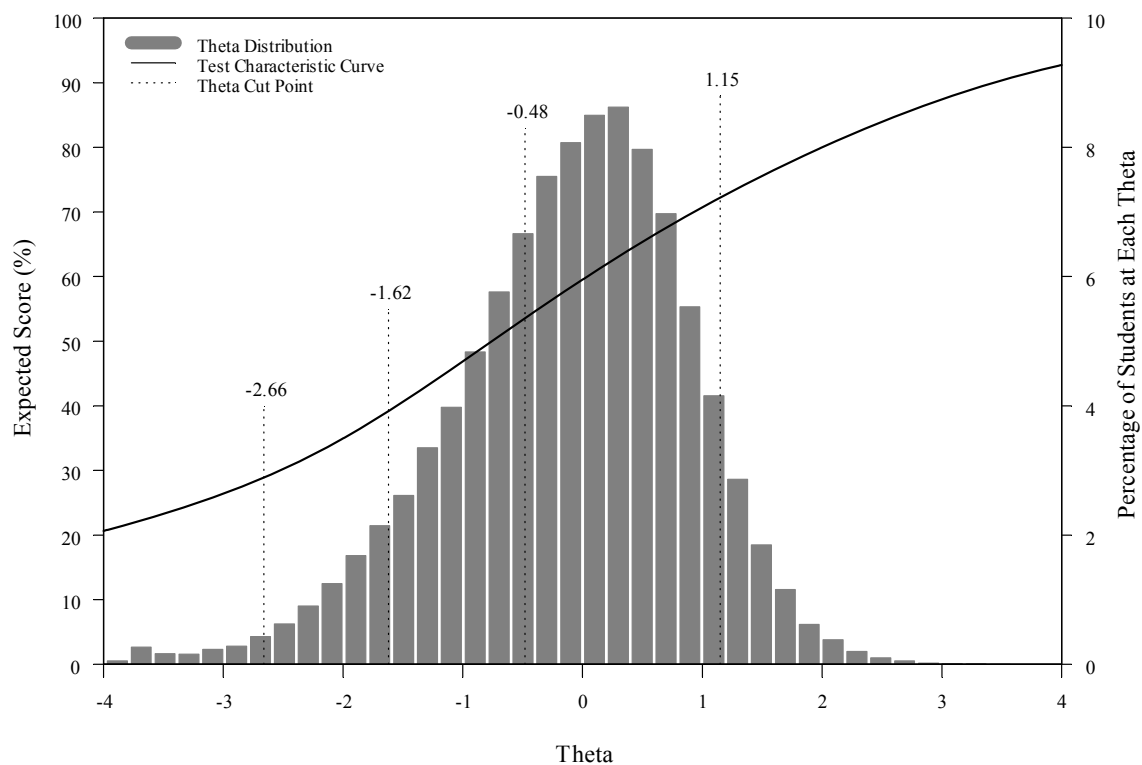


Figure 7.2 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (English)

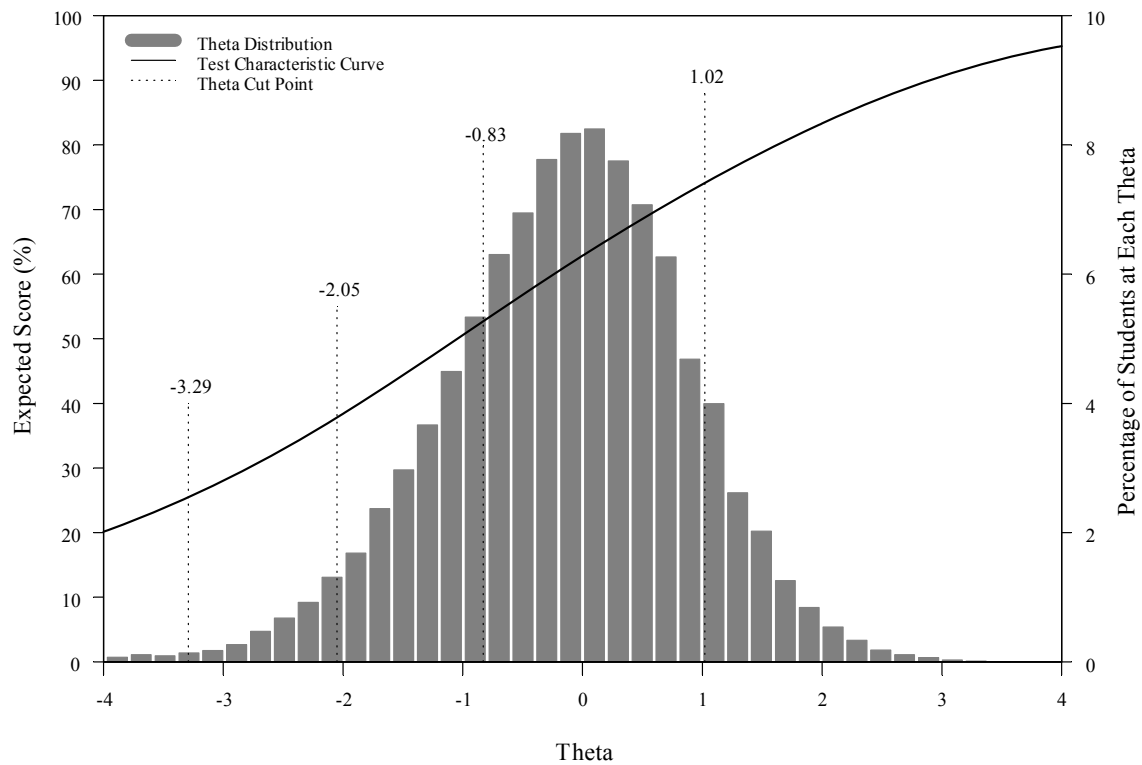


Figure 7.3 Test Characteristic Curve and Distribution of Thetas for Reading: Primary Division (French)

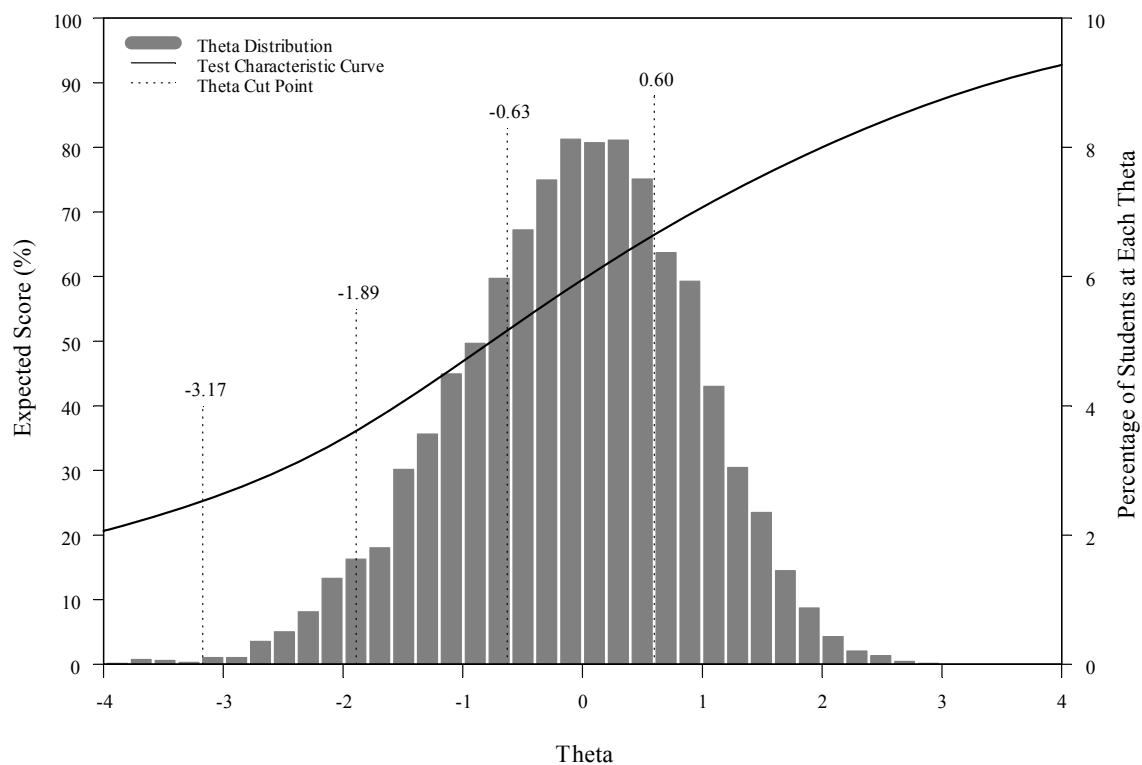


Figure 7.4 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (French)

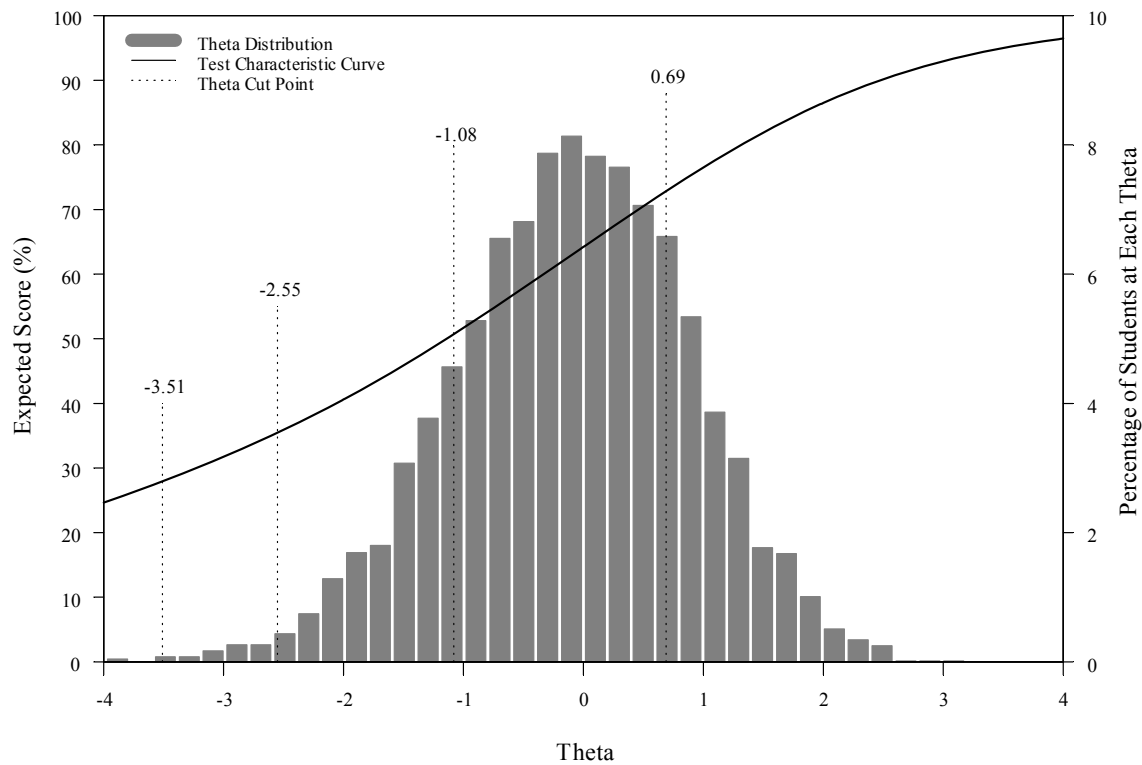


Figure 7.5 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (English)

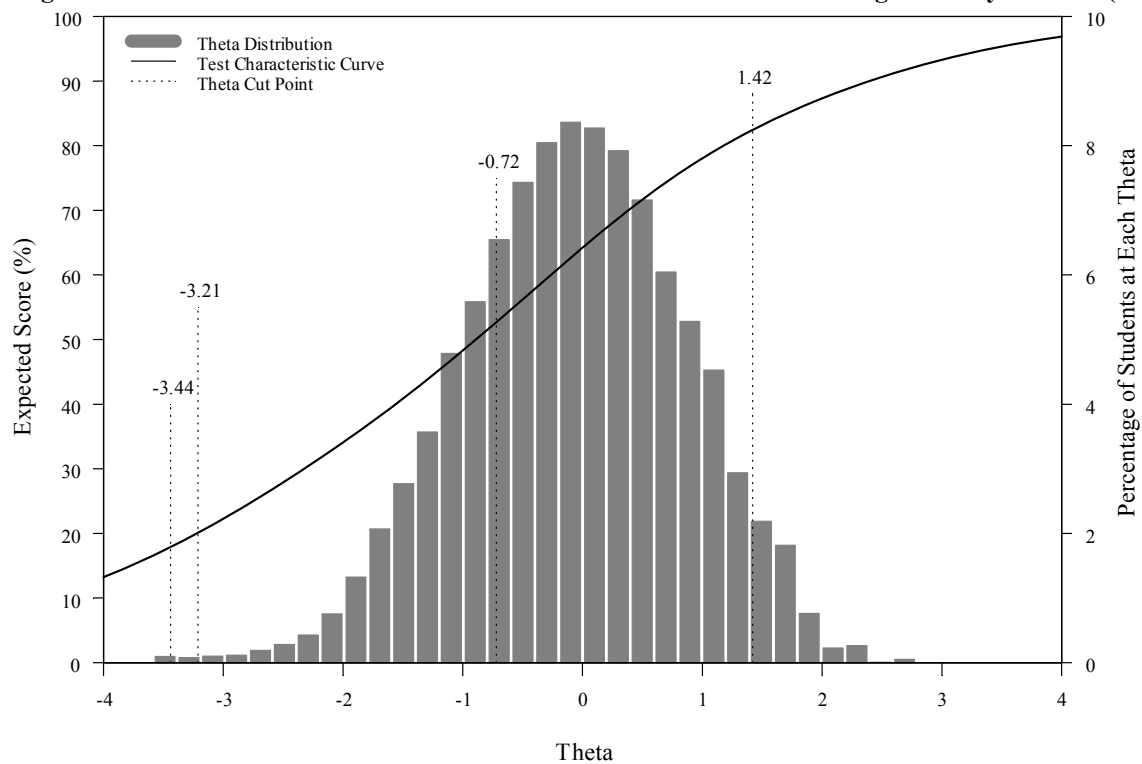


Figure 7.6 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (English)

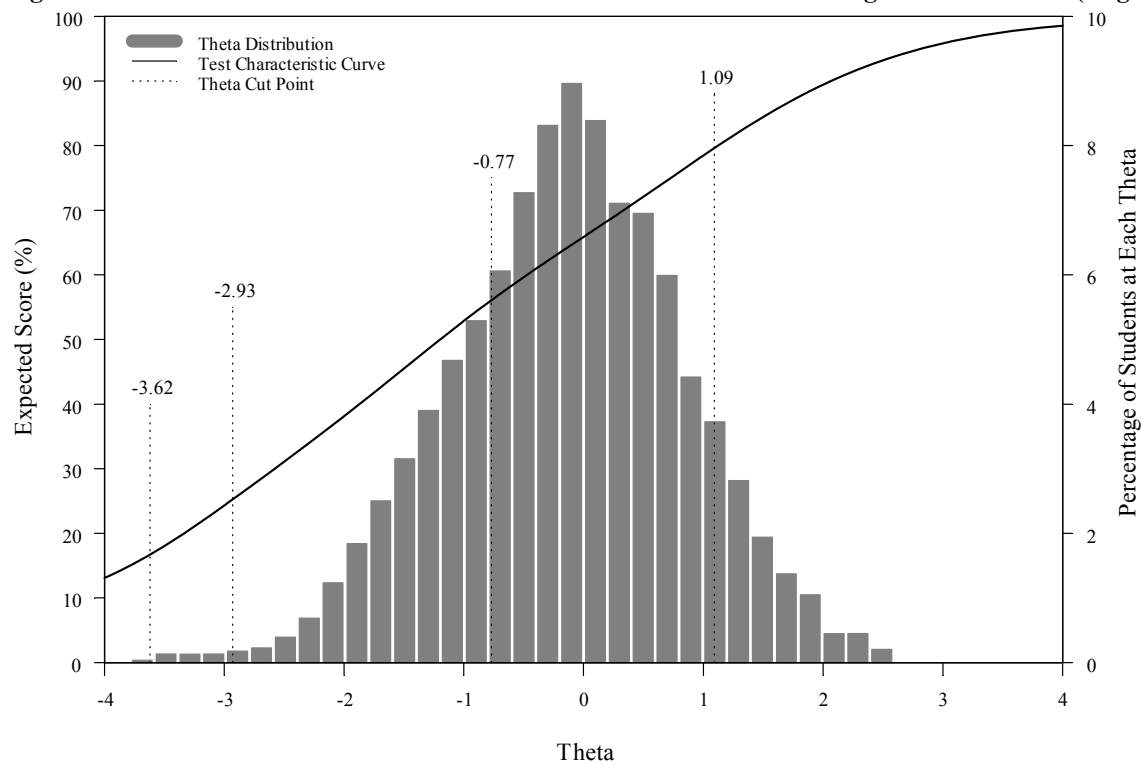


Figure 7.7 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (French)

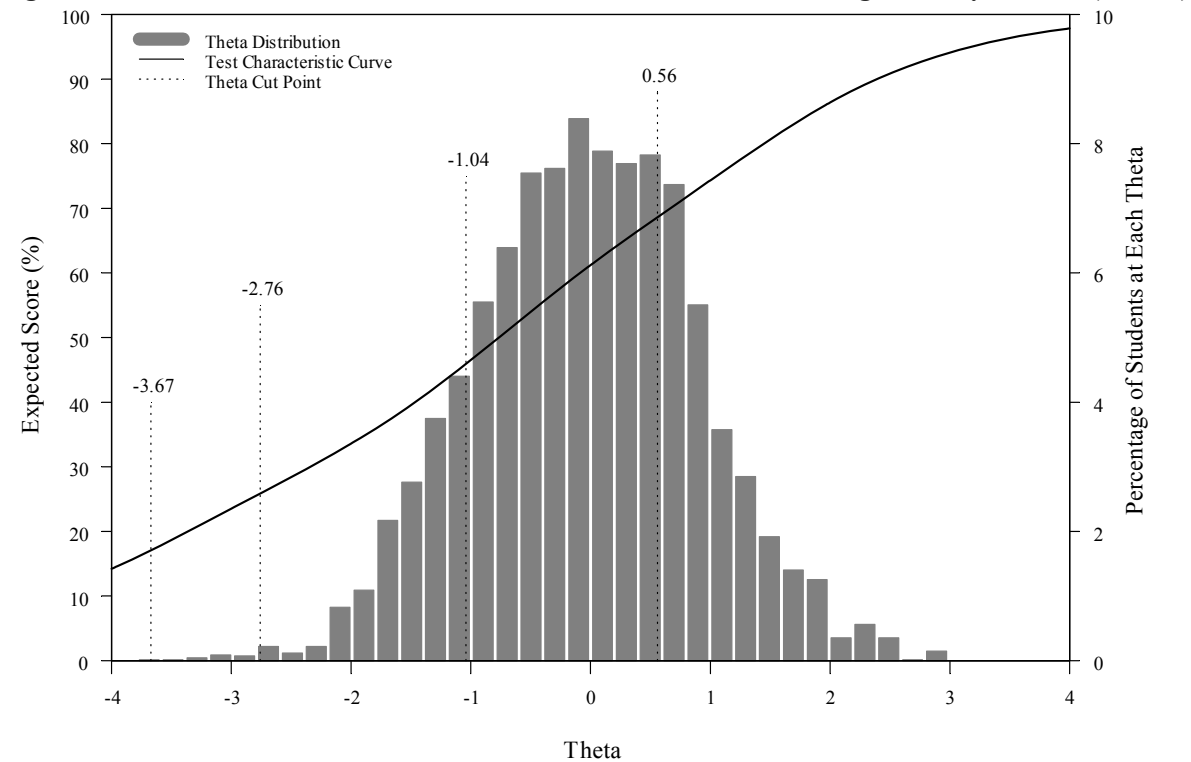


Figure 7.8 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (French)

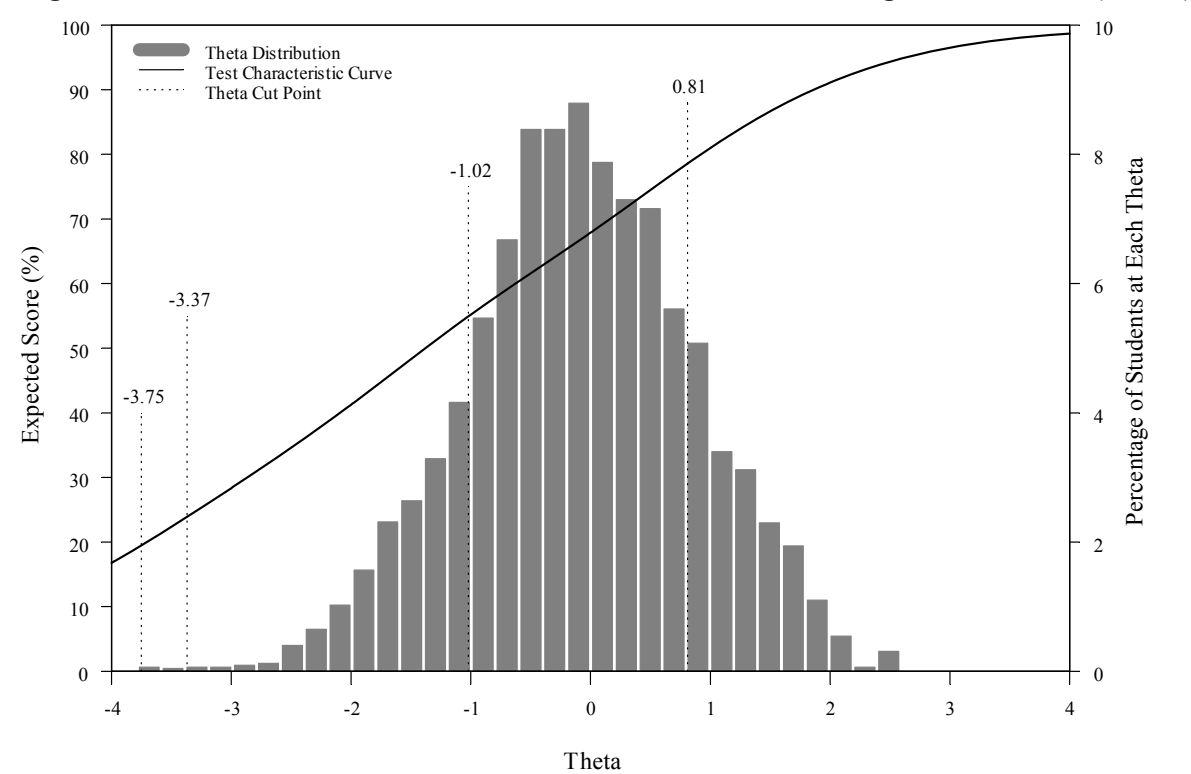


Figure 7.9 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (English)

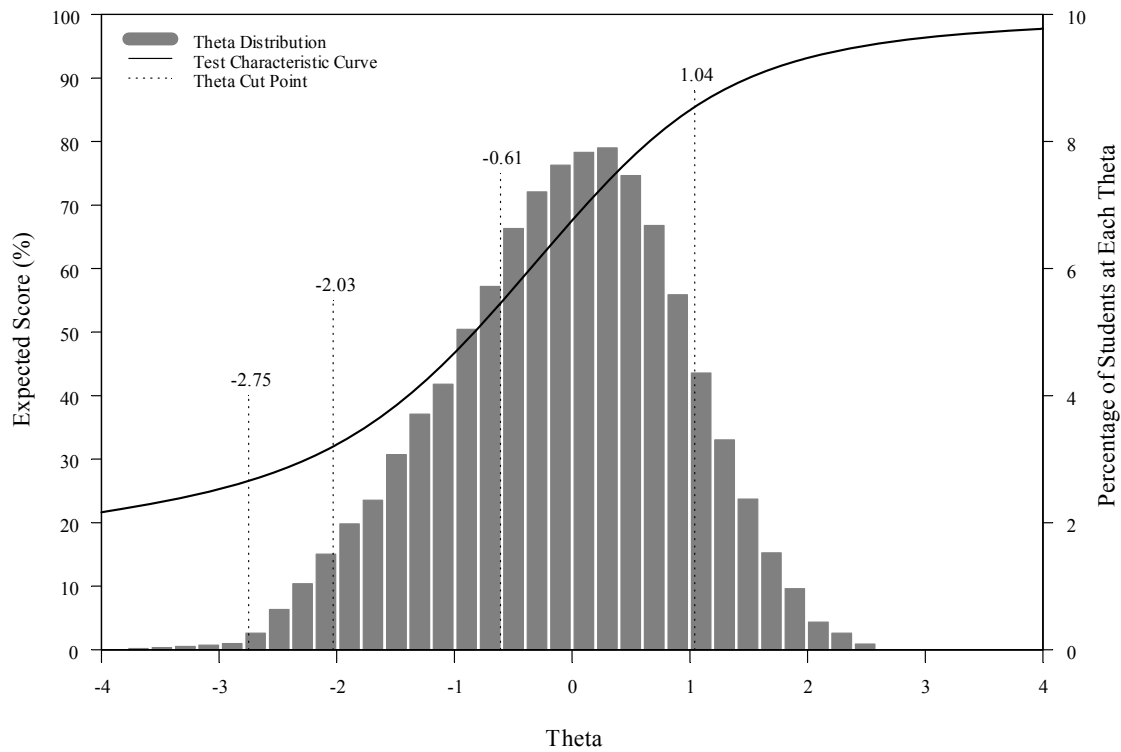


Figure 7.10 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (English)

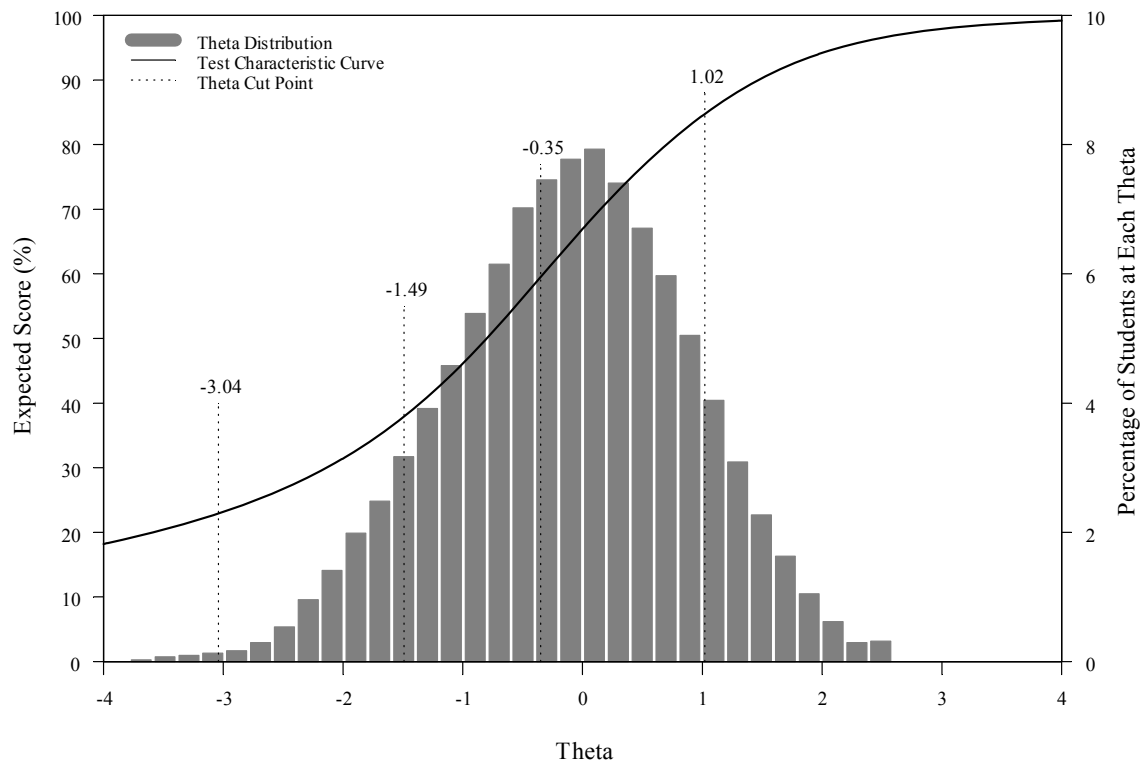


Figure 7.11 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (French)

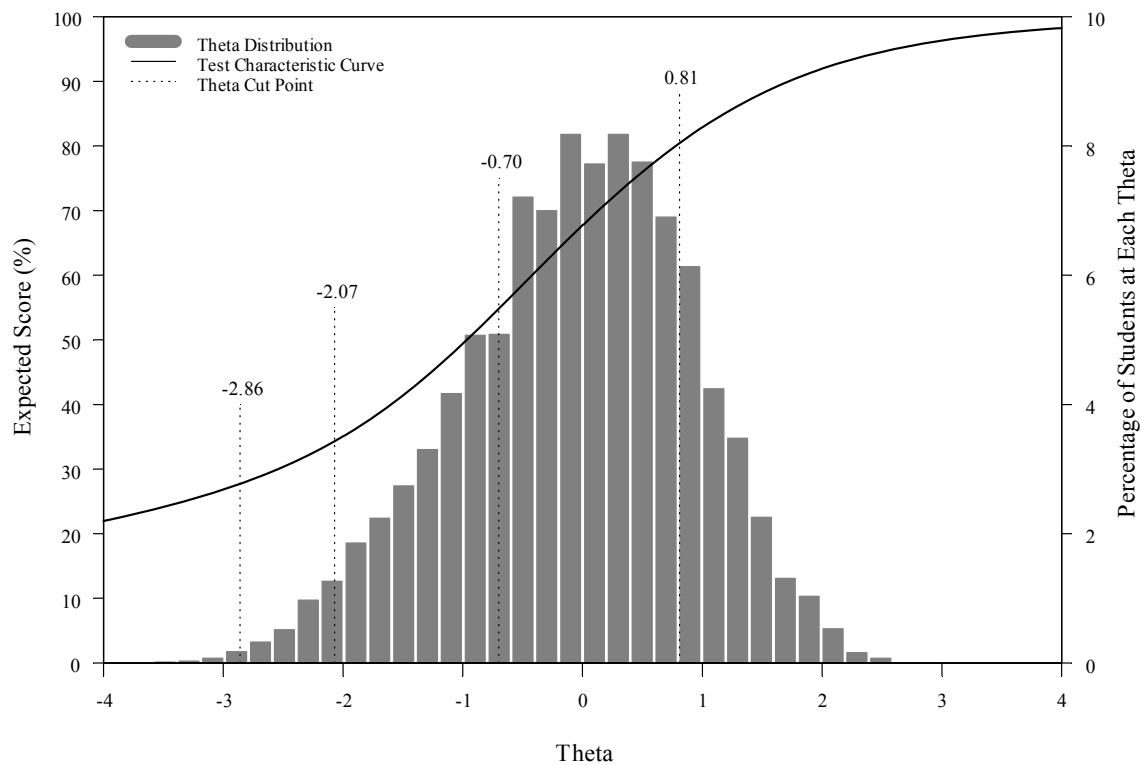


Figure 7.12 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (French)

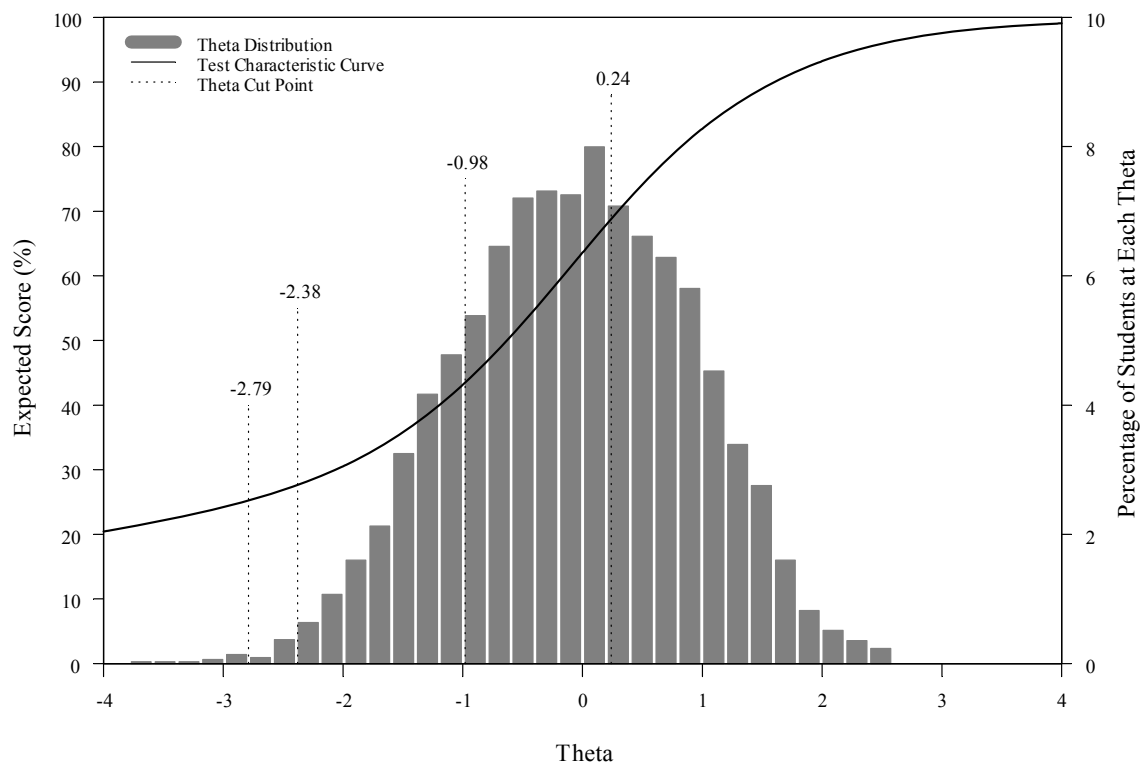


Figure 7.13 Test Information Function for Reading: Primary Division (English)

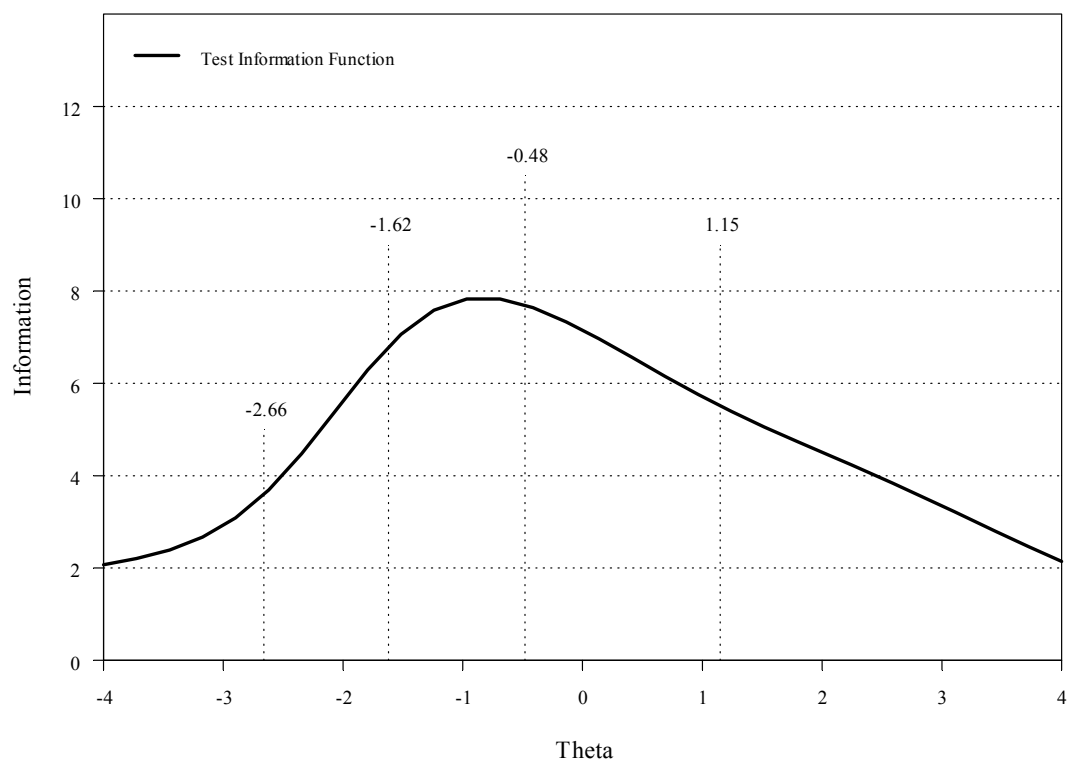


Figure 7.14 Test Information Function for Reading: Junior Division (English)

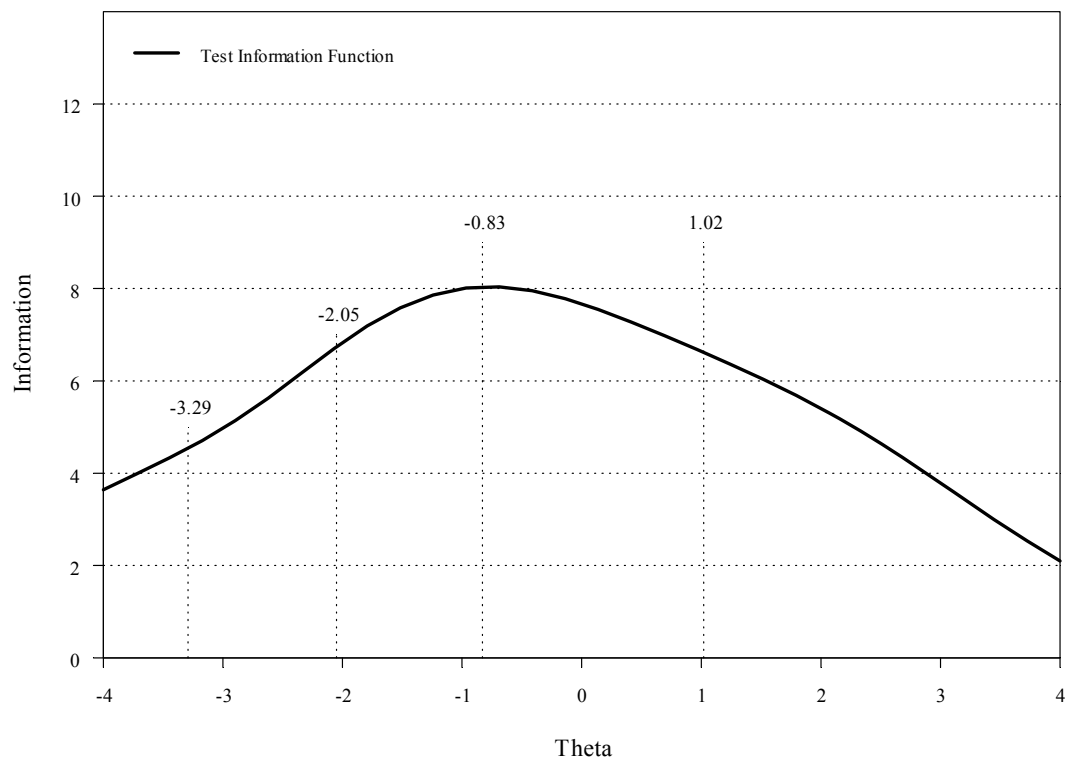


Figure 7.15 Test Information Function for Reading: Primary Division (French)

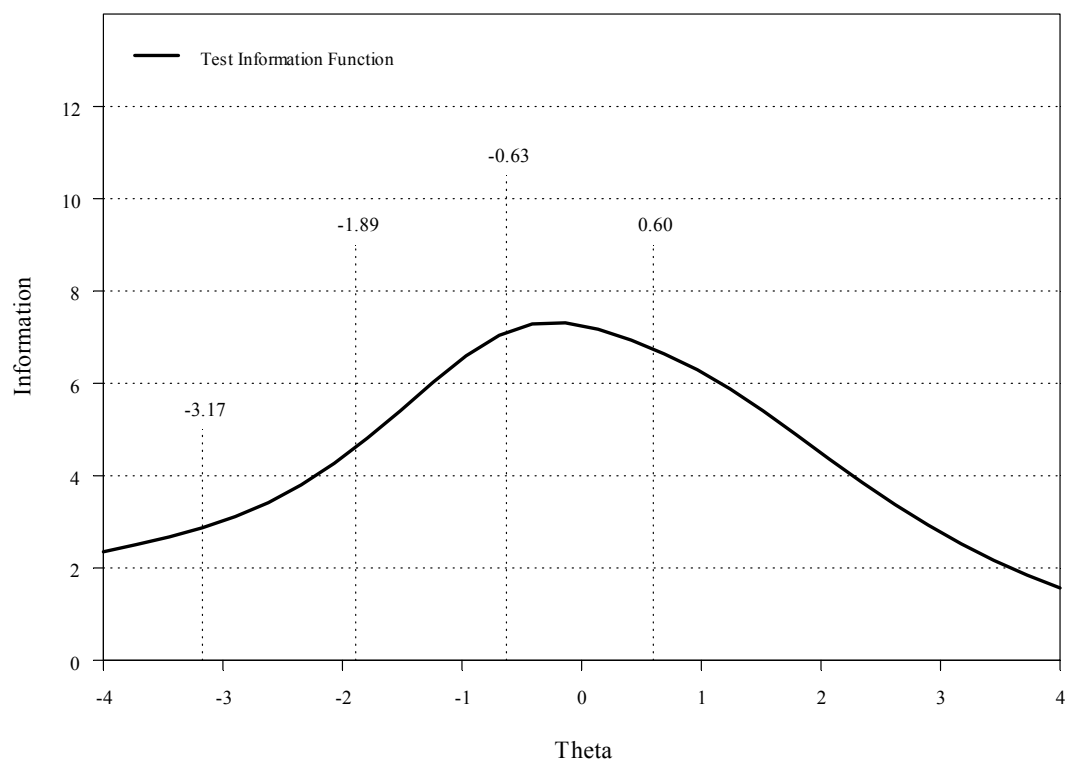


Figure 7.16 Test Information Function for Reading: Junior Division (French)

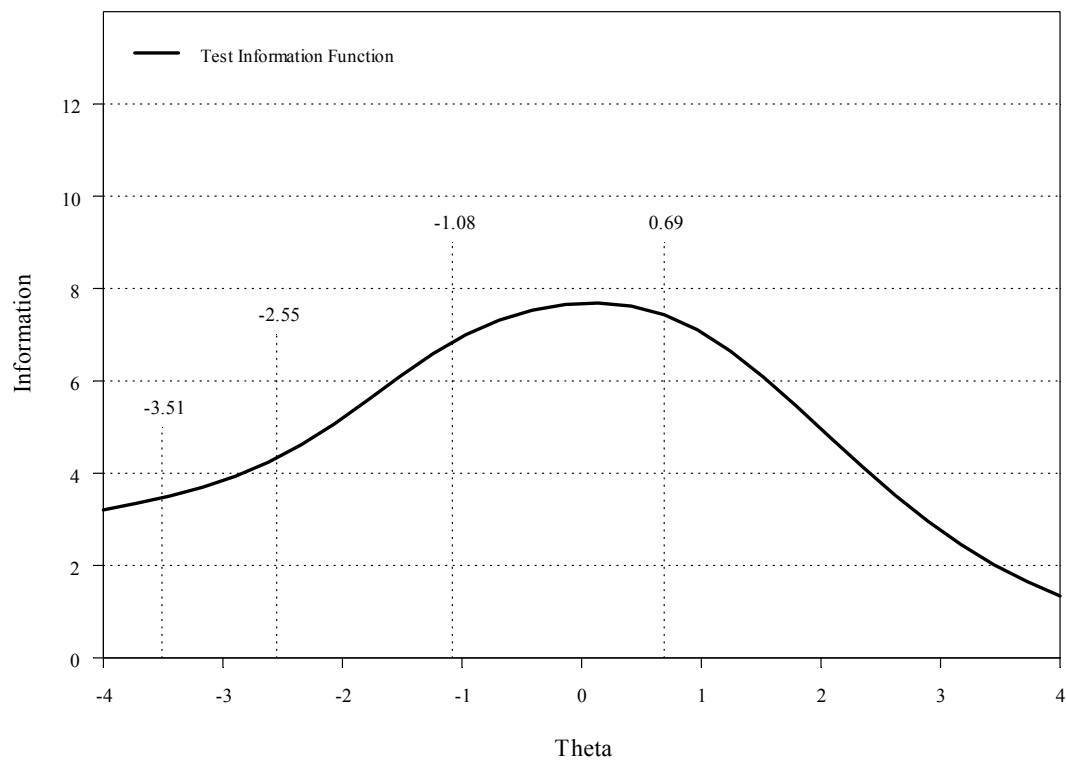


Figure 7.17 Test Information Function for Writing: Primary Division (English)

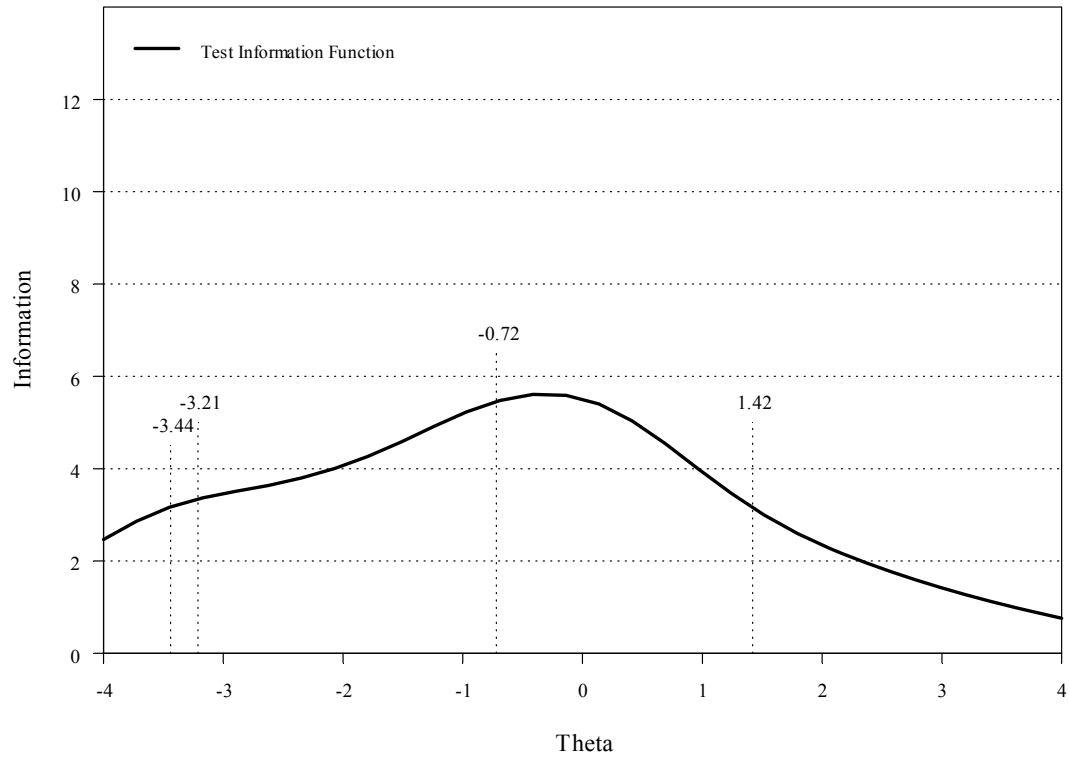


Figure 7.18 Test Information Function for Writing: Junior Division (English)

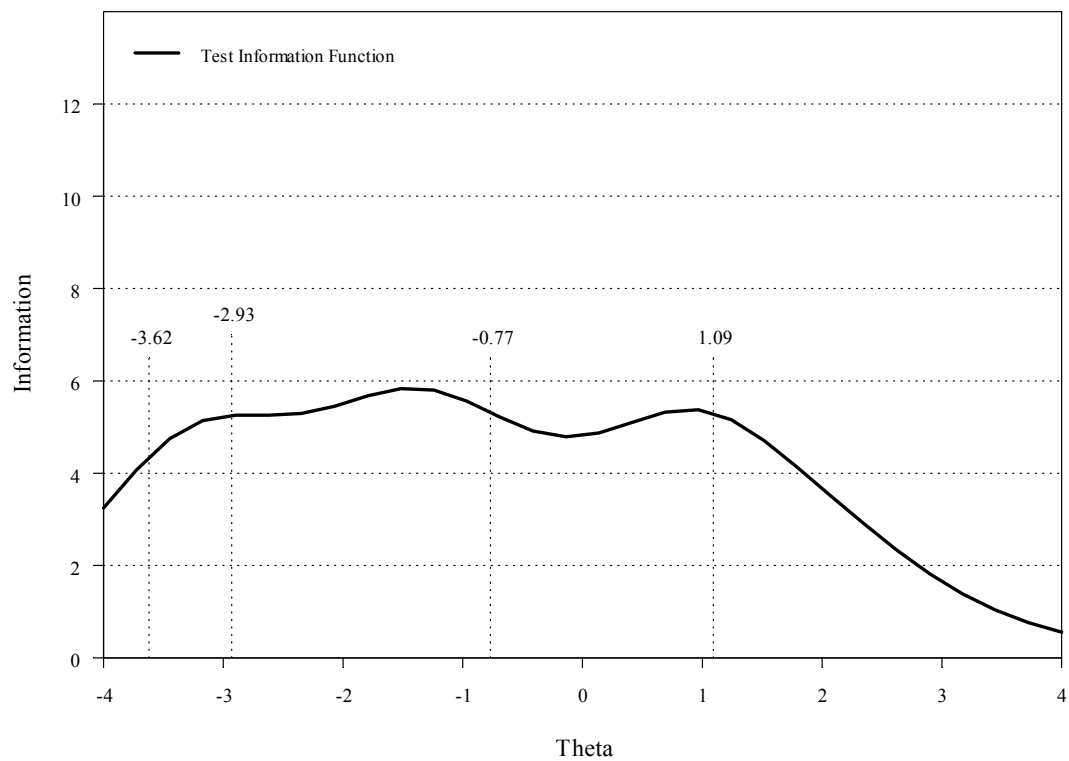


Figure 7.19 Test Information Function for Writing: Primary Division (French)

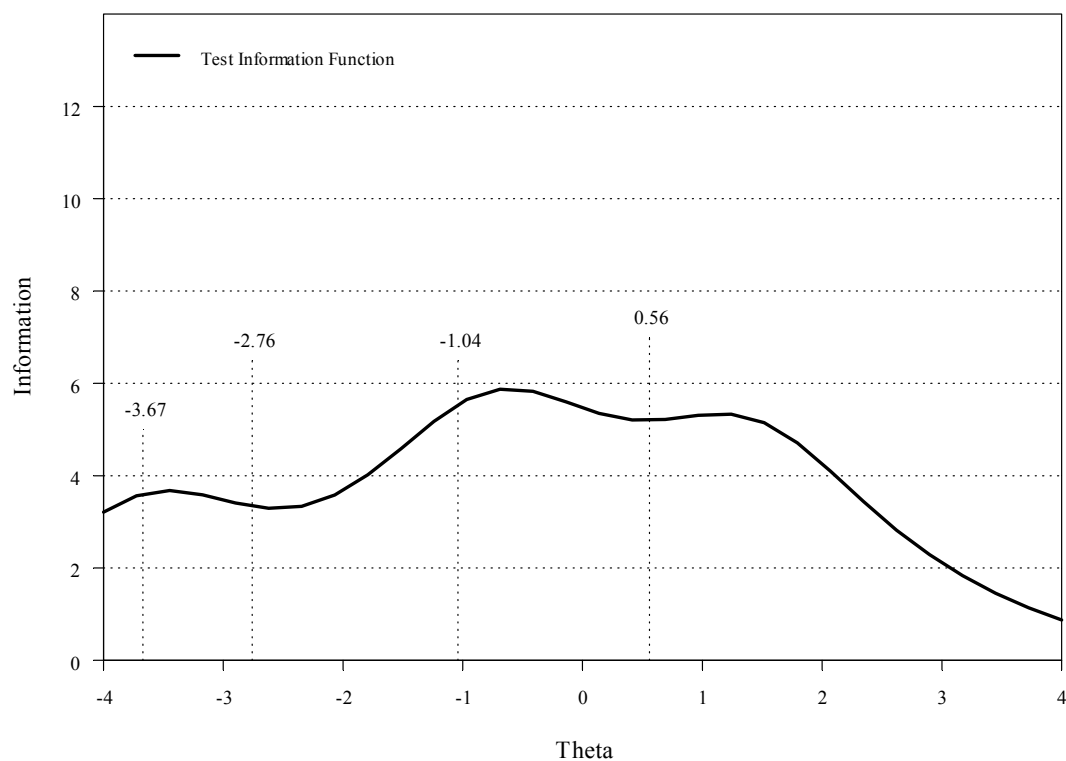


Figure 7.20 Test Information Function for Writing: Junior Division (French)

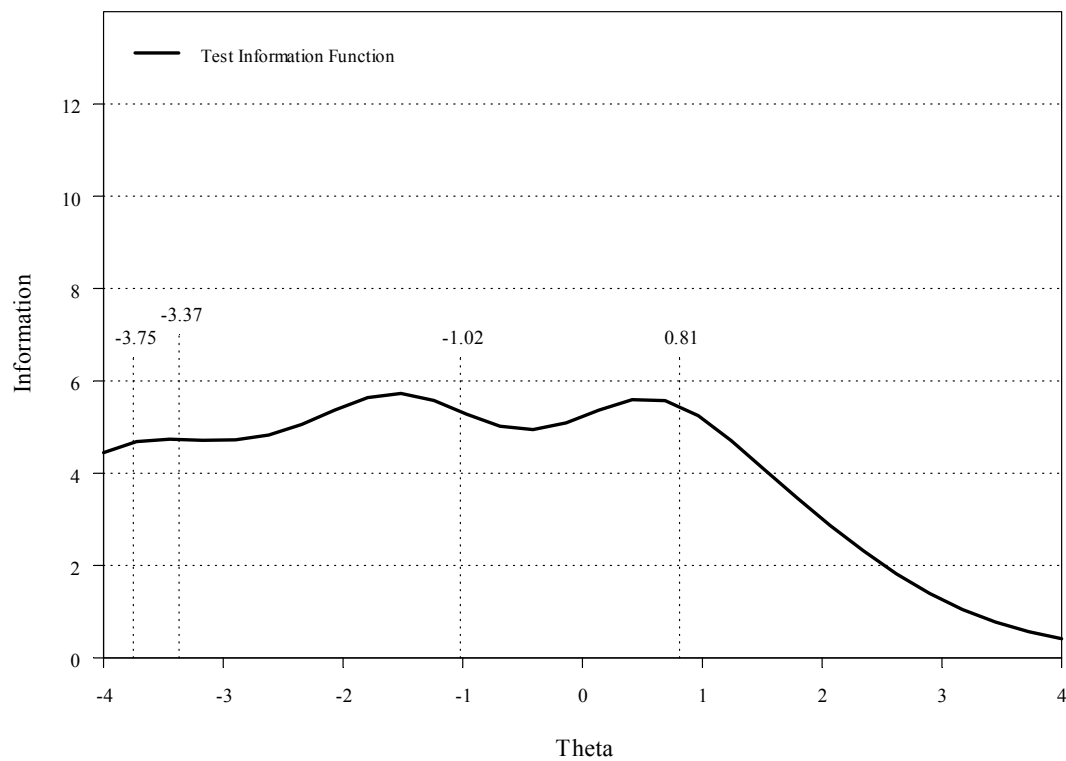


Figure 7.21 Test Information Function for Mathematics: Primary Division (English)

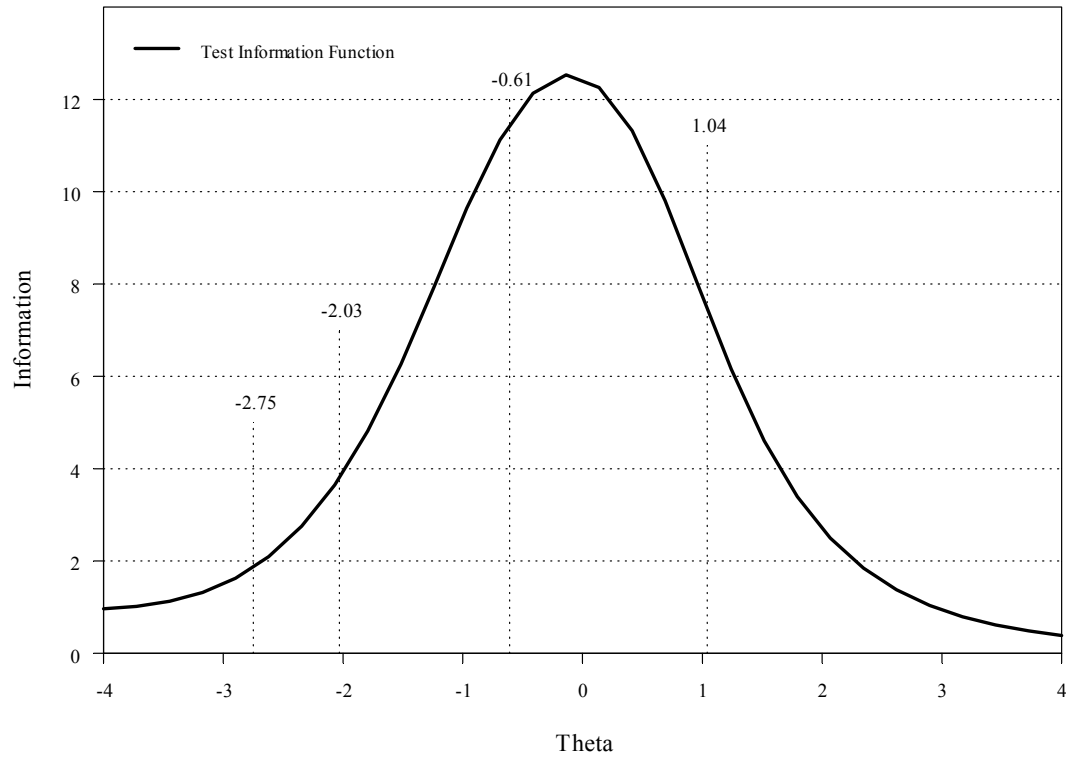


Figure 7.22 Test Information Function for Mathematics: Junior Division (English)

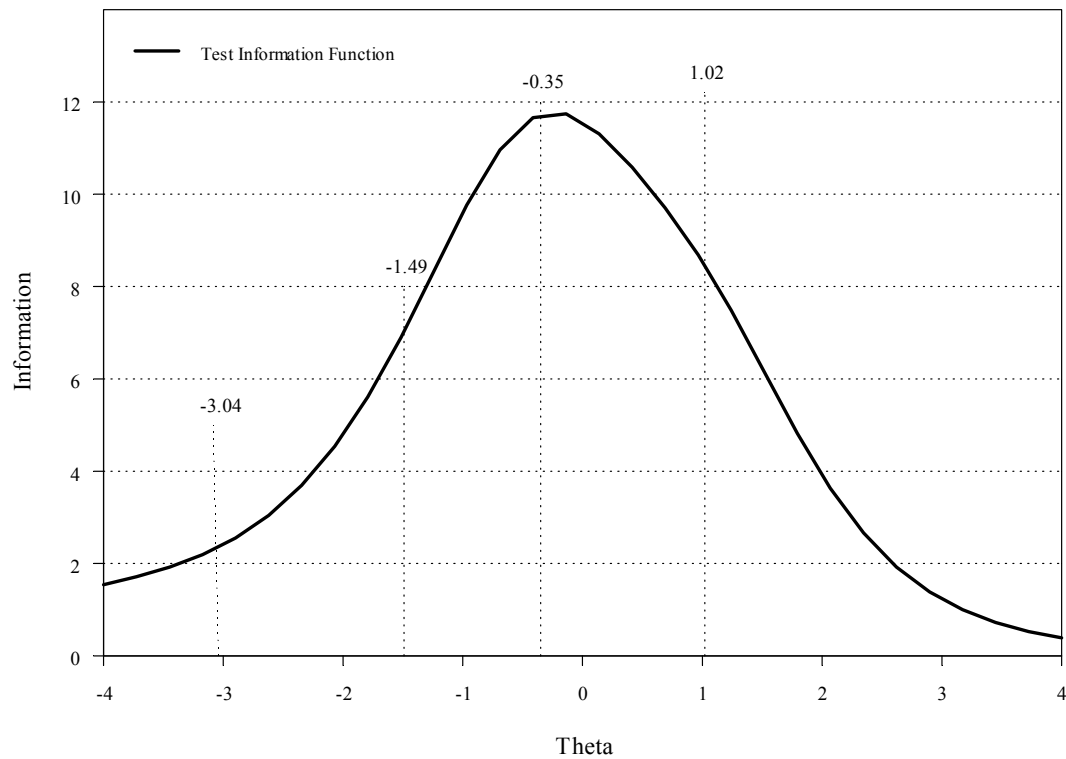


Figure 7.23 Test Information Function for Mathematics: Primary Division (French)

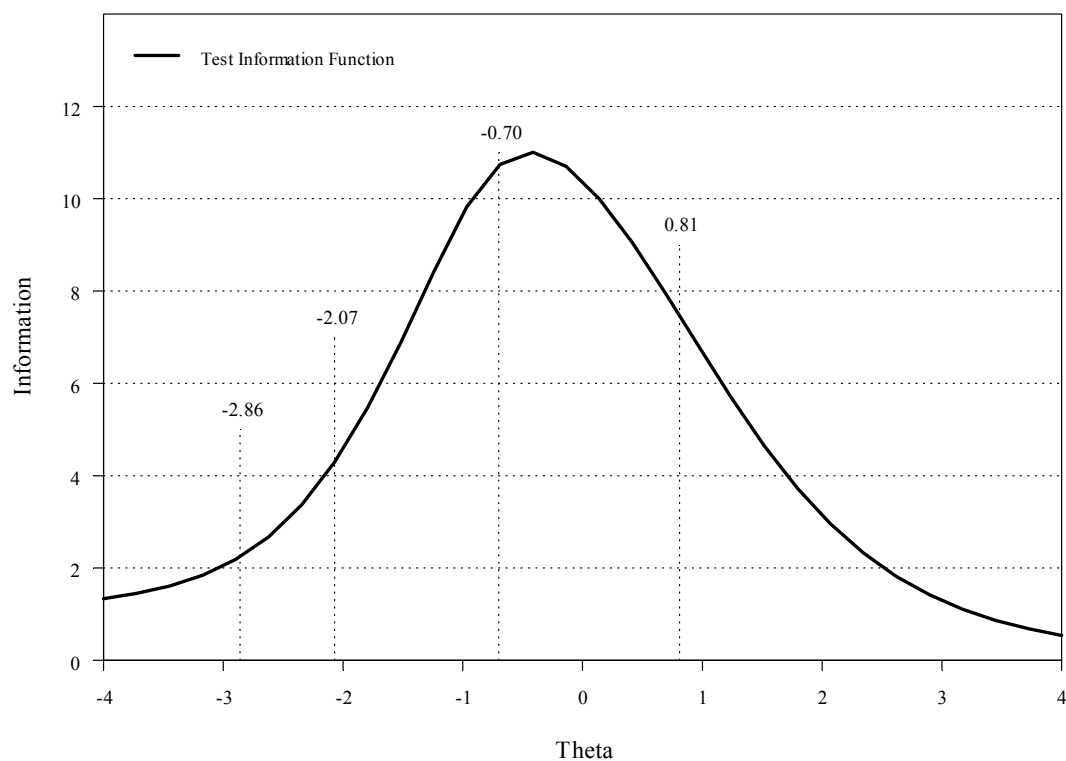
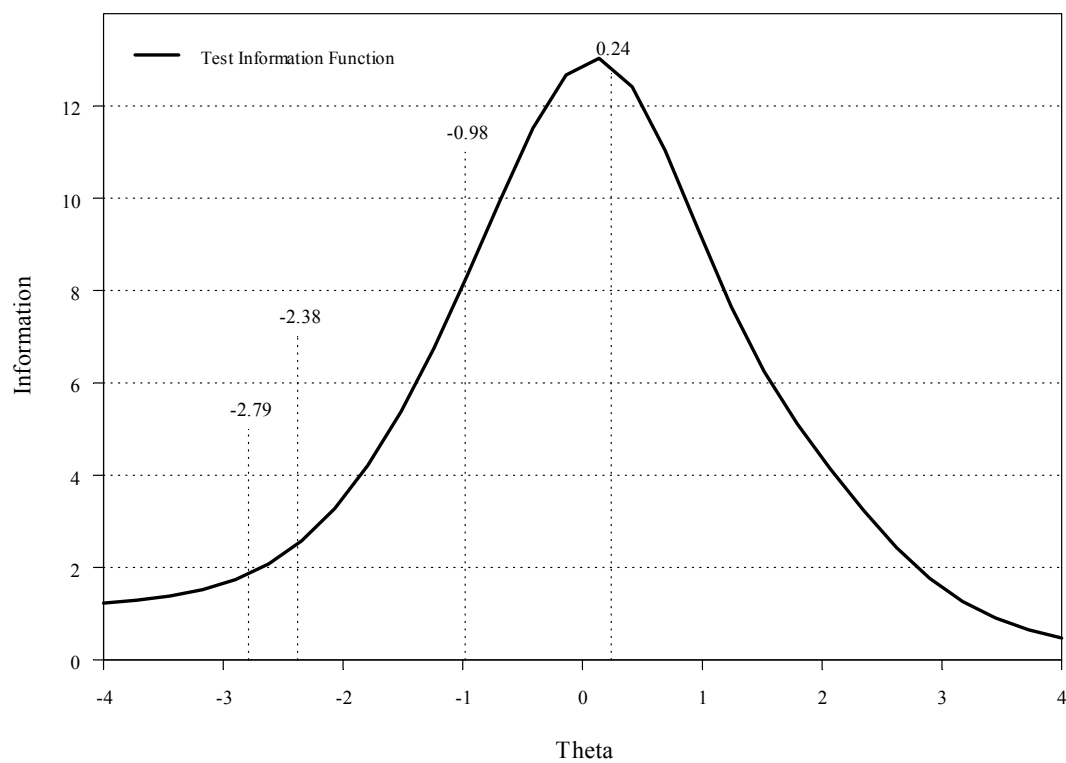


Figure 7.24 Test Information Function for Mathematics: Junior Division (French)



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

Assessment	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	0.08	0.13	-2.74
		Max.	0.93	0.50	1.08	1.31
		Mean	0.66	0.35**	0.62	-0.69
		SD	0.16	0.09	0.24	0.92
Junior Reading (English)	36	Min.	0.45	0.10	0.18	-4.10
		Max.	0.98	0.54	1.03	2.37
		Mean	0.70	0.35	0.60	-0.92
		SD	0.16	0.16	0.18	1.21
Primary Reading (French)	36	Min.	0.39	0.16	0.27	-3.62
		Max.	0.93	0.49	1.04	0.98
		Mean	0.66	0.35	0.60	-0.77
		SD	0.14	0.08	0.19	0.91
Junior Reading (French)	36	Min.	0.41	0.12	0.19	-2.36
		Max.	0.94	0.55	1.04	1.02
		Mean	0.66	0.35	0.58	-0.78
		SD	0.12	0.11	0.20	0.87

Note. SD = standard deviation.

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

** The mean item-total correlation was obtained by first transforming the item-total correlation for each item by Fisher's z and then back-transforming the resulting average z to the correlation metric.

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

Assessment	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*	
			Item Difficulty	Item-Total Correlation	Slope	Location
Primary Writing (English)	14	Min.	0.47	0.21	0.32	-1.71
		Max.	0.81	0.59	1.04	0.69
		Mean	0.64	0.42**	0.64	-0.70
		SD	0.11	0.13	0.18	0.69
Junior Writing (English)	14	Min.	0.40	0.24	0.41	-2.45
		Max.	0.90	0.61	1.15	1.28
		Mean	0.67	0.44	0.71	-0.99
		SD	0.13	0.14	0.25	0.92
Primary Writing (French)	14	Min.	0.44	0.22	0.41	-1.96
		Max.	0.83	0.58	1.14	0.30
		Mean	0.68	0.44	0.74	-0.97
		SD	0.14	0.13	0.20	0.69
Junior Writing (French)	14	Min.	0.29	0.18	0.35	-2.34
		Max.	0.89	0.62	1.21	2.12
		Mean	0.67	0.42	0.71	-0.93
		SD	0.15	0.17	0.26	1.14
Primary Mathematics (English)	36	Min.	0.45	0.29	0.27	-2.57
		Max.	0.89	0.73	1.33	0.59
		Mean	0.65	0.44	0.74	-0.70
		SD	0.11	0.11	0.29	0.84
Junior Mathematics (English)	36	Min.	0.32	0.24	0.29	-2.30
		Max.	0.90	0.83	1.20	1.22
		Mean	0.63	0.42	0.73	-0.60
		SD	0.15	0.14	0.21	0.96
Primary Mathematics (French)	36	Min.	0.47	0.23	0.32	-2.35
		Max.	0.90	0.63	1.41	0.60
		Mean	0.67	0.42	0.73	-0.77
		SD	0.12	0.10	0.28	0.81
Junior Mathematics (French)	36	Min.	0.26	0.08	0.32	-2.94
		Max.	0.91	0.69	1.28	1.98
		Mean	0.61	0.43	0.76	-0.49
		SD	0.15	0.14	0.26	1.05

Note. SD = standard deviation.

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

** The mean item-total correlation was obtained by first transforming the item-total correlation for each item by Fisher's z and then back-transforming the resulting average z to the correlation metric.

The Grade 9 Assessment of Mathematics

Classical Test Theory (CTT) Analysis

Table 7.4 presents descriptive statistics, Cronbach's alpha estimates of test score reliability and the SEM for the English-language and French-language applied and academic versions of the Grade 9 mathematics assessment. In 2010–2011 the mean percentages ranged from 51.11% (French, spring) to 57.38% (English, spring) for applied mathematics. The mean percentages ranged from 58.19% (French, spring) to 67.36% (English, spring) for academic mathematics.

Cronbach's alpha estimates ranged from 0.83 to 0.85 for applied mathematics and from 0.85 to 0.87 for academic mathematics. The corresponding SEMs were likewise similar, ranging from 7.06 to 7.82% of the possible maximum score for applied mathematics and 6.82% to 7.29% of the possible maximum score for academic mathematics. The obtained reliability coefficients and SEMs are acceptable, which 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

Assessment	No. of Items	Item Type		Possible Max. Score	No. of Students	Min.	Max.	Mean	SD	Alpha	SEM
		MC	OR								
Applied, Winter (English)	31	24	7	52	19 007	1	52	29.02	10.09	0.85	3.91
Applied, Spring (English)	31	24	7	52	22 986	0	52	29.84	8.68	0.83	3.57
Academic, Winter (English)	31	24	7	52	43 182	0	52	32.98	9.73	0.86	3.67
Academic, Spring (English)	31	24	7	52	55 299	3	52	35.03	9.19	0.85	3.53
Applied, Winter (French)	31	24	7	52	363	6	49	28.12	9.80	0.85	3.74
Applied, Spring (French)	31	24	7	52	1 058	6	52	26.58	9.02	0.83	3.67
Academic, Winter (French)	31	24	7	52	1 354	6	52	31.78	10.02	0.87	3.65
Academic, Spring (French)	31	24	7	52	2 744	1	52	30.26	9.97	0.86	3.71

Note. MC = multiple choice; OR = open response; 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 the student responses. The descriptive statistics reported in Table 7.5 refer to the total population of students. The mean student ability scores range from -0.14 to 0.04 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 parameter estimates for individual items are presented in Appendix 7.1.

Table 7.5 Descriptive Statistics of IRT Scores: Grade 9 Mathematics

Assessment	No. of Students	Min.	Max.	Mean	SD
Applied, Winter (English)	19 007	-3.24	2.77	-0.14	0.99
Applied, Spring (English)	22 986	-3.49	2.93	-0.01	0.96
Academic, Winter (English)	43 182	-3.44	2.34	-0.09	0.94
Academic, Spring (English)	55 299	-3.44	2.33	0.04	0.95
Applied, Winter (French)	363	-2.63	2.34	0.04	0.96
Applied, Spring (French)	1 058	-2.86	2.97	-0.11	0.94
Academic, Winter (French)	1 354	-2.76	2.55	0.01	0.97
Academic, Spring (French)	2 744	-3.35	2.60	-0.04	0.94

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 the graphs. The TCCs for all the winter and spring administrations either overlap or are very close to each other. This indicates that the winter and spring applied and academic versions of the assessment had the same difficulty level. The TIFs, which are displayed in Figures 7.29–7.32, indicate that the English applied version provided most of its information around the 2/3 cut points while the English academic, French academic and French applied versions provided most of their information between the 2/3 and 3/4 cut points.

Figure 7.25 Test Characteristic Curves and Distributions of Thetas: Grade 9 Applied Math (English)

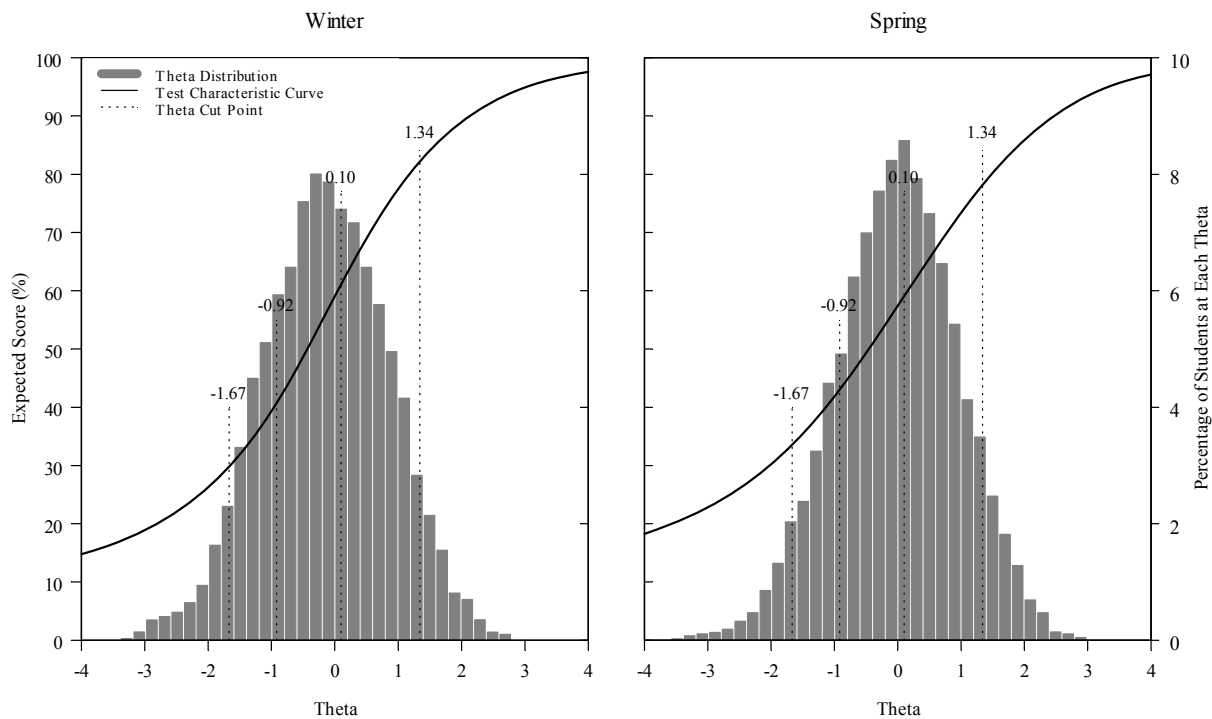


Figure 7.26 Test Characteristic Curves and Distributions of Thetas: Grade 9 Academic Math (English)

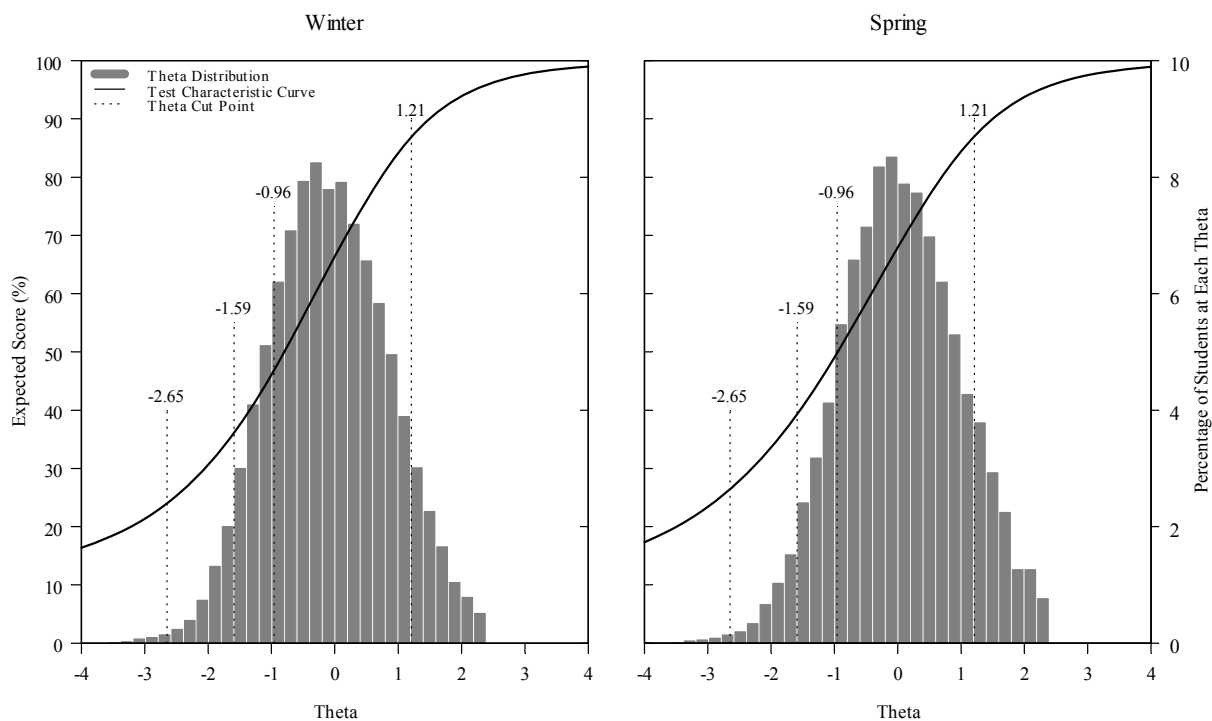


Figure 7.27 Test Characteristic Curves and Distributions of Thetas: Grade 9 Applied Math (French)

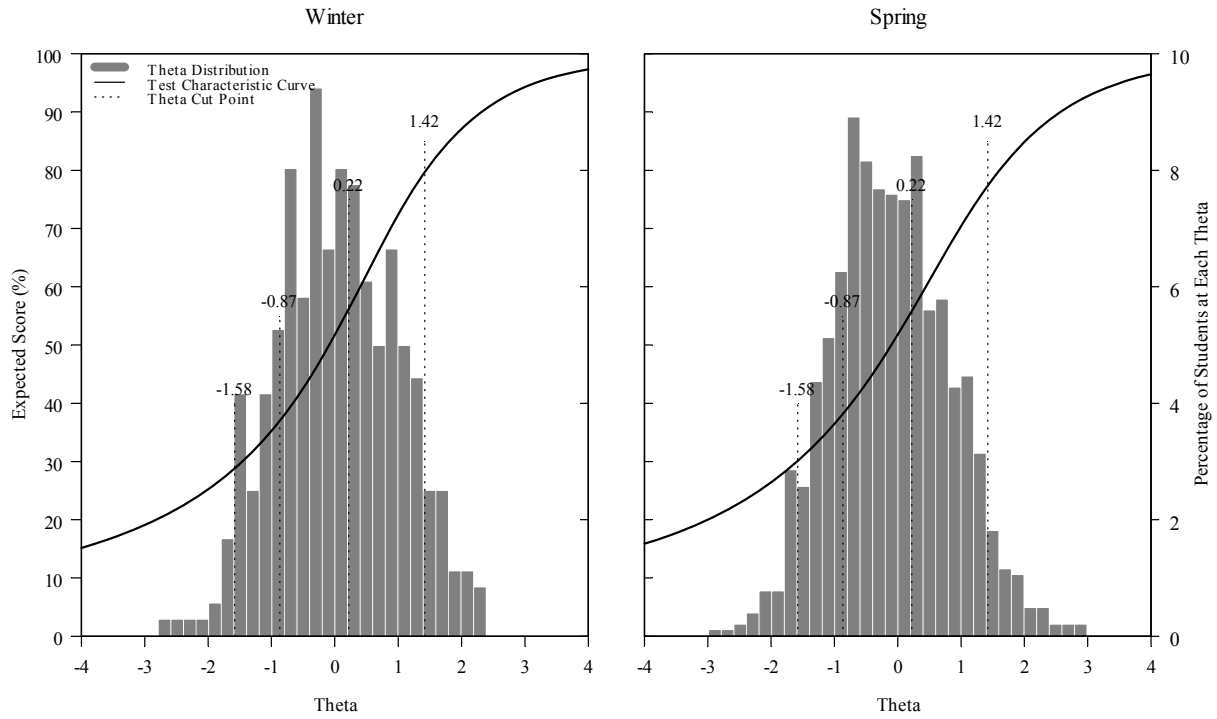


Figure 7.28 Test Characteristic Curves and Distributions of Thetas: Grade 9 Academic Math (French)

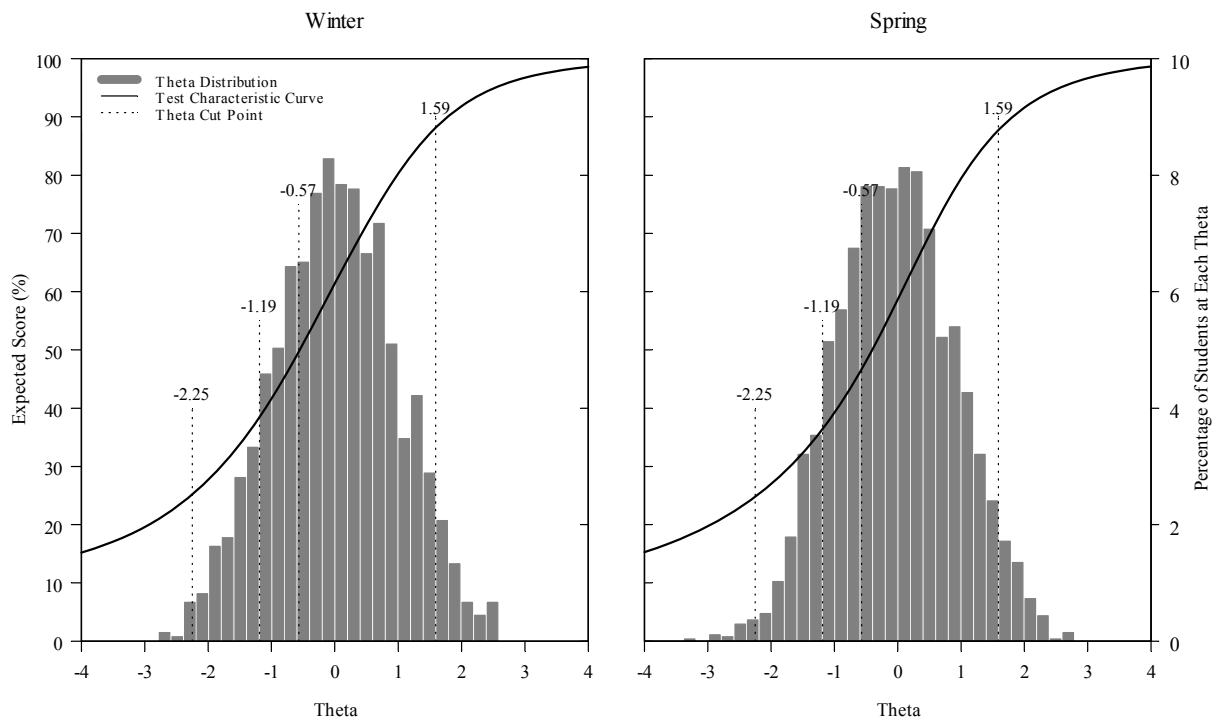


Figure 7.29 Test Information Functions: Grade 9 Applied Math (English)

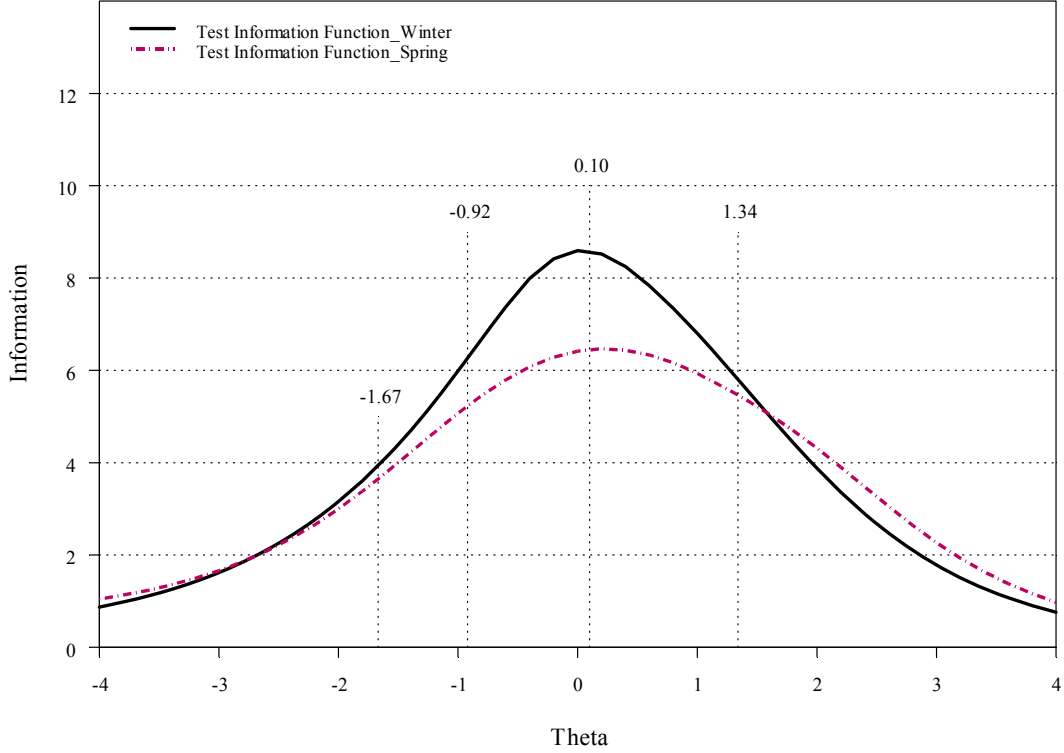


Figure 7.30 Test Information Functions: Grade 9 Academic Math (English)

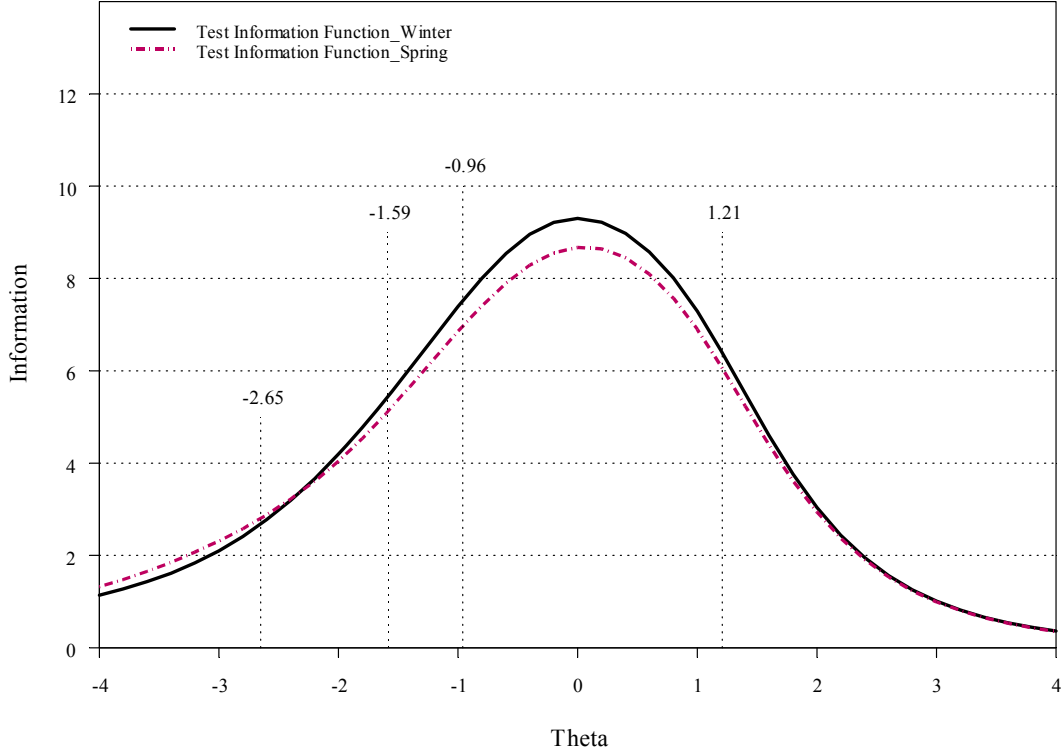


Figure 7.31 Test Information Functions: Grade 9 Applied Math (French)

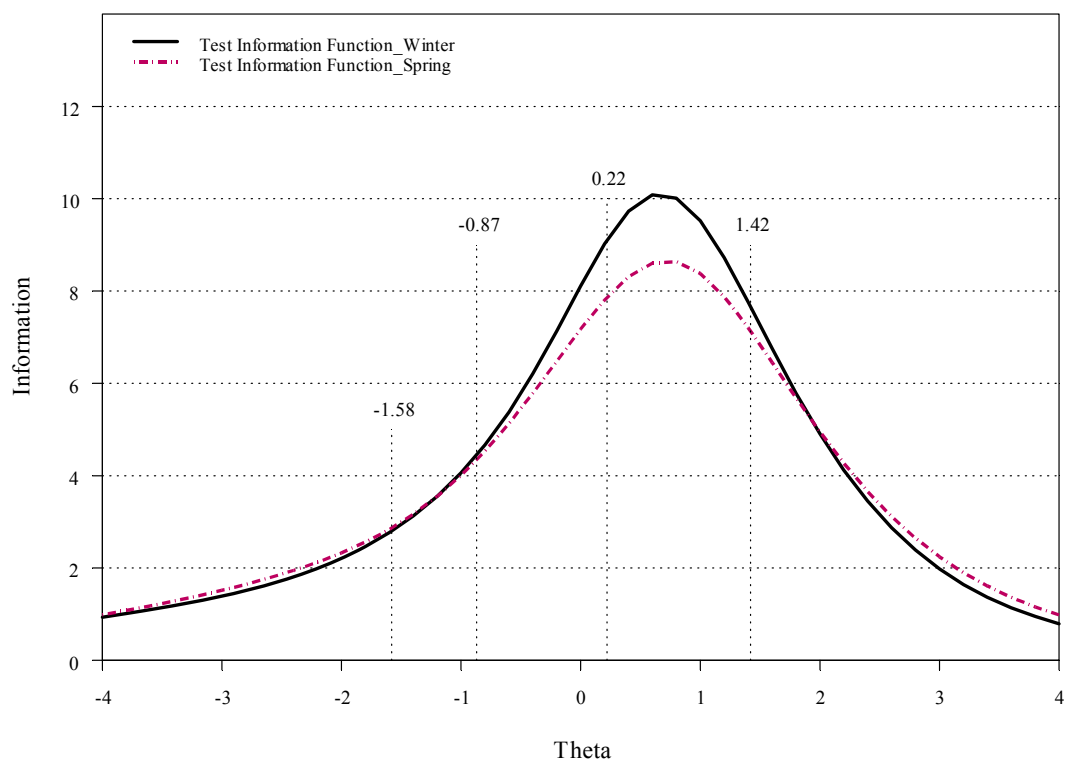
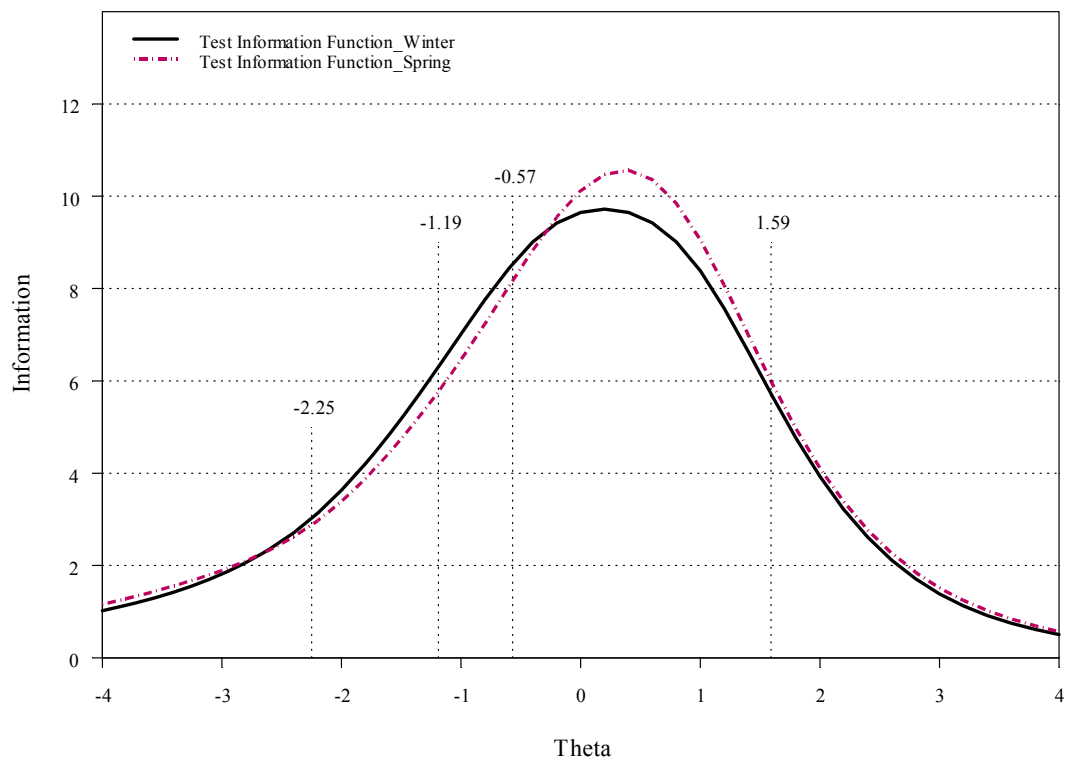


Figure 7.32 Test Information Functions: Grade 9 Academic Math (French)



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 on the Grade 9 mathematics assessment. Both classical item statistics and 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 parameter estimate, and the item-total correlation with the slope estimate. As expected, there is an inverse relationship between the CTT item difficulty estimates and the IRT location parameter estimates, and there is a positive relationship between the CTT item-total correlation estimates and the IRT slope parameter estimates. The item difficulty and location parameter estimates are in an acceptable range. Likewise, the point-biserial correlation coefficients are, for the most part, within an acceptable range, though values less than 0.20 are not ideal and indicate possible flaws in test items. The 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

Assessment	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*	
			Item Difficulty	Item-Total Correlation	Location	Slope
Applied, Winter (English)	31	Min.	29.86	0.18	-2.27	0.30
		Max.	87.44	0.71	1.85	1.38
		Mean	55.92	0.39	-0.08	0.62
		SD	14.34	0.14	0.98	0.22
Applied, Spring (English)	31	Min.	20.55	0.17**	-2.78	0.28
		Max.	87.38	0.64	2.43	1.03
		Mean	57.17	0.36	-0.06	0.57
		SD	16.62	0.13	1.22	0.17
Academic, Winter (English)	31	Min.	34.53	0.23	-1.84	0.34
		Max.	82.15	0.70	0.91	1.29
		Mean	62.71	0.41	-0.48	0.67
		SD	11.95	0.12	0.72	0.23
Academic, Spring (English)	31	Min.	38.37	0.22	-2.46	0.33
		Max.	88.71	0.67	0.91	1.29
		Mean	67.59	0.41	-0.68	0.67
		SD	12.62	0.11	0.88	0.25
Applied, Winter (French)	31	Min.	33.06	0.16	-1.56	0.29
		Max.	73.28	0.69	1.60	1.31
		Mean	53.68	0.40	0.20	0.65
		SD	9.97	0.14	0.74	0.29
Applied, Spring (French)	31	Min.	31.39	0.17	-1.78	0.29
		Max.	78.26	0.62	1.98	1.31
		Mean	51.22	0.37	0.20	0.61
		SD	13.11	0.13	0.96	0.26
Academic, Winter (French)	31	Min.	38.63	0.17	-1.85	0.31
		Max.	87.08	0.74	1.02	1.30
		Mean	60.47	0.43	-0.25	0.70
		SD	13.54	0.13	0.85	0.26
Academic, Spring (French)	31	Min.	37.21	0.18	-1.85	0.31
		Max.	86.70	0.76	1.36	1.33
		Mean	58.73	0.42	-0.22	0.68
		SD	13.41	0.15	0.85	0.25

Note. SD = standard deviation.

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

** The mean item-total correlation was obtained by first transforming the item-total correlation for each item by Fisher's z and then back-transforming the resulting average z to the correlation metric.

The Ontario Secondary School Literacy Test (OSSLT)

Classical Test Theory (CTT) Analysis

Table 7.7 presents descriptive statistics, Cronbach's alpha estimates of test-score reliability and the SEM for the first-time eligible students who wrote the English-language and French-language OSSLT. The test means (as percentages) for first-time eligible students are 76.9% for English-language students and 78.3% for French-language students.

Cronbach's alpha estimates are 0.90 and 0.84 for English and French, respectively. The corresponding SEMs are 4.0% and 4.1% of the possible maximum score. The obtained reliability coefficients and SEMs are acceptable and indicate that test scores from these assessments are at a satisfactory level of precision.

Table 7.7 Test Descriptive Statistics, Reliability and Standard Error of Measurement: OSSLT (First-Time Eligible Students)

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	143 246	4.0	81.0	62.3	10.07	0.90	3.26
French	47	39	4	2	2	81	5 458	14.0	81.0	63.4	8.39	0.84	3.34

Note. MC = multiple choice; OR = open response (reading); SW = short writing; LW = long writing; N = number of students; R = Cronbach's 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 all 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 parameter estimates for individual items are presented in Appendix 7.1.

Table 7.8 Descriptive Statistics for IRT Scores: OSSLT (First-Time Eligible Students)

Language	No. of Students	Min.	Max.	Mean	SD
English	143 246	-3.78	2.82	0.11	0.92
French	5 458	-3.60	2.76	0.18	0.80

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 tests are shown in Figures 7.35 and 7.36, respectively. The theta cut point for assigning students to the successful and unsuccessful levels of performance is marked on each plot.

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

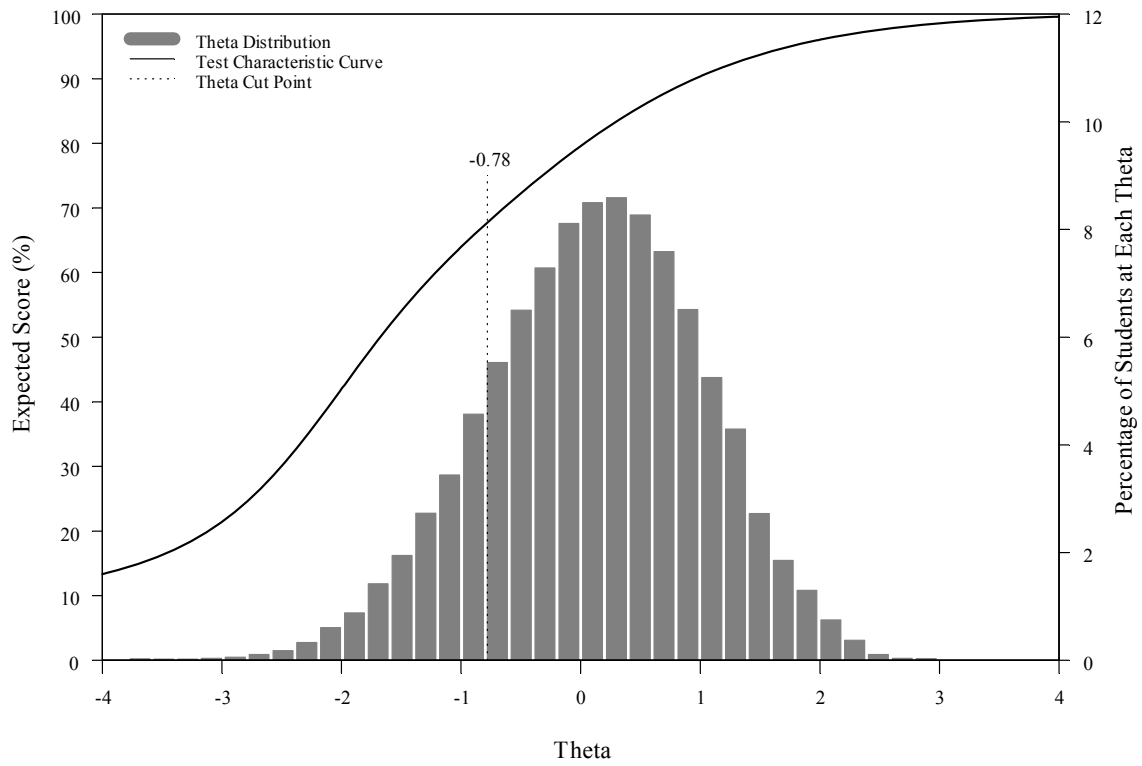


Figure 7.34 Test Characteristic Curve and Distribution of Theta: OSSLT (French)

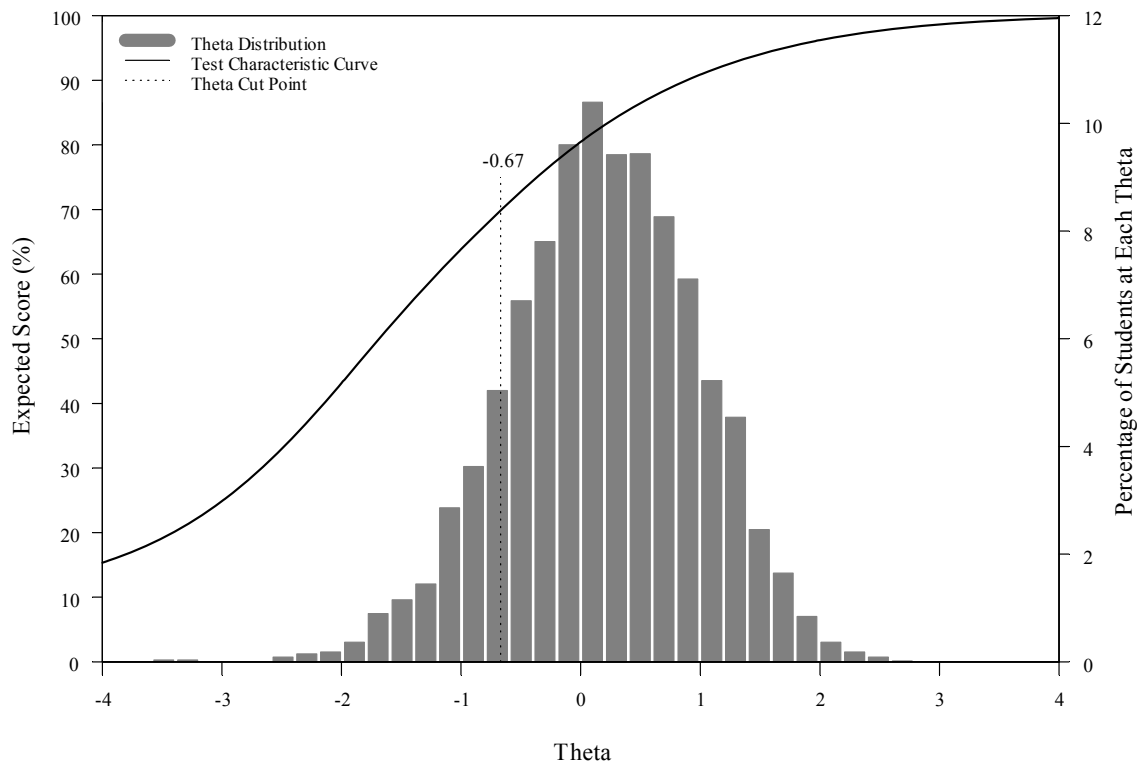


Figure 7.35 Test Information Function: OSSLT (English)

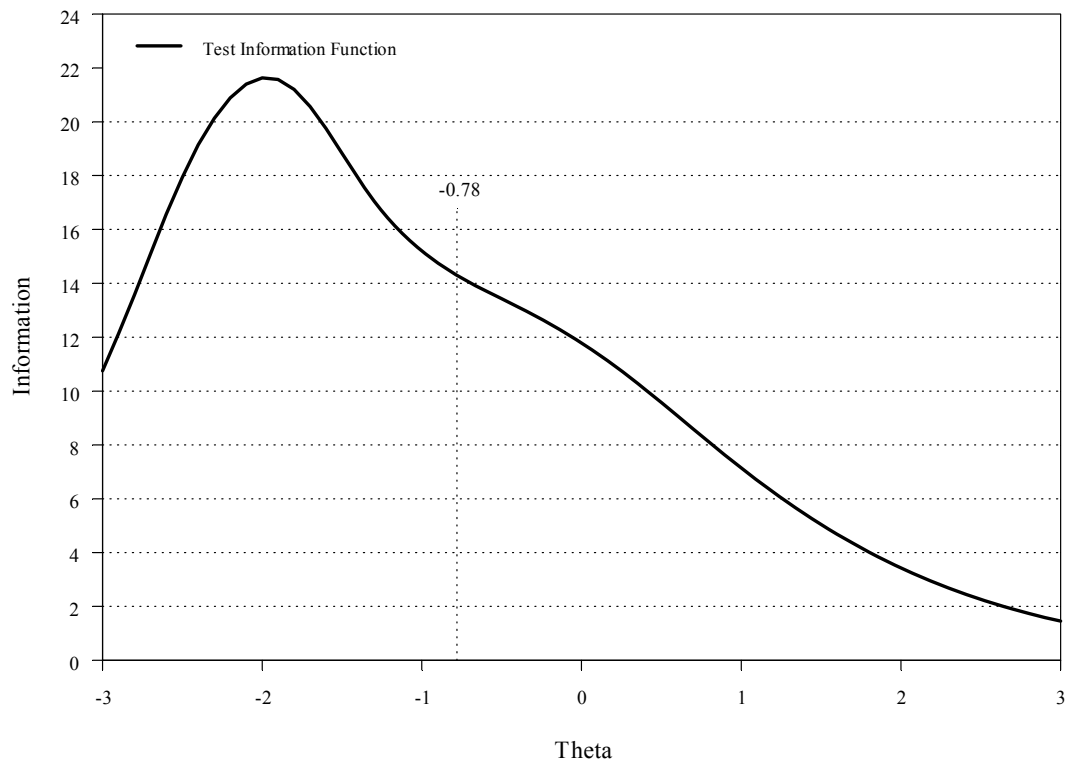
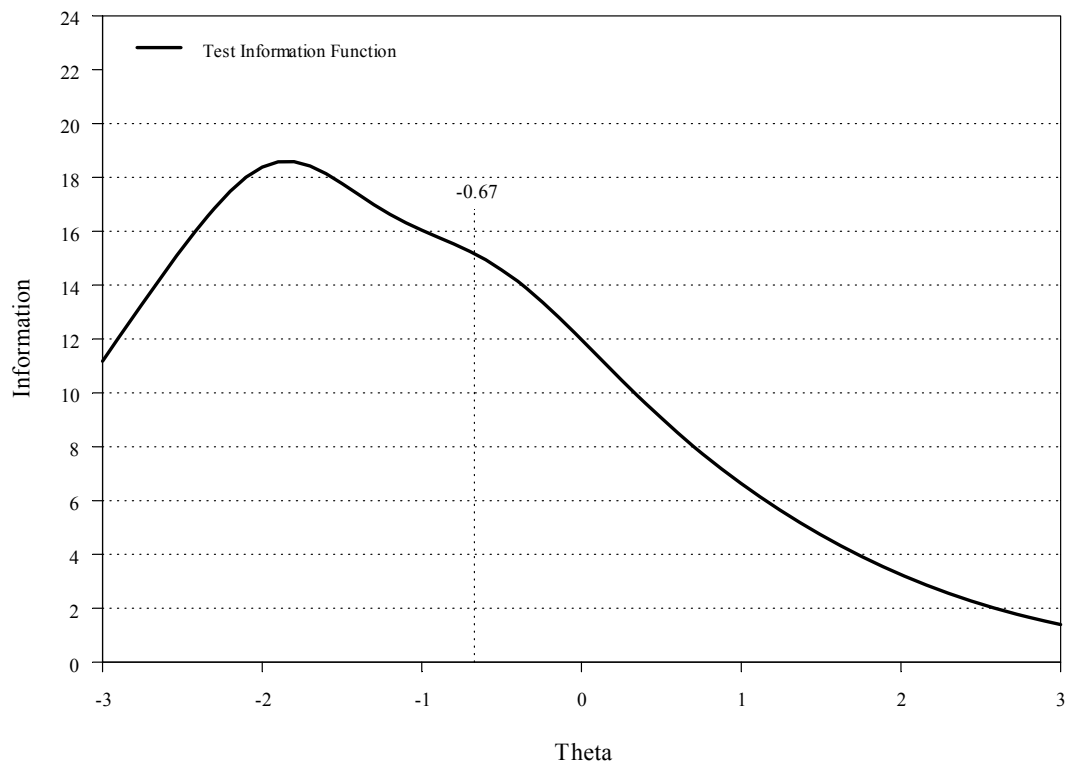


Figure 7.36 Test Information Function: OSSLT (French)



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 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 estimate, and the item-total correlation with the slope estimate. 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 correlations and the IRT slope parameter estimates. However, the low values for the minimum point-biserial correlations suggest that some of the items did not reach the desired level (0.20). 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.	0.39	0.08	-2.90
		Max.	0.99	0.60	1.26
		Mean	0.75**	0.36	-1.04
		SD	0.13	0.09	0.97
French	47	Min.	0.44	0.10	-2.89
		Max.	0.98	0.59	1.15
		Mean	0.77	0.29	-1.05
		SD	0.14	0.11	1.01

Note. SD = standard deviation.

* 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.

** The mean item-total correlation was obtained by first transforming the item-total correlation for each item by Fisher's z and then back-transforming the resulting average z to the correlation metric.

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., girls and boys, second language learners, or "SLLs," in English- or French-language schools). If an item is more difficult for one subgroup than for another, the item may be measuring something other than what it intends. However, it is important to recognize that DIF-flagged items may measure actual differences in relevant knowledge or skill (i.e., 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 the possible sources and interpretations of differences in achievement.

EQAO examined the 2011 assessments for gender- and second-language-learner-based (SLL-based) DIF using the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959) for multiple-choice items and Mantel's (1963) extension of the MH procedure, in conjunction with the standardized mean difference (SMD) (Dorans, 1989), for open-response items. In all analyses, males and non-SLLs were the reference group, and females and SLLs were the focal, or studied, group.

The MH test statistic was proposed as a method for detecting DIF by Holland and Thayer (1988). It examines whether an item shows DIF through the log of the ratio of the odds of a correct response for the reference group to the odds of a correct response for the focal group. With this procedure, examinees responding to a multiple-choice item are matched using the observed total score. The data for each item can be arranged in a $2 \times 2 \times K$ contingency table (see Table 7.10 for a slice of such a contingency table), where K is the number of possible total-score categories. The group of examinees is classified into two categories: the focal group and the reference group, and the item response is classified as correct or incorrect.

Table 7.10 2×2 Contingency Table for a MC Item for the k th Total-Test Score Category

Group	Item score		Total
	Correct = 1	Incorrect = 0	
Reference group	n_{11k}	n_{12k}	n_{1+k}
Focal group	n_{21k}	n_{22k}	n_{2+k}
Total group	n_{+1k}	n_{+2k}	n_{++k}

An effect-size measure of DIF for a MC item is obtained as the MH odds ratio:

$$\alpha_{MH} = \frac{\sum_{k=1}^K n_{11k} n_{22k} / n_{++k}}{\sum_{k=1}^K n_{12k} n_{21k} / n_{++k}} \quad (3)$$

The MH odds ratio was transformed to the delta scale in Equation 4 (used at Educational Testing Service, or ETS), and the ETS guidelines (Zieky, 1993) for interpreting the delta effect sizes were used to classify items into three categories of DIF magnitude, as shown in Table 7.11.

$$\Delta_{MH} = -2.35 \cdot \alpha_{MH} \quad (4)$$

Table 7.11 DIF Classification Rules for MC Items

Category	Description	Criterion
A	No or nominal DIF	Δ_{MH} not significantly different from 0, or $ \Delta_{MH} < 1$
B	Moderate DIF	Δ_{MH} significantly different from 0 and $1 \leq \Delta_{MH} < 1.5$, or Δ_{MH} significantly different from 0 and $ \Delta_{MH} \geq 1$ and $ \Delta_{MH} $ not significantly different from 1
C	Strong DIF	$ \Delta_{MH} $ significantly greater than 1 and $ \Delta_{MH} \geq 1.5$

For OR items, the SMD between the reference and focal groups was used in conjunction with the MH approach. The SMD compares the means of the reference and focal groups, adjusting for the differences in the distribution of the reference and focal group members across the values of the matching variable. The SMD has the following form:

$$SMD = \sum_k p_{Fk} m_{Rk} - \sum_k p_{Fk} m_{Fk}, \quad (5)$$

where $p_{Fk} = n_{F+k} / n_{F++}$ is the proportion of the focal group members who are at the k^{th} level of the matching variable, $m_{Fk} = 1/[n_{Fk} (\sum_t y_t \cdot n_{Ftk})]$ is the mean item score of the focal group members at the k^{th} level and m_{Rk} is the analogous value for the reference group. The SMD is divided by the item standard deviation of the total group to obtain an effect size value for the SMD, and these effect sizes, in conjunction with Mantel's (1963) extension of the MH chi-square (MH χ^2), are used to classify OR items into three categories of DIF magnitude, as shown in Table 7.12.

Table 7.12 DIF Classification Rules for OR Items

Category	Description	Criterion
A	No or nominal DIF	MH χ^2 not significantly different from 0 or $ \text{Effect size} \leq .17$
B	Moderate DIF	MH χ^2 significantly different from 0 and $.17 < \text{Effect size} \leq .25$
C	Strong DIF	MH χ^2 significantly different from 0 and $ \text{Effect size} > .25$

For each assessment, except for the French-language Grade 9 version, two random samples of 2000 examinees were selected from the population of students who took the test under standard administration conditions (i.e., with no accommodations).² The samples were stratified according to gender or second-language-learner (SLL) status. The term “second language learner” is used to represent English language learners for the English-language assessments and students in the ALF/PANA program for the French-language assessments. 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-level or C-level DIF in both samples were considered DIF items.

The item-level results are provided in Appendix 7.1. The results in each table are from two random samples and include the value of Δ for multiple-choice items, an effect size for open-response items and the significance level and the severity of DIF. Negative estimates of Δ and effect size indicate that the girls outperformed the boys or that the SLLs outperformed the non-SLLs; positive Δ and effect-size estimates indicate that the boys outperformed the girls or that the non-SLLs outperformed the SLLs.

The Primary- and Junior-Division Assessments

For the reading, writing and mathematics components of the 2011 primary- and junior-division 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 the two samples are reported in Tables 7.13 and 7.14, respectively. The numbers in the “boys” and “girls” columns indicate the number of DIF items favouring boys and girls.

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

Table 7.13 Number of B-Level Gender-Based DIF Items: Primary and Junior Assessments

Component	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)	4 (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	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)	0 (MC) 0 (OR)
Math <i>k</i> = 36	1 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	3 (MC) 0 (OR)	1 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)

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

Table 7.14 Number of C-Level Gender-Based DIF Items: Primary and Junior Assessments

Component	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)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (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)	0 (MC) 0 (OR)	0 (MC) 0 (OR)

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

Of the 344 items comprising the primary- and junior-division assessments, 14 items were found to have gender-based DIF. The majority of the DIF items had B-level DIF (13), and only one item had C-level DIF. Of the 13 multiple-choice items showing B- or C-level DIF, 12 favoured the boys, while one favoured the girls. The one open-response DIF item favoured the girls. 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 (5).

The summaries for SLL-based DIF are reported in Tables 7.15 and 7.16. Seven out of the 344 items across all the assessments showed SLL-based DIF, with six items having B-level DIF and one item having C-level DIF. Of the six multiple-choice items showing B- or C-level DIF, five favoured non-SLL students and one favoured SLL students. The one open-response DIF item favoured SLL students. The reading component of the English-language junior-division assessment was found to have the largest number of SLL-based DIF items (two).

Table 7.15 Number of B-Level SLL-Based DIF Items: Primary and Junior Assessments

Component	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) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 1 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)
Writing <i>k</i> = 14	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Math <i>k</i> = 35*	0 (MC) 0 (OR)	0 (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)

Note. SLL = Second-language learner; MC = multiple choice; OR = open response.

*The total number of items for the mathematics component of the English-language primary assessment was 36, but one item was dropped.

Table 7.16 Number of C-Level SLL-Based DIF Items: Primary and Junior Assessments

Component	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) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (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> = 35*	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)

Note. SLL = Second language learner; MC = multiple choice; OR = open response.

*The total number of items for the mathematics component of the English-language primary assessment was 36, but one item was dropped.

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 bias in the content of these items, they were not removed from the scoring process.

The Grade 9 Mathematics Assessment

The gender- and SLL-based DIF (favouring boys or girls or SLL or non-SLL) results for the academic and applied versions of the Grade 9 assessment are provided in Tables 7.17–7.20. 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 due to the small number of participating students. The number of participating French-language SLL students was also too small to conduct SLL-based DIF analysis.

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

Assessment	English		French	
	Boys	Girls	Boys	Girls
Applied, Winter <i>k</i> = 31	0 (MC)	0 (MC)	3 (MC)	2 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Applied, Spring <i>k</i> = 31	2 (MC)	1 (MC)	2 (MC)	1 (MC)
	0 (OR)	1 (OR)	0 (OR)	1 (OR)
Academic, Winter <i>k</i> = 31	2 (MC)	0 (MC)	0 (MC)	2 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Spring <i>k</i> = 31	1 (MC)	0 (MC)	5 (MC)	1 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)

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

Table 7.18 Number of C-Level Gender-Based DIF Items: Grade 9 Applied and Academic Mathematics

Assessment	English		French	
	Boys	Girls	Boys	Girls
Applied, Winter <i>k</i> = 31	0 (MC)	0 (MC)	1 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Applied, Spring <i>k</i> = 31	0 (MC)	0 (MC)	1 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Winter <i>k</i> = 31	0 (MC)	0 (MC)	1 (MC)	0 (MC)
	0 (OR)	0 (OR)	1 (OR)	0 (OR)
Academic, Spring <i>k</i> = 31	0 (MC)	0 (MC)	0 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)

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

Of the 248 items across eight Grade 9 math assessments, 22 multiple-choice items and two open-response items showed B-level gender-based DIF. Of the 22 B-level gender-based multiple-choice DIF items, fifteen favoured boys and seven favoured girls. The two B-level gender-based open-response items all favoured girls. Four items showed C-level gender-based DIF. All the items favoured boys. Overall, more multiple-choice items were found to favour boys, and more open-response items were found to favour girls.

Across Grade 9 math assessments, six to 12 multiple-choice items were used in both versions of the winter and spring administrations. Altogether five common items showed gender-based DIF in either winter or spring administration, and one item showed gender-based DIF across administrations.

Table 7.19 Number of B-Level SLL-Based DIF Items: Grade 9 Applied and Academic Mathematics

Assessment	Applied		Academic	
	Non-SLLs	SLLs	Non-SLLs	SLLs
Winter k = 31	0 (MC)	2 (MC)	1 (MC)	1 (MC)
	1 (OR)	0 (OR)	0 (OR)	0 (OR)
Spring k = 31	2 (MC)	0 (MC)	1 (MC)	6 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)

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

Table 7.20 Number of C-Level SLL-Based DIF Items: Grade 9 Applied and Academic Mathematics

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

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

There are 14 B-level SLL-based DIF items. Of these items, 13 are multiple-choice, with four items favouring non-SLL students, and one is open response, favouring non-SLL students. Ten items (nine multiple-choice and one open-response) showed C-level SLL-based DIF. Of the nine multiple-choice items, two favoured non-SLL students and seven favoured SLL students. The one open-response item favoured non-SLL students. Nine common items showed SLL-based DIF, with two items showing DIF in both administrations. Overall, a larger number of SLL-based DIF was found to favour SLL students, with a larger number of DIF items for the academic course.

All Grade 9 assessment items identified as B-level or C-level DIF items were reviewed by the assessment team. Since the reviewers did not identify any apparent bias in the content of these items, they were not removed from the scoring process.

The OSSLT

Gender-based DIF results for the OSSLT are presented in Tables 7.21 and 7.22 for B-level and C-level items, respectively.

Table 7.21 Number of B-Level Gender-Based DIF Items: OSSLT

English (<i>k</i> = 51)		French (<i>k</i> = 51)	
Males	Females	Males	Females
6 (MC)	1 (LW)	5 (MC)	2 (LW)

Note. MC = multiple choice; LW=long writing.

Table 7.22 Number of C-Level Gender-Based DIF Items: OSSLT

English (<i>k</i> = 51)		French (<i>k</i> = 51)	
Males	Females	Males	Females
2 (MC)	0	1 (MC)	0

Note. MC = multiple choice.

There were seven B-level DIF items on the English-language version of the OSSLT. Six multiple-choice items favoured the males, and one long-writing item exhibited B-level DIF favouring females for topic development and use of conventions. Two C-level DIF multiple-choice items favoured the males.

There were seven B-level DIF items on the French-language version of the OSSLT. Five multiple-choice items favoured the males, and two long-writing items exhibited B-level DIF favouring females: one long-writing item favoured the females for topic development and use of conventions, and another long-writing item favoured the females for use of conventions. One C-level DIF multiple-choice item favoured the males.

DIF analysis was not conducted for SLL students taking the French-language version of the OSSLT due to the small number of students in this group. For the English-language version of the OSSLT (see Table 7.23), one multiple-choice item and one open-response reading item exhibited B-level DIF favouring SLLs. One long-writing item exhibited B-level DIF favouring SLLs for topic development. Four multiple-choice items exhibited B-level DIF favouring non-SLLs.

One multiple-choice item and one open-response reading item exhibited C-level DIF favouring SLLs. Six multiple-choice items exhibited C-level DIF favouring non-SLLs.

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 bias in the content of these items, they were not removed from the scoring process.

Table 7.23 Number of SLL-Based DIF Items: OSSLT (English)

English (<i>k</i> = 51)			
B-Level DIF		C-Level DIF	
Non-SLLs	SLLs	Non-SLLs	SLLs
4 (MC)	1 (MC) 1 (OR) 1 (LW)	6 (MC)	1 (MC) 1 (OR)

Note. MC = multiple choice; OR = open response; LW=long writing.

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 who lack enough evidence to achieve Level 1 are placed at NE1. (Students without data and exempted students are not included in the calculation of results for participating students.) Through the equating process described in Chapter 5, four theta values are identified that define the cut points between adjacent levels (NE1 and Level 1, Level 1 and Level 2, Level 2 and Level 3 and Level 3 and Level 4). In the case of the OSSLT, EQAO reports only two levels of achievement: successful and unsuccessful. Thus, the OSSLT has one cut score.

Two issues that arise when students are placed into categories based on assessment scores are accuracy and consistency.

Accuracy

The term “accuracy” refers to the extent to which classifications based on observed student scores agree with classifications based on true scores. While observed scores include measurement error, true scores do not. Thus, classification decisions based on true scores are true or correct classifications. In contrast, classification decisions based on observed scores or derived from observed scores are not errorless. Since the errors may be positive, zero or negative, an observed score may be too low, just right or too high. This is illustrated in Table 7.24 for classifications in two adjacent categories (0 and 1).

Table 7.24 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 1.00.

Consistency

The term “consistency” refers to the extent to which classifications based on observed student scores on one form of the assessment agree with the classifications of the observed scores of the same students on a parallel form. In contrast to accuracy, neither set of observed scores on the two interchangeable tests is errorless. Some student scores on one test will be higher than their scores on the second test. For other students, their scores will be equal, and for other students still, their scores on the first test will be lower than their scores on the second. The differences,

when they occur, may be so large that they lead to different or inconsistent classifications. The classification based on the observed score could be lower, the same as or higher than the classification based on the second score. This is illustrated in Table 7.25 for classifications in two adjacent categories (0 and 1).

Table 7.25 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 & Lewis (1995) is used by EQAO because it yields estimates of both accuracy and consistency and allows for 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:

$$\tau_p = \frac{\xi_f(X) - X_{\min}}{X_{\max} - X_{\min}}, \quad (6)$$

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.

Decision consistency is estimated using the joint distribution of reported performance-level classifications on the current test form and performance-level classifications on the alternate or 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

$$\tilde{n} = \frac{(\hat{\mu}_X - X_{\min})(X_{\max} - \hat{\mu}_X) - r_{XX} \hat{\sigma}_X^2}{\hat{\sigma}_X^2 (1 - r_{XX})}, \quad (7)$$

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' :

$$X' = \tilde{n} \frac{X - X_{\min}}{X_{\max} - X_{\min}}. \quad (8)$$

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 $\tilde{n}$. 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.

The Primary and Junior Assessments

The classification indices for the primary- and junior-division assessments are presented in Table 7.26. The table includes the overall classification indices (i.e., across the five achievement levels) and the indices for the cut point at the provincial standard (i.e., classifying students into those who met the provincial standard and those who did not, using the Level 2-Level 3 cut). As expected, the indices for overall classification are lower than those for the provincial standard.

Table 7.26 Classification Accuracy and Consistency Indices: Primary- and Junior-Division Assessments

Assessment	Overall Accuracy	Overall Consistency	Accuracy at the Provincial Standard Cut	Consistency at the Provincial Standard Cut
Primary Reading (English)	0.79	0.72	0.90	0.86
Junior Reading (English)	0.83	0.75	0.91	0.88
Primary Reading (French)	0.77	0.68	0.90	0.86
Junior Reading (French)	0.83	0.76	0.93	0.89
Primary Writing (English)	0.85	0.79	0.88	0.83
Junior Writing (English)	0.82	0.75	0.89	0.84
Primary Writing (French)	0.79	0.72	0.90	0.86
Junior Writing (French)	0.81	0.73	0.91	0.86
Primary Mathematics (English)	0.81	0.73	0.90	0.85
Junior Mathematics (English)	0.79	0.70	0.90	0.86
Primary Mathematics (French)	0.82	0.74	0.91	0.88
Junior Mathematics (French)	0.82	0.75	0.92	0.89

The Grade 9 Assessment of Mathematics

The classification indices for the Grade 9 assessment are presented in Table 7.27. As is the case for the primary and junior assessments, the overall classification indices are lower than those for the provincial standard.

Table 7.27 Classification Accuracy and Consistency Indices: Grade 9 Mathematics

Assessment	Overall Accuracy	Overall Consistency	Accuracy at the Provincial Standard Cut	Consistency at the Provincial Standard Cut
Applied, Winter (English)	0.68	0.57	0.87	0.84
Applied, Spring (English)	0.68	0.57	0.87	0.82
Academic, Winter (English)	0.83	0.76	0.91	0.89
Academic, Spring (English)	0.83	0.76	0.92	0.89
Applied, Winter (French)	0.68	0.58	0.89	0.84
Applied, Spring (French)	0.66	0.56	0.89	0.84
Academic, Winter (French)	0.80	0.73	0.91	0.87
Academic, Spring (French)	0.81	0.74	0.90	0.86

The OSSLT

The classification indices for the English-language and French-language versions of the test are presented in Table 7.28. They indicate high accuracy and consistency for both versions.

Table 7.28 Classification Accuracy and Consistency Indices: OSSLT

Assessment	Accuracy (Successful or Unsuccessful)	Consistency (Successful or Unsuccessful)
English	0.93	0.90
French	0.93	0.89

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.
- Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. *Applied Measurement in Education*, 2, 217-233.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer, and H. I. Brown (Eds.), *Test Validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179-197.
- Mantel, N. (1963). Chi-square tests with one degree of freedom; Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, 58, 690-700.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.) *Differential item functioning* (pp. 337-347). Hillsdale, NJ: Erlbaum.

CHAPTER 8: VALIDITY EVIDENCE

Introduction

Each of the previous chapters in this report contributes important information to the validity argument by addressing one or more of the following aspects of the EQAO assessments: test development, test alignment, test administration, scoring, equating, item analyses, reliability, achievement levels and reporting. The goal of the present chapter is to build the validity argument for the EQAO assessments by tying together the information presented in the previous chapters, as well as introducing new, relevant information.

The Purposes of EQAO Assessments

EQAO assessments have the following general purposes: to report on student achievement and demonstrate the quality and accountability of Ontario's education system, to provide information to students (and their parents) on their achievement of the expectations at selected points in their education, to provide information to be used in improvement planning and to measure improvement in student achievement over time.

To meet these purposes, the EQAO annually conducts four province-wide assessments in both English and French languages: 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). These assessments measure how well students are achieving selected expectations as outlined in *The Ontario Curriculum*. The OSSLT is a graduation requirement and has been designed to ensure that students who graduate from Ontario high schools have achieved the minimum reading and writing skills defined in *The Ontario Curriculum* by the end of Grade 9.

Every year, the results are provided at the individual student, school, school board and provincial levels.

Conceptual Framework for the Validity Argument

In the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 1999), validity is defined as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (p. 9). The closely related term “validation” is viewed as the process of “developing a scientifically sound validity argument” and of accumulating evidence “to support the intended interpretation of test scores and their relevance to the proposed use” (p. 9). As suggested by Kane (2006), “The test developer is expected to make a case for the validity of the proposed interpretations and uses, and it is appropriate to talk about their efforts to validate the claims being made” (p. 17).

The above references (AERA et al., 1999; Kane, 2006) provide a framework for describing sources of evidence that should be considered when constructing a validity argument. These sources of evidence include: test content and response processes, internal structures, relationships to other variables and consequences of testing. These sources are not considered to be distinct types of validity. Instead, each contributes to a body of evidence about the validity of score interpretations and the actions taken on the basis of these interpretations. The usefulness of these different types of evidence may vary from test to test. A sound validity argument should integrate all the available evidence relevant to the technical quality and utility of a testing system.

Validity Evidence Based on the Content of the Assessments and the Assessment Processes

Test Specifications for EQAO Assessments

To fulfill the test purposes, the test specifications for EQAO assessments are based on curriculum content at the respective grades, in keeping with the *Principles for Fair Student Assessment Practices for Education in Canada* (1993). The Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions, measure how well elementary-school students at Grades 3 and 6 have met the reading, writing and mathematics curriculum expectations as outlined in *The Ontario Curriculum, Grades 1–8: Language* (revised 2006) and *The Ontario Curriculum, Grades 1–8: Mathematics* (revised 2005). The Grade 9 Assessment of Mathematics measures how well students have met the expectations for the ninth grade as outlined in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005). The OSSLT 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 test specifications are used in item development so that the number and types of items, as well as the coverage of expectations, are consistent across years. These specifications are presented in the EQAO framework documents, which define the construct measured by each assessment, identify the curriculum expectations covered by the assessment and present the target distribution of questions across content and cognitive domains. The curriculum expectations covered by the assessments are limited to those that can be measured by paper-and-pencil tests.

Appropriateness of Test Questions

EQAO ensures the appropriateness of the test questions to the age and grade of the students through the following two procedures in test development:

- involving Ontario educators as item writers and reviewers and
- field-testing all items prior to including them as operational items.

EQAO recruits and trains experienced Ontario educators as item writers and reviewers. The item-writing committee for each assessment consists of 10 to 20 educators who are selected because of their expert knowledge and recent classroom experience, familiarity with *The Ontario Curriculum*, expertise and experience in using scoring rubrics, written communication skills and experience in writing instructional or assessment materials for students. Workshops are conducted for training these item writers. After EQAO education officers review the items, item writers conduct cognitive labs in their own classes to try out the items. The results of the item tryouts help EQAO education officers review, revise and edit the items again.

EQAO also selects Ontario educators to serve on Assessment Development and Sensitivity Committees, based on their familiarity with *The Ontario Curriculum*, knowledge of and recent classroom experience in literacy education, mathematics education, experience with equity issues in education and experience with large-scale assessments. All items are reviewed by these committees. The goal of the Assessment Development Committee is to ensure that the items on EQAO assessments measure literacy and mathematics expectations in *The Ontario Curriculum*. The goal of the Sensitivity Committee is to ensure that these items are appropriate, fair and accessible to the broadest range of students in Ontario.

New items, except for the long-writing prompts on the primary- and junior-division assessments and the OSSLT, are field tested each year, as non-scored items embedded within the operational tests, before they are used as operational items. Each field-test item is answered by a representative sample of students. This field testing ensures that items selected for future operational assessments are psychometrically sound and fair for all students. The items selected for the operational assessments match the blueprint and have desirable psychometric properties.

Due to the amount of time required to field test long-writing prompts, these prompts are piloted only periodically, outside of the administration of the operational assessments.

Quality Assurance in Administration

EQAO has established quality assurance procedures to ensure both consistency and fairness in test administration and accuracy of results. These procedures include: external quality assurance monitors (visiting a random sample of schools to monitor whether EQAO guidelines are being followed), database analyses (examining the possibility of collusion between students and unusual changes in school performance) and examination of class sets of student booklets from a random sample of schools (looking for evidence of possible irregularities in the administration of assessments). EQAO also requires school boards to conduct thorough investigations of any reports of possible irregularities in the administration procedures.

Scoring of Open-Response Items

To ensure accurate and reliable results, EQAO follows rigorous procedures when scoring open-response items. All open-response items are scored by trained scorers. For consistency across items and years, EQAO uses generic rubrics to develop specific scoring rubrics for each open-response item included in each year's operational form. These item-specific scoring rubrics, together with anchors, are the key tools for scoring the open-response items. The anchors are chosen and validated by educators from across the province during range-finding. EQAO accesses the knowledge of subject experts from the Ontario education system in the process of preparing training materials for scorers. A range-finding committee, consisting of 8 to 25 selected Ontario educators, is formed to make recommendations on training materials. EQAO education officers then consider the recommendations and make final decisions for the development of these materials.

To ensure consistent scoring, scorers are trained to use the rubrics and anchors. Following training, scorers must pass a qualifying test before they begin scoring student responses. EQAO also conducts daily reviews of scorer validity and interrater reliability and provides additional training where indicated. Scorers failing to meet validity expectations may be dismissed.

Field-test items are scored using the same scoring requirements as those for the operational items. Scorers for field-test items are selected from the scorers of operational items to ensure accurate and consistent scoring of both. The results for the field-test items are used to select the items for the operational test for the next year.

For the items that are used for equating, it is essential to have accurate and consistent scoring across two consecutive years. To eliminate any possible changes in scoring across two years and to ensure the consistency of provincial standards, the student field-test responses to the open-response equating items from the previous year are rescored during the scoring of the current operational responses.

Scoring validity is assessed by examining the agreement between the scores assigned by the scorers and those assigned by an expert panel. EQAO has established targets for exact agreement and exact-plus-adjacent agreement. For the primary and junior assessments, the EQAO target of 95% exact-plus-adjacent agreement was met for all but one item in the reading components, for all items in the writing components and for all but one item in the mathematics components. The aggregate exact-plus-adjacent validity estimates for the items in each component ranged from 98.2 to 100%. For Grade 9 mathematics, the EQAO target of 95% exact-plus-adjacent agreement was met for all items, and the aggregate validity estimates ranged from 98.8 to 99.8%. For the

OSSLT, the EQAO target of 95% exact-plus-adjacent agreement was met for the four reading items on both the English- and French-language tests. This target was also met for all short-writing and all long-writing prompts. The aggregate validity estimates ranged from 96.7 to 99.9% (see Appendix 4.1).

In addition, student responses to multiple-choice items are captured by double-key entry for the primary assessments and by optical-scan forms for the other assessments. EQAO also conducts a quality assurance check to ensure that fields are captured with a 99.9% accuracy rate.

Equating

The fixed-common-item-parameter (FCIP) procedure 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. EQAO uses state-of-the-art equating procedures to ensure comparability of results across years. A small number of field-test items are embedded in each operational form in positions that are not revealed to the students. For more details on the equating process, see Chapter 5.

These equating procedures enable EQAO to monitor changes in student achievement over time. Research conducted by EQAO on model selection (Xie, 2007) and on equating methods (Pang, Madera, Radwan & Zhang, 2010) showed that both the current IRT models and the FCIP equating method used by EQAO are appropriate and function well with the EQAO assessments. To ensure that analyses are correctly completed, the analyses conducted by EQAO staff are replicated by a qualified external contractor.

Validity Evidence Based on the Test Constructs and Internal Structure

Test Dimensionality

An underlying assumption of IRT models for score interpretation is that there is a unidimensional structure underlying each assessment. A variation of the parallel analysis procedure was conducted for selected 2009 and 2010 EQAO operational assessments, and the results show that, although two or three dimensions were identified for the assessments, there is one dominant factor in each assessment (Zhang, Pang, Xu, Gu, Radwan, Madera, 2011). These results indicate that the IRT models would probably be robust with respect to the dimensionality of the assessments. This conclusion was also supported by EQAO research on the appropriateness of the IRT models used to calibrate assessment items, which included an examination of dimensionality (Xie, 2007).

Technical Quality of the Assessments

When selecting items for the operational assessment forms, the goal is to have items with p -values within the 0.25 to 0.95 range and item-to-total-test correlations of 0.20 or higher. To meet the requirements of the test blueprints, it is sometimes necessary to include a small number of items with statistics outside these ranges. For each assessment, a target test information function (TIF) also guides the construction of its new operational test form. Based on the pool of operational items from previous assessments, a target TIF was developed for each assessment by taking test length and item format into consideration. The use of target TIFs reduces the potential of drift across years and of perpetuating test weaknesses from one year to the next, and helps to meet and maintain the desired level of precision at critical points on the score scale.

To assess the precision of the scores for the EQAO assessments, a variety of test statistics are computed, including Cronbach's alpha reliability coefficient, the standard error of measurement, test characteristic curves, test information functions, differential item functioning statistics and classification accuracy and consistency. Overall, the results of these measures indicate that satisfactory levels of precision have been obtained. The reliability coefficients ranged from 0.80 to 0.89 for the primary and junior assessments, 0.83 to 0.87 for Grade 9 mathematics, and 0.84 to 0.90 for the OSSLT. The classification accuracy for students who were at or above the provincial standard for the primary, junior and Grade 9 assessments and who were successful on the OSSLT ranged from 0.87 to 0.93, indicating that about 90% of students were correctly classified.

As discussed above, a number of factors contributed to this level of precision: the quality of the individual assessment items, the accuracy and consistency of scoring and the interrelationships among the items. All items on the EQAO assessments are directly linked to expectations in the curriculum. For the operational assessments, EQAO selects items that are of an appropriate range of difficulty and that discriminate between students with high and low levels of achievement. As described above, a number of practices maintain and improve accuracy and consistency in scoring.

To further ensure that the assessments are well-designed and conducted according to current best practices, an External Psychometric Expert Panel (PEP) meets twice a year with officials from EQAO. The PEP responds to questions from EQAO staff and reviews the item and test statistics for all operational forms, the psychometric procedures used by EQAO and all the research projects on psychometric issues.

Validity Evidence Based on External Assessment Data

Linkages to International Assessment Programs

EQAO commissioned research to compare Ontario provincial reading assessment content and standards for the primary and junior assessments with those of the Progress in International Reading Literacy Study (PIRLS) in Grade 4 (Peterson, 2007; Simon, Dionne, Simoneau & Dupuis, 2008). The conclusion of these studies was that the constructs, benchmarks and performance levels for the EQAO and PIRLS assessments were sufficiently similar to allow for reasonable comparisons of the overall findings and trends in student performance. The expectations corresponding to the high international benchmark (for PIRLS) and Level 3 (Ontario provincial standard) were comparable. EQAO is presently conducting research to examine literacy skills by linking performance on the OSSLT with performance on the Programme for International Student Assessment (PISA) 2009 reading assessment. Both assessments were administered to the same group of students between April and May 2009.

The analysis completed to date shows that the standard for a successful result on the OSSLT is comparable to the standard for Level 2 achievement on PISA, which is the achievement benchmark at which students begin to demonstrate the kind of knowledge and skills needed to use reading competencies effectively. The basic literacy competency defined for the OSSLT is consistent with this description of Level 2 literacy in PISA. The percentage of students achieving at or above Level 2 on PISA is slightly higher than the percentage of successful students on the OSSLT.

Validity Evidence Supporting Appropriate Interpretations of Results

Setting Standards

During the first administrations of the EQAO assessments in Grades 3 and 6, teachers assigned student achievement levels to each student based on an evaluation of the student's body of work in a number of content and cognitive domains. A panel of educators reviewed the students' work and selected anchor papers, which were assigned to each achievement level. These anchor papers represented the quality of work expected at each achievement level, based on the expert opinion of the panel. Since 2004, these standards have been maintained through equating.

When the Grade 9 Assessment of Mathematics and the OSSLT were introduced, standard-setting panels were convened to set cut points for each reporting category. A modified Angoff approach was used to set the cut points.

A second standard-setting session was conducted for the OSSLT in 2006, when a single literacy score was calculated to replace the separate reading and writing scores that had been used up to that point. For OSSLT, the purpose of the standard-setting session was to apply the standards that had already been set for writing and reading separately to the combined test. EQAO also conducted a linking study by creating a pseudo test from the 2004 items that resembled the structure, content and length of the 2006 test. A scaling-for-comparability analysis, using common items across the two years, was conducted to place the scores of the two tests on a common scale. This analysis used a fixed common-item-parameter non-equivalent group design. The decision on the cut point for the 2006 test was informed by both the standard-setting session and the scaling-for-comparability analysis.

A second standard-setting session for Grade 9 applied mathematics was conducted in 2007, when there was a substantial change to the provincial curriculum. This process established a new standard for this assessment.

Reporting

EQAO employs a number of strategies to promote the appropriate interpretation of reported results. The Individual Student Report (ISR) presents student achievement according to levels that have been defined for the curriculum and used by teachers in determining report card marks. The ISR for the OSSLT identifies areas where a student has performed well and where a student should improve. The ISRs for the primary, junior and Grade 9 assessments include school, school board and provincial results that provide an external referent to further interpret individual student results. The ISR for the OSSLT includes the median scale score for the school and province.

EQAO provides interpretation guides and workshops on the appropriate uses of assessment results in school improvement planning. The workshops are conducted by the members of the Outreach Team. These members have intimate knowledge of the full assessment process and the final results. As well, EQAO provides school success stories that are shared with all schools in Ontario as a way of suggesting how school-based personnel can use the assessment results to improve student learning. EQAO also provides information to the media and the public on appropriate uses of the assessment results for schools. In particular, EQAO emphasizes that EQAO results must be interpreted in conjunction with a wide range of available information concerning student achievement and school success.

According to feedback collected by the Outreach Team and teacher responses on questionnaires, educators are finding the EQAO results useful.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 17-64). Westport, CT: American Council on Education and Praeger Publishers.
- Pang, X., Madera, E., Radwan, N. & Zhang, S. (2010). *A comparison of four test equating methods*. Retrieved November 8, 2011, from http://www.eqao.com/Research/pdf/E/Equating_Crp_cftem_ne_0410.pdf.
- Peterson, S.S. (2007). *Linking Ontario provincial student assessment standards with those of the Progress in International Reading Literacy Study (PIRLS), 2006*. Retrieved November 8, 2011, from http://www.eqao.com/Research/pdf/E/StandardsStudyReport_PIRLS2006E.pdf.
- Simon, M., Dionne, A., Simoneau, M. & Dupuis, J. (2008). *Comparison des normes établies pour les évaluations provinciales en Ontario avec celles du Programme international de recherche en lecture scolaire (PIRLS), 2006*. Retrieved November 8, 2011, from http://www.eqao.com/Research/pdf/F/StandardsStudyReport_PIRLS2006F.pdf.
- Working Group and Joint Advisory Committee. (1993). *Principles for fair student assessment practices for education in Canada*. Retrieved November 8, 2011, from http://www2.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf.
- Xie, Y. (2007). *Model selection for the analysis of EQAO assessment data*. Unpublished paper.
- Zhang, S., Pang, X., Xu, Y., Gu, Z., Radwan, N., & Madera, E. (2011). *Multidimensional item response theory (MIRT) for subscale scoring*. Unpublished paper.

APPENDIX 4.1: SCORING VALIDITY AND INTERRATER RELIABILITY

This appendix presents validity and interrater reliability estimates for the scoring of all open-response questions.

Validity: the Primary and Junior Assessments

Table 4.1.1 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
10772	1	5	1 658	97.7	71.2	26.5	18.0	8.5	2.4
10767	1	6	1 493	99.0	85.8	13.2	9.2	4.0	1.1
10934	1	11	1 344	99.8	78.4	21.4	12.2	9.2	0.2
10929	1	12	1 525	99.9	83.7	16.2	8.1	8.1	0.1
10916	1	18	1 167	100.0	90.1	10.0	3.7	6.3	0.0
10922	1	19	1 311	99.3	73.5	25.8	17.9	7.9	0.7
12065	2	11	1 167	99.7	93.7	6.0	5.7	0.3	0.3
12066	2	12	1 362	99.7	91.3	8.4	4.6	3.8	0.3
10879	2	22	1 331	99.0	72.9	26.1	20.1	6.0	1.1
10877	2	23	1 462	99.5	86.7	12.8	8.5	4.3	0.5
Aggregate			13 820	99.3	82.4	17.0	11.0	6.0	0.7

Table 4.1.2 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
10625	1	5	1 656	99.6	84.4	15.2	4.6	10.6	0.4
10627	1	6	1 833	98.4	76.8	21.6	20.0	1.6	1.6
10616	1	11	1 840	98.6	76.8	21.8	13.0	8.8	1.5
10620	1	12	2 164	98.9	77.8	21.1	12.5	8.6	1.1
10518	1	18	1 834	99.0	81.5	17.5	8.7	8.8	1.0
10523	1	19	1 892	99.4	79.0	20.4	10.1	10.3	0.7
12321	2	11	1 700	99.9	84.0	15.9	13.8	2.1	0.1
12320	2	12	1 651	99.7	80.1	19.6	15.0	4.6	0.3
10589	2	22	1 696	99.1	78.2	20.9	14.4	6.5	0.9
10587	2	23	1 789	99.8	77.6	22.2	9.6	12.6	0.2
Aggregate			18 055	99.2	79.5	19.7	12.2	7.5	0.8

Table 4.1.3 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
11063	1	5	145	100.0	93.1	6.9	1.4	5.5	0.0
11062	1	6	98	100.0	76.5	23.5	4.1	19.4	0.0
12239	1	11	125	100.0	83.2	16.8	0.8	16.0	0.0
12240	1	12	127	100.0	91.3	8.7	0.8	7.9	0.0
11193	1	18	128	99.2	92.2	7.0	7.0	0.0	0.8
11194	1	19	127	100.1	96.9	3.2	1.6	1.6	0.0
12203	2	11	130	100.0	96.9	3.1	0.8	2.3	0.0
12204	2	12	105	100.0	93.3	6.7	4.8	1.9	0.0
12251	2	22	112	100.0	94.6	5.4	0.0	5.4	0.0
12252	2	23	112	99.1	82.1	17.0	12.5	4.5	0.9
Aggregate			1209	99.8	90.4	9.5	3.2	6.2	0.2

Table 4.1.4 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
11379	1	5	163	100.0	92.0	8.0	3.1	4.9	0.0
11380	1	6	109	100.0	69.7	30.3	13.8	16.5	0.0
11617	1	11	133	100.0	71.4	28.6	26.3	2.3	0.0
11619	1	12	135	99.2	68.1	31.1	5.2	25.9	0.7
11423	1	18	164	99.3	96.3	3.0	0.0	3.0	0.6
11424	1	19	109	100.0	90.8	9.2	6.4	2.8	0.0
11475	2	11	60	100.0	95.0	5.0	5.0	0.0	0.0
11476	2	12	61	91.8	67.2	24.6	23.0	1.6	8.2
11405	2	22	89	100.0	91.0	9.0	4.5	4.5	0.0
11407	2	23	119	99.2	90.8	8.4	2.5	5.9	0.8
Aggregate			1142	99.3	83.8	15.5	8.2	7.4	0.7

Table 4.1.5 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
17369 T	2	24	1716	100.0	88.5	11.5	6.4	5.1	0.0
17369 V	2	24	1777	99.9	90.2	9.7	3.4	6.4	0.1
Aggregate Long Writing			3493	99.9	89.4	10.6	4.9	5.8	0.1
12173 T	1	13	1343	99.5	81.8	17.6	8.9	8.8	0.5
12173 V	1	13	1355	100.0	83.2	16.8	5.8	10.9	0.0
11050 T	2	17	1509	98.9	75.7	23.1	19.9	3.2	1.1
11050 V	2	17	1614	99.5	86.6	12.9	6.2	6.8	0.5
Aggregate Short Writing			5821	99.5	81.9	17.5	10.3	7.3	0.5
Aggregate All Items			9314	99.6	84.7	14.9	8.3	6.7	0.4

Table 4.1.6 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
17733 T	2	24	2277	99.6	76.9	22.8	8.5	14.3	0.4
17733 V	2	24	2264	99.9	81.8	18.1	5.6	12.5	0.1
Aggregate Long Writing			4541	99.7	79.3	20.5	7.1	13.4	0.3
10703 T	1	13	1826	99.7	83.8	15.8	7.0	8.9	0.3
10703 V	1	13	1939	99.9	82.5	17.5	2.7	14.7	0.1
10708 T	2	17	1673	99.4	82.8	16.6	11.9	4.7	0.6
10708 V	2	17	1677	99.8	86.1	13.7	5.8	7.9	0.2
Aggregate Short Writing			7115	99.7	83.8	16.0	6.7	9.3	0.3
Aggregate All Items			11656	99.7	82.0	17.7	6.8	10.9	0.3

Table 4.1.7 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
17714_T	2	24	200	99.0	83.0	16.0	8.0	8.0	1.0
17714_V	2	24	174	100.0	90.2	9.8	4.0	5.7	0.0
Aggregate Long Writing			374	99.5	86.3	13.1	6.1	6.9	0.5
11242_T	1	13	168	100.0	76.8	23.2	6.5	16.7	0.0
11242_V	1	13	139	100.0	83.5	16.5	6.5	10.1	0.0
11241_T	2	17	138	100.0	77.5	22.5	15.9	6.5	0.0
11241_V	2	17	82	100.0	89.0	11.0	6.1	4.9	0.0
Aggregate Short Writing			527	100.0	80.6	19.4	8.9	10.4	0.0
Aggregate All Items			901	99.8	83.0	16.8	7.8	9.0	0.2

Table 4.1.8 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
18331_T	2	24	123	98.4	83.7	14.6	10.6	4.1	1.6
18331_V	2	24	155	100.0	81.9	18.1	7.1	11.0	0.0
Aggregate Long Writing			278	99.3	82.7	16.5	8.6	7.9	0.7
11596_T	1	13	238	99.2	84.0	15.1	8.8	6.3	0.8
11596_V	1	13	139	100.0	87.1	12.9	3.6	9.4	0.0
11589_T	2	17	192	99.5	84.4	15.1	7.3	7.8	0.5
11589_V	2	17	160	99.4	82.5	16.9	5.6	11.3	0.6
Aggregate Short Writing			729	99.5	84.4	15.1	6.7	8.4	0.5
Aggregate All Items			1007	99.4	83.9	15.5	7.2	8.2	0.6

Table 4.1.9 Validity Estimates for Mathematics: Primary Division (English)

Item Code	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
13132	7	945	99.9	96.1	3.8	0.0	3.8	0.1
EQME10064	8	848	99.2	90.3	8.8	8.6	0.2	0.8
EQME10132	9	1058	100.0	95.7	4.3	3.1	1.2	0.0
10845	10	850	99.4	96.7	2.7	2.4	0.4	0.6
10963	26	878	99.7	95.4	4.2	2.8	1.4	0.3
10791	27	996	98.9	92.1	6.8	3.3	3.5	1.1
10964	28	1120	99.7	93.7	6.1	1.5	4.6	0.3
10848	29	749	99.5	93.5	6.0	5.1	0.9	0.5
Aggregate		7444	99.5	94.2	5.4	3.2	2.1	0.5

Table 4.1.10 Validity Estimates for Mathematics: Junior Division (English)

Item Code	Sequence	No. of Scores	% Exact Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
11393	7	644	97.7	89.4	8.2	2.3	5.9	2.3
11348	8	912	99.8	98.2	1.5	1.2	0.3	0.2
11455	9	828	100.0	90.6	9.4	6.3	3.1	0.0
11454	10	824	99.6	91.5	8.1	0.2	7.9	0.4
11408	26	999	99.7	95.0	4.7	2.3	2.4	0.3
11452	27	1252	99.8	92.8	7.0	2.9	4.2	0.2
11453	28	1007	98.8	76.7	22.1	1.3	20.9	1.2
11404	29	1173	99.7	83.2	16.5	1.2	15.3	0.3
Aggregate		7639	99.5	89.5	10.0	2.2	7.8	0.5

Table 4.1.11 Validity Estimates for Mathematics: Primary Division (French)

Item Code	Sequence	No. of Scores	% Exact Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
13105	7	62	100.0	96.8	3.2	0.0	3.2	0.0
13095	8	63	100.0	96.8	3.2	0.0	3.2	0.0
13313	9	55	100.0	100.0	0.0	0.0	0.0	0.0
13265	10	54	100.0	100.0	0.0	0.0	0.0	0.0
OQMF31569	26	86	100.0	80.2	19.8	0.0	19.8	0.0
13219	27	114	99.1	86.8	12.3	11.4	0.9	0.9
13341	28	75	100.0	84.0	16.0	13.3	2.7	0.0
13326	29	103	99.0	84.5	14.6	9.7	4.9	1.0
Aggregate		612	99.7	89.5	10.1	5.4	4.7	0.3

Table 4.1.12 Validity Estimates for Mathematics: Junior Division (French)

Item Code	Sequence	No. of Scores	% Exact Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
11660	7	86	98.8	89.5	9.3	2.3	7.0	1.2
13328	8	62	100.0	91.9	8.1	8.1	0.0	0.0
13362	9	55	100.0	92.7	7.3	1.8	5.5	0.0
13319	10	71	100.0	95.8	4.2	1.4	2.8	0.0
11511	26	76	100.0	92.1	7.9	6.6	1.3	0.0
11642	27	73	100.0	80.8	19.2	15.1	4.1	0.0
12574	28	83	94.0	72.3	21.7	16.9	4.8	6.0
13364	29	59	98.3	79.7	18.6	15.3	3.4	1.7
Aggregate		565	98.8	86.6	12.2	8.5	3.7	1.2

Validity: the Grade 9 Assessment of Mathematics (Academic and Applied)

Table 4.1.13 Validity Estimates for Grade 9 Applied Mathematics (English)

Administration	Item Code	Sequence	No. of Scores	% Exact Plus Adjnt.	% Exact	% Adjnt.	% Adjnt.-Low	% Adjnt.-High	% Non-Adjnt.
Winter	10252	NR	189	100.0	87.8	12.2	9.5	2.6	0.0
	10110	1NR	157	100.0	93.0	7.0	5.1	1.9	0.0
	EQME31847	28	190	100.0	88.9	11.1	6.8	4.2	0.0
	10212	NR	197	99.5	96.4	3.0	0.0	3.0	0.5
	10115	36	261	99.6	93.9	5.7	2.3	3.4	0.4
	10250	NR	271	99.6	91.1	8.5	6.3	2.2	0.4
	10164	NR	248	99.2	73.0	26.2	6.0	20.2	0.8
Aggregate			1513	99.7	88.8	10.8	5.1	5.8	0.3
Spring	10140	NR	157	99.4	94.3	5.1	3.8	1.3	0.6
	10251	26	173	100.0	95.4	4.6	1.7	2.9	0.0
	14847	NR	174	100.0	92.5	7.5	2.9	4.6	0.0
	10116	37	149	100.0	91.3	8.7	6.7	2.0	0.0
	10141	11	287	99.7	89.5	10.1	2.8	7.3	0.3
	10165	27	290	99.3	85.9	13.4	11.7	1.7	0.7
	10167	NR	311	100.0	91.6	8.4	6.1	2.3	0.0
Aggregate			1541	99.7	90.9	8.8	5.5	3.3	0.3
Aggregate Across Administrations			3054	99.7	89.9	9.8	5.3	4.5	0.3

Table 4.1.14 Validity Estimates for Grade 9 Academic Mathematics (English)

Administration	Item Code	Sequence	No. of Scores	% Exact Plus Adjnt.	% Exact	% Adjnt.	% Adjnt.-Low	% Adjnt.-High	% Non-Adjnt.
Winter	EQME31894	13	340	97.1	83.2	13.8	7.6	6.2	2.9
	10289	NR	277	100.0	96.8	3.2	2.9	0.4	0.0
	10297	NR	335	99.1	80.9	18.2	11.9	6.3	0.9
	10295	NR	331	98.8	85.2	13.6	7.9	5.7	1.2
	10186	NR	488	98.6	91.2	7.4	3.3	4.1	1.4
	10273	22	485	99.0	81.0	17.9	5.4	12.6	1.0
	10276	3NR	483	100.0	97.5	2.5	0.6	1.9	0.0
Aggregate			2739	98.9	88.1	10.8	5.3	5.5	1.1
Spring	10316	6	657	99.8	91.0	8.8	5.3	3.5	0.2
	10322	NR	635	99.4	87.2	12.1	8.2	3.9	0.6
	10231	23	654	99.8	93.9	6.0	2.3	3.7	0.2
	EQME31917	31	658	97.9	89.4	8.5	3.8	4.7	2.1
	10232	NR	823	99.6	96.5	3.2	1.8	1.3	0.4
	10321	14	760	98.9	88.6	10.4	8.7	1.7	1.1
	10291	NR	704	98.6	88.4	10.2	6.0	4.3	1.4
Aggregate			4891	99.2	90.8	8.3	5.1	3.2	0.8
Aggregate Across Administrations			7630	99.1	89.9	9.2	5.2	4.0	0.9

Table 4.1.15 Validity Estimates for Grade 9 Applied Mathematics (French)

Administration	Item Code	Sequence	No. of Scores	% Exact Plus Adjnt.	% Exact	% Adjnt.	% Adjnt.-Low	% Adjnt.-High	% Non-Adjnt.
Winter	CMS10037	NR	5	100.0	100.0	0.0	0.0	0.0	0.0
	OQMF31838	29	4	100.0	100.0	0.0	0.0	0.0	0.0
	CMS9759	NR	6	100.0	100.0	0.0	0.0	0.0	0.0
	OQMF31768	19	9	100.0	100.0	0.0	0.0	0.0	0.0
	OQMF31839	NR	12	100.0	100.0	0.0	0.0	0.0	0.0
	OQMF31764	8	13	100.0	92.3	7.7	7.7	0.0	0.0
	OQMF31791	NR	18	100.0	88.9	11.1	11.1	0.0	0.0
Aggregate			67	100.0	95.5	4.5	4.5	0.0	0.0
Spring	CMS9718	NR	52	100.0	94.2	5.8	0.0	5.8	0.0
	CMS9626	NR	64	100.0	98.4	1.6	0.0	1.6	0.0
	OQMF31796	18	61	100.0	83.6	16.4	14.8	1.6	0.0
	OQMF21395	NR	83	100.0	94.0	6.0	6.0	0.0	0.0
	OQMF31841	3NR	159	100.0	96.2	3.8	3.8	0.0	0.0
	CMS9995	9	170	100.0	97.6	2.4	2.4	0.0	0.0
	OQMF21378	NR	129	100.0	95.3	4.7	1.6	3.1	0.0
Aggregate			718	100.0	95.1	4.9	3.6	1.3	0.0
Aggregate Across Administrations			785	100.0	95.2	4.8	3.7	1.1	0.0

Table 4.1.16 Validity Estimates for Grade 9 Academic Mathematics (French)

Administration	Item Code	Sequence	No. of Scores	% Exact Plus Adjnt.	% Exact	% Adjnt.	% Adjnt.-Low	% Adjnt.-High	% Non-Adjnt.
Winter	OQMF10361	NR	65	98.5	98.5	0.0	0.0	0.0	1.5
	OQMF20904	7	56	100.0	100.0	0.0	0.0	0.0	0.0
	CMS9970	22	56	100.0	87.5	12.5	8.9	3.6	0.0
	CMS9674	15	73	100.0	98.6	1.4	1.4	0.0	0.0
	OQMF20917	NR	118	100.0	93.2	6.8	1.7	5.1	0.0
	13514	3NR	103	98.1	86.4	11.7	2.9	8.7	1.9
	OQMF31900	NR	106	100.0	98.1	1.9	0.0	1.9	0.0
Aggregate			577	99.5	94.3	5.2	1.9	3.3	0.5
Spring	CMS9968	NR	363	100.0	94.8	5.2	3.6	1.7	0.0
	CMS10066	NR	306	99.7	94.8	4.9	3.3	1.6	0.3
	CMS9970	22	235	100.0	92.3	7.7	6.0	1.7	0.0
	OQMF20911	8	267	98.5	97.0	1.5	1.5	0.0	1.5
	OQMF20902	NR	215	100.0	93.5	6.5	3.3	3.3	0.0
	OQMF31874	29	208	100.0	91.3	8.7	2.4	6.3	0.0
	CMS9969	16	226	98.7	87.2	11.5	9.3	2.2	1.3
Aggregate			1820	99.6	93.3	6.3	4.1	2.2	0.4
Aggregate Across Administrations			2397	99.5	93.5	6.0	3.5	2.5	0.5

Validity: the Ontario Secondary School Literacy Test

Table 4.1.17 Validity Estimates for Reading: OSSLT (English)

Item Code	Section	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
13056_23	I	6	7 869	98.7	84.0	14.7	8.7	6.0	1.3
11715_102	V	7	8 467	99.4	76.9	22.5	9.0	13.5	0.6
10333_30	IX	6	7 967	97.5	76.3	21.2	12.4	8.9	2.5
10334_30	IX	7	8 059	99.3	83.9	15.4	8.5	6.9	0.7
Aggregate			32 362	98.8	80.2	18.5	9.6	8.9	1.2

Table 4.1.18 Validity Estimates for Writing: OSSLT (English)

Item Code	Section	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
18340_T	IV	1	20 904	96.7	63.3	33.4	15.7	17.7	3.3
18340_V	IV	1	20 910	98.6	71.0	27.5	11.0	16.5	1.4
18339_T	VI	1	23 962	97.3	64.7	32.6	11.3	21.3	2.7
18339_V	VI	1	23 962	98.2	71.5	26.8	12.1	14.7	1.8
Aggregate Long Writing			89 738	97.7	67.6	30.0	12.4	17.6	2.3
13110_T&V	III	1	8 821	99.2	81.0	18.2	4.3	14.0	0.8
13173_T&V	X	1	7 777	99.8	87.5	12.3	5.2	7.1	0.2
Aggregate Short Writing			16 598	99.5	84.0	15.5	4.7	10.8	0.5
Aggregate All Items			106 336	98.0	70.2	27.8	11.2	16.5	2.0

Table 4.1.19 Validity Estimates for Reading: OSSLT (French)

Item Code	Section	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
10494_42	I	6	250	100.0	96.4	3.6	1.6	2.0	0.0
10460_40	V	7	187	99.5	92.0	7.5	4.8	2.7	0.5
10402_34	IX	6	193	100.0	92.2	7.8	4.1	3.6	0.0
10404_34	IX	7	191	100.0	83.8	16.2	6.3	9.9	0.0
Aggregate			821	99.9	91.5	8.4	4.0	4.4	0.1

Table 4.1.20 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_030T	IV	1	836	97.6	70.6	27.0	10.9	16.1	2.4
2É3_030V	IV	1	836	98.6	72.2	26.3	12.4	13.9	1.4
2É2_033T	VI	1	1074	97.0	65.5	31.6	16.9	14.6	3.0
2É2_033V	VI	1	1074	99.0	67.6	31.4	14.2	17.1	1.0
Aggregate Long Writing			3820	98.0	68.7	29.4	13.9	15.5	2.0
10498 T&V	III	1	351	97.7	77.8	19.9	8.8	11.1	2.3
10499 T&V	X	1	371	95.7	75.7	19.9	3.8	16.2	4.3
Aggregate Short Writing			722	96.7	76.7	19.9	6.2	13.7	3.3
Aggregate All Items			4542	97.8	69.9	27.9	12.7	15.2	2.2

Interrater Reliability: the Primary and Junior Assessments**Table 4.1.21 Interrater Reliability Estimates for Reading: Primary Division (English)**

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
10772	1	5	1100	94.5	56.6	37.8	5.5
10767	1	6	1183	96.4	68.9	27.5	3.6
10934	1	11	759	95.9	75.4	20.6	4.1
10929	1	12	691	97.1	63.1	34.0	2.9
10916	1	18	1126	97.3	66.9	30.5	2.7
10922	1	19	1036	98.1	67.3	30.8	1.9
12065	2	11	795	99.4	78.9	20.5	0.6
12066	2	12	697	96.3	78.8	17.5	3.7
10879	2	22	853	97.3	66.6	30.7	2.7
10877	2	23	858	98.1	76.6	21.6	1.9
Aggregate			9098	97.0	69.2	27.8	3.0

Table 4.1.22 Interrater Reliability Estimates for Reading: Junior Division (English)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
10625	1	5	909	96.7	68.8	27.9	3.3
10627	1	6	822	95.9	63.4	32.5	4.1
10616	1	11	1120	98.5	62.5	36.0	1.5
10620	1	12	1093	97.3	63.1	34.1	2.7
10518	1	18	1044	97.9	63.4	34.5	2.1
10523	1	19	884	98.3	70.5	27.8	1.7
12321	2	11	1151	97.3	62.6	34.7	2.7
12320	2	12	1230	97.7	68.0	29.7	2.3
10589	2	22	858	98.7	66.7	32.1	1.3
10587	2	23	835	99.0	63.8	35.2	1.0
Aggregate			9946	97.7	65.2	32.5	2.3

Table 4.1.23 Interrater Reliability Estimates for Reading: Primary Division (French)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
11063	1	5	207	100.0	92.3	7.7	0.0
11062	1	6	230	99.6	90.4	9.1	0.4
12239	1	11	181	100.0	98.9	1.1	0.0
12240	1	12	181	100.0	94.5	5.5	0.0
11193	1	18	147	98.6	77.6	21.1	1.4
11194	1	19	147	95.2	69.4	25.9	4.8
12203	2	11	104	98.1	88.5	9.6	1.9
12204	2	12	115	93.9	83.5	10.4	6.1
12251	2	22	113	99.1	93.8	5.3	0.9
12252	2	23	114	99.1	85.1	14.0	0.9
Aggregate			1539	98.6	88.1	10.5	1.4

Table 4.1.24 Interrater Reliability Estimates for Reading: Junior Division (French)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
11379	1	5	64	98.4	57.8	40.6	1.6
11380	1	6	91	93.4	53.8	39.6	6.6
11617	1	11	57	100.0	66.7	33.3	0.0
11619	1	12	83	97.6	55.4	42.2	2.4
11423	1	18	93	96.8	59.1	37.6	3.2
11424	1	19	93	93.5	53.8	39.8	6.5
11475	2	11	60	98.3	86.7	11.7	1.7
11476	2	12	60	100.0	91.7	8.3	0.0
11405	2	22	152	100.0	82.2	17.8	0.0
11407	2	23	139	98.6	82.0	16.5	1.4
Aggregate			892	97.6	69.6	28.0	2.4

Table 4.1.25 Interrater Reliability Estimates for Writing: Primary Division (English)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
17369 T	2	24	2014	98.0	68.5	29.6	2.0
17369 V	2	24	2010	98.0	76.3	21.8	1.9
Aggregate Long Writing			4051	98.0	72.3	25.7	2.0
12173 T	1	13	1014	95.8	74.2	21.6	4.2
12173 V	1	13	928	97.2	78.4	18.8	2.8
11050 T	2	17	967	97.0	66.2	30.8	3.0
11050 V	2	17	841	97.6	74.0	23.7	2.4
Aggregate Short Writing			3750	96.9	73.1	23.7	3.1
Aggregate All Items			7801	97.5	72.7	24.8	2.5

Table 4.1.26 Interrater Reliability Estimates for Writing: Junior Division (English)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
17733_T	2	24	2213	97.8	63.6	34.2	2.2
17733_V	2	24	2214	98.7	70.0	28.7	1.3
Aggregate Long Writing			4427	98.2	66.8	31.4	1.8
10703_T	1	13	1341	94.8	63.8	30.9	5.2
10703_V	1	13	1284	98.2	66.3	31.9	1.8
10708_T	2	17	1308	98.5	75.9	22.6	1.5
10708_V	2	17	1310	98.9	80.4	18.5	1.1
Aggregate Short Writing			5243	97.6	71.6	26.0	2.4
Aggregate All Items			9670	97.9	69.4	28.5	2.1

Table 4.1.27 Interrater Reliability Estimates for Writing: Primary Division (French)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
17714_T	2	24	69	98.7	71.0	26.1	2.9
17714_V	2	24	81	98.0	79.0	21.0	0.0
Aggregate Long Writing			150	98.7	75.3	23.3	1.3
11242_T	1	13	70	97.1	61.4	35.7	2.9
11242_V	1	13	84	100.0	81.1	19.0	0.0
11241_T	2	17	50	100.0	64.0	36.0	0.0
11241_V	2	17	79	100.0	73.4	26.6	0.0
Aggregate Short Writing			283	99.3	71.0	28.3	0.7
Aggregate All Items			433	99.1	72.5	26.6	0.9

Table 4.1.28 Interrater Reliability Estimates for Writing: Junior Division (French)

Item Code	Booklet	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
18331_T	2	24	132	96.2	62.9	33.3	3.8
18331_V	2	24	116	94.8	62.1	32.8	5.2
Aggregate Long Writing			248	95.6	62.5	33.1	4.4
11596_T	1	13	66	100.0	78.8	21.2	0.0
11596_V	1	13	115	99.1	70.4	28.7	0.9
11589_T	2	17	118	97.5	55.1	42.4	2.5
11589_V	2	17	134	97.8	62.7	35.1	2.2
Aggregate Short Writing			433	98.4	65.1	33.3	1.6
Aggregate All Items			681	97.4	64.2	33.2	2.6

Table 4.1.29 Interrater Reliability Estimates for Mathematics: Primary Division (English)

Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
13132	7	2 026	97.8	91.2	6.6	2.2
EQME10064	8	2 077	96.7	69.1	27.6	3.3
EQME10132	9	1 971	97.2	75.2	22.0	2.8
10845	10	2 075	97.3	88.1	9.1	2.7
10963	26	1 656	99.5	88.3	11.2	0.5
10791	27	1 597	96.4	85.6	10.8	3.6
10964	28	1 537	97.3	83.2	14.1	2.7
10848	29	1 720	98.4	83.4	14.9	1.6
Aggregate		14 659	97.5	82.8	14.7	2.5

Table 4.1.30 Interrater Reliability Estimates for Mathematics: Junior Division (English)

Item Code	Sequence	No. of Pairs	% Exact Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
11393	7	2 050	96.3	80.6	15.7	3.7
11348	8	1 917	98.4	86.1	12.3	1.6
11455	9	1 960	96.2	80.3	15.9	3.8
11454	10	1 961	96.8	82.0	14.7	3.2
11408	26	2 121	98.4	84.9	13.5	1.6
11452	27	1 997	97.1	87.0	10.1	2.9
11453	28	2 116	98.1	86.2	11.9	1.9
11404	29	2 035	99.1	83.8	15.3	0.9
Aggregate		16 157	97.6	83.9	13.7	2.4

Table 4.1.31 Interrater Reliability Estimates for Mathematics: Primary Division (French)

Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
13105	7	164	98.8	82.9	15.9	1.2
13095	8	163	96.9	93.3	3.7	3.1
13313	9	167	97.0	81.4	15.6	3.0
13265	10	167	97.6	87.4	10.2	2.4
OQMF31569	26	148	98.0	77.7	20.3	2.0
13219	27	141	87.2	62.4	24.8	12.8
13341	28	152	96.1	77.6	18.4	3.9
13326	29	147	95.2	71.4	23.8	4.8
Aggregate		1249	96.0	79.8	16.3	4.0

Table 4.1.32 Interrater Reliability Estimates for Mathematics: Junior Division (French)

Item Code	Sequence	No. of Pairs	% Exact Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
11660	7	102	99.0	96.1	2.9	1.0
13328	8	113	98.2	86.7	11.5	1.8
13362	9	116	96.6	79.3	17.2	3.4
13319	10	108	98.1	75.0	23.1	1.9
11511	26	113	95.6	87.6	8.0	4.4
11642	27	116	94.0	81.9	12.1	6.0
12574	28	109	95.4	73.4	22.0	4.6
13364	29	120	90.0	80.8	9.2	10.0
Aggregate		897	95.8	82.5	13.3	4.2

Interrater Reliability: the Grade 9 Assessment of Mathematics (Academic and Applied)

Table 4.1.33 Interrater Reliability Estimates for Grade 9 Applied Mathematics (English)

Administration	Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
Winter	10252	NR	259	96.5	79.5	17.0	3.5
	10110	10	267	97.8	73.8	24.0	2.2
	EQME31847	28	257	94.9	68.5	26.5	5.1
	10212	NR	251	98.4	87.6	10.8	1.6
	10115	36	237	98.7	92.0	6.8	1.3
	10250	NR	227	99.6	98.2	1.3	0.4
	10164	NR	242	97.5	88.4	9.1	2.5
Aggregate			1740	97.6	83.6	14.0	2.4
Spring	10140	NR	240	95.4	76.7	18.8	4.6
	10251	26	234	98.7	83.3	15.4	1.3
	14847	NR	234	99.1	84.6	14.5	0.9
	10116	37	249	99.2	84.7	14.5	0.8
	10141	11	253	98.0	89.7	8.3	2.0
	10165	27	252	96.8	76.2	20.6	3.2
	10167	NR	247	98.0	87.4	10.5	2.0
Aggregate			1709	97.9	83.3	14.6	2.1
Aggregate Across Administrations			3449	97.7	83.4	14.3	2.3

Table 4.1.34 Interrater Reliability Estimates for Grade 9 Academic Mathematics (English)

Administration	Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
Winter	EQME31894	13	521	95.6	71.0	24.6	4.4
	10289	NR	549	97.6	88.0	9.7	2.4
	10297	NR	521	97.5	79.8	17.7	2.5
	10295	NR	528	96.0	81.4	14.6	4.0
	10186	NR	566	95.9	87.1	8.8	4.1
	10273	22	557	97.5	63.9	33.6	2.5
	10276	30	561	99.1	93.0	6.1	0.9
Aggregate			3803	97.1	80.7	16.3	2.9
Spring	10316	6	934	97.0	82.4	14.6	3.0
	10322	NR	941	96.4	69.6	26.8	3.6
	10231	23	923	99.0	82.6	16.5	1.0
	EQME31917	31	928	98.3	89.0	9.3	1.7
	10232	NR	686	99.6	94.8	4.8	0.4
	10321	14	708	96.8	83.5	13.3	3.2
	10291	NR	728	96.7	76.8	19.9	3.3
Aggregate			5848	97.7	82.3	15.4	2.3
Aggregate Across Administrations			9651	97.4	81.5	15.8	2.6

Table 4.1.35 Interrater Reliability Estimates for Grade 9 Applied Mathematics (French)

Administration	Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
Winter	CMS10037	NR	13	100.0	100.0	0.0	0.0
	OQMF31838	29	15	100.0	86.7	13.3	0.0
	CMS9759	NR	12	100.0	91.7	8.3	0.0
	OQMF31768	19	8	100.0	100.0	0.0	0.0
	OQMF31839	NR	16	100.0	75.0	25.0	0.0
	OQMF31764	8	16	100.0	93.8	6.3	0.0
	OQMF31791	NR	14	100.0	78.6	21.4	0.0
Aggregate			94	100.0	88.3	11.7	0.0
Spring	CMS9718	NR	48	97.9	77.1	20.8	2.1
	CMS9626	NR	46	100.0	84.8	15.2	0.0
	OQMF31796	18	48	97.9	83.3	14.6	2.1
	OQMF21395	NR	43	97.7	95.3	2.3	2.3
	OQMF31841	30	35	100.0	91.4	8.6	0.0
	CMS9995	9	34	100.0	97.1	2.9	0.0
	OQMF21378	NR	36	100.0	97.2	2.8	0.0
Aggregate			290	99.0	88.6	10.3	1.0
Aggregate Across Administrations			384	99.5	88.5	11.0	0.5

Table 4.1.36 Interrater Reliability Estimates for Grade 9 Academic Mathematics (French)

Administration	Item Code	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
Winter	OQMF10361	NR	50	96.0	82.0	14.0	4.0
	OQMF20904	7	49	100.0	83.7	16.3	0.0
	CMS9970	22	49	95.9	79.6	16.3	4.1
	CMS9674	15	45	93.3	75.6	17.8	6.7
	OQMF20917	NR	51	100.0	86.3	13.7	0.0
	13514	30	52	88.5	80.8	7.7	11.5
	OQMF31900	NR	52	100.0	88.5	11.5	0.0
Aggregate			348	96.3	82.5	13.8	3.7
Spring	CMS9968	NR	101	99.0	95.0	4.0	1.0
	CMS10066	NR	111	95.5	87.4	8.1	4.5
	CMS9970	22	117	99.1	72.6	26.5	0.9
	OQMF20911	8	114	94.7	78.9	15.8	5.3
	OQMF20902	NR	77	98.7	81.8	16.9	1.3
	OQMF31874	29	86	96.5	79.1	17.4	3.5
	CMS9969	16	76	98.7	86.8	11.8	1.3
Aggregate			682	97.4	82.8	14.5	2.6
Aggregate Across Administrations			1030	96.8	82.7	14.2	3.2

Interrater Reliability: the Ontario Secondary School Literacy Test (OSSLT)

Table 4.1.37 Interrater Reliability Estimates for Reading: OSSLT (English)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
13056 23	I	6	187 202	97.6	67.1	30.5	2.4
11715 102	V	7	187 202	97.0	64.0	33.0	3.0
10333 30	IX	6	187 202	97.3	62.2	35.0	2.7
10334 30	IX	7	187 202	98.2	64.6	33.5	1.8
Aggregate			748 808	97.5	64.5	33.0	2.5

Table 4.1.38 Interrater Reliability Estimates for Writing: OSSLT (English)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
18340 T	IV	1	187 201	90.7	45.6	45.1	9.3
18340 V	IV	1	187 201	97.5	61.2	36.3	2.5
18339 T	VI	1	187 202	92.8	48.1	44.7	7.2
18339 V	VI	1	187 202	97.6	57.5	40.1	2.4
Aggregate Long Writing			748 806	94.7	53.1	41.5	5.3
13110 T&V	III	1	187 202	97.8	66.7	31.2	2.2
13173 T&V	X	1	187 203	98.0	67.5	30.6	2.0
Aggregate Short Writing			374 405	97.9	67.1	30.9	2.1
Aggregate All Items			1 123 211	95.7	57.8	38.0	4.3

Table 4.1.39 Interrater Reliability Estimates for Reading: OSSLT (French)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
10494 42	I	6	6 957	96.1	74.8	21.4	3.9
10460 40	V	7	6 957	95.2	75.1	20.0	4.8
10402 34	IX	6	6 954	96.1	76.9	19.2	3.9
10404 34	IX	7	6 954	96.3	60.5	35.8	3.7
Aggregate			27 822	95.9	71.8	24.1	4.1

Table 4.1.40 Interrater Reliability Estimates for Writing: OSSLT (French)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
2E3 030T	IV	1	6 957	82.3	38.5	43.9	17.7
2E3 030V	IV	1	6 957	96.4	60.6	35.8	3.6
2E2 033T	VI	1	6 954	87.1	43.8	43.3	12.9
2E2 033V	VI	1	6 954	95.5	55.2	40.3	4.5
Aggregate Long Writing			27 822	90.3	49.5	40.8	9.7
10498 T&V	III	1	6 957	95.6	63.8	31.8	4.4
10499 T&V	X	1	6 954	95.9	70.6	25.3	4.1
Aggregate Short Writing			13 911	95.8	67.2	28.6	4.2
Aggregate All Items			41 733	92.1	55.4	36.7	7.9

APPENDIX 7.1: SCORE DISTRIBUTIONS AND ITEM STATISTICS FOR THE 2011 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 and students who are second language learners. For the French-language versions of the Grade 9 Assessment of Mathematics and the OSSLT, DIF analysis for SLLs was not conducted due to the small number of students in the French-language SLL population.

Classical item statistics and IRT item parameter estimates are combined into tables for each assessment: Tables 7.1.1–7.1.24 for the primary- and junior assessments, Tables 7.1.49–7.1.64 for the Grade 9 Assessment of Mathematics and Tables 7.1.7 –7.1.82 for the OSSLT. The distribution of score points and the step parameter estimates are also provided for open-response items.

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

The 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
10769	1	1	R2.0	I	1	0.88	0.40	-1.74	0.93
10773	1	2	R1.0	C	4	0.89	0.42	-1.81	1.00
10771	1	3	R3.0	C	2	0.90	0.38	-2.11	0.88
10770	1	4	R3.0	I	2	0.71	0.12	-2.74	0.13
10772	1	5	R1.0	C	4*	0.50 (2.01)	0.39	-0.59	0.33
10767	1	6	R1.0	C	4*	0.54 (2.15)	0.53	-0.54	0.50
10931	1	7	R3.0	E	1	0.86	0.21	-2.63	0.38
10932	1	8	R3.0	I	4	0.75	0.30	-1.25	0.50
10925	1	9	R1.0	I	1	0.58	0.20	0.04	0.32
10935	1	10	R2.0	C	3	0.56	0.38	0.09	0.83
10934	1	11	R1.0	C	4*	0.42 (1.69)	0.43	-0.07	0.47
10929	1	12	R2.0	I	4*	0.47 (1.87)	0.52	-0.02	0.56
10919	1	14	R3.0	E	1	0.56	0.26	0.15	0.43
10914	1	15	R1.0	I	3	0.58	0.35	0.01	0.67
10924	1	16	R2.0	C	1	0.86	0.46	-1.52	1.08
10920	1	17	R3.0	C	2	0.59	0.33	-0.04	0.60
10916	1	18	R1.0	E	4*	0.60 (2.39)	0.50	-1.45	0.35
10922	1	19	R1.0	C	4*	0.43 (1.70)	0.45	0.20	0.46
12053	2	1	R2.0	I	3	0.86	0.41	-1.67	0.84
12056	2	2	R3.0	I	4	0.77	0.44	-0.94	0.95
12057	2	3	R1.0	I	3	0.83	0.40	-1.35	0.86
12058	2	4	R1.0	C	3	0.77	0.48	-0.92	1.05
12059	2	5	R3.0	I	3	0.62	0.43	-0.24	0.88
12060	2	6	R1.0	I	1	0.53	0.29	0.34	0.59
12061	2	7	R2.0	C	4	0.53	0.25	0.45	0.44
12062	2	8	R3.0	C	2	0.81	0.41	-1.37	0.74
12063	2	9	R1.0	E	4	0.75	0.34	-1.02	0.61
12064	2	10	R2.0	C	3	0.36	0.24	1.31	0.70
12065	2	11	R1.0	C	4*	0.54 (2.17)	0.37	-0.63	0.38
12066	2	12	R2.0	I	4*	0.51 (2.05)	0.44	-0.33	0.38
10876	2	18	R1.0	I	3	0.76	0.37	-1.07	0.70
10878	2	19	R2.0	I	2	0.79	0.36	-1.26	0.69
10880	2	20	R3.0	I	3	0.61	0.27	-0.23	0.41
10881	2	21	R1.0	I	2	0.51	0.35	0.34	0.72
10879	2	22	R2.0	C	4*	0.45 (1.78)	0.48	-0.01	0.52
10877	2	23	R1.0	C	4*	0.49 (1.94)	0.45	-0.16	0.46

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. C = connections; E = explicit; I = implicit; R = reading; () = Mean score for OR items.

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

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
10772	1	5	% of Students	0.70	1.23	27.79	41.56	24.34	4.37

			Parameters			-6.13	-1.06	1.13	3.70
10767	1	6	% of Students	0.94	1.25	22.79	37.90	31.74	5.40
			Parameters			-4.17	-1.17	0.38	2.81
10934	1	11	% of Students	0.99	0.70	44.04	39.66	12.34	2.28
			Parameters			-5.21	-0.07	1.93	3.07
10929	1	12	% of Students	2.03	1.10	29.10	47.68	17.57	2.52
			Parameters			-3.60	-0.89	1.43	2.99
10916	1	18	% of Students	0.63	1.19	30.87	15.30	30.48	21.54
			Parameters			-6.26	0.71	-1.20	0.97
10922	1	19	% of Students	1.42	1.45	40.26	42.68	12.55	1.65
			Parameters			-4.53	-0.24	2.00	3.56
12065	2	11	% of Students	0.63	0.76	9.66	63.18	22.20	3.56
			Parameters			-4.46	-3.40	1.82	3.51
12066	2	12	% of Students	1.42	1.32	26.92	36.97	29.11	4.26
			Parameters			-4.77	-0.85	0.61	3.69
10879	2	22	% of Students	0.83	1.56	34.44	47.75	13.52	1.90
			Parameters			-4.40	-0.63	1.82	3.18
10877	2	23	% of Students	1.17	1.37	24.20	52.74	17.88	2.63
			Parameters			-4.27	-1.37	1.70	3.30

Note. The total number of students is 115 950.

Table 7.1.3 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
10623	1	1	R1.0	E	3	0.98	0.19	-4.10	0.62
10626	1	2	R2.0	C	2	0.71	0.43	-0.75	0.83
10631	1	3	R3.0	I	3	0.72	0.22	-1.51	0.30
10624	1	4	R1.0	I	2	0.88	0.32	-2.06	0.65
10625	1	5	R1.0	C	4*	0.57 (2.27)	0.55	-0.97	0.60
10627	1	6	R2.0	I	4*	0.56 (2.25)	0.56	-0.78	0.56
10614	1	7	R1.0	E	3	0.96	0.21	-3.58	0.64
10618	1	8	R2.0	C	3	0.61	0.39	-0.22	0.68
10619	1	9	R3.0	C	4	0.44	0.12	2.37	0.20
10615	1	10	R1.0	I	1	0.82	0.34	-1.62	0.62
10616	1	11	R1.0	C	4*	0.50 (1.98)	0.53	-0.53	0.53
10620	1	12	R1.0	C	4*	0.44 (1.76)	0.51	-0.14	0.56
10516	1	14	R1.0	C	2	0.90	0.34	-2.30	0.69
10515	1	15	R1.0	I	4	0.61	0.34	-0.17	0.58
10519	1	16	R2.0	I	3	0.85	0.43	-1.64	0.89
10522	1	17	R3.0	I	2	0.65	0.42	-0.40	0.81
10518	1	18	R1.0	C	4*	0.58 (2.32)	0.52	-0.73	0.58
10523	1	19	R1.0	C	4*	0.48 (1.90)	0.57	-0.05	0.62
12313	2	1	R2.0	I	3	0.51	0.35	0.35	0.69
12311	2	2	R1.0	I	3	0.79	0.30	-1.77	0.45
12319	2	3	R1.0	I	1	0.76	0.24	-1.48	0.40
12314	2	4	R1.0	E	3	0.83	0.41	-1.39	0.89
12315	2	5	R3.0	E	4	0.68	0.35	-0.63	0.57
12316	2	6	R1.0	I	1	0.92	0.38	-2.04	1.03
12310	2	7	R1.0	I	4	0.48	0.10	2.02	0.18
12317	2	8	R1.0	E	2	0.83	0.36	-1.59	0.69
12318	2	9	R2.0	I	2	0.76	0.44	-0.93	0.91
12312	2	10	R3.0	C	1	0.75	0.33	-1.23	0.53
12321	2	11	R1.0	I	4*	0.53 (2.12)	0.51	-0.82	0.59
12320	2	12	R2.0	I	4*	0.44 (1.76)	0.52	-0.02	0.55
10585	2	18	R1.0	E	3	0.75	0.30	-1.32	0.46
10591	2	19	R1.0	I	1	0.58	0.30	-0.05	0.48
10588	2	20	R2.0	I	3	0.77	0.32	-1.55	0.49
10590	2	21	R3.0	C	3	0.65	0.37	-0.44	0.64
10589	2	22	R2.0	I	4*	0.54 (2.14)	0.52	-0.45	0.60
10587	2	23	R1.0	C	4*	0.55 (2.20)	0.51	-0.58	0.59

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. C = connections; E = explicit; I = implicit; R = reading. () = Mean score for OR items.

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

Table 7.1.4 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
10625	1	5	% of Students	0.32	0.33	13.18	51.05	28.98	6.14
			Parameters			-4.58	-2.17	0.60	2.29
10627	1	6	% of Students	0.96	0.67	21.00	36.36	32.72	8.29
			Parameters			-4.18	-1.22	0.15	2.12
10616	1	11	% of Students	0.80	0.39	29.76	42.94	22.09	4.02
			Parameters			-4.89	-0.92	0.96	2.74
10620	1	12	% of Students	1.19	0.49	39.49	42.27	14.03	2.53
			Parameters			-4.59	-0.40	1.56	2.85
10518	1	18	% of Students	0.37	1.08	10.23	48.36	35.17	4.79
			Parameters			-3.64	-2.38	0.35	2.74
10523	1	19	% of Students	1.39	2.33	28.58	44.22	20.48	3.00
			Parameters			-3.20	-0.86	1.03	2.82
12321	2	11	% of Students	0.28	0.11	19.17	52.19	24.98	3.28
			Parameters			-5.49	-1.60	0.91	2.91
12320	2	12	% of Students	2.12	0.57	40.55	37.30	16.50	2.96
			Parameters			-3.93	-0.22	1.25	2.84
10589	2	22	% of Students	1.17	0.69	15.37	53.83	25.17	3.77
			Parameters			-3.58	-1.86	0.91	2.74
10587	2	23	% of Students	1.10	0.35	12.68	54.48	27.34	4.04
			Parameters			-3.72	-2.12	0.79	2.72

Note. The total number of students is 128 687.

Table 7.1.5 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
11055	1	1	B	I	3	0.78	0.24	-1.83	0.38
11060	1	2	C	L	2	0.83	0.34	-1.61	0.65
11052	1	3	A	L	2	0.86	0.33	-1.85	0.66
11056	1	4	C	I	1	0.63	0.37	-0.30	0.68
11063	1	5	A	L	4*	0.56 (2.25)	0.34	-1.56	0.32
11062	1	6	A	L	4*	0.56 (2.24)	0.38	-1.07	0.27
12235	1	7	C	E	2	0.92	0.20	-3.62	0.40
12238	1	8	C	L	3	0.79	0.27	-1.46	0.51
12236	1	9	A	I	2	0.56	0.34	0.16	0.68
12237	1	10	B	L	4	0.69	0.42	-0.54	0.86
12239	1	11	A	I	4*	0.57 (2.28)	0.49	-0.85	0.49
12240	1	12	A	L	4*	0.59 (2.34)	0.50	-0.92	0.50
11189	1	14	B	L	2	0.74	0.40	-0.87	0.80
11191	1	15	C	I	1	0.37	0.36	0.98	0.94
11187	1	16	A	I	1	0.46	0.24	0.92	0.54
11190	1	17	C	E	2	0.71	0.38	-0.69	0.74
11193	1	18	A	I	4*	0.51 (2.05)	0.38	-1.21	0.33
11194	1	19	A	L	4*	0.59 (2.34)	0.45	-0.93	0.41
12193	2	1	A	E	2	0.87	0.37	-1.64	0.86
12194	2	2	A	I	4	0.75	0.31	-1.11	0.55
12195	2	3	B	L	1	0.59	0.34	-0.05	0.64
12202	2	4	C	L	4	0.42	0.30	0.96	0.70
12196	2	5	C	I	2	0.78	0.34	-1.27	0.62
12197	2	6	C	I	4	0.42	0.30	0.89	0.81
12198	2	7	C	L	3	0.70	0.44	-0.53	0.95
12199	2	8	C	I	1	0.61	0.21	-0.18	0.31
12201	2	9	B	I	4	0.76	0.35	-1.04	0.65
12200	2	10	C	L	2	0.75	0.33	-0.94	0.65
12203	2	11	A	I	4*	0.52 (2.09)	0.49	-0.91	0.47
12204	2	12	A	L	4*	0.65 (2.61)	0.49	-1.50	0.44
12247	2	18	B	I	4	0.69	0.47	-0.48	1.04
12248	2	19	A	I	3	0.78	0.37	-1.11	0.73
12249	2	20	C	I	4	0.59	0.31	0.03	0.58
12250	2	21	A	I	1	0.65	0.35	-0.39	0.62
12251	2	22	A	I	4*	0.52 (2.09)	0.46	-0.81	0.50
12252	2	23	B	L	4*	0.49 (1.96)	0.39	-0.46	0.40

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. A = comprehension; B = organization; C = vocabulary and linguistic conventions; L = connections; E = explicit; I = implicit; () = Mean score for OR items.

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

Table 7.1.6 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
11063	1	5	% of Students	0.34	0.12	7.62	62.66	24.47	4.78
			Parameters			-7.12	-4.43	1.75	3.56
11062	1	6	% of Students	1.50	0.75	22.74	34.00	30.95	10.05
			Parameters			-6.16	-1.30	0.31	2.90
12239	1	11	% of Students	0.56	0.46	13.31	48.99	29.84	6.84
			Parameters			-4.38	-2.15	0.68	2.47
12240	1	12	% of Students	0.80	0.34	12.12	47.74	29.92	9.07
			Parameters			-4.09	-2.27	0.60	2.08
11193	1	18	% of Students	0.43	0.19	26.40	45.57	22.64	4.78
			Parameters			-8.40	-1.29	1.42	3.45
11194	1	19	% of Students	1.24	0.62	9.43	51.07	27.92	9.71
			Parameters			-3.72	-3.03	0.92	2.12
12203	2	11	% of Students	0.41	0.28	25.12	46.65	19.78	7.76
			Parameters			-5.64	-1.25	1.29	1.96
12204	2	12	% of Students	0.64	0.31	6.59	42.29	30.53	19.64
			Parameters			-3.98	-3.42	0.33	1.05
12251	2	22	% of Students	0.44	0.22	19.33	54.80	20.61	4.59
			Parameters			-5.53	-1.73	1.40	2.64
12252	2	23	% of Students	1.10	0.15	21.24	60.64	13.81	3.06
			Parameters			-5.42	-1.93	2.46	3.06

Note. The total number of students is 6766.

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
11373	1	1	B	L	1	0.59	0.13	0.11	0.20
11376	1	2	C	I	3	0.51	0.27	0.55	0.48
11375	1	3	C	I	4	0.68	0.40	-0.57	0.71
11371	1	4	A	I	4	0.84	0.29	-1.94	0.52
11379	1	5	A	I	4*	0.62 (2.47)	0.42	-2.09	0.44
11380	1	6	A	L	4*	0.67 (2.66)	0.43	-2.36	0.39
11610	1	7	A	E	4	0.62	0.35	-0.27	0.60
11613	1	8	B	L	1	0.59	0.30	-0.10	0.47
11611	1	9	A	I	4	0.43	0.40	0.71	0.96
11615	1	10	C	L	2	0.63	0.28	-0.47	0.41
11617	1	11	A	L	4*	0.57 (2.26)	0.51	-1.11	0.55
11619	1	12	A	L	4*	0.55 (2.18)	0.57	-0.73	0.61
11415	1	14	A	E	4	0.64	0.40	-0.33	0.74
11420	1	15	B	I	3	0.68	0.30	-0.63	0.50
11422	1	16	C	I	1	0.79	0.24	-1.80	0.39
11417	1	17	A	I	2	0.61	0.42	-0.21	0.81
11423	1	18	A	I	4*	0.53 (2.10)	0.52	-0.93	0.56
11424	1	19	A	L	4*	0.64 (2.56)	0.54	-1.56	0.53
11474	2	1	C	L	2	0.71	0.30	-0.87	0.51
11462	2	2	A	I	2	0.86	0.41	-1.42	1.04
11469	2	3	B	I	1	0.84	0.30	-1.99	0.52
11467	2	4	A	I	4	0.39	0.34	1.02	0.75
11470	2	5	C	I	2	0.56	0.29	0.16	0.52
11443	2	6	A	E	3	0.50	0.33	0.50	0.58
11461	2	7	A	E	4	0.93	0.30	-2.23	0.89
11466	2	8	A	I	4	0.59	0.14	-0.02	0.19
11472	2	9	C	I	2	0.78	0.34	-1.31	0.60
11463	2	10	A	I	4	0.76	0.44	-0.94	0.93
11475	2	11	A	I	4*	0.64 (2.56)	0.54	-1.39	0.61
11476	2	12	A	L	4*	0.59 (2.36)	0.47	-1.21	0.54
11397	2	18	A	I	1	0.76	0.33	-1.11	0.61
11403	2	19	C	L	1	0.59	0.15	0.01	0.24
11398	2	20	A	L	3	0.61	0.41	-0.16	0.82
11399	2	21	B	I	4	0.76	0.38	-1.01	0.71
11405	2	22	A	I	4*	0.60 (2.41)	0.52	-0.98	0.52
11407	2	23	B	L	4*	0.64 (2.54)	0.46	-1.51	0.40

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. A = comprehension; B = organization; C = vocabulary and linguistic conventions; L = connections; E = explicit; I = implicit; () = Mean score for OR items.

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

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
11379	1	5	% of Students	0.05	0.03	5.29	49.01	39.30	6.33
			Parameters			-7.74	-3.83	0.22	2.98
11380	1	6	% of Students	0.09	0.03	3.00	45.11	34.39	17.38
			Parameters			-5.87	-5.13	0.18	1.39
11617	1	11	% of Students	0.22	0.23	13.03	54.84	23.39	8.29
			Parameters			-5.04	-2.26	1.00	1.87
11619	1	12	% of Students	0.99	0.26	23.55	40.00	26.30	8.89
			Parameters			-4.21	-1.09	0.54	1.83
11423	1	18	% of Students	0.31	0.08	23.44	47.82	22.54	5.81
			Parameters			-5.75	-1.22	0.99	2.25
11424	1	19	% of Students	0.34	0.02	12.62	33.16	38.77	15.09
			Parameters			-5.62	-1.86	-0.31	1.54
11475	2	11	% of Students	0.30	0.06	7.10	40.34	40.17	12.03
			Parameters			-4.56	-2.62	-0.13	1.76
11476	2	12	% of Students	0.39	0.02	6.62	56.65	29.57	6.76
			Parameters			-4.69	-3.19	0.74	2.32
11405	2	22	% of Students	1.17	0.25	10.10	45.64	31.26	11.58
			Parameters			-3.62	-2.45	0.41	1.76
11407	2	23	% of Students	0.85	0.11	7.26	43.76	32.43	15.59
			Parameters			-4.40	-3.46	0.30	1.53

Note. The total number of students is 6442.

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
12173_T	1	13	W2.0	T	4*	0.50 (2.01)	0.47	-0.02	0.58
12173_V	1	13	W3.0	V	3*	0.68 (2.05)	0.55	-1.28	0.80
11041	2	13	W1.0	T	3	0.68	0.21	0.85	0.33
11049	2	14	W2.0	T	4	0.81	0.30	-1.71	0.55
11047	2	15	W3.0	V	2	0.47	0.27	0.69	0.50
11039	2	16	W1.0	T	2	0.71	0.35	-0.90	0.57
11050_T	2	17	W2.0	T	4*	0.48 (1.90)	0.52	0.04	0.61
11050_V	2	17	W3.0	V	3*	0.70 (2.10)	0.58	-1.45	0.87
17369_T	2	24	W2.0	T	4*	0.60 (2.41)	0.54	-0.63	0.70
17369_V	2	24	W3.0	V	3*	0.71 (2.13)	0.59	-1.38	1.04
11031	2	25	W3.0	V	3	0.79	0.37	-1.28	0.69
11043	2	26	W3.0	V	2	0.57	0.35	0.02	0.66
11024	2	27	W2.0	T	2	0.66	0.33	-0.54	0.56
11030	2	28	W1.0	T	1	0.66	0.31	-0.54	0.51

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. T = content; V = conventions; W = writing; () = Mean score for OR items.

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

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
12173_T	1	13	% of Students	1.10	2.09	16.20	59.27	18.90	2.43
			Parameters			-2.80	-1.76	1.43	3.06
12173_V	1	13	% of Students	1.10	0.22	27.60	36.01	35.07	
			Parameters			-3.53	-0.67	0.35	
17369_T	2	24	% of Students	0.65	1.86	9.17	42.11	37.54	8.67
			Parameters			-2.48	-2.10	0.07	1.98
17369_V	2	24	% of Students	0.65	0.20	20.24	43.82	35.09	
			Parameters			-3.47	-1.10	0.42	
11050_T	2	17	% of Students	0.48	2.52	28.02	47.26	19.74	1.98
			Parameters			-3.35	-0.94	1.19	3.26
11050_V	2	17	% of Students	0.48	0.36	26.79	34.28	38.09	
			Parameters			-3.83	-0.73	0.22	

Note. The total number of students is 116 331.

Table 7.1.11 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
10703_T	1	13	W2.0	T	4*	0.58 (2.33)	0.56	-0.61	0.64
10703_V	1	13	W3.0	V	3*	0.68 (2.05)	0.58	-1.44	0.82
10707	2	13	W2.0	T	2	0.70	0.31	-0.93	0.47
10716	2	14	W3.0	V	3	0.90	0.34	-2.45	0.71
10699	2	15	W1.0	T	4	0.79	0.39	-1.45	0.66
10694	2	16	W3.0	V	3	0.61	0.31	-0.28	0.50
10708_T	2	17	W2.0	T	4*	0.55 (2.21)	0.58	-0.50	0.82
10708_V	2	17	W3.0	V	3*	0.67 (2.00)	0.61	-1.19	1.15
17733_T	2	24	W2.0	T	4*	0.54 (2.17)	0.61	-0.52	1.04
17733_V	2	24	W3.0	V	3*	0.63 (1.90)	0.60	-1.11	1.13
10721	2	25	W2.0	T	2	0.82	0.28	-2.33	0.41
10714	2	26	W1.0	T	1	0.68	0.35	-0.75	0.54
10705	2	27	W3.0	V	3	0.78	0.32	-1.59	0.51
10723	2	28	W2.0	T	2	0.40	0.24	1.28	0.52

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. T = content; V = conventions; W = writing; () = Mean score for OR items.

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

Table 7.1.12 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
10703_T	1	13	% of Students	0.81	1.65	10.59	45.19	35.34	6.43
			Parameters			-2.79	-2.19	0.21	2.33
10703_V	1	13	% of Students	0.81	0.31	20.80	50.52	27.55	
			Parameters			-3.74	-1.31	0.72	
17733_T	2	24	% of Students	0.68	0.43	14.82	54.05	25.61	4.41
			Parameters			-3.24	-1.57	0.62	2.10
17733_V	2	24	% of Students	0.68	0.17	26.05	55.60	17.50	
			Parameters			-3.47	-0.97	1.11	
10708_T	2	17	% of Students	0.62	1.00	15.56	47.74	30.69	4.40
			Parameters			-3.22	-1.57	0.44	2.36
10708_V	2	17	% of Students	0.62	0.27	17.07	63.41	18.62	
			Parameters			-3.25	-1.40	1.08	

Note. The total number of students is 128 827.

Table 7.1.13 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
11242_T	1	13	A	T	4*	0.44 (1.74)	0.54	0.30	0.73
11242_V	1	13	C	V	3*	0.60 (1.81)	0.58	-0.96	1.05
13015	2	13	A	T	3	0.67	0.34	-0.48	0.63
11255	2	14	B	T	1	0.83	0.27	-1.96	0.49
11280	2	15	C	V	2	0.79	0.35	-1.24	0.73
11275	2	16	A	T	1	0.83	0.35	-1.48	0.73
11241_T	2	17	A	T	4*	0.44 (1.77)	0.58	0.25	0.76
11241_V	2	17	C	V	3*	0.61 (1.84)	0.53	-1.09	0.90
17714_T	2	24	A	T	4*	0.52 (2.09)	0.58	-0.54	0.77
17714_V	2	24	C	V	3*	0.62 (1.86)	0.57	-0.96	1.14
11279	2	25	B	T	3	0.81	0.36	-1.35	0.70
13016	2	26	A	T	2	0.70	0.33	-0.75	0.56
11258	2	27	C	V	1	0.81	0.22	-1.94	0.41
11282	2	28	C	V	2	0.83	0.38	-1.39	0.80

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. A = comprehension; B = organization; C = vocabulary and linguistic conventions; T = content; V = conventions; () = Mean score for OR items.

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

Table 7.1.14 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
11242_T	1	13	% of Students	0.40	2.63	37.97	42.85	14.68	1.46
			Parameters			-3.04	-0.36	1.41	3.18
11242_V	1	13	% of Students	0.40	0.12	33.76	49.89	15.84	
			Parameters			-3.66	-0.54	1.33	
17714_T	2	24	% of Students	0.49	0.44	24.62	44.95	23.21	6.28
			Parameters			-3.94	-0.99	0.78	2.00
17714_V	2	24	% of Students	0.49	0.16	26.45	59.59	13.31	
			Parameters			-3.55	-0.83	1.50	
11241_T	2	17	% of Students	0.40	3.39	35.15	43.63	15.33	2.10
			Parameters			-2.77	-0.43	1.34	2.86
11241_V	2	17	% of Students	0.40	0.21	30.84	52.80	15.75	
			Parameters			-4.04	-0.69	1.46	

Note. The total number of students is 6773.

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
11596_T	1	13	A	T	4*	0.62 (2.46)	0.59	-0.93	0.90
11596_V	1	13	C	V	3*	0.71 (2.13)	0.62	-1.66	1.21
11553	2	13	A	T	1	0.89	0.30	-2.21	0.66
11572	2	14	C	V	2	0.30	0.18	-2.12	0.60
12538	2	15	A	T	3	0.53	0.24	-0.45	0.39
11554	2	16	B	T	2	0.66	0.23	-0.67	0.45
11589_T	2	17	A	T	4*	0.56 (2.24)	0.58	-0.73	0.92
11589_V	2	17	C	V	3*	0.72 (2.16)	0.58	-1.60	1.08
18331_T	2	24	A	T	4*	0.68 (2.71)	0.56	-1.26	0.71
18331_V	2	24	C	V	3*	0.72 (2.17)	0.58	-1.76	0.92
11561	2	25	C	V	1	0.65	0.35	-0.41	0.63
11569	2	26	A	T	2	0.70	0.29	-0.94	0.45
11557	2	27	B	T	1	0.89	0.27	-2.34	0.59
13030	2	28	C	V	3	0.74	0.31	-1.08	0.53

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. A = comprehension; B = organization; C = vocabulary and linguistic conventions; T = content; V = conventions; () = Mean score for OR items.

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

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
11596_T	1	13	% of Students	0.11	0.58	10.59	40.30	39.01	9.42
			Parameters			-3.54	-1.85	-0.60	1.73
11596_V	1	13	% of Students	0.11	0.08	16.39	53.78	29.65	
			Parameters			-4.22	-1.37	0.26	
18331_T	2	24	% of Students	0.33	0.54	4.26	33.10	46.21	15.56
			Parameters			-2.97	-2.96	-0.23	1.43
18331_V	2	24	% of Students	0.33	0.05	13.25	55.64	30.74	
			Parameters			-4.18	-1.73	0.14	
11589_T	2	17	% of Students	0.23	0.23	13.14	53.75	26.99	5.66
			Parameters			-3.86	-1.68	0.00	2.02
11589_V	2	17	% of Students	0.23	0.08	12.72	57.90	29.07	
			Parameters			-3.81	-1.67	0.34	

Note. The total number of students is 6440.

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
EQME31605	1	PV1	AP	P	4	69.67	0.41	-0.63	0.84
10691	2	NV3	AP	N	4	51.70	0.38	0.29	0.99
EQME10049	3	NV3	PS	N	3	64.20	0.48	-0.26	1.27
10814	4	NV3	AP	N	3	66.60	0.41	-0.40	0.91
10689	5	NV2	PS	N	4	72.14	0.43	-0.66	1.00
13153	6	NV1	AP	N	3	73.26	0.41	-0.77	0.87
13132	7	NV2	AP	N	4*	81.25 (3.25)	0.56	-2.30	0.43
EQME10064	8	GV2	PS	G	4*	44.50 (1.78)	0.35	-0.20	0.27
EQME10132	9	GV3	AP	G	4*	68.25 (2.73)	0.44	-2.52	0.27
10845	10	NV3	PS	N	4*	57.50 (2.30)	0.52	-1.35	0.34
EQME10124	11	MV1	KU	M	3	56.15	0.38	0.08	0.83
13164	12	PV2	KU	P	2	58.80	0.28	0.05	0.54
10957	13	PV1	KU	P	1	73.34	0.36	-0.83	0.75
EQME31590	14	MV1	AP	M	2	74.37	0.44	-0.75	1.01
10729	15	DV2	KU	D	3	56.24	0.33	0.11	0.58
EQME20830	16	GV3	AP	G	2	64.50	0.29	-0.35	0.54
10782	17	PV1	PS	P	3	53.25	0.31	0.31	0.71
EQME31603	18	MV1	AP	M	4	82.02	0.37	-1.40	0.75
EQME10028	19	PV1	KU	P	2	89.04	0.38	-1.95	0.88
EQME10003	20	DV3	AP	D	2	51.18	0.38	0.31	0.97
EQME10016	21	MV2	KU	M	3	73.72	0.50	-0.72	1.23
EQME20875	22	NV3	KU	N	3	75.74	0.42	-0.93	0.84
EQME31501	23	MV2	AP	M	2	71.20	0.46	-0.62	1.05
EQME20831	24	MV2	KU	M	1	64.08	0.40	-0.34	0.74
EQME10100	25	GV2	AP	G	2	82.40	0.35	-1.45	0.71
10963	26	MV2	PS	M	4*	58.50 (2.34)	0.62	-1.08	0.53
10791	27	DV3	AP	D	4*	71.50 (2.86)	0.51	-2.57	0.30
10964	28	PV1	AP	P	4*	67.75 (2.71)	0.55	-1.81	0.36
10848	29	DV2	PS	D	4*	57.00 (2.28)	0.61	-1.14	0.50
EQME10116	30	GV1	KU	G	1	70.65	0.34	-0.84	0.62
10887	31	MV2	PS	M	2	44.77	0.41	0.53	1.33
EQME31552	32	MV1	KU	M	4	68.04	0.31	-0.65	0.55
13136	33	DV1	AP	D	2	49.18	0.47	0.29	1.18
10850	34	NV1	KU	N	2	77.86	0.36	-1.27	0.67
13092	35	PV1	AP	P	3	46.75	0.35	0.59	0.87
13091	36	GV2	KU	G	4	60.41	0.32	-0.15	0.55

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. KU = knowledge and understanding; AP = application; PS = thinking; N = number sense and numeration; M = measurement; G = geometry and spatial sense; P = patterning and algebra; D = data management and probability; () = Mean score for OR items.

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

**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
13132	7	% of Students	0.54	0.19	16.55	7.13	8.26	67.33
		Parameters			-5.76	0.04	-0.68	-2.80
EQME10064	8	% of Students	1.42	0.42	46.76	25.17	23.84	2.38
		Parameters			-7.88	1.09	0.31	5.69
EQME10132	9	% of Students	0.57	0.11	19.27	22.75	20.57	36.73
		Parameters			-8.22	-0.94	0.02	-0.96
10845	10	% of Students	1.72	0.61	37.39	20.06	8.41	31.81
		Parameters			-5.78	0.73	1.48	-1.85
10963	26	% of Students	1.06	0.52	32.83	19.63	22.31	23.65
		Parameters			-4.75	0.02	-0.09	0.49
10791	27	% of Students	0.71	0.12	26.15	11.09	9.71	52.22
		Parameters			-8.15	0.90	0.12	-3.14
10964	28	% of Students	1.24	0.28	24.09	18.18	14.58	41.63
		Parameters			-5.82	-0.27	0.24	-1.38
10848	29	% of Students	1.16	0.23	41.32	14.29	13.86	29.14
		Parameters			-5.18	0.78	0.12	-0.30

Note. The total number of students is 120 381.

Table 7.1.19 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
EQME21007	1	PV1	KU	P	3	83.96	0.35	-1.53	0.83
11374	2	MV2	PS	M	2	32.36	0.23	1.22	1.20
11412	3	GV3	AP	G	3	76.01	0.30	-1.07	0.66
11304	4	DV2	AP	D	4	66.61	0.36	-0.52	0.72
11313	5	NV2	PS	N	1	44.87	0.31	0.62	0.99
EQME21026	6	PV1	AP	P	3	82.69	0.35	-1.48	0.77
11393	7	NV2	PS	N	4*	69.25 (2.77)	0.49	-2.17	0.29
11348	8	NV1	AP	N	4*	44.25 (1.77)	0.50	-0.40	0.37
11455	9	DV3	AP	D	4*	65.25 (2.61)	0.55	-1.58	0.35
11454	10	GV3	PS	G	4*	67.00 (2.68)	0.56	-1.42	0.43
11175	11	MV1	KU	M	3	64.83	0.33	-0.42	0.67
EQME10290	12	DV2	KU	D	4	65.53	0.36	-0.44	0.74
11300	13	GV3	PS	G	3	47.12	0.29	0.58	0.76
11344	14	PV2	KU	P	2	88.67	0.30	-2.30	0.63
13033	15	MV2	AP	M	1	59.51	0.40	-0.13	0.89
EQME21097	16	MV2	PS	M	3	42.25	0.30	0.74	1.02
11312	17	NV1	KU	N	2	43.13	0.31	0.70	0.93
11447	18	DV1	AP	D	3	82.95	0.31	-1.73	0.61
3459676	19	MV2	AP	M	4	64.21	0.39	-0.32	0.94
EQME10231	20	NV1	PS	N	3	59.55	0.35	-0.10	0.80
EQME21095	21	DV3	PS	D	2	67.43	0.35	-0.54	0.75
13041	22	PV2	AP	P	2	89.77	0.35	-2.07	0.82
11299	23	GV2	KU	G	3	72.24	0.37	-0.79	0.77
11367	24	MV2	KU	M	4	53.74	0.36	0.16	0.80
EQME10187	25	PV1	AP	P	2	75.91	0.31	-1.26	0.54
11408	26	DV1	AP	D	4*	66.50 (2.66)	0.57	-1.45	0.44
11452	27	MV2	PS	M	4*	54.00 (2.16)	0.62	-0.57	0.52
11453	28	GV1	AP	G	4*	70.50 (2.82)	0.62	-1.39	0.49
11404	29	PV1	PS	P	4*	66.50 (2.66)	0.68	-1.22	0.59
11339	30	GV1	PS	G	2	37.45	0.29	1.03	0.88
11384	31	PV2	PS	P	4	38.58	0.29	0.96	0.96
11347	32	DV3	AP	D	3	72.44	0.38	-0.77	0.85
11293	33	NV2	AP	N	1	75.78	0.44	-0.90	1.04
11338	34	MV2	AP	M	3	50.27	0.28	0.46	0.62
EQME21037	35	NV3	KU	N	2	78.57	0.39	-1.14	0.86
11362	36	NV2	AP	N	3	65.84	0.41	-0.43	0.92

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. KU = knowledge and understanding; AP = application; PS = thinking; N = number sense and numeration; M = measurement; G = geometry and spatial sense; P = patterning and algebra; D = data management and probability; () = Mean score for OR items.

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

Table 7.1.20 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

11393	7	% of Students	1.02	0.33	23.74	16.68	13.00	45.24
		Parameters			-6.63	-0.06	0.31	-2.28
11348	8	% of Students	4.99	1.65	48.92	22.22	4.93	17.29
		Parameters			-3.97	0.99	2.69	-1.32
11455	9	% of Students	1.97	0.42	28.55	12.18	19.33	37.55
		Parameters			-5.33	0.73	-0.91	-0.80
11454	10	% of Students	1.38	0.67	19.16	17.50	31.83	29.47
		Parameters			-4.44	-0.64	-1.01	0.39
11408	26	% of Students	1.51	0.54	14.37	32.31	18.36	32.92
		Parameters			-4.00	-1.91	0.53	-0.41
11452	27	% of Students	3.78	1.35	35.12	15.85	26.07	17.84
		Parameters			-3.25	0.40	-0.48	1.04
11453	28	% of Students	1.97	0.94	21.91	9.83	21.02	44.33
		Parameters			-3.84	0.16	-1.22	-0.68
11404	29	% of Students	2.08	0.5	27.66	15.21	10.69	43.86
		Parameters			-3.76	-0.17	0.14	-1.09

Note. The total number of students in Primary Reading (English) is 128 374.

Table 7.1.21 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
13102	1	NA1	CC	N	3	87.96	0.28	-2.09	0.62
10993	2	NA2	HP	N	1	61.83	0.42	-0.16	1.05
12651	3	NA1	MA	N	3	63.21	0.30	-0.28	0.56
13281	4	NA3	MA	N	2	77.63	0.43	-0.90	1.04
13280	5	AA2	HP	A	2	78.04	0.35	-1.10	0.74
11141	6	NA4	CC	N	4	89.59	0.34	-1.90	0.85
13105	7	NA2	MA	N	4*	59.00 (2.36)	0.48	-1.73	0.40
13095	8	TA1	MA	T	4*	66.00 (2.64)	0.52	-1.38	0.37
13313	9	GA1	HP	G	4*	55.00 (2.20)	0.49	-1.16	0.36
13265	10	MA2	HP	M	4*	70.75 (2.83)	0.56	-1.69	0.43
13252	11	MA4	CC	M	1	77.48	0.41	-0.89	1.02
11001	12	AA1	CC	A	2	71.38	0.33	-0.72	0.67
13342	13	GA1	MA	G	4	56.01	0.22	0.26	0.53
10995	14	MA1	HP	M	2	46.71	0.32	0.60	0.87
13239	15	NA3	HP	N	2	51.30	0.37	0.36	0.89
13251	16	AA3	MA	A	3	67.97	0.45	-0.40	1.10
13254	17	GA2	MA	G	4	56.33	0.32	0.16	0.67
11123	18	TA2	MA	T	1	63.88	0.31	-0.23	0.68
13311	19	TA1	CC	T	3	81.53	0.36	-1.35	0.72
13103	20	MA2	CC	M	3	66.87	0.34	-0.51	0.61
12606	21	AA2	CC	A	3	49.73	0.34	0.45	0.76
13257	22	NA4	MA	N	2	87.99	0.41	-1.54	1.10
11032	23	AA3	CC	A	3	73.16	0.46	-0.65	1.10
12639	24	TA1	HP	T	2	77.85	0.52	-0.85	1.41
OQMF31545	25	MA3	CC	M	3	84.69	0.38	-1.41	0.89
OQMF31569	26	GA1	MA	G	4*	74.00 (2.96)	0.52	-2.35	0.38
13219	27	NA4	HP	N	4*	59.75 (2.39)	0.46	-1.37	0.39
13341	28	TA1	HP	T	4*	64.25 (2.57)	0.49	-2.07	0.32
13326	29	AA1	HP	A	4*	51.75 (2.07)	0.46	-0.82	0.33
12642	30	MA2	MA	M	3	53.29	0.41	0.22	1.02
10994	31	MA1	MA	M	4	77.44	0.43	-0.96	0.93
11272	32	TA1	MA	T	3	56.64	0.26	0.15	0.49
11112	33	GA1	HP	G	4	48.18	0.31	0.57	0.81
12647	34	NA2	MA	N	3	69.95	0.42	-0.51	0.99
10988	35	NA3	CC	N	3	75.95	0.33	-1.08	0.61
12641	36	GA1	CC	G	2	62.54	0.32	-0.22	0.61

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. CC = knowledge and understanding; MA = application; HP = thinking; N = number sense and numeration; M = measurement; G = geometry and spatial sense; A = patterning and algebra; T = data management and probability; () = Mean score for OR items.

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

**Table 7.1.22 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
13105	7	% of Students	0.35	0.04	17.58	49.46	10.71	21.84
		Parameters			-6.59	-2.04	2.27	-0.54
13095	8	% of Students	2.18	0.31	16.58	25.54	25.11	30.27
		Parameters			-4.08	-1.35	-0.12	0.04
13313	9	% of Students	1.18	0.19	37.63	15.02	31.55	14.43
		Parameters			-6.47	1.16	-1.15	1.84
13265	10	% of Students	1.05	0.18	15.89	22.21	19.85	40.83
		Parameters			-4.77	-1.19	-0.10	-0.68
OQMF31569	26	% of Students	0.46	0.04	12.28	16.09	33.29	37.83
		Parameters			-6.65	-1.30	-1.44	0.00
13219	27	% of Students	1.06	0.21	8.39	60.27	10.51	19.57
		Parameters			-4.09	-3.61	2.64	-0.41
13341	28	% of Students	0.6	0.22	22.23	27.41	17.89	31.64
		Parameters			-7.37	-0.92	0.69	-0.67
13326	29	% of Students	2.25	0.5	32.28	33.90	17.76	13.31
		Parameters			-5.20	-0.46	1.30	1.08

Note. The total number of students is 6 785.

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
11583	1	NA2	MA	N	3	84.29	0.36	-1.40	0.91
13344	2	GA2	CC	G	1	85.01	0.28	-1.85	0.59
13021	3	MA2	HP	M	3	25.90	0.08	1.98	1.28
11573	4	TA1	MA	T	1	82.55	0.31	-1.46	0.68
13352	5	NA3	MA	N	4	54.69	0.33	0.20	0.69
13355	6	MA1	MA	M	1	52.86	0.46	0.18	1.22
11660	7	MA2	MA	M	4*	74.50 (2.98)	0.50	-2.33	0.32
13328	8	AA2	MA	A	4*	67.25 (2.69)	0.60	-1.21	0.49
13362	9	NA1	HP	N	4*	57.75 (2.31)	0.58	-1.09	0.45
13319	10	GA2	HP	G	4*	55.00 (2.20)	0.59	-0.95	0.54
13348	11	GA2	CC	G	3	37.36	0.21	1.37	0.67
13322	12	MA3	CC	M	3	73.37	0.36	-0.78	0.79
13361	13	MA1	HP	M	4	58.06	0.46	-0.01	1.18
11530	14	TA1	HP	T	2	48.09	0.31	0.60	0.66
11604	15	GA1	CC	G	3	76.48	0.40	-0.88	0.93
11576	16	TA2	HP	T	3	55.54	0.37	0.14	0.87
13353	17	AA2	HP	A	2	59.06	0.39	-0.03	0.87
13316	18	AA1	MA	A	4	73.24	0.39	-0.79	0.82
OQMF20842	19	NA1	CC	N	3	60.21	0.44	-0.09	1.18
13354	20	NA2	CC	N	3	91.13	0.24	-2.94	0.51
11568	21	AA2	MA	A	2	57.41	0.42	0.04	1.03
11649	22	MA3	MA	M	3	37.55	0.33	1.02	0.90
11544	23	GA1	MA	G	2	72.17	0.35	-0.77	0.69
11525	24	AA1	CC	A	4	52.29	0.36	0.32	0.84
13333	25	MA2	CC	M	2	50.44	0.32	0.45	0.91
11511	26	AA1	HP	A	4*	72.50 (2.90)	0.54	-1.91	0.37
11642	27	TA1	HP	T	4*	46.00 (1.84)	0.47	-1.42	0.32
12574	28	GA1	MA	G	4*	59.75 (2.39)	0.56	-1.45	0.39
13364	29	NA3	MA	N	4*	63.50 (2.54)	0.57	-1.41	0.40
13359	30	TA2	CC	T	4	72.81	0.39	-0.74	0.83
13335	31	NA1	MA	N	2	54.00	0.40	0.18	0.87
13358	32	GA2	HP	G	4	47.14	0.32	0.61	0.75
13357	33	NA3	HP	N	2	52.24	0.44	0.24	1.16
11482	34	GA1	HP	G	3	64.98	0.34	-0.39	0.64
11486	35	TA2	MA	T	3	59.43	0.41	-0.04	0.99
13312	36	MA1	CC	M	4	77.57	0.36	-1.01	0.78

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. CC = knowledge and understanding; MA = application; HP = thinking; N = number sense and numeration; M = measurement; G = geometry and spatial sense; A = patterning and algebra; T = data management and probability; () = Mean score for OR items.

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

**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
11660	7	% of Students	0.73	0.26	21.06	14.32	6.65	56.98
		Parameters			-6.61	-0.13	1.18	-3.77
13328	8	% of Students	1.98	0.59	25.15	6.72	31.70	33.85
		Parameters			-3.97	0.88	-2.02	0.25
13362	9	% of Students	1.59	0.44	32.82	25.54	11.72	27.89
		Parameters			-4.70	-0.13	1.07	-0.59
13319	10	% of Students	0.61	0.28	31.98	25.65	28.92	12.56
		Parameters			-5.11	-0.29	-0.01	1.60
11511	26	% of Students	1.25	0.31	18.82	17.93	11.81	49.88
		Parameters			-5.28	-0.67	0.39	-2.07
11642	27	% of Students	0.7	0.25	58.83	11.77	11.70	16.75
		Parameters			-8.60	2.71	0.26	-0.04
12574	28	% of Students	1.07	0.48	29.87	28.02	9.26	31.30
		Parameters			-5.66	-0.38	1.65	-1.39
13364	29	% of Students	1.87	0.47	26.61	27.08	2.41	41.56
		Parameters			-4.70	-0.61	3.54	-3.88

Note. The total number of students is 6 425.

Differential Item Functioning (DIF) Analysis Results

The gender-based and SLL-based DIF results are provided for the primary- and junior-division assessments in Tables 7.1.25a–7.1.48b. Results are presented for two random samples of 2000 examinees. Each table indicates the value of Δ for multiple choice items or effect size for open response items, and the significance level for each item. The DIF is also provided for those items that had a significant level of DIF, with at least a B- or C-level effect size. The results for items with B or C level DIF in both samples are presented in bold type. For gender-based DIF, negative values of Δ for multiple choice items and negative effect size values for open response items indicate that the girls outperformed the boys; positive values of Δ for multiple choice items and positive effect sizes for open response items indicate that the boys outperformed the girls. For SLL-based DIF, negative values of Δ for multiple choice items and negative effect sizes for open response items indicate that the SLLs outperformed the non-SLLs; positive values of Δ for multiple choice items and positive effect sizes for open response items indicate that the non-SLLs outperformed the SLLs.

Table 7.1.25a Gender-Based DIF Statistics for MC Items: Primary Reading (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
10769	1	1	0.11	-0.40	0.62		0.25	-0.27	0.77	
10773	1	2	-0.64	-1.18	-0.10		-0.77	-1.30	-0.23	
10771	1	3	0.05	-0.53	0.63		-0.27	-0.83	0.30	
10770	1	4	-0.01	-0.34	0.32		0.52	0.19	0.86	
10931	1	7	0.13	-0.30	0.56		0.90	0.47	1.33	
10932	1	8	0.16	-0.20	0.52		0.23	-0.13	0.59	
10925	1	9	-0.13	-0.44	0.18		0.08	-0.23	0.39	
10935	1	10	0.45	0.12	0.79		0.22	-0.11	0.55	
10919	1	14	0.84	0.53	1.16		0.45	0.14	0.76	
10914	1	15	0.36	0.03	0.68		0.23	-0.09	0.55	
10924	1	16	-0.42	-0.90	0.06		0.01	-0.48	0.51	
10920	1	17	0.29	-0.03	0.61		0.53	0.20	0.86	
12053	2	1	-0.48	-0.96	-0.01		0.00	-0.48	0.48	
12056	2	2	-0.09	-0.49	0.30		0.09	-0.31	0.49	
12057	2	3	0.20	-0.25	0.64		-0.02	-0.46	0.41	
12058	2	4	0.43	0.02	0.84		0.61	0.20	1.01	
12059	2	5	0.26	-0.08	0.61		0.35	0.00	0.69	
12060	2	6	0.21	-0.11	0.52		0.39	0.08	0.70	
12061	2	7	-0.51	-0.81	-0.20		-0.23	-0.54	0.07	
12062	2	8	0.14	-0.28	0.55		-0.10	-0.52	0.33	
12063	2	9	-0.04	-0.40	0.32		-0.23	-0.59	0.14	
12064	2	10	-0.22	-0.54	0.10		-0.58	-0.90	-0.27	
10876	2	18	-0.36	-0.74	0.01		0.12	-0.26	0.50	
10878	2	19	-0.28	-0.68	0.11		0.34	-0.05	0.73	
10880	2	20	0.29	-0.02	0.61		0.09	-0.23	0.40	
10881	2	21	-0.18	-0.50	0.13		-0.30	-0.63	0.02	

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

Table 7.1.25b Gender-Based DIF Statistics for OR Items: Primary Reading (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10772	1	5	-0.03	0.05		-0.00	0.31	
10767	1	6	-0.06	0.03		-0.06	0.11	
10934	1	11	-0.02	0.60		-0.08	0.02	
10929	1	12	-0.10	0.00		-0.08	0.01	
10916	1	18	0.05	0.00		0.01	0.17	
10922	1	19	-0.06	0.25		-0.04	0.71	
12065	2	11	-0.09	0.03		-0.14	0.00	
12066	2	12	-0.02	0.13		-0.06	0.00	
10879	2	22	-0.04	0.66		-0.02	0.61	
10877	2	23	0.04	0.01		-0.01	0.45	

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

Table 7.1.26a Gender-Based DIF Statistics for MC Items: Junior Reading (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
10623	1	1	-0.20	-1.29	0.88		-0.55	-1.51	0.40	

10626	1	2	0.03	-0.35	0.41		0.06	-0.31	0.42	
10631	1	3	-0.28	-0.62	0.06		-0.25	-0.58	0.08	
10624	1	4	-0.42	-0.91	0.06		-0.57	-1.05	-0.10	
10614	1	7	0.80	-0.02	1.62		0.42	-0.42	1.26	
10618	1	8	1.06	0.72	1.41	B+	1.15	0.81	1.49	B+
10619	1	9	0.13	-0.17	0.43		-0.25	-0.56	0.05	
10615	1	10	0.61	0.18	1.04		0.34	-0.08	0.76	
10516	1	14	0.37	-0.18	0.91		0.13	-0.38	0.64	
10515	1	15	0.73	0.40	1.06		0.91	0.58	1.24	
10519	1	16	-0.27	-0.74	0.21		-0.17	-0.63	0.29	
10522	1	17	1.00	0.64	1.35		1.12	0.76	1.47	B+
12313	2	1	-0.15	-0.47	0.17		-0.14	-0.47	0.18	
12311	2	2	-0.05	-0.43	0.34		0.53	0.14	0.91	
12319	2	3	0.26	-0.10	0.62		-0.19	-0.55	0.17	
12314	2	4	0.04	-0.41	0.49		0.57	0.13	1.01	
12315	2	5	1.34	0.99	1.69	B+	1.58	1.23	1.93	C+
12316	2	6	0.12	-0.50	0.75		-0.28	-0.89	0.33	
12310	2	7	-0.08	-0.38	0.22		-0.22	-0.51	0.08	
12317	2	8	0.37	-0.07	0.80		0.33	-0.11	0.76	
12318	2	9	0.29	-0.12	0.69		0.43	0.03	0.83	
12312	2	10	-0.29	-0.66	0.08		-0.08	-0.45	0.28	
10585	2	18	0.25	-0.11	0.62		0.01	-0.35	0.36	
10591	2	19	1.31	0.98	1.63	B+	1.64	1.30	1.97	C+
10588	2	20	1.14	0.75	1.53	B+	1.13	0.75	1.52	B+
10590	2	21	1.16	0.81	1.50	B+	0.64	0.30	0.98	

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

Table 7.1.26b Gender-Based DIF Statistics for OR Items: Junior Reading (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10625	1	5	-0.20	0.00	B-	-0.20	0.00	B-
10627	1	6	-0.16	0.00		-0.18	0.00	B-
10616	1	11	-0.13	0.00		-0.09	0.00	
10620	1	12	-0.02	0.01		-0.01	0.95	
10518	1	18	-0.06	0.05		-0.05	0.21	
10523	1	19	0.02	0.29		-0.04	0.08	
12321	2	11	-0.12	0.00		-0.08	0.01	
12320	2	12	-0.08	0.00		-0.08	0.02	
10589	2	22	-0.12	0.00		-0.15	0.00	
10587	2	23	-0.11	0.00		-0.08	0.04	

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

Table 7.1.27a Gender-Based DIF Statistics for MC Items: Primary Reading (French)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11055	1	1	-0.43	-0.81	-0.06		-0.45	-0.82	-0.08	
11060	1	2	-0.11	-0.53	0.30		-0.61	-1.03	-0.18	
11052	1	3	0.08	-0.37	0.53		-0.32	-0.77	0.14	
11056	1	4	0.22	-0.12	0.56		0.17	-0.16	0.51	
12235	1	7	-0.65	-1.20	-0.10		-0.43	-0.96	0.10	
12238	1	8	-0.44	-0.82	-0.06		-0.45	-0.83	-0.08	
12236	1	9	0.51	0.19	0.83		0.43	0.11	0.76	
12237	1	10	0.49	0.13	0.85		0.29	-0.07	0.65	
11189	1	14	0.58	0.20	0.95		0.58	0.20	0.97	
11191	1	15	-0.17	-0.50	0.16		-0.09	-0.42	0.24	
11187	1	16	-0.06	-0.37	0.25		-0.14	-0.45	0.17	
11190	1	17	-0.06	-0.42	0.29		-0.10	-0.46	0.26	
12193	2	1	-0.92	-1.42	-0.43		-0.84	-1.32	-0.36	
12194	2	2	-0.18	-0.54	0.19		-0.20	-0.56	0.16	
12195	2	3	-0.05	-0.38	0.27		-0.01	-0.34	0.31	
12202	2	4	-0.22	-0.54	0.09		-0.28	-0.59	0.04	
12196	2	5	0.86	0.47	1.25		0.48	0.10	0.86	
12197	2	6	0.14	-0.18	0.46		0.11	-0.21	0.43	
12198	2	7	-0.40	-0.77	-0.04		-0.51	-0.87	-0.15	
12199	2	8	-0.04	-0.35	0.27		0.03	-0.28	0.35	
12201	2	9	0.02	-0.36	0.39		0.02	-0.36	0.39	
12200	2	10	-0.17	-0.54	0.19		-0.19	-0.56	0.17	
12247	2	18	0.62	0.24	1.00		0.54	0.16	0.91	
12248	2	19	1.12	0.72	1.51	B+	0.89	0.50	1.28	
12249	2	20	0.45	0.13	0.77		0.15	-0.17	0.47	
12250	2	21	0.50	0.17	0.84		0.42	0.09	0.76	

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

Table 7.1.27b Gender-Based DIF Statistics for OR Items: Primary Reading (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11063	1	5	-0.07	0.09		-0.07	0.08	
11062	1	6	0.02	0.27		0.06	0.04	
12239	1	11	-0.11	0.00		-0.10	0.00	
12240	1	12	-0.12	0.00		-0.10	0.00	
11193	1	18	0.11	0.00		0.11	0.00	
11194	1	19	0.03	0.03		0.06	0.03	
12203	2	11	-0.01	0.04		-0.05	0.22	
12204	2	12	-0.05	0.12		-0.01	0.10	
12251	2	22	-0.06	0.23		-0.02	0.71	
12252	2	23	-0.10	0.01		-0.07	0.09	

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

Table 7.1.28a Gender-Based DIF Statistics for MC Items: Junior Reading (French)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11373	1	1	-0.26	-0.57	0.04		-0.15	-0.46	0.15	
11376	1	2	0.47	0.16	0.78		0.42	0.11	0.73	
11375	1	3	-0.47	-0.82	-0.12		-0.46	-0.81	-0.10	
11371	1	4	-0.17	-0.59	0.25		0.22	-0.20	0.64	
11610	1	7	0.07	-0.26	0.40		0.09	-0.24	0.42	
11613	1	8	0.30	-0.02	0.62		0.40	0.08	0.73	
11611	1	9	0.74	0.40	1.07		0.49	0.15	0.82	
11615	1	10	-0.78	-1.10	-0.46		-0.77	-1.09	-0.44	
11415	1	14	0.11	-0.24	0.45		0.01	-0.33	0.35	
11420	1	15	0.03	-0.31	0.36		0.05	-0.29	0.38	
11422	1	16	0.82	0.44	1.20		0.46	0.08	0.84	
11417	1	17	0.01	-0.33	0.35		0.29	-0.05	0.64	
11474	2	1	0.54	0.20	0.89		0.59	0.24	0.94	
11462	2	2	-0.17	-0.65	0.31		0.11	-0.37	0.59	
11469	2	3	-0.18	-0.61	0.25		0.02	-0.41	0.45	
11467	2	4	0.78	0.45	1.11		0.83	0.49	1.16	
11470	2	5	0.57	0.25	0.89		0.38	0.07	0.70	
11443	2	6	-0.01	-0.33	0.31		0.15	-0.17	0.47	
11461	2	7	-0.10	-0.72	0.53		0.02	-0.60	0.65	
11466	2	8	0.28	-0.02	0.59		0.19	-0.11	0.50	
11472	2	9	0.04	-0.35	0.43		0.22	-0.17	0.60	
11463	2	10	0.64	0.24	1.05		0.47	0.07	0.88	
11397	2	18	2.26	1.86	2.66	C+	2.13	1.74	2.52	C+
11403	2	19	0.00	-0.31	0.30		-0.21	-0.51	0.10	
11398	2	20	0.40	0.06	0.74		0.06	-0.28	0.40	
11399	2	21	0.60	0.22	0.99		0.55	0.16	0.93	

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

Table 7.1.28b Gender-Based DIF Statistics for OR Items: Junior Reading (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11379	1	5	-0.14	0.00		-0.09	0.00	
11380	1	6	-0.17	0.00		-0.16	0.00	
11617	1	11	-0.11	0.00		-0.11	0.00	
11619	1	12	-0.18	0.00	B-	-0.16	0.00	
11423	1	18	-0.01	0.06		0.01	0.38	
11424	1	19	-0.14	0.00		-0.14	0.00	
11475	2	11	-0.05	0.16		-0.03	0.08	
11476	2	12	-0.06	0.18		-0.05	0.05	
11405	2	22	-0.10	0.00		-0.13	0.00	
11407	2	23	-0.04	0.13		-0.13	0.00	

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

Table 7.1.29a SLL-Based DIF Statistics for MC Items: Primary Reading (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
10769	1	1	1.34	0.85	1.82	B+	0.71	0.25	1.17	
10773	1	2	0.35	-0.13	0.83		0.31	-0.19	0.81	
10771	1	3	1.43	0.92	1.93	B+	1.44	0.92	1.96	B+
10770	1	4	0.85	0.53	1.16		0.98	0.66	1.29	
10931	1	7	0.02	-0.39	0.43		0.14	-0.28	0.55	
10932	1	8	-0.13	-0.48	0.23		-0.02	-0.37	0.34	
10925	1	9	-0.17	-0.47	0.14		-0.08	-0.39	0.23	
10935	1	10	0.88	0.56	1.21		0.68	0.35	1.01	
10919	1	14	0.00	-0.31	0.31		0.21	-0.10	0.53	
10914	1	15	0.18	-0.14	0.50		0.27	-0.05	0.59	
10924	1	16	-0.23	-0.72	0.25		-0.09	-0.54	0.37	
10920	1	17	0.81	0.50	1.13		0.87	0.55	1.18	
12053	2	1	0.50	0.05	0.94		0.19	-0.26	0.64	
12056	2	2	0.63	0.24	1.01		0.14	-0.24	0.53	
12057	2	3	-0.05	-0.47	0.38		-0.19	-0.62	0.24	
12058	2	4	0.27	-0.13	0.66		0.31	-0.08	0.70	
12059	2	5	1.06	0.73	1.40	B+	0.99	0.65	1.32	
12060	2	6	-0.11	-0.42	0.20		0.03	-0.29	0.35	
12061	2	7	0.01	-0.30	0.32		-0.06	-0.37	0.25	
12062	2	8	-0.42	-0.82	-0.02		0.02	-0.39	0.43	
12063	2	9	-0.06	-0.41	0.29		-0.16	-0.52	0.20	
12064	2	10	0.34	0.02	0.66		0.37	0.04	0.69	
10876	2	18	0.18	-0.18	0.54		0.42	0.06	0.78	
10878	2	19	-0.04	-0.41	0.33		0.05	-0.33	0.44	
10880	2	20	0.14	-0.18	0.45		-0.48	-0.80	-0.16	
10881	2	21	0.89	0.57	1.21		0.82	0.50	1.14	

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

Table 7.1.29b SLL-Based DIF Statistics for OR Items: Primary Reading (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10772	1	5	-0.05	0.17		-0.02	0.18	
10767	1	6	-0.01	0.05		-0.02	0.16	
10934	1	11	-0.09	0.00		-0.02	0.46	
10929	1	12	-0.09	0.01		-0.07	0.01	
10916	1	18	-0.15	0.00		-0.13	0.00	
10922	1	19	-0.03	0.64		-0.02	0.66	
12065	2	11	-0.04	0.39		-0.06	0.20	
12066	2	12	0.02	0.00		0.11	0.00	
10879	2	22	-0.06	0.29		-0.03	0.67	
10877	2	23	-0.16	0.00		-0.06	0.15	

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

Table 7.1.30a SLL-Based DIF Statistics for MC Items: Junior Reading (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
10623	1	1	-0.11	-0.92	0.70		0.05	-0.83	0.93	

10626	1	2	0.43	0.07	0.78		0.36	0.01	0.71	
10631	1	3	1.25	0.92	1.58	B+	1.39	1.07	1.72	B+
10624	1	4	0.33	-0.11	0.77		0.40	-0.04	0.83	
10614	1	7	-0.28	-1.00	0.44		0.71	-0.10	1.52	
10618	1	8	0.68	0.35	1.01		0.74	0.41	1.08	
10619	1	9	0.06	-0.25	0.36		-0.27	-0.58	0.03	
10615	1	10	0.40	0.00	0.80		0.10	-0.30	0.50	
10516	1	14	0.36	-0.13	0.85		0.32	-0.16	0.81	
10515	1	15	0.61	0.29	0.93		0.49	0.16	0.81	
10519	1	16	-0.05	-0.50	0.39		0.15	-0.28	0.59	
10522	1	17	0.69	0.35	1.03		0.74	0.40	1.07	
12313	2	1	0.18	-0.15	0.50		0.06	-0.27	0.39	
12311	2	2	-0.11	-0.50	0.27		0.03	-0.35	0.40	
12319	2	3	0.38	0.04	0.73		0.33	-0.02	0.67	
12314	2	4	0.05	-0.37	0.48		0.14	-0.28	0.56	
12315	2	5	1.77	1.44	2.10	C+	1.96	1.64	2.29	C+
12316	2	6	0.46	-0.08	1.00		0.25	-0.29	0.79	
12310	2	7	0.18	-0.12	0.48		0.10	-0.20	0.40	
12317	2	8	0.24	-0.17	0.65		0.24	-0.17	0.65	
12318	2	9	0.77	0.39	1.15		1.09	0.72	1.47	B+
12312	2	10	0.75	0.39	1.10		0.49	0.14	0.84	
10585	2	18	-0.58	-0.95	-0.21		-0.76	-1.12	-0.40	
10591	2	19	-0.33	-0.65	0.00		-0.50	-0.83	-0.17	
10588	2	20	-0.02	-0.39	0.35		0.00	-0.37	0.36	
10590	2	21	-0.20	-0.54	0.13		0.18	-0.16	0.51	

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

Table 7.1.30b SLL-Based DIF Statistics for OR Items: Junior Reading (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10625	1	5	-0.01	0.73		-0.03	0.80	
10627	1	6	0.01	0.11		-0.02	0.06	
10616	1	11	-0.04	0.39		-0.02	0.08	
10620	1	12	-0.00	0.15		-0.04	0.43	
10518	1	18	-0.09	0.00		-0.03	0.01	
10523	1	19	-0.09	0.00		-0.05	0.02	
12321	2	11	-0.18	0.00	B-	-0.19	0.00	B-
12320	2	12	-0.15	0.00		-0.15	0.00	
10589	2	22	-0.11	0.00		-0.07	0.02	
10587	2	23	0.02	0.64		-0.01	0.55	

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

Table 7.1.31a SLL-Based DIF Statistics for MC Items: Primary Reading (French)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11055	1	1	0.21	-0.27	0.70		-0.14	-0.62	0.34	
11060	1	2	0.08	-0.46	0.61		0.42	-0.13	0.97	
11052	1	3	0.17	-0.40	0.74		0.13	-0.46	0.71	
11056	1	4	0.16	-0.26	0.59		0.48	0.06	0.91	
12235	1	7	0.23	-0.47	0.93		-0.10	-0.79	0.59	
12238	1	8	-0.09	-0.58	0.39		-0.03	-0.52	0.47	
12236	1	9	-0.01	-0.43	0.41		0.09	-0.33	0.51	
12237	1	10	0.22	-0.25	0.69		0.04	-0.43	0.50	
11189	1	14	0.31	-0.16	0.79		0.41	-0.08	0.90	
11191	1	15	0.65	0.20	1.10		0.78	0.34	1.23	
11187	1	16	-0.01	-0.42	0.40		0.39	-0.02	0.79	
11190	1	17	0.59	0.13	1.05		0.57	0.12	1.03	
12193	2	1	-0.23	-0.82	0.37		-0.22	-0.83	0.39	
12194	2	2	0.87	0.41	1.34		0.74	0.28	1.20	
12195	2	3	-0.40	-0.82	0.03		-0.18	-0.60	0.25	
12202	2	4	0.03	-0.39	0.45		-0.01	-0.44	0.42	
12196	2	5	-0.10	-0.60	0.39		-0.25	-0.75	0.24	
12197	2	6	0.39	-0.03	0.82		0.36	-0.06	0.79	
12198	2	7	-0.30	-0.77	0.17		-0.25	-0.72	0.21	
12199	2	8	0.81	0.40	1.21		0.77	0.37	1.18	
12201	2	9	-0.31	-0.79	0.17		-0.15	-0.63	0.33	
12200	2	10	0.27	-0.21	0.74		-0.34	-0.81	0.12	
12247	2	18	-0.43	-0.92	0.06		-0.55	-1.03	-0.07	
12248	2	19	-0.57	-1.09	-0.05		-0.95	-1.46	-0.44	
12249	2	20	-0.14	-0.55	0.28		-0.02	-0.43	0.40	
12250	2	21	0.15	-0.28	0.58		0.14	-0.29	0.57	

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

Table 7.1.31b SLL-Based DIF Statistics for OR Items: Primary Reading (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11063	1	5	-0.06	0.54		-0.11	0.16	
11062	1	6	-0.06	0.46		-0.07	0.11	
12239	1	11	-0.08	0.17		-0.10	0.07	
12240	1	12	0.04	0.72		0.05	0.10	
11193	1	18	-0.00	0.49		-0.01	0.54	
11194	1	19	0.06	0.30		0.06	0.02	
12203	2	11	-0.06	0.19		-0.08	0.09	
12204	2	12	-0.02	0.01		-0.04	0.01	
12251	2	22	-0.01	0.47		-0.03	0.62	
12252	2	23	0.04	0.43		0.03	0.28	

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

Table 7.1.32a SLL-Based DIF Statistics for MC Items: Junior Reading (French)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11373	1	1	0.85	0.33	1.38		0.64	0.11	1.16	
11376	1	2	-0.24	-0.78	0.30		-0.11	-0.66	0.43	
11375	1	3	-0.32	-0.94	0.29		-0.42	-1.04	0.20	
11371	1	4	-0.21	-0.96	0.54		-0.59	-1.32	0.14	
11610	1	7	-0.16	-0.74	0.42		-0.35	-0.93	0.23	
11613	1	8	0.56	0.01	1.10		0.73	0.18	1.28	
11611	1	9	-0.40	-0.99	0.19		-0.97	-1.56	-0.38	
11615	1	10	-0.56	-1.12	0.00		-0.17	-0.73	0.38	
11415	1	14	-0.69	-1.30	-0.08		-0.43	-1.04	0.19	
11420	1	15	-0.86	-1.47	-0.26		-1.05	-1.65	-0.46	B-
11422	1	16	-0.47	-1.12	0.17		-0.40	-1.04	0.23	
11417	1	17	0.09	-0.52	0.69		-0.08	-0.67	0.52	
11474	2	1	-0.65	-1.27	-0.02		-0.87	-1.48	-0.26	
11462	2	2	-0.14	-0.98	0.71		-0.38	-1.21	0.45	
11469	2	3	0.74	-0.03	1.50		0.30	-0.43	1.03	
11467	2	4	-0.28	-0.87	0.30		-0.31	-0.89	0.28	
11470	2	5	-0.52	-1.07	0.02		-0.17	-0.71	0.37	
11443	2	6	0.16	-0.39	0.71		-0.13	-0.69	0.42	
11461	2	7	0.48	-0.64	1.60		-0.02	-1.09	1.04	
11466	2	8	-0.29	-0.83	0.25		-0.04	-0.57	0.49	
11472	2	9	1.44	0.79	2.10	B+	1.65	0.99	2.30	B+
11463	2	10	0.59	-0.11	1.29		-0.27	-0.97	0.43	
11397	2	18	-0.91	-1.59	-0.23		-0.64	-1.32	0.04	
11403	2	19	-0.08	-0.62	0.46		0.00	-0.54	0.55	
11398	2	20	0.29	-0.29	0.87		0.17	-0.42	0.76	
11399	2	21	0.30	-0.37	0.97		0.12	-0.52	0.77	

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

Table 7.1.32b SLL-Based DIF Statistics for OR Items: Junior Reading (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11379	1	5	-0.01	0.88		0.04	0.69	
11380	1	6	0.11	0.20		0.06	0.33	
11617	1	11	0.07	0.05		0.01	0.15	
11619	1	12	0.06	0.02		0.02	0.02	
11423	1	18	0.13	0.04		0.15	0.01	
11424	1	19	0.01	0.71		0.06	0.06	
11475	2	11	-0.06	0.64		-0.05	0.68	
11476	2	12	0.02	0.62		0.01	0.65	
11405	2	22	-0.03	0.32		-0.02	0.31	
11407	2	23	0.03	0.28		0.01	0.69	

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

Table 7.1.33a Gender-Based DIF Statistics for MC Items: Primary Writing (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11041	2	13	0.17	0.50	-0.17		-0.05	0.29	-0.39	
11049	2	14	-0.20	0.23	-0.62		-0.10	0.32	-0.51	
11047	2	15	-0.06	0.26	-0.39		-0.32	0.01	-0.64	
11039	2	16	-0.64	-0.27	-1.01		-0.31	0.07	-0.68	
11031	2	25	0.11	0.52	-0.30		-0.10	0.31	-0.51	
11043	2	26	0.20	0.54	-0.14		-0.23	0.11	-0.58	
11024	2	27	-1.06	-0.71	-1.42	B+	-0.62	-0.26	-0.97	
11030	2	28	-0.06	0.28	-0.41		0.16	0.50	-0.18	

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

Table 7.1.33b Gender-Based DIF Statistics for OR Items: Primary Writing (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
12173_T	1	13	-0.05	0.08		0.00	0.11	
12173_V	1	13	0.02	0.71		0.03	0.35	
11050_T	2	17	0.02	0.34		-0.01	0.13	
11050_V	2	17	0.05	0.01		0.03	0.11	
17369_T	2	24	0.01	0.28		0.00	0.70	
11369_V	2	24	0.10	0.00		0.05	0.05	

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

Table 7.1.34a Gender-Based DIF Statistics for MC Items: Junior Writing (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
10707	2	13	-0.32	0.05	-0.68		-0.54	-0.18	-0.90	
10716	2	14	-0.17	0.40	-0.73		-0.73	-0.15	-1.32	
10699	2	15	-0.05	0.38	-0.49		-0.26	0.17	-0.68	
10694	2	16	-0.43	-0.08	-0.77		-0.57	-0.23	-0.91	
10721	2	25	-0.33	0.10	-0.76		-0.36	0.07	-0.78	
10714	2	26	-0.61	-0.25	-0.97		-0.81	-0.44	-1.17	
10705	2	27	-0.02	0.37	-0.42		-0.09	0.32	-0.50	
10723	2	28	-0.63	-0.30	-0.96		-0.20	0.13	-0.52	

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

Table 7.1.34b Gender-Based DIF Statistics for OR Items: Junior Writing (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10703_T	1	13	0.01	0.65		0.03	0.15	
10703_V	1	13	0.10	0.00		0.13	0.00	
10708_T	2	17	-0.03	0.63		-0.02	0.07	
10708_V	2	17	0.05	0.10		0.05	0.00	
17733_T	2	24	0.01	0.03		0.01	0.10	
17733_V	2	24	0.08	0.00		0.07	0.00	

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

Table 7.1.35a Gender-Based DIF Statistics for MC Items: Primary Writing (French)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
13015	2	13	-0.13	0.36	-0.63		-0.84	-0.34	-1.34	
11255	2	14	-0.00	0.61	-0.61		-0.59	0.00	-1.19	
11280	2	15	0.08	0.66	-0.49		-0.53	0.06	-1.11	
11275	2	16	-0.95	-0.31	-1.60		-0.78	-0.15	-1.41	
11279	2	25	-0.14	0.46	-0.75		-0.07	0.53	-0.67	
13016	2	26	-0.40	0.12	-0.92		-0.37	0.14	-0.87	
11258	2	27	-0.57	-0.01	-1.14		-0.07	0.50	-0.65	
11282	2	28	0.13	0.78	-0.53		-0.09	0.52	-0.71	

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

Table 7.1.35b Gender-Based DIF Statistics for OR Items: Primary Writing (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11242_T	1	13	0.06	0.69		0.04	0.25	
11242_V	1	13	0.04	0.00		0.07	0.06	
11241_T	2	17	-0.06	0.29		0.01	0.02	
11241_V	2	17	0.04	0.73		0.05	0.41	
17714_T	2	24	0.02	0.74		0.05	0.47	
17714_V	2	24	0.07	0.01		0.09	0.01	

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

Table 7.1.36a Gender-Based DIF Statistics for MC Items: Junior Writing (French)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11553	2	13	-0.07	0.65	-0.80		-1.11	0.34	-1.88	B+
11572	2	14	-0.34	0.17	-0.85		-0.01	0.47	-0.49	
12538	2	15	-0.19	0.27	-0.65		-0.27	0.19	-0.73	
11554	2	16	-0.36	0.12	-0.84		-0.58	-0.10	1.06	
11561	2	25	-0.75	-0.24	-1.26		-1.00	-0.49	-1.51	B+
11569	2	26	-0.90	-0.39	-1.41		-0.11	0.40	-0.62	
11557	2	27	-0.24	0.46	-0.95		-0.87	-0.11	-1.64	
13030	2	28	-1.50	-0.95	-2.04	B+	-1.05	-0.52	-1.59	B+

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

Table 7.1.36b Gender-Based DIF Statistics for OR Items: Junior Writing (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11596_T	1	13	0.10	0.02		-0.01	0.82	
11596_V	1	13	0.15	0.00		0.12	0.00	
11589_T	2	17	0.06	0.10		0.02	0.59	
11589_V	2	17	0.05	0.02		0.06	0.12	
18331_T	2	24	0.00	0.06		0.06	0.04	
18331_V	2	24	0.08	0.00		0.11	0.00	

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

Table 7.1.37a SLL-Based DIF Statistics for MC Items: Primary Writing (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11041	2	13	-0.09	0.24	-0.42		0.14	0.47	-0.19	
11049	2	14	0.04	0.44	-0.36		0.13	0.52	-0.26	
11047	2	15	0.07	0.39	-0.25		-0.07	0.24	-0.39	
11039	2	16	0.15	0.51	-0.20		0.31	0.67	-0.04	
11031	2	25	0.74	1.13	0.34		0.14	0.53	-0.24	
11043	2	26	0.34	0.67	0.01		0.22	0.55	-0.11	
11024	2	27	0.58	0.91	0.24		0.71	1.04	0.37	
11030	2	28	0.51	0.85	0.17		0.10	0.44	-0.23	

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

Table 7.1.37b SLL-Based DIF Statistics for OR Items: Primary Writing (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
12173_T	1	13	-0.03	0.75		0.00	0.83	
12173_V	1	13	-0.06	0.07		-0.07	0.04	
11050_T	2	17	0.03	0.65		-0.03	0.17	
11050_V	2	17	-0.04	0.33		-0.02	0.15	
17369_T	2	24	0.02	0.52		0.01	0.01	
11369_V	2	24	-0.09	0.00		-0.02	0.54	

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

Table 7.1.38a SLL-Based DIF Statistics for MC Items: Junior Writing (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
10707	2	13	0.85	1.19	0.51		0.96	1.29	0.62	
10716	2	14	1.03	1.54	0.52	B-	1.35	1.86	0.84	B-
10699	2	15	0.37	0.77	-0.02		0.29	0.69	-0.11	
10694	2	16	-0.31	0.02	-0.64		-0.08	0.24	-0.41	
10721	2	25	0.21	0.60	-0.19		0.26	0.66	-0.14	
10714	2	26	0.03	0.38	-0.31		-0.18	0.16	-0.53	
10705	2	27	0.00	0.39	-0.38		-0.36	0.03	-0.74	
10723	2	28	0.00	0.33	-0.32		0.32	0.65	0.00	

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

Table 7.1.38b SLL-Based DIF Statistics for OR Items: Junior Writing (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10703_T	1	13	0.00	0.01		-0.02	0.01	
10703_V	1	13	-0.03	0.31		-0.03	0.05	
10708_T	2	17	-0.01	0.00		-0.01	0.00	
10708_V	2	17	-0.06	0.00		-0.02	0.02	
17733_T	2	24	0.01	0.05		-0.03	0.04	
17733_V	2	24	-0.02	0.19		-0.04	0.01	

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

Table 7.1.39a SLL-Based DIF Statistics for MC Items: Primary Writing (French)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
13015	2	13	0.03	0.49	-0.43		-0.10	0.37	-0.56	
11255	2	14	0.47	1.05	-0.11		-0.06	0.50	-0.61	
11280	2	15	-0.31	0.23	-0.85		-0.47	0.07	-1.01	
11275	2	16	0.79	1.38	0.20		0.37	0.95	-0.22	
11279	2	25	0.50	1.06	-0.06		0.48	1.05	-0.09	
13016	2	26	0.61	1.09	0.13		0.52	0.99	0.04	
11258	2	27	0.08	0.60	-0.45		0.20	0.74	-0.33	
11282	2	28	0.78	1.38	0.18		0.02	0.59	-0.56	

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

Table 7.1.39b SLL-Based DIF Statistics for OR Items: Primary Writing (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11242_T	1	13	0.06	0.37		0.05	0.63	
11242_V	1	13	-0.06	0.35		-0.02	0.26	
11241_T	2	17	0.01	0.87		-0.02	0.33	
11241_V	2	17	-0.01	0.82		0.02	0.96	
17714_T	2	24	-0.06	0.31		-0.01	0.94	
17714_V	2	24	-0.11	0.00		-0.09	0.06	

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

Table 7.1.40a SLL-Based DIF Statistics for MC Items: Junior Writing (French)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11553	2	13	0.24	1.02	-0.53		0.22	1.04	-0.59	
11572	2	14	-0.41	0.13	-0.95		-0.02	0.51	-0.56	
12538	2	15	-0.48	0.02	-0.98		-0.47	0.04	-0.98	
11554	2	16	0.90	1.42	0.38		0.43	0.95	-0.09	
11561	2	25	-0.19	0.35	-0.72		-0.13	0.42	-0.68	
11569	2	26	0.23	0.77	-0.32		-0.45	0.11	-1.00	
11557	2	27	-0.44	0.35	-1.22		-0.33	0.49	-1.15	
13030	2	28	0.07	0.63	-0.49		0.34	0.93	-0.25	

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

Table 7.1.40b SLL-Based DIF Statistics for OR Items: Junior Writing (French)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11596_T	1	13	0.09	0.06		0.04	0.49	
11596_V	1	13	0.07	0.23		0.09	0.01	
11589_T	2	17	-0.06	0.47		-0.08	0.31	
11589_V	2	17	-0.10	0.02		-0.06	0.30	
18331_T	2	24	0.00	0.08		0.04	0.46	
18331_V	2	24	-0.03	0.79		0.04	0.48	

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

Table 7.1.41a Gender-Based DIF Statistics for MC Items: Primary Mathematics (English)

Item Code	Sequence	Sample 1				Sample 2			
		Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
EQME31605	1	-0.47	-0.84	-0.10		-0.29	-0.66	0.08	
10691	2	0.03	-0.30	0.36		0.06	-0.27	0.39	
EQME10049	3	0.16	-0.20	0.52		0.29	-0.08	0.66	
10814	4	0.14	-0.21	0.48		0.12	-0.23	0.47	
10689	5	0.58	0.20	0.95		0.64	0.26	1.01	
13153	6	0.55	0.17	0.92		0.42	0.05	0.80	
EQME10124	11	1.01	0.67	1.34	B+	0.70	0.37	1.03	
13164	12	-0.16	-0.48	0.15		0.00	-0.32	0.31	
10957	13	-0.25	-0.62	0.11		-0.44	-0.81	-0.08	
EQME31590	14	0.31	-0.08	0.70		0.25	-0.15	0.64	
10729	15	-0.61	-0.93	-0.29		-0.17	-0.49	0.14	
EQME20830	16	0.27	-0.06	0.59		0.13	-0.20	0.45	
10782	17	-0.20	-0.51	0.12		-0.01	-0.32	0.31	
EQME31603	18	-0.60	-1.03	-0.18		-0.54	-0.96	-0.11	
EQME10028	19	0.77	0.25	1.29		1.13	0.58	1.68	B+
EQME10003	20	0.38	0.05	0.71		0.24	-0.09	0.57	
EQME10016	21	0.36	-0.05	0.77		0.17	-0.23	0.58	
EQME20875	22	-0.12	-0.50	0.27		0.19	-0.19	0.57	
EQME31501	23	-0.35	-0.72	0.02		-0.49	-0.87	-0.11	
EQME20831	24	0.08	-0.26	0.42		0.01	-0.34	0.35	
EQME10100	25	-0.50	-0.92	-0.08		-0.62	-1.05	-0.19	
EQME10116	30	0.00	-0.34	0.35		-0.07	-0.42	0.27	
10887	31	1.05	0.70	1.39	B+	1.49	1.14	1.83	B+
EQME31552	32	0.55	0.21	0.88		0.64	0.31	0.98	
13136	33	0.22	-0.13	0.57		0.37	0.02	0.72	
10850	34	0.45	0.07	0.84		0.88	0.49	1.26	
13092	35	0.04	-0.28	0.37		-0.01	-0.33	0.31	
13091	36	-0.72	-1.04	-0.40		-0.37	-0.70	-0.05	

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

Table 7.1.41b Gender-Based DIF Statistics for OR Items: Primary Mathematics (English)

Item Code	Sequence	Sample 1			Sample 2		
		Effect Size	p -value	DIF Level	Effect Size	p -value	DIF Level
13132	7	0.12	0.00		0.09	0.00	
EQME10064	8	-0.09	0.04		-0.11	0.00	
EQME10132	9	0.01	0.35		0.01	0.00	
10845	10	0.04	0.10		0.05	0.00	
10963	26	-0.11	0.00		-0.10	0.00	
10791	27	-0.10	0.00		-0.12	0.00	
10964	28	-0.04	0.04		0.06	0.03	
10848	29	0.02	0.44		-0.02	0.58	

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

Table 7.1.42a Gender-Based DIF Statistics for MC Items: Junior Mathematics (English)

Item Code	Sequence	Sample 1				Sample 2			
		Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
EQME21007	1	0.79	0.35	1.23		0.80	0.34	1.25	
11374	2	-0.09	-0.42	0.25		0.03	-0.31	0.38	
11412	3	1.02	0.65	1.40	B+	0.98	0.60	1.35	
11304	4	-0.08	-0.43	0.26		-0.29	-0.63	0.05	
11313	5	-0.11	-0.44	0.22		-0.16	-0.48	0.16	
EQME21026	6	-0.41	-0.83	0.02		-0.84	-1.28	-0.41	
11175	11	0.65	0.32	0.99		0.64	0.30	0.97	
EQME10290	12	-0.53	-0.87	-0.19		-0.47	-0.81	-0.13	
11300	13	-0.48	-0.80	-0.16		-0.40	-0.72	-0.09	
11344	14	-0.30	-0.82	0.21		-0.79	-1.29	-0.29	
13033	15	0.02	-0.33	0.37		0.09	-0.25	0.43	
EQME21097	16	-0.19	-0.51	0.14		-0.08	-0.41	0.25	
11312	17	0.76	0.43	1.09		0.94	0.61	1.27	
11447	18	-0.83	-1.25	-0.41		-0.65	-1.07	-0.22	
3459676	19	-0.35	-0.70	0.00		-0.37	-0.71	-0.02	
EQME10231	20	0.63	0.30	0.96		0.64	0.31	0.98	
EQME21095	21	0.44	0.09	0.79		0.33	-0.02	0.68	
13041	22	0.22	-0.34	0.77		-0.12	-0.66	0.42	
11299	23	0.71	0.34	1.07		0.56	0.19	0.92	
11367	24	1.29	0.95	1.62	B+	1.20	0.87	1.54	B+
EQME10187	25	0.46	0.09	0.83		0.37	0.01	0.73	
11339	30	-0.03	-0.36	0.30		0.15	-0.18	0.48	
11384	31	-0.21	-0.54	0.11		-0.09	-0.41	0.23	
11347	32	0.03	-0.34	0.40		0.07	-0.30	0.44	
11293	33	0.26	-0.14	0.67		0.25	-0.15	0.65	
11338	34	0.15	-0.16	0.47		0.24	-0.08	0.55	
EQME21037	35	-0.30	-0.71	0.11		-0.46	-0.86	-0.06	
11362	36	-0.53	-0.89	-0.18		-0.50	-0.86	-0.15	

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

Table 7.1.42b Gender-Based DIF Statistics for OR Items: Junior Mathematics (English)

Item Code	Sequence	Sample 1			Sample 2		
		Effect Size	p -value	DIF Level	Effect Size	p -value	DIF Level
11393	7	0.09	0.00		0.11	0.00	
11348	8	-0.03	0.00		-0.01	0.00	
11455	9	-0.03	0.12		-0.02	0.23	
11454	10	-0.06	0.07		-0.11	0.00	
11408	26	0.05	0.00		0.06	0.00	
11452	27	-0.09	0.00		-0.06	0.01	
11453	28	-0.07	0.01		-0.06	0.02	
11404	29	0.00	0.00		-0.01	0.11	

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

Table 7.1.43a Gender-Based DIF Statistics for MC Items: Primary Mathematics (French)

Item Code	Sequence	Sample 1				Sample 2			
		Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
13102	1	-0.64	-1.13	-0.16		-0.62	-1.09	-0.14	
10993	2	0.94	0.58	1.29		0.77	0.42	1.12	
12651	3	0.20	-0.13	0.53		0.24	-0.08	0.57	
13281	4	1.38	0.96	1.79	B+	1.34	0.92	1.76	B+
13280	5	0.77	0.37	1.17		0.74	0.35	1.14	
11141	6	0.00	-0.52	0.53		-0.17	-0.70	0.35	
13252	11	0.66	0.25	1.06		0.79	0.38	1.20	
11001	12	-0.31	-0.66	0.04		-0.49	-0.84	-0.14	
13342	13	-0.17	-0.48	0.14		-0.09	-0.40	0.22	
10995	14	0.18	-0.14	0.50		0.29	-0.04	0.61	
13239	15	0.37	0.04	0.70		0.64	0.31	0.98	
13251	16	0.64	0.27	1.02		0.92	0.54	1.29	
13254	17	0.26	-0.06	0.58		0.20	-0.12	0.52	
11123	18	0.58	0.25	0.91		0.59	0.26	0.92	
13311	19	-1.01	-1.42	-0.59	B-	-1.17	-1.59	-0.75	B-
13103	20	0.06	-0.28	0.40		0.19	-0.14	0.53	
12606	21	-0.40	-0.72	-0.08		-0.28	-0.60	0.04	
13257	22	1.57	1.03	2.11	C+	1.16	0.64	1.68	B+
11032	23	0.60	0.21	1.00		0.62	0.22	1.01	
12639	24	-0.62	-1.06	-0.17		-0.80	-1.25	-0.35	
OQMF31545	25	-0.43	-0.89	0.03		-0.18	-0.64	0.27	
12642	30	0.76	0.41	1.10		0.65	0.31	0.98	
10994	31	1.09	0.68	1.49	B+	1.09	0.68	1.50	B+
11272	32	0.12	-0.20	0.43		0.27	-0.04	0.59	
11112	33	0.23	-0.09	0.54		0.05	-0.27	0.37	
12647	34	0.68	0.31	1.05		0.35	-0.02	0.72	
10988	35	-0.18	-0.55	0.18		0.13	-0.24	0.50	
12641	36	-0.27	-0.60	0.06		-0.23	-0.55	0.10	

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

Table 7.1.43b Gender-Based DIF Statistics for OR Items: Primary Mathematics (French)

Item Code	Sequence	Sample 1			Sample 2		
		Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
13105	7	-0.01	0.13		-0.03	0.05	
13095	8	-0.15	0.00		-0.17	0.00	B-
13313	9	0.03	0.03		0.05	0.00	
13265	10	0.02	0.33		0.03	0.19	
OQMF31569	26	-0.04	0.20		-0.04	0.26	
13219	27	-0.05	0.11		-0.01	0.04	
13341	28	-0.12	0.00		-0.17	0.00	
13326	29	-0.13	0.00		-0.13	0.00	

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

Table 7.1.44a Gender-Based DIF Statistics for MC Items: Junior Mathematics (French)

Item Code	Sequence	Sample 1				Sample 2			
		Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
11583	1	-0.11	-0.56	0.33		-0.05	-0.50	0.40	
13344	2	-0.62	-1.06	-0.17		-0.54	-0.98	-0.11	
13021	3	-0.19	-0.53	0.15		-0.17	-0.51	0.17	
11573	4	0.05	-0.36	0.46		-0.10	-0.51	0.32	
13352	5	-0.47	-0.79	-0.15		-0.66	-0.98	-0.34	
13355	6	0.56	0.21	0.91		0.55	0.20	0.89	
13348	11	-0.07	-0.39	0.24		-0.03	-0.35	0.29	
13322	12	1.22	0.84	1.59	B+	1.17	0.80	1.54	B+
13361	13	-0.11	-0.46	0.24		-0.37	-0.73	-0.02	
11530	14	0.50	0.18	0.81		0.45	0.14	0.77	
11604	15	-0.38	-0.77	0.01		-0.22	-0.61	0.17	
11576	16	-0.19	-0.52	0.14		-0.35	-0.68	-0.02	
13353	17	-0.03	-0.36	0.30		-0.06	-0.40	0.27	
13316	18	0.75	0.37	1.12		0.84	0.46	1.21	
OQMF20842	19	0.02	-0.33	0.37		0.32	-0.03	0.68	
13354	20	-0.14	-0.68	0.40		0.00	-0.54	0.54	
11568	21	-0.43	-0.77	-0.09		-0.45	-0.79	-0.11	
11649	22	1.07	0.74	1.41	B+	0.97	0.64	1.30	
11544	23	0.58	0.22	0.93		0.87	0.51	1.24	
11525	24	-0.38	-0.70	-0.05		-0.35	-0.68	-0.03	
13333	25	0.72	0.39	1.04		0.88	0.55	1.20	
13359	30	-0.41	-0.77	-0.04		-0.33	-0.69	0.04	
13335	31	-0.16	-0.49	0.17		-0.19	-0.52	0.14	
13358	32	-0.21	-0.52	0.11		-0.34	-0.66	-0.03	
13357	33	0.18	-0.15	0.52		0.14	-0.20	0.48	
11482	34	-0.54	-0.88	-0.20		-0.69	-1.03	-0.36	
11486	35	0.26	-0.07	0.60		0.12	-0.22	0.46	
13312	36	0.26	-0.13	0.64		0.30	-0.09	0.68	

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

Table 7.1.44b Gender-Based DIF Statistics for OR Items: Junior Mathematics (French)

Item Code	Sequence	Sample 1			Sample 2		
		Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11660	7	0.01	0.01		-0.02	0.10	
13328	8	-0.04	0.03		-0.04	0.00	
13362	9	0.04	0.29		0.08	0.00	
13319	10	-0.06	0.01		-0.08	0.02	
11511	26	-0.03	0.46		0.01	0.34	
11642	27	-0.04	0.00		-0.06	0.00	
12574	28	-0.04	0.01		-0.04	0.00	
13364	29	0.05	0.02		0.05	0.00	

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

Table 7.1.45a SLL-Based DIF Statistics for MC Items: Primary Mathematics (English)

Item Code	Sequence	Sample 1				Sample 2			
		Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
EQME31605	1	-0.40	-0.77	-0.03		-0.20	-0.56	0.16	
10691	2	-0.25	-0.58	0.08		-0.32	-0.65	0.02	
EQME10049	3	0.55	0.18	0.92		0.41	0.05	0.77	
10814	4	-0.41	-0.76	-0.06		-0.42	-0.76	-0.07	
10689	5	0.05	-0.33	0.42		0.18	-0.20	0.55	
13153	6	0.66	0.29	1.03		0.27	-0.10	0.63	
EQME10124	11	-0.44	-0.77	-0.11		-0.38	-0.71	-0.05	
13164	12	-0.13	-0.45	0.18		-0.41	-0.73	-0.10	
10957	13	0.47	0.12	0.82		0.47	0.12	0.83	
EQME31590	14	0.26	-0.12	0.64		0.07	-0.31	0.45	
10729	15	-0.02	-0.34	0.30		0.15	-0.17	0.47	
EQME20830	16	0.01	-0.31	0.33		0.45	0.13	0.76	
10782	17	-0.07	-0.38	0.25		0.00	-0.32	0.32	
EQME31603	18	0.33	-0.08	0.75		0.21	-0.21	0.63	
EQME10028	19	-0.25	-0.76	0.26		-0.33	-0.87	0.20	
EQME10003	20	0.20	-0.13	0.53		0.41	0.09	0.74	
EQME10016	21	0.37	-0.03	0.77		0.36	-0.04	0.77	
EQME20875	22	-0.11	-0.49	0.28		-0.13	-0.52	0.26	
EQME31501	23	0.82	0.45	1.19		0.54	0.17	0.90	
EQME20831	24	0.30	-0.04	0.64		0.25	-0.09	0.58	
EQME10100	25	0.06	-0.35	0.47		-0.08	-0.49	0.33	
EQME10116	30	-0.82	-1.17	-0.46		-0.59	-0.94	-0.24	
10887	31	-0.11	-0.46	0.23		0.07	-0.28	0.41	
EQME31552	32	0.26	-0.07	0.59		0.22	-0.11	0.54	
13136	33	0.02	-0.32	0.37		0.10	-0.25	0.45	
10850	34	-0.06	-0.44	0.32		0.16	-0.21	0.54	
13092	35	0.33	0.00	0.66		0.03	-0.29	0.36	
13091	36	-0.24	-0.56	0.09		-0.06	-0.38	0.26	

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

Table 7.1.45b SLL-Based DIF Statistics for OR Items: Primary Mathematics (English)

Item Code	Sequence	Sample 1			Sample 2		
		Effect Size	<i>p</i> -value	DIF Level	Effect Size	<i>p</i> -value	DIF Level
13132	7	0.01	0.85		0.01	0.71	
EQME10064	8	0.00	0.66		-0.03	0.17	
EQME10132	9	0.00	0.94		0.02	0.59	
10845	10	0.07	0.00		0.00	0.00	
10963	26	-0.02	0.01		0.00	0.51	
10791	27	0.00	0.56		-0.04	0.59	
10964	28	-0.01	0.23		0.00	0.83	
10848	29	0.05	0.06		0.01	0.82	

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

Table 7.1.46a SLL-Based DIF Statistics for MC Items: Junior Mathematics (English)

Item Code	Sequence	Sample 1				Sample 2			
		Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
EQME21007	1	0.35	-0.08	0.78		-0.02	-0.45	0.41	
11374	2	0.45	0.11	0.80		0.24	-0.10	0.59	
11412	3	0.16	-0.20	0.52		0.19	-0.17	0.55	
11304	4	0.69	0.34	1.03		0.66	0.32	1.00	
11313	5	0.28	-0.05	0.60		0.31	-0.02	0.64	
EQME21026	6	0.14	-0.28	0.57		-0.16	-0.59	0.27	
11175	11	0.08	-0.25	0.41		0.09	-0.25	0.42	
EQME10290	12	-0.53	-0.87	-0.19		-0.46	-0.81	-0.12	
11300	13	-0.37	-0.69	-0.04		-0.42	-0.74	-0.10	
11344	14	-0.63	-1.14	-0.13		-0.73	-1.25	-0.22	
13033	15	-0.54	-0.88	-0.19		-0.73	-1.07	-0.38	
EQME21097	16	-0.06	-0.39	0.28		0.12	-0.21	0.45	
11312	17	-0.06	-0.39	0.27		0.17	-0.16	0.49	
11447	18	-0.09	-0.51	0.33		0.09	-0.33	0.50	
3459676	19	-0.90	-1.24	-0.55		-0.48	-0.82	-0.13	
EQME10231	20	-0.07	-0.40	0.26		0.05	-0.28	0.39	
EQME21095	21	0.45	0.11	0.79		0.33	-0.01	0.67	
13041	22	-0.96	-1.48	-0.45		-0.33	-0.87	0.21	
11299	23	0.65	0.30	1.01		0.68	0.33	1.03	
11367	24	0.06	-0.27	0.39		0.03	-0.30	0.36	
EQME10187	25	0.30	-0.05	0.66		0.61	0.25	0.97	
11339	30	0.19	-0.14	0.52		0.38	0.05	0.72	
11384	31	-0.78	-1.11	-0.45		-0.87	-1.21	-0.54	
11347	32	1.15	0.80	1.50	B+	1.24	0.89	1.60	B+
11293	33	0.14	-0.25	0.53		0.08	-0.32	0.48	
11338	34	-0.37	-0.68	-0.05		-0.66	-0.98	-0.34	
EQME21037	35	0.02	-0.37	0.41		0.29	-0.11	0.69	
11362	36	-0.33	-0.68	0.02		-0.39	-0.75	-0.04	

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

Table 7.1.46b SLL-Based DIF Statistics for OR Items: Junior Mathematics (English)

Item Code	Sequence	Sample 1			Sample 2		
		Effect Size	<i>p</i> -value	DIF Level	Effect Size	<i>p</i> -value	DIF Level
11393	7	-0.05	0.17		-0.07	0.02	
11348	8	0.10	0.00		0.11	0.00	
11455	9	-0.08	0.00		-0.11	0.00	
11454	10	-0.02	0.03		-0.02	0.11	
11408	26	0.00	0.60		0.02	0.79	
11452	27	0.01	0.04		-0.01	0.47	
11453	28	-0.01	0.81		-0.01	0.78	
11404	29	0.02	0.96		0.01	0.73	

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

Table 7.1.47a SLL-Based DIF Statistics for MC Items: Primary Mathematics (French)

Item Code	Sequence	Sample 1				Sample 2			
		Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
13102	1	0.11	-0.49	0.70		0.48	-0.12	1.09	
10993	2	-0.06	-0.51	0.39		-0.10	-0.55	0.36	
12651	3	-0.50	-0.94	-0.07		-0.43	-0.86	0.01	
13281	4	-0.50	-1.02	0.03		-0.24	-0.78	0.30	
13280	5	0.27	-0.22	0.77		0.32	-0.18	0.83	
11141	6	-0.16	-0.84	0.52		-0.42	-1.09	0.25	
13252	11	-0.67	-1.19	-0.15		-0.26	-0.79	0.26	
11001	12	0.07	-0.38	0.51		-0.02	-0.47	0.43	
13342	13	-0.06	-0.47	0.34		-0.05	-0.45	0.36	
10995	14	-0.41	-0.83	0.01		-0.20	-0.62	0.22	
13239	15	0.09	-0.34	0.51		-0.17	-0.60	0.26	
13251	16	-0.03	-0.51	0.44		0.00	-0.47	0.48	
13254	17	0.13	-0.28	0.55		0.22	-0.20	0.64	
11123	18	-0.68	-1.12	-0.25		-0.49	-0.93	-0.05	
13311	19	-0.25	-0.78	0.28		-0.37	-0.90	0.16	
13103	20	0.03	-0.41	0.47		-0.04	-0.48	0.40	
12606	21	-0.38	-0.80	0.05		-0.41	-0.84	0.01	
13257	22	0.11	-0.56	0.78		-0.01	-0.66	0.64	
11032	23	-0.31	-0.81	0.20		-0.10	-0.63	0.42	
12639	24	0.01	-0.57	0.58		-0.13	-0.69	0.43	
OQMF31545	25	-0.20	-0.77	0.38		-0.19	-0.78	0.40	
12642	30	0.59	0.15	1.03		0.76	0.32	1.21	
10994	31	-0.44	-0.96	0.08		0.15	-0.38	0.68	
11272	32	0.20	-0.21	0.60		0.09	-0.32	0.50	
11112	33	0.11	-0.31	0.54		0.30	-0.13	0.72	
12647	34	-0.15	-0.63	0.32		-0.20	-0.67	0.27	
10988	35	-0.38	-0.86	0.10		-0.34	-0.82	0.14	
12641	36	0.66	0.24	1.08		0.54	0.12	0.97	

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

Table 7.1.47b SLL-Based DIF Statistics for OR Items: Primary Mathematics (French)

Item Code	Sequence	Sample 1			Sample 2		
		Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
13105	7	-0.04	0.29		0.01	0.49	
13095	8	-0.11	0.01		-0.06	0.09	
13313	9	-0.10	0.04		-0.09	0.04	
13265	10	0.02	0.59		0.01	0.09	
OQMF31569	26	0.04	0.61		0.04	0.49	
13219	27	0.03	0.21		0.06	0.00	
13341	28	0.01	0.49		0.01	0.92	
13326	29	-0.02	0.12		-0.02	0.02	

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

Table 7.1.48a SLL-Based DIF Statistics for MC Items: Junior Mathematics (French)

Item Code	Sequence	Sample 1				Sample 2			
		Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
11583	1	-0.20	-0.99	0.59		0.00	-0.80	0.80	
13344	2	0.37	-0.39	1.13		0.23	-0.52	0.98	
13021	3	0.51	-0.10	1.12		0.48	-0.13	1.09	
11573	4	0.07	-0.67	0.82		-0.40	-1.11	0.32	
13352	5	-0.39	-0.96	0.18		-0.26	-0.83	0.32	
13355	6	-0.04	-0.66	0.58		-0.28	-0.89	0.33	
13348	11	-0.06	-0.62	0.50		0.01	-0.55	0.57	
13322	12	0.30	-0.35	0.96		-0.21	-0.87	0.45	
13361	13	-0.34	-0.96	0.28		-0.37	-0.99	0.25	
11530	14	0.52	-0.04	1.08		0.80	0.24	1.36	
11604	15	-0.13	-0.81	0.55		0.24	-0.44	0.93	
11576	16	-0.10	-0.68	0.49		0.08	-0.49	0.65	
13353	17	-0.08	-0.66	0.50		0.02	-0.56	0.61	
13316	18	-0.37	-1.05	0.31		-0.58	-1.26	0.09	
OQMF20842	19	-0.87	-1.50	-0.24		-0.85	-1.47	-0.23	
13354	20	1.12	0.20	2.03	B+	0.51	-0.37	1.39	
11568	21	0.30	-0.30	0.90		0.43	-0.17	1.03	
11649	22	-0.59	-1.18	-0.01		-0.46	-1.03	0.12	
11544	23	-0.13	-0.76	0.50		0.06	-0.58	0.70	
11525	24	-0.28	-0.85	0.30		0.13	-0.44	0.71	
13333	25	-0.32	-0.88	0.24		-0.51	-1.07	0.05	
13359	30	0.08	-0.56	0.72		0.09	-0.56	0.73	
13335	31	-0.05	-0.63	0.53		-0.32	-0.91	0.27	
13358	32	-0.31	-0.88	0.25		0.29	-0.28	0.86	
13357	33	-0.26	-0.87	0.35		0.00	-0.61	0.60	
11482	34	-0.55	-1.15	0.04		-0.31	-0.91	0.29	
11486	35	-0.04	-0.64	0.57		-0.23	-0.82	0.37	
13312	36	-0.46	-1.14	0.22		-0.44	-1.13	0.25	

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

Table 7.1.48b SLL-Based DIF Statistics for OR Items: Junior Mathematics (French)

Item Code	Sequence	Sample 1			Sample 2		
		Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
11660	7	-0.02	0.18		-0.12	0.01	
13328	8	-0.11	0.06		-0.12	0.02	
13362	9	0.04	0.81		0.05	0.58	
13319	10	0.07	0.07		0.07	0.09	
11511	26	-0.01	0.68		0.00	0.49	
11642	27	-0.15	0.00		0.00	0.06	
12574	28	-0.10	0.28		-0.05	0.63	
13364	29	-0.02	0.73		-0.04	0.78	

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

The Grade 9 Assessment of Mathematics
Classical Item Statistics and IRT Item Parameters

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

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
10080	2	NA1.04	AP	N	3	71.13	0.33	0.62	-0.72
10151	4	NA1.06	PS	N	1	35.21	0.33	1.00	1.14
EQME31830	5	NA2.01	KU	N	3	41.50	0.26	0.60	1.12
EQME31750	6	NA2.02	KU	N	4	61.87	0.34	0.71	-0.26
EQME31791	8	NA2.04	AP	N	2	66.84	0.34	0.64	-0.45
10110	10	NA1.05	PS	N	4*	63.11 (2.52)	0.53	0.37	-1.03
EQME31774	13	LR2.01	KU	R	3	81.85	0.30	0.64	-1.43
EQME10532	14	LR2.02	AP	R	4	45.75	0.30	0.65	0.78
EQME10513	16	LR2.03	AP	R	1	50.09	0.22	0.42	0.53
EQME10659	18	LR3.04	KU	R	4	88.96	0.23	0.56	-2.27
10197	20	LR3.05	AP	R	4	60.21	0.31	0.60	-0.06
EQME31778	21	LR4.02	PS	R	1	39.77	0.25	0.61	1.25
EQME31858	22	LR4.04	PS	R	3	69.46	0.35	0.63	-0.58
EQME31859	23	LR4.05	KU	R	3	85.87	0.29	0.69	-1.78
10134	24	LR4.06	AP	R	4	64.36	0.29	0.50	-0.35
EQME31847	28	LR4.02	PS	R	4*	43.95 (1.76)	0.45	0.34	0.34
EQME10587	31	MG2.02	KU	M	2	48.48	0.32	0.68	0.58
EQME10596	32	MG2.03	PS	M	2	50.98	0.38	0.96	0.37
10248	34	MG3.01	KU	M	3	55.15	0.26	0.52	0.29
EQME31761	35	MG3.02	AP	M	3	37.94	0.17	0.53	1.42
10115	36	MG2.03	PS	M	4*	62.30 (2.49)	0.56	0.51	-0.70
EQME21490	0	LR3.04	PS	R	3	46.09	0.38	0.95	0.60
10236	0	NA1.03	KU	N	3	74.47	0.30	0.57	-0.98
10135	0	MG2.03	AP	M	2	62.26	0.50	1.38	-0.15
10158	0	MG2.05	PS	M	2	29.94	0.19	0.77	1.85
10150	0	NA1.04	PS	N	3	47.38	0.30	0.67	0.67
10196	0	LR1.01	AP	R	4	53.22	0.35	0.71	0.31
10252	0	LR3.01	PS	R	4*	60.20 (2.41)	0.49	0.30	-0.96
10212	0	MG3.02	AP	M	4*	46.24 (1.85)	0.53	0.41	0.00
10250	0	NA2.08	AP	N	4*	61.44 (2.46)	0.59	0.42	-0.65
10164	0	LR1.01	AP	R	4*	67.97 (2.72)	0.52	0.35	-1.31

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. KU = knowledge and understanding; AP = application; PS = thinking; N = number sense and algebra; M = measurement and geometry; R = linear relations; NR = not released; () = Mean score for OR items.

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

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
10110	10	% of Students	3.77	1.29	23.07	21.02	16.06	34.79
		Parameters			-3.28	-0.34	0.35	-0.85
10115	36	% of Students	4.43	1.15	18.42	19.25	34.74	22.02
		Parameters			-2.44	-0.62	-0.76	1.00
10164	NR	% of Students	3.94	0.54	25.90	9.46	13.55	46.60
		Parameters			-3.83	1.13	-0.75	-1.80
10212	NR	% of Students	10.82	1.01	34.96	22.12	18.62	12.47
		Parameters			-2.22	0.44	0.50	1.28
10250	NR	% of Students	10.34	2.28	22.61	11.07	13.80	39.90
		Parameters			-1.67	0.56	-0.37	-1.14
10252	NR	% of Students	7.10	1.39	28.41	15.78	8.45	38.88
		Parameters			-3.13	0.79	1.22	-2.70
EQME31847	28	% of Students	10.73	0.99	34.01	27.07	21.17	6.04
		Parameters			-2.43	0.23	0.70	2.87

Note. The total number of students is 16 715. NR = not released.

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

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQME10623	1	NA1.03	KU	N	4	81.32	0.28	-1.43	0.56
10081	3	NA1.04	PS	N	2	59.38	0.33	0.08	0.62
EQME31750	6	NA2.02	KU	N	4	66.93	0.39	-0.26	0.71
14775	7	NA2.02	AP	N	2	43.69	0.18	1.49	0.39
EQME31811	9	NA2.06	KU	N	1	37.40	0.23	1.52	0.61
10141	11	NA2.04	AP	N	4*	76.84 (3.07)	0.48	-1.86	0.41
10085	12	LR1.01	AP	R	2	79.27	0.23	-1.58	0.42
EQME31774	13	LR2.01	KU	R	3	82.88	0.30	-1.43	0.64
14806	15	LR2.02	AP	R	1	49.35	0.20	1.02	0.35
EQME10513	16	LR2.03	AP	R	1	55.67	0.24	0.53	0.42
10153	17	LR3.03	PS	R	4	26.33	0.23	1.88	1.03
14817	19	LR3.05	KU	R	3	67.22	0.27	-0.47	0.48
EQME31858	22	LR4.04	PS	R	3	69.94	0.33	-0.58	0.63
EQME31859	23	LR4.05	KU	R	3	88.59	0.29	-1.78	0.69
14832	25	LR4.06	PS	R	2	65.88	0.32	-0.32	0.59
10251	26	LR2.02	AP	R	4*	73.52 (2.94)	0.37	-2.78	0.28
10165	27	LR3.04	AP	R	4*	50.47 (2.02)	0.48	-0.18	0.34
13008	29	MG1.03	AP	M	4	47.39	0.31	0.76	0.67
EQME31820	30	MG2.01	PS	M	2	20.90	0.17	2.43	0.58
10202	33	MG2.03	PS	M	2	38.05	0.28	1.28	0.75
EQME31761	35	MG3.02	AP	M	3	42.43	0.21	1.42	0.53
10116	37	MG3.02	AP	M	4*	47.18 (1.89)	0.54	-0.05	0.39
EQME10508	0	LR3.01	AP	R	2	50.35	0.34	0.56	0.76
10240	0	NA1.06	AP	N	1	51.19	0.29	0.60	0.56
EQME21446	0	MG3.01	KU	M	3	74.57	0.36	-0.77	0.71
10103	0	LR4.03	AP	R	4	61.29	0.33	-0.03	0.65
10104	0	MG2.05	KU	M	4	66.97	0.43	-0.30	0.91
10127	0	NA1.05	PS	N	4	59.58	0.33	0.06	0.61
10140	0	NA1.04	PS	N	4*	54.00 (2.16)	0.51	-0.83	0.39
14847	0	LR4.04	PS	R	4*	68.67 (2.75)	0.55	-1.19	0.50
10167	0	MG2.02	PS	M	4*	40.15 (1.61)	0.42	0.23	0.37

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. KU = knowledge and understanding; AP = application; PS = thinking; N = number sense and algebra; M = measurement and geometry; R = linear relations; NR = not released; () = Mean score for OR items.

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

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
10116	37	% of Students	13.05	1.56	31.78	25.03	7.39	21.18
		Parameters			-1.73	0.24	2.14	-0.84
10140	NR	% of Students	1.98	0.49	39.23	19.04	18.34	20.93
		Parameters			-4.88	0.83	0.26	0.45
10141	11	% of Students	0.92	0.11	17.37	4.84	26.72	50.03
		Parameters			-5.20	1.11	-2.68	-0.66
10165	27	% of Students	9.99	1.87	21.81	38.50	8.28	19.56
		Parameters			-1.65	-1.14	2.87	-0.80
10167	NR	% of Students	5.91	0.60	43.53	39.45	3.90	6.62
		Parameters			-3.52	0.15	4.14	0.13
10251	26	% of Students	0.31	0.12	5.50	35.80	16.09	42.18
		Parameters			-6.34	-4.54	1.58	-1.80
14847	NR	% of Students	1.82	0.15	12.66	27.57	24.30	33.50
		Parameters			-3.35	-1.54	0.05	0.07

Note. The total number of students is 20 806. NR = not released.

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

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
10214	1	NA1.02	KU	N	4	59.68	0.45	-0.06	1.08
EQME31972	2	NA1.03	AP	N	2	65.59	0.35	-0.39	0.73
10300	4	NA2.03	AP	N	2	67.78	0.31	-0.58	0.55
EQME10392	5	NA2.07	PS	N	4	34.77	0.42	0.91	1.29
10302	7	LR1.01	PS	R	3	55.48	0.26	0.24	0.48
EQME31941	10	LR2.04	KU	R	3	63.48	0.40	-0.22	0.79
EQME31973	11	LR3.02	AP	R	3	72.06	0.27	-0.90	0.50
EQME31971	12	LR3.03	KU	R	3	81.06	0.40	-1.14	0.97
EQME31894	13	LR2.05	AP	R	4*	61.45 (2.46)	0.54	-1.33	0.40
10177	15	AG1.02	KU	G	3	70.45	0.22	-0.97	0.38
EQME10334	17	AG2.01	PS	G	2	41.17	0.33	0.78	0.94
EQME31912	18	AG2.03	AP	G	2	56.16	0.33	0.06	0.66
EQME31995	19	AG3.01	KU	G	1	50.14	0.40	0.40	0.95
EQME31883	20	AG3.02	AP	G	1	54.36	0.39	0.13	0.87
10286	21	AG3.04	PS	G	3	82.19	0.30	-1.51	0.61
10273	22	AG2.01	AP	G	4*	61.51 (2.46)	0.60	-0.70	0.55
EQME31958	25	MG2.02	AP	M	1	54.90	0.37	0.16	0.77
10182	27	MG2.06	PS	M	2	47.47	0.32	0.63	0.65
EQME31948	28	MG3.01	KU	M	1	63.96	0.34	-0.36	0.74
EQME10325	29	MG3.01	AP	M	4	67.97	0.38	-0.53	0.78
10276	30	MG2.03	AP	M	4*	80.10 (3.20)	0.46	-1.84	0.48
EQME21249	0	NA2.06	KU	N	1	65.83	0.27	-0.48	0.49
10312	0	MG2.01	KU	M	4	73.32	0.32	-0.87	0.62
EQME21204	0	MG2.02	PS	M	3	60.68	0.43	-0.12	0.93
EQME10437	0	AG1.03	AP	G	1	82.41	0.34	-1.30	0.79
10283	0	LR2.02	PS	R	2	58.47	0.30	0.02	0.56
10179	0	LR2.03	AP	R	1	69.34	0.34	-0.59	0.67
10289	0	NA2.02	AP	N	4*	71.82 (2.87)	0.46	-1.60	0.34
10297	0	MG3.01	PS	M	4*	66.74 (2.67)	0.56	-1.14	0.40
10295	0	AG3.05	PS	G	4*	71.50 (2.86)	0.52	-1.31	0.34
10186	0	LR3.03	AP	R	4*	41.54 (1.66)	0.51	-0.22	0.40

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. KU = knowledge and understanding; AP = application; PS = thinking; G = analytic geometry; M = measurement and geometry; N = number sense and algebra; R = linear relations; NR = not released; () = Mean score for OR items.

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

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
10186	NR	% of Students	5.21	2.67	59.98	8.37	5.66	18.11
		Parameters			-3.58	2.77	0.89	-0.97
10273	22	% of Students	4.64	0.69	16.92	26.66	28.60	22.49
		Parameters			-2.32	-1.08	-0.15	0.74
10276	30	% of Students	1.03	0.10	5.61	8.05	42.12	43.08
		Parameters			-3.43	-1.44	-2.54	0.05
10289	NR	% of Students	2.11	0.59	15.00	13.30	30.30	38.70
		Parameters			-4.03	-0.43	-1.69	-0.25
10295	NR	% of Students	6.30	0.41	18.96	9.55	11.16	53.62
		Parameters			-2.75	0.59	-0.51	-2.57
10297	NR	% of Students	4.92	0.84	23.58	16.32	6.63	47.71
		Parameters			-3.03	-0.02	1.16	-2.66
EQME31894	13	% of Students	1.33	0.30	30.85	16.65	21.84	29.04
		Parameters			-5.22	0.39	-0.46	-0.02

Note. The total number of students is 42 294. NR = not released.

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

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
EQME31972	2	NA1.03	AP	N	2	68.42	0.37	-0.39	0.73
EQME21271	3	NA2.01	KU	N	1	82.11	0.20	-2.02	0.39
EQME10392	5	NA2.07	PS	N	4	38.57	0.43	0.91	1.29
10316	6	NA2.08	AP	N	4*	73.64 (2.95)	0.50	-1.55	0.39
10219	8	LR2.01	PS	R	3	67.99	0.34	-0.40	0.66
EQME10381	9	LR2.03	AP	R	4	87.80	0.21	-2.40	0.46
EQME31941	10	LR2.04	KU	R	3	64.82	0.38	-0.22	0.79
EQME31973	11	LR3.02	AP	R	3	73.53	0.28	-0.90	0.50
EQME31971	12	LR3.03	KU	R	3	84.01	0.39	-1.14	0.97
10321	14	LR3.01	PS	R	4*	48.32 (1.93)	0.49	-0.35	0.33
EQME21221	16	AG1.03	AP	G	2	70.77	0.37	-0.50	0.77
EQME10334	17	AG2.01	PS	G	2	45.47	0.36	0.78	0.94
EQME31912	18	AG2.03	AP	G	2	61.13	0.33	0.06	0.66
EQME31995	19	AG3.01	KU	G	1	51.61	0.39	0.40	0.95
EQME31883	20	AG3.02	AP	G	1	58.88	0.40	0.13	0.87
10231	23	AG3.04	PS	G	4*	64.41 (2.58)	0.49	-1.06	0.33
10311	24	MG1.01	KU	M	4	67.11	0.41	-0.29	0.89
EQME31958	25	MG2.02	AP	M	1	57.85	0.37	0.16	0.77
EQME10471	26	MG2.03	PS	M	2	73.01	0.37	-0.64	0.75
EQME31948	28	MG3.01	KU	M	1	69.28	0.37	-0.36	0.74
EQME10325	29	MG3.01	AP	M	4	71.94	0.38	-0.53	0.78
EQME31917	31	MG3.01	PS	M	4*	77.27 (3.09)	0.49	-1.46	0.35
10180	0	AG3.03	PS	G	1	58.19	0.44	0.12	1.07
10278	0	NA1.01	KU	N	4	81.82	0.23	-1.74	0.45
EQME21223	0	LR1.04	PS	R	3	65.11	0.27	-0.31	0.47
EQME10477	0	NA2.02	AP	N	4	79.11	0.34	-1.03	0.72
10313	0	MG2.06	PS	M	4	66.36	0.42	-0.26	0.90
EQME10321	0	AG1.02	KU	G	3	86.90	0.22	-2.25	0.47
10322	0	AG2.04	AP	G	4*	52.64 (2.11)	0.58	-0.29	0.56
10232	0	MG2.06	AP	M	4*	89.24 (3.57)	0.37	-2.46	0.35
10291	0	LR1.04	AP	R	4*	66.64 (2.67)	0.49	-1.16	0.33

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. KU = knowledge and understanding; AP = application; PS = thinking; G = analytic geometry; M = measurement and geometry; N = number sense and algebra; R = linear relations; NR = not released; () = Mean score for OR items.

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

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
10231	23	% of Students	4.26	0.59	18.41	25.91	15.91	34.93
		Parameters			-3.14	-1.00	0.87	-0.96
10232	NR	% of Students	0.90	0.16	3.25	7.42	14.19	74.07
		Parameters			-3.20	-2.29	-1.56	-2.77
10291	NR	% of Students	3.68	0.71	19.14	17.80	22.91	35.77
		Parameters			-3.44	-0.30	-0.48	-0.41
10316	6	% of Students	1.26	0.69	12.81	17.81	23.56	43.87
		Parameters			-3.87	-1.11	-0.60	-0.63
10321	14	% of Students	6.19	2.50	45.05	13.57	9.68	23.01
		Parameters			-3.42	1.99	0.87	-0.84
10322	NR	% of Students	3.75	0.47	26.64	35.78	21.11	12.26
		Parameters			-2.81	-0.61	0.83	1.45
EQME31917	31	% of Students	4.29	0.34	9.76	15.30	12.54	57.77
		Parameters			-2.24	-1.37	0.11	-2.33

Note. The total number of students is 54 381. NR = not released.

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

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQMF21301	1	PNA1	CC	N	2	57.86	0.14	0.78	0.36
CMS9575	2	PNA1	MA	N	4	75.25	0.31	-0.86	0.57
OQMF31746	3	PNA2	MA	N	1	60.20	0.36	0.05	0.72
OQMF21318	5	PNA2	CC	N	3	44.15	0.25	1.14	0.56
OQMF31759	6	PNA3	MA	N	4	58.19	0.31	0.37	0.55
OQMF31764	8	PNA1	MA	N	4*	68.06 (2.72)	0.50	-0.97	0.31
CMS9884	11	PRA1	CC	R	4	43.48	0.31	1.10	0.66
OQMF21346	12	PRA1	HP	R	2	58.19	0.35	0.38	0.64
OQMF31778	13	PRA2	CC	R	4	58.86	0.43	0.53	1.18
13475	14	PRA2	MA	R	4	43.48	0.48	0.82	1.31
CMS9746	15	PRA3	CC	R	1	48.83	0.36	0.69	0.93
OQMF21371	16	PRA3	MA	R	3	60.20	0.46	0.11	1.04
OQMF31768	19	PRA3	HP	R	4*	49.75 (1.99)	0.45	0.11	0.29
OQMF21386	20	PMA1	HP	M	2	33.11	0.23	1.60	0.88
CMS9747	21	PMA2	CC	M	3	49.83	0.45	0.56	1.03
OQMF31814	22	PMA2	HP	M	3	43.48	0.31	1.07	0.68
OQMF31824	24	PMA3	MA	M	2	62.88	0.23	-0.12	0.37
OQMF21416	25	PMA3	HP	M	3	43.81	0.24	1.27	0.38
OQMF21427	27	PMA4	MA	M	2	73.24	0.22	-1.56	0.37
OQMF21433	28	PMA4	HP	M	1	53.51	0.33	0.50	0.81
OQMF31838	29	PMA1	MA	M	4*	56.94 (2.28)	0.57	-0.44	0.43
OQMF31752	0	PNA2	CC	N	1	46.82	0.32	0.84	0.77
CMS10021	0	PNA3	HP	N	2	65.55	0.29	-0.24	0.52
OQMF21399	0	PMA2	MA	M	1	57.86	0.37	0.24	0.71
OQMF31770	0	PRA1	MA	R	1	51.51	0.47	0.46	1.28
OQMF31783	0	PRA3	HP	R	4	60.20	0.32	0.12	0.58
OQMF31850	0	PMA4	CC	M	3	67.22	0.30	-0.31	0.56
CMS10037	0	PNA3	HP	N	4*	60.79 (2.43)	0.54	-0.71	0.40
CMS9759	0	PRA2	MA	R	4*	50.59 (2.02)	0.54	-0.42	0.41
OQMF31839	0	PMA2	HP	M	4*	60.62 (2.42)	0.53	-0.73	0.42
OQMF31791	0	PRA1	MA	R	4*	46.40 (1.86)	0.51	-0.14	0.40

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. CC = knowledge and understanding; MA = application; HP = thinking; N = number sense and algebra; M = measurement and geometry; R = linear relations; NR = not released; () = Mean score for OR items.

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

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
CMS10037	NR	% of Students	4.01	0.67	25.75	17.39	26.09	26.09
		Parameters			-3.28	0.27	-0.43	0.62
CMS9759	NR	% of Students	3.68	1.34	43.81	17.39	11.37	22.41
		Parameters			-3.70	1.22	0.97	-0.16
OQMF31764	8	% of Students	6.02	2.01	14.72	19.06	13.38	44.82
		Parameters			-1.89	-0.84	0.70	-1.84
OQMF31768	19	% of Students	14.72	0.33	20.74	30.10	18.39	15.72
		Parameters			-1.11	-0.82	1.34	1.02
OQMF31791	NR	% of Students	4.68	0.67	41.81	28.09	11.37	13.38
		Parameters			-3.57	0.55	1.78	0.67
OQMF31838	29	% of Students	5.35	1.34	27.42	21.07	21.07	23.75
		Parameters			-2.67	0.12	0.24	0.54
OQMF31839	NR	% of Students	3.01	0.67	18.39	33.11	21.40	23.41
		Parameters			-3.10	-1.15	0.78	0.54

Note. The total number of students is 299. NR = not released.

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

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQMF21301	1	PNA1	CC	N	2	49.13	0.21	0.78	0.36
OQMF31746	3	PNA2	MA	N	1	58.36	0.37	0.05	0.72
CMS9974	4	PNA2	CC	N	2	69.20	0.31	-0.61	0.59
OQMF21318	5	PNA2	CC	N	3	42.68	0.24	1.14	0.56
OQMF31759	6	PNA3	MA	N	4	52.60	0.30	0.37	0.55
CMS10023	7	PNA3	HP	N	1	35.52	0.30	1.21	0.92
CMS9995	9	PNA3	HP	N	4*	31.69 (1.27)	0.52	0.71	0.41
13474	10	PRA1	MA	R	1	74.05	0.25	-1.09	0.47
OQMF21346	12	PRA1	HP	R	2	51.56	0.31	0.38	0.64
OQMF31778	13	PRA2	CC	R	4	44.29	0.38	0.53	1.18
13475	14	PRA2	MA	R	4	36.79	0.43	0.82	1.31
OQMF21371	16	PRA3	MA	R	3	56.63	0.41	0.11	1.04
OQMF31784	17	PRA3	HP	R	3	47.64	0.28	0.73	0.55
OQMF31796	18	PRA2	MA	R	4*	32.21 (1.29)	0.39	1.04	0.36
OQMF21386	20	PMA1	HP	M	2	31.72	0.22	1.60	0.88
OQMF31811	23	PMA2	MA	M	2	55.94	0.39	0.14	0.83
OQMF21416	25	PMA3	HP	M	3	46.02	0.19	1.27	0.38
OQMF31847	26	PMA4	CC	M	1	62.17	0.21	-0.23	0.36
OQMF21427	27	PMA4	MA	M	2	78.32	0.21	-1.56	0.37
OQMF21433	28	PMA4	HP	M	1	48.79	0.38	0.50	0.81
OQMF31841	30	PMA3	HP	M	4*	60.99 (2.44)	0.50	-1.12	0.33
CMS10031	0	PRA3	CC	R	2	74.63	0.28	-1.02	0.54
OQMF31807	0	PMA2	CC	M	2	44.52	0.17	1.48	0.34
OQMF31806	0	PMA2	HP	M	3	43.14	0.33	0.88	0.69
CMS9887	0	PRA1	CC	R	2	60.21	0.33	-0.04	0.72
CMS9700	0	PNA1	MA	N	2	31.49	0.19	1.98	0.62
OQMF31826	0	PMA3	MA	M	3	60.21	0.35	-0.04	0.74
CMS9718	0	PRA1	MA	R	4*	59.31 (2.37)	0.55	-0.94	0.57
CMS9626	0	PNA1	MA	N	4*	52.65 (2.11)	0.47	-0.64	0.35
OQMF21395	0	PMA1	MA	M	4*	53.58 (2.14)	0.56	-0.38	0.40

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. CC = knowledge and understanding; MA = application; HP = thinking; N = number sense and algebra; M = measurement and geometry; R = linear relations; NR = not released; () = Mean score for OR items.

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

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
CMS9626	NR	% of Students	2.77	1.96	34.95	22.72	20.18	17.42
		Parameters			-4.10	0.42	0.32	0.80
CMS9718	NR	% of Students	0.58	0.81	13.61	50.40	15.57	19.03
		Parameters			-3.52	-1.91	1.24	0.43
CMS9995	9	% of Students	21.91	5.54	41.75	16.15	5.88	8.77
		Parameters			-1.01	1.45	1.99	0.40
OQMF21378	NR	% of Students	3.00	0.46	13.96	16.49	14.76	51.33
		Parameters			-3.79	-0.93	0.00	-2.41
OQMF21395	NR	% of Students	10.73	1.96	23.76	26.99	9.69	26.87
		Parameters			-1.65	-0.51	1.59	-0.95
OQMF31796	18	% of Students	11.19	2.42	54.90	22.84	6.34	2.31
		Parameters			-2.73	1.51	2.66	2.72
OQMF31841	30	% of Students	3.69	1.15	30.91	17.07	9.80	37.37
		Parameters			-4.08	0.65	0.95	-2.02

Note. The total number of students is 867. NR = not released.

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

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
OQMF31944	2	DNA2	CC	N	1	65.11	0.32	-0.36	0.50
OQMF10395	4	DNA3	CC	N	3	83.50	0.31	-1.58	0.57
OQMF31905	5	DNA4	MA	N	4	53.96	0.33	0.19	0.59
OQMF20949	6	DNA4	HP	N	4	50.26	0.39	0.50	0.94
OQMF20904	7	DNA3	MA	N	4*	59.23 (2.37)	0.51	-0.94	0.42
OQMF31886	10	DRA1	MA	R	4	87.49	0.39	-1.37	1.09
CMS9682	11	DRA2	CC	R	3	45.29	0.32	0.84	0.72
OQMF31868	12	DRA2	MA	R	1	65.71	0.47	-0.33	0.86
OQMF31930	13	DRA3	HP	R	1	38.96	0.45	0.86	1.30
OQMF31949	14	DRA3	MA	R	2	78.52	0.17	-1.85	0.31
CMS9674	15	DRA1	MA	R	4*	42.60 (1.70)	0.54	0.10	0.40
CMS10056	18	DGA1	MA	G	2	45.29	0.31	0.83	0.79
CMS9958	19	DGA1	HP	G	3	46.04	0.41	0.64	0.98
OQMF31932	20	DGA2	HP	G	4	40.54	0.34	1.02	0.94
OQMF31911	21	DGA2	CC	G	3	54.71	0.39	0.39	0.79
CMS9970	22	DGA1	MA	G	4*	62.98 (2.52)	0.62	-0.54	0.55
OQMF21011	24	DMA2	MA	M	4	70.01	0.44	-0.45	0.73
OQMF21018	25	DMA3	CC	M	1	60.29	0.45	-0.04	0.92
OQMF31957	26	DMA3	HP	M	3	48.53	0.33	0.56	0.65
OQMF21024	27	DMA4	CC	M	1	72.19	0.28	-0.86	0.47
OQMF31917	28	DMA4	MA	M	1	44.69	0.29	1.00	0.73
13514	30	DMA4	HP	M	4*	51.85 (2.07)	0.53	-0.27	0.36
CMS9669	0	DMA1	HP	M	3	68.05	0.41	-0.39	0.77
CMS9660	0	DRA1	HP	R	3	68.50	0.49	-0.32	1.19
CMS9657	0	DNA1	MA	N	2	67.14	0.26	-0.49	0.46
OQMF10391	0	DNA3	MA	N	2	73.70	0.33	-0.82	0.61
OQMF31952	0	DMA2	CC	M	2	49.81	0.39	0.52	0.92
CMS10054	0	DGA1	CC	G	1	70.99	0.42	-0.53	0.84
OQMF10361	0	DMA3	HP	M	4*	76.58 (3.06)	0.50	-1.60	0.39
OQMF20917	0	DNA4	MA	N	4*	77.83 (3.11)	0.46	-1.71	0.32
OQMF31900	0	DRA3	HP	R	4*	66.20 (2.65)	0.62	-0.82	0.52

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. CC = knowledge and understanding; MA = application; HP = thinking; G = analytic geometry; M = measurement and geometry; N = number sense and algebra; R = linear relations; () = Mean score for OR items.

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

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
13514	30	% of Students	7.46	3.32	32.03	17.03	19.37	20.80
		Parameters			-2.43	0.82	0.01	0.53
CMS9674	15	% of Students	9.27	5.43	41.90	19.52	6.10	17.78
		Parameters			-2.07	1.08	2.12	-0.72
CMS9970	22	% of Students	6.25	3.77	16.73	15.15	27.51	30.60
		Parameters			-1.88	-0.19	-0.52	0.45
OQMF10361	NR	% of Students	1.28	1.21	10.63	12.58	26.68	47.63
		Parameters			-3.37	-0.98	-1.40	-0.64
OQMF20904	7	% of Students	0.98	0.68	22.53	28.33	32.18	15.30
		Parameters			-4.61	-0.75	-0.08	1.67
OQMF20917	NR	% of Students	2.19	1.28	8.74	15.90	16.80	55.09
		Parameters			-2.75	-1.75	-0.34	-1.98
OQMF31900	NR	% of Students	4.82	1.21	19.52	15.60	21.33	37.53
		Parameters			-2.37	-0.30	-0.42	-0.19

Note. The total number of students is 1 327. NR = not released.

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

Item Code	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
CMS9769	1	DNA1	MA	N	2	62.18	0.28	-0.16	0.50
OQMF31944	2	DNA2	CC	N	1	65.46	0.26	-0.36	0.50
OQMF31865	3	DNA3	MA	N	4	38.87	0.24	1.36	0.60
OQMF10395	4	DNA3	CC	N	3	82.22	0.26	-1.58	0.57
OQMF31905	5	DNA4	MA	N	4	58.08	0.33	0.19	0.59
OQMF20911	8	DNA4	MA	N	4*	47.97 (1.92)	0.55	-0.22	0.37
CMS10049	9	DRA1	HP	R	3	51.19	0.32	0.48	0.68
OQMF31886	10	DRA1	MA	R	4	87.27	0.38	-1.37	1.09
OQMF31868	12	DRA2	MA	R	1	67.72	0.37	-0.33	0.86
OQMF31949	14	DRA3	MA	R	2	76.23	0.18	-1.85	0.31
CMS9969	16	DRA3	HP	R	4*	49.62 (1.98)	0.54	-0.12	0.50
OQMF20982	17	DGA1	CC	G	3	75.07	0.34	-0.87	0.70
CMS9958	19	DGA1	HP	G	3	45.80	0.38	0.64	0.98
OQMF31932	20	DGA2	HP	G	4	37.78	0.36	1.02	0.94
OQMF31911	21	DGA2	CC	G	3	50.70	0.37	0.39	0.79
CMS9970	22	DGA1	MA	G	4*	60.00 (2.40)	0.63	-0.54	0.55
CMS9813	23	DMA1	HP	M	3	39.85	0.47	0.75	1.33
OQMF21011	24	DMA2	MA	M	4	67.46	0.35	-0.45	0.73
OQMF21018	25	DMA3	CC	M	1	60.83	0.42	-0.04	0.92
OQMF31957	26	DMA3	HP	M	3	51.07	0.31	0.56	0.65
OQMF31917	28	DMA4	MA	M	1	41.02	0.29	1.00	0.73
OQMF31874	29	DMA2	HP	M	4*	69.78 (2.79)	0.50	-1.54	0.40
CMS9806	0	DRA3	HP	R	2	56.16	0.31	0.21	0.60
OQMF20950	0	DNA4	HP	N	4	38.49	0.41	0.88	1.16
CMS9947	0	DRA2	CC	R	2	62.30	0.31	-0.16	0.56
OQMF21008	0	DMA2	CC	M	4	71.56	0.27	-0.86	0.48
OQMF10331	0	DGA1	MA	G	3	58.49	0.45	0.04	1.05
CMS9694	0	DMA4	CC	M	3	80.41	0.28	-1.41	0.56
CMS9968	0	DRA1	MA	R	4*	61.75 (2.47)	0.56	-1.03	0.39
CMS10066	0	DMA4	HP	M	4*	55.54 (2.22)	0.64	-0.54	0.54
OQMF20902	0	DNA3	MA	N	4*	60.65 (2.43)	0.52	-0.94	0.48

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. CC = knowledge and understanding; MA = application; HP = thinking; G = analytic geometry; M = measurement and geometry; N = number sense and algebra; R = linear relations; NR = not released; () = Mean score for OR items.

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

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
CMS10066	NR	% of Students	3.80	2.94	37.36	13.45	11.90	30.55
		Parameters			-2.72	0.78	0.25	-0.46
CMS9968	NR	% of Students	2.37	2.00	35.56	8.06	12.73	39.28
		Parameters			-3.95	1.79	-0.70	-1.28
CMS9969	16	% of Students	4.33	0.83	27.80	38.00	21.47	7.57
		Parameters			-2.81	-0.65	0.94	2.05
CMS9970	22	% of Students	6.70	1.47	22.90	16.50	25.61	26.82
		Parameters			-1.88	-0.19	-0.52	0.45
OQMF20902	NR	% of Students	1.51	0.60	14.46	42.03	21.51	19.89
		Parameters			-3.43	-1.83	0.85	0.66
OQMF20911	8	% of Students	9.38	5.01	37.25	16.50	5.84	26.03
		Parameters			-2.08	1.10	1.85	-1.73
OQMF31874	29	% of Students	1.28	0.30	16.16	21.13	23.80	37.33
		Parameters			-4.45	-1.00	-0.34	-0.35

Note. The total number of students is 2 655. NR = not released.

Differential Item Functioning (DIF) Analysis Results

The gender- and SLL-based DIF results for the Grade 9 Assessment of Mathematics are provided in Tables 7.1.65–7.1.76. Only for the English-language test are the results from two random samples of 2000 examinees. For the French-language test, gender-based DIF analysis was conducted for one sample of students, and SLL-based DIF analysis was not conducted. In both cases, DIF analysis was limited by the relatively small population of students who wrote the French-language test. DIF for multiple choice (MC) items and open-response (OR) items are presented in separate tables. Each table for MC DIF items contains the value of Δ , lower limit and upper limit of the confidence band, the category of B- and C-level DIF items. Each table for OR DIF includes the effect size, p-value of significance level, and the category of B- and C- DIF items.

Table 7.1.65a Gender-Based DIF Statistics for MC Items: Grade 9 Applied Mathematics, Winter (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
EQME31774	111xx	1	0.72	0.17	1.28		0.66	0.11	1.21	
EQME31791	111xx	2	-0.83	-1.31	-0.36		-0.86	-1.34	-0.39	
10151	111xx	4	0.43	-0.04	0.91		0.63	0.15	1.10	
10197	111xx	5	-0.18	-0.63	0.26		-0.38	-0.82	0.07	
10248	111xx	6	0.53	0.09	0.96		0.47	0.03	0.90	
EQME10596	111xx	7	0.27	-0.20	0.73		0.28	-0.17	0.74	
EQME21490	111xx	12	0.77	0.31	1.23		1.23	0.77	1.70	B+
EQME10587	111xx	14	0.11	-0.34	0.55		-0.09	-0.54	0.35	
10236	111xx	15	-0.21	-0.72	0.31		0.12	-0.38	0.63	
EQME31830	111xx	16	-0.57	-1.03	-0.11		-0.07	-0.52	0.37	
10134	111xx	18	0.16	-0.29	0.61		0.25	-0.21	0.71	
EQME31858	111xx	19	0.15	-0.33	0.64		-0.06	-0.53	0.42	
EQME10659	111xx	20	0.31	-0.34	0.97		0.05	-0.61	0.71	
EQME10513	111xx	21	0.10	-0.33	0.54		-0.14	-0.58	0.29	
10135	111xx	22	0.00	-0.50	0.50		-0.02	-0.52	0.48	
10158	111xx	23	0.02	-0.44	0.49		0.06	-0.42	0.54	
10150	111xx	24	0.27	-0.17	0.72		-0.07	-0.51	0.38	
10196	111xx	25	0.12	-0.33	0.58		0.28	-0.18	0.73	
EQME31778	111xx	30	1.28	0.84	1.73		1.63	1.18	2.08	C+
EQME31859	111xx	31	0.17	-0.43	0.78		-0.11	-0.72	0.50	
EQME31750	111xx	32	-0.62	-1.09	-0.16		-0.78	-1.25	-0.32	
EQME31761	111xx	33	0.29	-0.15	0.74		0.02	-0.42	0.47	
10080	111xx	34	0.12	-0.37	0.60		0.61	0.13	1.09	
EQME10532	111xx	35	0.40	-0.04	0.85		0.68	0.24	1.13	

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

Table 7.1.65b Gender-Based DIF Statistics for OR Items: Grade 9 Applied Mathematics, Winter (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10252	111xx	8	0.07	0.02		0.13	0.00	
10110	111xx	9	0.14	0.00		0.17	0.00	
EQME31847	111xx	10	0.02	0.65		-0.03	0.46	
10212	111xx	11	-0.01	0.98		-0.06	0.06	
10115	111xx	26	-0.10	0.01		-0.05	0.03	
10250	111xx	28	-0.13	0.00		-0.13	0.00	
10164	111xx	29	0.01	0.62		-0.03	0.10	

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

Table 7.1.66a Gender-Based DIF Statistics for MC Items: Grade 9 Applied Mathematics, Spring (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
EQME31774	112xx	1	1.23	0.64	1.82	B+	1.03	0.44	1.63	B+
EQME31811	112xx	2	0.04	-0.42	0.49		-0.73	-1.17	-0.28	

10153	112xx	4	1.03	0.53	1.53	B+	1.10	0.59	1.61	B+
EQME10508	112xx	5	1.22	0.77	1.68		0.95	0.51	1.40	
EQME10623	112xx	6	1.46	0.91	2.01		0.63	0.07	1.19	
EQME31820	112xx	7	0.16	-0.38	0.69		0.19	-0.33	0.72	
10240	112xx	12	1.38	0.93	1.82		0.79	0.35	1.24	
EQME21446	112xx	14	0.63	0.12	1.15		0.25	-0.26	0.77	
14832	112xx	15	0.33	-0.14	0.80		0.33	-0.14	0.80	
10103	112xx	16	0.51	0.05	0.98		0.20	-0.26	0.66	
10202	112xx	18	0.96	0.50	1.41		0.87	0.42	1.32	
EQME31858	112xx	19	0.98	0.49	1.47		0.60	0.12	1.09	
13008	112xx	20	-0.38	-0.83	0.07		0.00	-0.45	0.45	
EQME10513	112xx	21	0.08	-0.36	0.52		-0.28	-0.72	0.16	
14817	112xx	22	0.74	0.28	1.21		0.54	0.08	1.00	
10104	112xx	23	-1.63	-2.14	-1.12	C-	-1.17	-1.67	-0.67	B-
10127	112xx	24	0.99	0.53	1.46		0.71	0.25	1.16	
10081	112xx	25	0.21	-0.25	0.66		0.30	-0.16	0.75	
14775	112xx	30	0.18	-0.25	0.61		0.53	0.09	0.96	
EQME31859	112xx	31	0.24	-0.42	0.91		-0.24	-0.91	0.44	
EQME31750	112xx	32	-0.64	-1.14	-0.15		-0.59	-1.08	-0.10	
EQME31761	112xx	33	0.42	-0.01	0.85		0.39	-0.05	0.84	
10085	112xx	34	0.06	-0.47	0.59		-0.01	-0.54	0.52	
14806	112xx	35	-0.16	-0.59	0.27		0.34	-0.09	0.77	

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

Table 7.1.66b Gender-Based DIF Statistics for OR Items: Grade 9 Applied Mathematics, Spring (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10140	112xx	8	-0.11	0.00		-0.05	0.00	
10251	112xx	9	-0.09	0.01		-0.11	0.02	
14847	112xx	10	-0.14	0.00		-0.19	0.00	
10116	112xx	11	0.10	0.00		0.14	0.00	
10141	112xx	26	-0.10	0.00		-0.14	0.00	
10165	112xx	28	-0.00	0.91		-0.05	0.42	
10167	112xx	29	-0.21	0.00	B-	-0.22	0.00	B-

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

Table 7.1.67a Gender-Based DIF Statistics for MC Items: Grade 9 Academic Mathematics, Winter (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
10214	121xx	1	0.63	0.15	1.11		1.00	0.51	1.49	
EQME21249	121xx	2	-0.28	-0.74	0.19		-0.45	-0.90	0.01	
10312	121xx	4	0.44	-0.06	0.94		0.33	-0.16	0.82	
EQME21204	121xx	5	-0.10	-0.58	0.38		-0.46	-0.94	0.03	
EQME31971	121xx	6	0.80	0.20	1.40		0.31	-0.27	0.90	
EQME10437	121xx	7	-1.02	-1.62	-0.41		-0.74	-1.32	-0.16	
10283	121xx	12	-0.01	-0.47	0.44		0.10	-0.35	0.55	
10302	121xx	14	0.18	-0.27	0.62		0.39	-0.04	0.83	

EQME31912	121xx	15	1.08	0.63	1.53		0.67	0.22	1.11	
EQME31948	121xx	16	-0.08	-0.54	0.39		0.12	-0.34	0.59	
EQME31972	121xx	18	-0.23	-0.69	0.23		-0.56	-1.03	-0.09	
10179	121xx	19	-0.14	-0.63	0.35		-0.12	-0.60	0.37	
EQME10392	121xx	20	0.05	-0.45	0.55		0.16	-0.34	0.66	
10182	121xx	21	0.13	-0.31	0.57		-0.41	-0.86	0.03	
EQME31941	121xx	22	0.79	0.30	1.28		0.64	0.17	1.11	
EQME10334	121xx	23	1.25	0.79	1.70	B+	1.18	0.71	1.66	B+
10177	121xx	24	-0.43	-0.91	0.05		-0.08	-0.56	0.39	
EQME31883	121xx	25	-0.80	-1.26	-0.33		-0.53	-1.00	-0.07	
EQME31973	121xx	30	1.08	0.58	1.58	B+	1.05	0.56	1.55	B+
10300	121xx	31	-0.13	-0.60	0.33		-0.48	-0.96	0.00	
EQME10325	121xx	32	0.10	-0.40	0.59		0.66	0.18	1.14	
10286	121xx	33	1.13	0.54	1.72		0.63	0.05	1.21	
EQME31995	121xx	34	-0.39	-0.85	0.07		0.01	-0.46	0.47	
EQME31958	121xx	35	-0.14	-0.60	0.31		0.02	-0.44	0.49	

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

Table 7.1.67b Gender-Based DIF Statistics for OR Items: Grade 9 Academic Mathematics, Winter (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
EQME31894	121xx	8	-0.04	0.78		-0.14	0.00	
10289	121xx	9	0.06	0.00		0.09	0.00	
10297	121xx	10	0.02	0.22		0.07	0.00	
10295	121xx	11	0.03	0.72		0.03	0.17	
10186	121xx	26	-0.04	0.70		0.03	0.27	
10273	121xx	28	-0.08	0.00		-0.16	0.00	
10276	121xx	29	-0.06	0.07		-0.04	0.01	

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

Table 7.1.68a Gender-Based DIF Statistics for MC Items: Grade 9 Academic Mathematics, Spring (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
EQME10381	122xx	1	0.43	-0.21	1.07		-0.07	-0.72	0.57	
10180	122xx	2	0.73	0.24	1.22		0.34	-0.14	0.82	
10278	122xx	4	-0.23	-0.79	0.32		-0.09	-0.66	0.47	
EQME10471	122xx	5	0.22	-0.30	0.73		0.14	-0.38	0.66	
EQME31971	122xx	6	0.20	-0.43	0.82		0.63	-0.02	1.27	
EQME21221	122xx	7	-0.53	-1.02	-0.04		-0.18	-0.68	0.32	
EQME21223	122xx	12	0.28	-0.18	0.73		0.82	0.36	1.28	
EQME10477	122xx	14	1.19	0.63	1.75		0.74	0.19	1.29	
EQME31912	122xx	15	0.98	0.52	1.45		1.11	0.64	1.57	
EQME31948	122xx	16	-0.24	-0.73	0.25		-0.14	-0.64	0.35	
EQME31972	122xx	18	-0.43	-0.91	0.05		-0.62	-1.11	-0.12	
EQME21271	122xx	19	-0.42	-0.95	0.12		-0.73	-1.31	-0.16	
EQME10392	122xx	20	0.21	-0.29	0.70		0.11	-0.37	0.59	
10311	122xx	21	0.27	-0.22	0.76		0.15	-0.35	0.64	
EQME31941	122xx	22	0.46	-0.01	0.93		1.13	0.63	1.62	
EQME10334	122xx	23	1.11	0.65	1.57	B+	1.21	0.74	1.67	B+
10219	122xx	24	0.33	-0.15	0.81		0.57	0.10	1.05	
EQME31883	122xx	25	-0.63	-1.10	-0.16		-0.22	-0.68	0.25	
EQME31973	122xx	30	1.11	0.61	1.61		0.68	0.19	1.18	
10313	122xx	31	0.32	-0.18	0.81		0.24	-0.26	0.75	
EQME10325	122xx	32	0.62	0.11	1.14		0.63	0.13	1.14	
EQME10321	122xx	33	-0.43	-1.07	0.20		-0.46	-1.10	0.18	
EQME31995	122xx	34	0.05	-0.41	0.51		0.10	-0.37	0.57	
EQME31958	122xx	35	0.30	-0.17	0.76		0.37	-0.09	0.83	

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

Table 7.1.68b Gender-Based DIF Statistics for OR Items: Grade 9 Academic Mathematics, Spring (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10316	122xx	8	-0.10	0.00		-0.10	0.00	
10322	122xx	9	-0.22	0.00		-0.15	0.00	
10231	122xx	10	0.00	0.04		-0.08	0.00	
EQME31917	122xx	11	-0.01	0.93		-0.01	0.79	
10232	122xx	26	-0.10	0.00		-0.07	0.09	
10321	122xx	28	0.07	0.09		0.12	0.00	
10291	122xx	29	-0.06	0.16		-0.14	0.00	

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

Table 7.1.69a Gender-Based DIF Statistics for MC Items: Grade 9 Applied Mathematics, Winter (French)

Item Code	Booklet	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
CMS9575	211xx	1	0.33	-0.87	1.53	
OQMF31752	211xx	2	-0.78	-1.88	0.33	

OQMF31778	211xx	4	-0.75	-1.87	0.37	
OQMF31759	211xx	5	0.22	-0.87	1.30	
OQMF21416	211xx	6	-0.48	-1.54	0.58	
CMS10021	211xx	7	1.12	0.01	2.23	B+
13475	211xx	12	0.64	-0.56	1.85	
OQMF21318	211xx	14	-1.07	-2.14	0.00	B-
OQMF21399	211xx	15	1.02	-0.10	2.14	B+
CMS9884	211xx	16	1.04	-0.05	2.13	B+
OQMF21427	211xx	18	0.72	-0.46	1.90	
OQMF31746	211xx	19	1.62	0.49	2.74	C+
OQMF21371	211xx	20	-1.02	-2.21	0.16	B-
OQMF21301	211xx	21	-0.04	-1.09	1.01	
OQMF21346	211xx	22	0.68	-0.43	1.79	
OQMF31770	211xx	23	-0.58	-1.76	0.61	
OQMF21433	211xx	24	-0.33	-1.43	0.77	
OQMF21386	211xx	25	0.43	-0.69	1.56	
OQMF31783	211xx	30	-0.19	-1.27	0.89	
CMS9747	211xx	31	-0.18	-1.35	0.99	
CMS9746	211xx	32	-0.94	-2.06	0.18	
OQMF31824	211xx	33	0.65	-0.41	1.71	
OQMF31850	211xx	34	-0.08	-1.21	1.06	
OQMF31814	211xx	35	0.53	-0.58	1.64	

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

Table 7.1.69b Gender-Based DIF Statistics for OR Items: Grade 9 Applied Mathematics, Winter (French)

Item Code	Booklet	Sequence	Sample 1		
			Effect Size	p-value	DIF Level
CMS10037	211xx	8	-0.06	0.53	
OQMF31838	211xx	9	-0.01	0.74	
CMS9759	211xx	10	-0.02	0.55	
OQMF31768	211xx	11	-0.11	0.17	
OQMF31839	211xx	26	-0.10	0.42	
OQMF31764	211xx	28	0.15	0.03	
OQMF31791	211xx	29	0.06	0.85	

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

Table 7.1.70a Gender-Based DIF Statistics for MC Items: Grade 9 Applied Mathematics, Spring (French)

Item Code	Booklet	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
CMS9974	212xx	1	-0.24	-0.91	0.43	
OQMF31847	212xx	2	-0.19	-0.81	0.43	
OQMF31778	212xx	4	0.52	-0.14	1.18	
OQMF31759	212xx	5	1.51	0.87	2.15	C+
OQMF21416	212xx	6	-0.18	-0.79	0.42	
CMS10031	212xx	7	0.38	-0.32	1.09	
13475	212xx	12	-0.39	-1.10	0.31	

OQMF21318	212xx	14	-0.22	-0.84	0.40	
OQMF31807	212xx	15	-0.03	-0.64	0.57	
CMS10023	212xx	16	0.10	-0.55	0.75	
OQMF21427	212xx	18	0.55	-0.17	1.27	
OQMF31746	212xx	19	0.06	-0.59	0.70	
OQMF21371	212xx	20	-0.52	-1.17	0.13	
OQMF21301	212xx	21	0.81	0.21	1.41	
OQMF21346	212xx	22	-0.64	-1.27	-0.01	
OQMF31806	212xx	23	0.62	-0.03	1.26	
OQMF21433	212xx	24	1.00	0.37	1.63	B+
OQMF21386	212xx	25	0.45	-0.21	1.11	
CMS9887	212xx	30	0.45	-0.19	1.09	
OQMF31784	212xx	31	0.09	-0.53	0.70	
CMS9700	212xx	32	1.09	0.43	1.75	B+
OQMF31826	212xx	33	-0.03	-0.67	0.62	
OQMF31811	212xx	34	0.14	-0.51	0.80	
13474	212xx	35	-1.21	-1.91	-0.52	B-

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

Table 7.1.70b Gender-Based DIF Statistics for OR Items: Grade 9 Applied Mathematics, Spring (French)

Item Code	Booklet	Sequence	Sample 1		
			Effect Size	p-value	DIF Level
CMS9718	212xx	8	-0.09	0.22	
CMS9626	212xx	9	0.19	0.00	B+
OQMF31796	212xx	10	-0.02	0.61	
OQMF21395	212xx	11	-0.08	0.50	
OQMF31841	212xx	26	-0.17	0.00	
CMS9995	212xx	28	-0.02	0.47	
OQMF21378	212xx	29	0.05	0.10	

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

Table 7.1.71a Gender-Based DIF Statistics for MC Items: Grade 9 Academic Mathematics, Winter (French)

Item Code	Booklet	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
OQMF10395	221xx	1	-0.70	-1.44	0.03	
OQMF31868	221xx	2	-0.76	-1.38	-0.13	
OQMF31905	221xx	4	0.63	0.08	1.18	
OQMF31930	221xx	5	0.68	0.08	1.28	
OQMF21018	221xx	6	0.62	0.03	1.21	
CMS9669	221xx	7	0.42	-0.20	1.03	
CMS10056	221xx	12	-0.21	-0.76	0.34	
OQMF31944	221xx	14	0.20	-0.37	0.77	
CMS9958	221xx	15	-0.72	-1.29	-0.15	
CMS9682	221xx	16	0.90	0.35	1.45	
OQMF21011	221xx	18	0.48	-0.16	1.12	
OQMF31949	221xx	19	0.61	-0.03	1.25	
OQMF31911	221xx	20	-0.18	-0.75	0.38	

OQMF21024	221xx	21	0.35	-0.25	0.95	
OQMF20949	221xx	22	0.87	0.30	1.44	
OQMF31932	221xx	23	-0.23	-0.79	0.33	
CMS9660	221xx	24	1.66	0.99	2.33	C+
CMS9657	221xx	25	0.24	-0.33	0.82	
OQMF10391	221xx	30	-1.12	-1.75	-0.50	B-
OQMF31952	221xx	31	0.07	-0.49	0.63	
CMS10054	221xx	32	-0.28	-0.92	0.35	
OQMF31917	221xx	33	0.58	0.03	1.12	
OQMF31886	221xx	34	-1.04	-1.90	-0.18	B-
OQMF31957	221xx	35	0.07	-0.48	0.62	

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

Table 7.1.71b Gender-Based DIF Statistics for OR Items: Grade 9 Academic Mathematics, Winter (French)

Item Code	Booklet	Sequence	Sample 1		
			Effect Size	p-value	DIF Level
OQMF10361	221xx	8	0.03	0.16	
OQMF20904	221xx	9	-0.13	0.00	
CMS9970	221xx	10	-0.12	0.00	
CMS9674	221xx	11	-0.17	0.00	
OQMF20917	221xx	26	0.29	0.00	C+
13514	221xx	28	-0.10	0.00	
OQMF31900	221xx	29	0.06	0.00	

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

Table 7.1.72a Gender-Based DIF Statistics for MC Items: Grade 9 Academic Mathematics, Spring (French)

Item Code	Booklet	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
OQMF10395	222xx	1	-0.63	-1.12	-0.15	
OQMF31868	222xx	2	-0.64	-1.05	-0.22	
OQMF31905	222xx	4	0.18	-0.20	0.57	
CMS9806	222xx	5	1.34	0.95	1.73	B+
OQMF21018	222xx	6	0.55	0.14	0.96	
OQMF20950	222xx	7	1.07	0.66	1.48	B+
CMS10049	222xx	12	1.08	0.70	1.47	B+
OQMF31944	222xx	14	-0.22	-0.61	0.18	
CMS9958	222xx	15	-0.04	-0.44	0.35	
CMS9947	222xx	16	0.54	0.15	0.93	
OQMF21011	222xx	18	-0.11	-0.52	0.30	
OQMF31949	222xx	19	0.52	0.09	0.95	
OQMF31911	222xx	20	-0.51	-0.90	-0.12	
CMS9769	222xx	21	1.07	0.68	1.46	B+
OQMF20982	222xx	22	-1.48	-1.92	-1.03	B-
OQMF31932	222xx	23	-0.37	-0.77	0.03	
OQMF21008	222xx	24	-0.16	-0.58	0.25	
CMS9813	222xx	25	1.50	1.07	1.92	B+
OQMF10331	222xx	30	-0.44	-0.85	-0.03	
OQMF31865	222xx	31	-0.30	-0.68	0.08	
CMS9694	222xx	32	0.61	0.13	1.08	
OQMF31917	222xx	33	0.76	0.38	1.15	
OQMF31886	222xx	34	0.04	-0.55	0.63	
OQMF31957	222xx	35	0.18	-0.20	0.56	

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

Table 7.1.72b Gender-Based DIF Statistics for OR Items: Grade 9 Academic Mathematics, Spring (French)

Item Code	Booklet	Sequence	Sample 1		
			Effect Size	p-value	DIF Level
CMS9968	222xx	8	0.09	0.00	
CMS10066	222xx	9	-0.14	0.00	
CMS9970	222xx	10	-0.01	0.65	
OQMF20911	222xx	11	-0.12	0.00	
OQMF20902	222xx	26	-0.15	0.00	
OQMF31874	222xx	28	-0.06	0.01	
CMS9969	222xx	29	0.14	0.00	

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

Table 7.1.73a SLL-Based DIF Statistics for MC Items: Grade 9 Applied Mathematics, Winter (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
EQME31774	111xx	1	0.91	0.43	1.40		0.84	0.36	1.32	
EQME31791	111xx	2	-0.20	-0.65	0.24		-0.32	-0.77	0.14	

10151	111xx	4	-1.20	-1.66	-0.74		-0.82	-1.28	-0.37	
10197	111xx	5	0.32	-0.10	0.73		0.48	0.06	0.90	
10248	111xx	6	-0.37	-0.78	0.04		-0.60	-1.02	-0.19	
EQME10596	111xx	7	0.26	-0.17	0.69		0.14	-0.29	0.58	
EQME21490	111xx	12	-0.55	-0.98	-0.11		-0.51	-0.95	-0.08	
EQME10587	111xx	14	0.25	-0.17	0.68		0.24	-0.19	0.66	
10236	111xx	15	-1.31	-1.80	-0.82	B-	-1.15	-1.64	-0.65	B-
EQME31830	111xx	16	-1.76	-2.18	-1.34	C-	-1.69	-2.12	-1.27	C-
10134	111xx	18	0.36	-0.05	0.78		0.57	0.14	1.00	
EQME31858	111xx	19	0.65	0.20	1.10		0.50	0.05	0.94	
EQME10659	111xx	20	-0.04	-0.65	0.56		-0.11	-0.71	0.49	
EQME10513	111xx	21	-0.46	-0.86	-0.06		-0.52	-0.92	-0.12	
10135	111xx	22	0.71	0.25	1.17		0.82	0.36	1.27	
10158	111xx	23	-0.20	-0.65	0.26		-0.30	-0.76	0.15	
10150	111xx	24	0.46	0.05	0.88		0.75	0.33	1.17	
10196	111xx	25	-1.12	-1.55	-0.68	B-	-1.48	-1.93	-1.03	B-
EQME31778	111xx	30	0.54	0.11	0.97		0.64	0.21	1.07	
EQME31859	111xx	31	-0.34	-0.89	0.20		-0.49	-1.04	0.06	
EQME31750	111xx	32	-0.54	-0.97	-0.10		-0.39	-0.83	0.05	
EQME31761	111xx	33	-0.16	-0.57	0.26		-0.41	-0.83	0.00	
10080	111xx	34	-0.18	-0.63	0.27		0.03	-0.42	0.47	
EQME10532	111xx	35	-0.65	-1.07	-0.23		-0.65	-1.07	-0.23	

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

Table 7.1.73b SLL-Based DIF Statistics for OR Items: Grade 9 Applied Mathematics, Winter (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10252	111xx	8	0.09	0.02		0.09	0.00	
10110	111xx	9	0.07	0.05		0.06	0.01	
EQME31847	111xx	10	0.20	0.00	B+	0.18	0.00	B+
10212	111xx	11	-0.17	0.00		-0.19	0.00	
10115	111xx	26	0.11	0.00		0.11	0.00	
10250	111xx	28	-0.11	0.00		-0.05	0.06	
10164	111xx	29	0.11	0.00		0.13	0.00	

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

Table 7.1.74a SLL-Based DIF Statistics for MC Items: Grade 9 Applied Mathematics, Spring (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
EQME31774	112xx	1	1.72	1.28	2.16	C+	1.87	1.43	2.31	C+
EQME31811	112xx	2	-3.01	-3.39	-2.62	C-	-3.00	-3.38	-2.61	C-
10153	112xx	4	-0.19	-0.61	0.24		0.21	-0.20	0.62	
EQME10508	112xx	5	0.22	-0.15	0.60		0.32	-0.06	0.69	
EQME10623	112xx	6	1.43	1.01	1.85	B+	1.82	1.40	2.25	C+
EQME31820	112xx	7	-0.66	-1.09	-0.22		-0.31	-0.74	0.12	
10240	112xx	12	0.69	0.32	1.05		0.69	0.32	1.06	
EQME21446	112xx	14	0.01	-0.41	0.42		-0.08	-0.50	0.34	
14832	112xx	15	0.19	-0.20	0.57		-0.01	-0.39	0.37	
10103	112xx	16	-0.46	-0.84	-0.08		-0.19	-0.57	0.19	
10202	112xx	18	0.85	0.47	1.24		0.92	0.54	1.31	
EQME31858	112xx	19	1.10	0.71	1.49	B+	1.06	0.67	1.45	B+
13008	112xx	20	-0.03	-0.40	0.34		-0.19	-0.56	0.18	
EQME10513	112xx	21	-0.58	-0.94	-0.22		-0.60	-0.96	-0.23	
14817	112xx	22	0.24	-0.14	0.61		0.35	-0.02	0.73	
10104	112xx	23	-1.53	-1.97	-1.10	C-	-1.62	-2.05	-1.18	C-
10127	112xx	24	2.05	1.67	2.42	C+	2.03	1.65	2.40	C+
10081	112xx	25	-0.03	-0.40	0.35		0.06	-0.31	0.43	
14775	112xx	30	0.25	-0.11	0.60		0.22	-0.13	0.58	
EQME31859	112xx	31	-0.22	-0.74	0.30		-0.14	-0.66	0.38	
EQME31750	112xx	32	0.03	-0.37	0.42		-0.19	-0.58	0.21	
EQME31761	112xx	33	-0.20	-0.56	0.17		-0.39	-0.75	-0.03	
10085	112xx	34	0.30	-0.12	0.72		0.45	0.03	0.86	
14806	112xx	35	0.32	-0.04	0.68		0.17	-0.19	0.52	

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

Table 7.1.74b SLL-Based DIF Statistics for OR Items: Grade 9 Applied Mathematics, Spring (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10140	112xx	8	0.04	0.45		0.00	0.55	
10251	112xx	9	0.23	0.00		0.17	0.00	
14847	112xx	10	0.08	0.02		0.06	0.02	
10116	112xx	11	-0.14	0.00		-0.08	0.01	
10141	112xx	26	-0.16	0.00		-0.13	0.00	
10165	112xx	28	-0.06	0.26		-0.07	0.22	
10167	112xx	29	-0.13	0.00		-0.12	0.00	

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

Table 7.1.75a SLL-Based DIF Statistics for MC Items: Grade 9 Academic Mathematics, Winter (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
10214	121xx	1	-0.23	-0.61	0.16		0.09	-0.29	0.47	
EQME21249	121xx	2	-1.57	-1.95	-1.18	C-	-1.56	-1.94	-1.18	C-

10312	121xx	4	-0.56	-0.96	-0.17		-0.31	-0.70	0.09	
EQME21204	121xx	5	-0.57	-0.95	-0.20		-0.80	-1.18	-0.42	
EQME31971	121xx	6	-0.12	-0.57	0.33		0.26	-0.20	0.71	
EQME10437	121xx	7	-1.50	-2.01	-1.00	C-	-1.50	-2.01	-0.99	C-
10283	121xx	12	1.16	0.82	1.51	B+	1.08	0.73	1.43	B+
10302	121xx	14	0.29	-0.05	0.64		0.44	0.10	0.78	
EQME31912	121xx	15	0.94	0.59	1.30		1.52	1.17	1.87	
EQME31948	121xx	16	-0.80	-1.18	-0.42		-0.94	-1.32	-0.57	
EQME31972	121xx	18	-1.74	-2.14	-1.34	C-	-1.62	-2.02	-1.22	C-
10179	121xx	19	-0.37	-0.75	0.01		0.09	-0.30	0.48	
EQME10392	121xx	20	-0.68	-1.07	-0.28		-0.66	-1.05	-0.27	
10182	121xx	21	-0.47	-0.82	-0.12		-0.52	-0.87	-0.17	
EQME31941	121xx	22	0.58	0.21	0.95		0.72	0.36	1.09	
EQME10334	121xx	23	0.52	0.15	0.88		0.46	0.09	0.83	
10177	121xx	24	-0.61	-0.98	-0.24		-0.57	-0.95	-0.20	
EQME31883	121xx	25	-1.09	-1.46	-0.72	B-	-1.16	-1.53	-0.79	B-
EQME31973	121xx	30	1.07	0.69	1.44		0.90	0.53	1.27	
10300	121xx	31	-0.20	-0.58	0.19		-0.39	-0.77	-0.01	
EQME10325	121xx	32	-0.31	-0.70	0.07		-0.52	-0.90	-0.14	
10286	121xx	33	1.27	0.85	1.69		0.99	0.58	1.41	
EQME31995	121xx	34	-0.83	-1.20	-0.47		-1.10	-1.47	-0.73	
EQME31958	121xx	35	-1.06	-1.42	-0.69		-0.62	-0.98	-0.26	

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

Table 7.1.75b SLL-Based DIF Statistics for OR Items: Grade 9 Academic Mathematics, Winter (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
EQME31894	121xx	8	0.02	0.17		0.06	0.04	
10289	121xx	9	0.06	0.23		0.05	0.04	
10297	121xx	10	0.03	0.58		0.02	0.31	
10295	121xx	11	0.19	0.00		0.17	0.00	
10186	121xx	26	-0.11	0.00		-0.07	0.03	
10273	121xx	28	0.04	0.02		0.00	0.39	
10276	121xx	29	0.09	0.00		0.13	0.00	

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

Table 7.1.76a SLL-Based DIF Statistics for MC Items: Grade 9 Academic Mathematics, Spring (English)

Item Code	Booklet	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
EQME10381	122xx	1	-0.37	-0.80	0.05		-0.34	-0.76	0.09	
10180	122xx	2	0.11	-0.21	0.43		0.17	-0.15	0.49	
10278	122xx	4	-1.45	-1.84	-1.06	B-	-1.32	-1.72	-0.93	B-
EQME10471	122xx	5	0.56	0.23	0.89		0.58	0.24	0.91	
EQME31971	122xx	6	0.36	-0.03	0.76		0.62	0.22	1.03	
EQME21221	122xx	7	-1.61	-1.96	-1.26	C-	-1.58	-1.93	-1.23	C-
EQME21223	122xx	12	0.68	0.39	0.97		0.71	0.42	1.00	
EQME10477	122xx	14	0.03	-0.33	0.40		-0.10	-0.46	0.26	
EQME31912	122xx	15	1.13	0.84	1.43	B+	1.13	0.84	1.43	B+
EQME31948	122xx	16	-1.28	-1.62	-0.94	B-	-1.37	-1.71	-1.03	B-
EQME31972	122xx	18	-1.46	-1.80	-1.12	B-	-1.35	-1.69	-1.01	B-
EQME21271	122xx	19	-1.30	-1.69	-0.91	B-	-1.27	-1.66	-0.88	B-
EQME10392	122xx	20	-0.57	-0.88	-0.25		-0.74	-1.06	-0.42	
10311	122xx	21	0.82	0.50	1.13		0.88	0.56	1.20	
EQME31941	122xx	22	0.35	0.05	0.66		0.66	0.35	0.97	
EQME10334	122xx	23	0.53	0.22	0.83		0.53	0.22	0.84	
10219	122xx	24	0.80	0.49	1.12		0.71	0.39	1.02	
EQME31883	122xx	25	-1.43	-1.74	-1.11	B-	-1.13	-1.45	-0.81	B-
EQME31973	122xx	30	0.88	0.57	1.20		0.98	0.67	1.30	
10313	122xx	31	0.03	-0.29	0.35		0.18	-0.15	0.51	
EQME10325	122xx	32	-0.27	-0.61	0.06		-0.25	-0.58	0.09	
EQME10321	122xx	33	-0.51	-0.93	-0.08		-0.63	-1.05	-0.21	
EQME31995	122xx	34	-1.08	-1.38	-0.77	B-	-1.25	-1.55	-0.94	B-
EQME31958	122xx	35	-0.98	-1.29	-0.67		-0.92	-1.23	-0.61	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring Non-SLLs.

Table 7.1.76b SLL-Based DIF Statistics for OR Items: Grade 9 Academic Mathematics, Spring (English)

Item Code	Booklet	Sequence	Sample 1			Sample 2		
			Effect Size	p-value	DIF Level	Effect Size	p-value	DIF Level
10316	122xx	8	-0.16	0.00		-0.17	0.00	
10322	122xx	9	0.14	0.00		0.12	0.00	
10231	122xx	10	0.29	0.00	C+	0.23	0.00	C+
EQME31917	122xx	11	-0.06	0.09		-0.11	0.00	
10232	122xx	26	0.13	0.00		0.15	0.00	
10321	122xx	28	-0.17	0.00		-0.13	0.00	
10291	122xx	29	0.19	0.00		0.21	0.00	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring Non-SLLs.

The 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
13052_23	I	1	R2	2	0.73	0.29	-0.68
13057_23	I	2	R1	3	0.95	0.30	-2.90
13058_23	I	3	R1	1	0.78	0.32	-1.06
13059_23	I	4	R2	1	0.80	0.40	-1.21
13061_23	I	5	R3	1	0.66	0.34	-0.26
13056_23	I	6	R3	3*	0.72 (2.17)	0.08	-1.63
13407	II	1	W2	3	0.80	0.34	-1.19
13423	II	2	W3	2	0.75	0.40	-0.91
13415	II	3	W3	2	0.39	0.32	1.26
13400	II	4	W1	2	0.69	0.37	-0.50
13110_T	III	1	W4	3*	0.84 (2.52)	0.30	-1.84
13110_V	III	1	W3	2*	0.99 (1.97)	0.17	-2.85
18340_T	IV	1	W4	6*	0.73 (4.37)	0.54	-1.04
18340_V	IV	1	W3	4*	0.90 (3.59)	0.51	-2.13
11709_102	V	1	R1	2	0.75	0.43	-0.87
11710_102	V	2	R3	4	0.54	0.35	0.47
11711_102	V	3	R2	1	0.94	0.35	-2.66
11712_102	V	4	R1	3	0.60	0.28	0.13
11713_102	V	5	R2	4	0.86	0.41	-1.81
11714_102	V	6	R2	3	0.56	0.27	0.34
11715_102	V	7	R2	3*	0.70 (2.09)	0.41	-0.86
18339_T	VI	1	W4	6*	0.77 (4.65)	0.60	-1.26
18339_V	VI	1	W3	4*	0.84 (3.36)	0.55	-1.81
13424	VII	1	W2	2	0.82	0.26	-1.30
13425	VII	2	W3	4	0.53	0.31	0.54
13401	VII	3	W1	4	0.87	0.22	-1.84
13375	VII	4	W3	2	0.51	0.30	0.61
10126_15	VIII	1	R1	3	0.95	0.23	-2.86
9743_15	VIII	2	R2	4	0.78	0.35	-1.05
13119_15	VIII	3	R1	2	0.67	0.37	-0.32
13118_15	VIII	4	R2	2	0.79	0.30	-1.09
13117_15	VIII	5	R2	1	0.67	0.37	-0.29

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. R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections; W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development; () = Mean score for open-response items.

¹Answer key for multiple-choice items and maximum score category for open-response items.

*Open-response items (OR, SW or LW).

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
13116_15	VIII	6	R3	3	0.70	0.29	-0.40
13115_15	VIII	7	R3	4	0.66	0.41	-0.32
13114_15	VIII	8	R2	2	0.79	0.34	-1.09
9736_15	VIII	9	R2	2	0.78	0.44	-1.07
9957_30	IX	1	R2	3	0.81	0.35	-1.32
9955_30	IX	2	R2	4	0.70	0.38	-0.55
9960_30	IX	3	R3	2	0.85	0.35	-1.61
9959_30	IX	4	R2	2	0.49	0.44	0.62
13141_30	IX	5	R2	1	0.87	0.31	-1.82
10333_30	IX	6	R2	3*	0.84 (2.25)	0.48	-1.58
10334_30	IX	7	R3	3*	0.76 (2.28)	0.46	-1.73
13173_T	X	1	W4	3*	0.80 (2.40)	0.34	-1.44
13173_V	X	1	W3	2*	0.98 (1.95)	0.27	-2.48
13370_238_24	XI	1	R2	1	0.84	0.33	-1.46
13374_238_24	XI	2	R2	1	0.79	0.35	-1.16
13369_238_24	XI	3	R2	3	0.52	0.41	0.50
13373_238_24	XI	4	R3	3	0.73	0.34	-0.73
13372_238_24	XI	5	R1	1	0.88	0.40	-1.93
13369_238_24	XI	6	R2	4	0.67	0.42	-0.39

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. R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections; W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development; () = Mean score for open-response items.

¹Answer key for multiple-choice items and maximum score category for open-response items.

*Open-response items (OR, SW or LW).

Table 7.1.78 Distribution of Score Points and Step Parameters for Open-Response Reading and Short-Writing Tasks: OSSLT (English)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1	Score 2	Score 3
13056_23	I	6	% of Students	0.36	0.31	0.01	20.34	40.63	38.35
			Parameters				-4.48	-0.89	0.49
13110_T	III	1	% of Students	0.32	0.91	0.03	2.83	38.48	57.44
			Parameters				-1.99	-3.27	-0.25
13110_V	III	1	% of Students	0.32	0.91	0.03	0.49	98.25	
			Parameters				-0.41	-5.30	
11715_102	V	7	% of Students	1.20	3.67	0.25	16.99	41.49	36.41
			Parameters				-1.98	-1.10	0.51
10333_30	IX	6	% of Students	1.70	0.94	0.03	4.43	31.52	61.38
			Parameters				-1.60	-2.59	-0.55
10334_30	IX	7	% of Students	0.25	0.64	0.00	18.52	32.26	48.33
			Parameters				-4.11	-0.95	-0.13
13173_T	X	1	% of Students	0.64	1.73	0.02	3.08	47.14	47.38
			Parameters				-1.41	-3.15	0.24
13173_V	X	1	% of Students	0.64	1.73	0.02	0.20	97.42	
			Parameters				0.98	-5.93	

Note. The total number of students is 143 178.

Table 7.1.79 Distribution of Score Points and Step Parameters for Long-Writing Tasks: OSSLT (English)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1.0	Score 1.5	Score 2.0	Score 2.5
18340_T	IV	1	% of Students	0.03	0.51	0.02	0.25	0.26	0.61	1.35
			Parameters				-2.37*		-1.36	-1.87
18340_V	IV	1	% of Students	0.00	0.51	0.02	0.19	0.21	1.06	3.57
			Parameters				-1.66*		-2.51	-2.59
18339_T	VI	1	% of Students	0.05	0.64	0.03	0.20	0.23	0.85	1.62
			Parameters				-1.94*		-1.90	-1.91
18339_V	VI	1	% of Students	0.00	0.64	0.03	0.24	0.25	1.80	6.67
			Parameters				-1.61*		-2.63	-2.53
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
18340_T	IV	1	% of Students	4.77	11.16	24.90	25.10	15.25	10.86	4.94
			Parameters	-2.25	-1.66	-1.48	-0.50	0.17	0.17	0.76
18340_V	IV	1	% of Students	14.44	32.02	48.01				
			Parameters	-2.69	-1.96	-1.36				
18339_T	VI	1	% of Students	4.38	7.61	13.58	19.70	21.25	18.95	10.89
			Parameters	-2.18	-1.57	-1.45	-1.10	-0.64	-0.27	0.31
18339_V	VI	1	% of Students	29.75	34.09	26.52				
			Parameters	-2.58	-1.04	-0.46				

Note. The total number of students is 143 178.

*Scores 1.0 and 1.5 have only one step parameter, since the two categories were collapsed.

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
10491 42	I	1	L1	2	0.86	0.30	-1.56
18336 42	I	2	L1	1	0.92	0.23	-2.11
10490 42	I	3	L3	1	0.58	0.27	0.30
10488 42	I	4	L2	1	0.68	0.31	-0.27
10492 42	I	5	L2	3	0.95	0.26	-2.55
10494 42	I	6	L3	3*	0.81 (2.42)	0.11	-2.28
10360	II	1	É1	2	0.79	0.29	-0.95
10346	II	2	É2	3	0.89	0.21	-1.70
10381	II	3	É3	1	0.79	0.29	-0.92
10480	II	4	É3	2	0.88	0.10	-1.65
10498 T	III	1	É4	3*	0.88 (2.64)	0.33	-1.91
10498 V	III	1	É3	2*	0.98 (1.97)	0.14	-2.77
2É3 030T	IV	1	É4	6*	0.76 (4.58)	0.57	-1.54
2É3 030V	IV	1	É3	4*	0.88 (3.52)	0.57	-2.43
10447 40	V	1	L1	2	0.81	0.23	-1.08
10453 40	V	2	L2	4	0.88	0.15	-1.62
10450 40	V	3	L1	3	0.63	0.14	0.07
10452 40	V	4	L2	2	0.76	0.29	-0.77
10451 40	V	5	L2	3	0.91	0.26	-2.01
10449 40	V	6	L3	2	0.85	0.37	-1.45
10460 40	V	7	L2	3*	0.68 (2.03)	0.35	-0.81
2É2 033T	VI	1	É4	6*	0.73 (4.38)	0.59	-1.12
2É2 033V	VI	1	É3	4*	0.83 (3.31)	0.59	-1.85
10482	VII	1	É3	4	0.82	0.18	-1.13
10348	VII	2	É2	2	0.55	0.43	0.38
10361	VII	3	É1	4	0.87	0.30	-1.61
10383	VII	4	É3	3	0.65	0.22	-0.08
10371 33	VIII	1	L2	1	0.71	0.33	-0.45
10365 33	VIII	2	L1	3	0.57	0.21	0.39
10367 33	VIII	3	L2	4	0.77	0.30	-0.84
10374 33	VIII	4	L2	2	0.54	0.24	0.57
10366 33	VIII	5	L1	1	0.82	0.32	-1.18
10379 33	VIII	6	L3	1	0.86	0.24	-1.52
10375 33	VIII	7	L2	3	0.91	0.23	-1.96
10376 33	VIII	8	L2	4	0.76	0.24	-0.74
10373 33	VIII	9	L3	2	0.65	0.25	-0.07

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. L1 = understanding explicitly; L2 = understanding implicitly; L3 = making connections; É1 = developing main idea; É2 = organizing information; É3 = using conventions; É4 = topic development; () = Mean score for open-response items.

¹Answer key for multiple-choice items and maximum score category for open-response items.

*Open-response items (OR, SW or LW).

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
10397_34	IX	1	L2	1	0.62	0.12	0.17
10399_34	IX	2	L2	3	0.55	0.25	0.46
10400_34	IX	3	L2	1	0.93	0.29	-2.24
10394_34	IX	4	L2	2	0.44	0.21	1.15
10395_34	IX	5	L3	3	0.64	0.32	-0.04
10402_34	IX	6	L2	3*	0.72 (2.17)	0.32	-1.44
10404_34	IX	7	L3	3*	0.85 (2.56)	0.40	-1.91
10499_T	X	1	É4	3*	0.88 (2.64)	0.35	-2.00
10499_V	X	1	É3	2*	0.98 (1.96)	0.22	-2.89
10337_32_11	XI	1	L1	1	0.83	0.31	-1.25
10342_32_11	XI	2	L2	2	0.84	0.24	-1.33
10339_32_11	XI	3	L2	1	0.89	0.29	-1.81
10336_32_11	XI	4	L2	4	0.55	0.27	0.49
10338_32_11	XI	5	L2	4	0.50	0.11	0.86
10344_32_11	XI	6	L3	2	0.76	0.27	-0.77

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. R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections; W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development; () = Mean score for open-response items.

¹Answer key for multiple-choice items and maximum score category for open-response items.

*Open-response items (OR, SW or LW).

Table 7.1.81 Distribution of Score Points and Step Parameters for Open-Response Reading and Short-Writing Tasks: OSSLT (French)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1.0	Score 2.0	Score 3.0
10494_42	I	6	% of Students	0.02	0.22	0.00	5.32	46.61	47.84
			Parameters				-4.64	-2.44	0.24
10498_T	III	1	% of Students	0.38	0.95	0.00	1.63	28.76	68.27
			Parameters				-2.04	-2.94	-0.75
10498_V	III	1	% of Students	0.38	0.95	0.00	0.55	98.11	
			Parameters				-1.49	-4.05	
10460_40	V	7	% of Students	0.79	3.76	0.02	22.34	38.40	34.70
			Parameters				-2.33	-0.64	0.54
10402_34	IX	6	% of Students	0.44	0.71	0.00	18.79	42.19	37.87
			Parameters				-3.83	-0.99	0.49
10404_34	IX	7	% of Students	0.20	0.97	0.00	3.48	33.06	62.28
			Parameters				-2.56	-2.69	-0.49
10499_T	X	1	% of Students	0.05	1.06	0.04	3.41	25.97	69.46
			Parameters				-2.71	-2.38	-0.89
10499_V	X	1	% of Students	0.05	1.06	0.04	1.54	97.31	
			Parameters				-1.98	-3.80	

Note. The total number of students is 5456.

Table 7.1.82 Distribution of Score Points and Step Parameters for Long-Writing Tasks: OSSLT (French)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1.0	Score 1.5	Score 2.0	Score 2.5
2É3_030T	IV	1	% of Students	0.00	0.11	0.00	0.18	0.29	0.66	1.67
			Parameters				-4.44*		-1.34	-1.97
2É3_030V	IV	1	% of Students	0.00	0.11	0.00	0.24	0.16	1.61	5.90
			Parameters				-3.67*		-2.46	-2.86
2É2_033T	VI	1	% of Students	0.04	0.51	0.02	0.27	0.35	1.17	2.80
			Parameters				-3.15*		-1.39	-1.84
2É2_033V	VI	1	% of Students	0.00	0.51	0.02	0.31	0.42	3.37	11.60
			Parameters				-2.66*		-2.50	-2.40
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_030T	IV	1	% of Students	4.77	11.13	17.61	21.44	12.12	17.80	12.23
			Parameters	-2.38	-1.86	-1.43	-0.95	-0.12	-0.86	-0.07
2É3_030V	IV	1	% of Students	20.05	29.45	42.47				
			Parameters	-2.51	-1.64	-1.43				
2É2_033T	VI	1	% of Students	7.79	12.24	16.57	20.73	16.24	13.51	7.75
			Parameters	-2.01	-1.22	-0.97	-0.73	-0.15	-0.09	0.33
2É2_033V	VI	1	% of Students	25.84	30.00	27.91				
			Parameters	-1.86	-1.02	-0.65				

Note. The total number of students is 5456.

*Scores 1.0 and 1.5 have only one step parameter, since the two categories were collapsed.

Differential Item Functioning (DIF) Analysis Results

The gender- and SLL-based DIF results for the OSSLT are provided in Tables 7.1.83–7.1.85.

Only for the English-language test are the results from two random samples of 2000 students.

For the French-language test, gender-based DIF analysis was conducted for one sample of students, and SLL-based DIF analysis was not conducted. In both cases, DIF analysis was limited by the relatively small population of students who wrote the French-language test. Each table indicates the value of Δ , the 95% confidence interval and the category of the effect size for each item for B- and C-level DIF items. For gender-based DIF, negative estimates of Δ indicate that the girls outperformed the boys; positive Δ estimates indicate that the boys outperformed the girls. For SLL-based DIF, negative estimates of Δ indicate that the SLLs outperformed the non-SLLs; positive Δ estimates indicate that the non-SLLs outperformed the SLLs.

Table 7.1.83a Gender-Based DIF Statistics for MC Items: OSSLT (English)

Item Code	Section	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
13052_23	I	1	-0.04	-0.39	0.31		0.12	-0.23	0.48	
13057_23	I	2	-0.24	-0.96	0.48		0.05	-0.64	0.74	
13058_23	I	3	1.36	0.97	1.75	B+	1.33	0.94	1.72	B+
13059_23	I	4	0.74	0.32	1.15		0.34	-0.07	0.75	
13061_23	I	5	1.61	1.27	1.96	C+	1.70	1.35	2.05	C+
13407	II	1	-0.65	-1.05	-0.24		-0.67	-1.06	-0.27	
13423	II	2	-0.51	-0.89	-0.12		-0.55	-0.93	-0.18	
13415	II	3	0.07	-0.26	0.41		0.38	0.04	0.71	
13400	II	4	0.06	-0.29	0.42		0.04	-0.32	0.39	
11709_102	V	1	0.34	-0.04	0.72		0.49	0.10	0.89	
11710_102	V	2	1.34	1.01	1.68	B+	0.95	0.62	1.28	
11711_102	V	3	0.90	0.24	1.57		0.49	-0.15	1.12	
11712_102	V	4	-0.33	-0.65	-0.01		-0.34	-0.66	-0.02	
11713_102	V	5	1.04	0.53	1.54	B+	0.75	0.28	1.22	
11714_102	V	6	0.35	0.04	0.66		0.53	0.21	0.84	
13424	VII	1	0.10	-0.32	0.51		0.03	-0.37	0.43	
13425	VII	2	0.51	0.19	0.83		0.31	-0.01	0.63	
13401	VII	3	-0.46	-0.91	0.00		-0.21	-0.67	0.25	
13375	VII	4	-0.32	-0.63	-0.01		0.06	-0.25	0.38	
10126_15	VIII	1	0.46	-0.28	1.21		0.03	-0.63	0.69	
9743_15	VIII	2	-0.39	-0.79	0.00		-0.27	-0.66	0.11	
13119_15	VIII	3	0.48	0.13	0.82		0.22	-0.12	0.57	
13118_15	VIII	4	-0.07	-0.46	0.32		-0.06	-0.44	0.32	
13117_15	VIII	5	1.11	0.76	1.46	B+	1.28	0.93	1.63	B+
13116_15	VIII	6	0.41	0.06	0.76		0.44	0.10	0.77	
13115_15	VIII	7	-0.17	-0.52	0.18		-0.10	-0.45	0.26	
13114_15	VIII	8	0.19	-0.20	0.58		-0.14	-0.52	0.23	
9736_15	VIII	9	1.58	1.17	2.00	C+	1.63	1.22	2.03	C+
9957_30	IX	1	0.23	-0.19	0.64		-0.09	-0.50	0.32	
9955_30	IX	2	-0.14	-0.49	0.22		0.02	-0.34	0.37	
9960_30	IX	3	-1.08	-1.53	-0.62	B-	-0.78	-1.23	-0.34	
9959_30	IX	4	-0.26	-0.60	0.08		-0.48	-0.82	-0.14	
13141_30	IX	5	0.98	0.50	1.46		0.92	0.45	1.39	
13370_238_24	XI	1	1.38	0.94	1.82	B+	1.18	0.75	1.60	B+
13374_238_24	XI	2	1.25	0.85	1.65	B+	1.59	1.18	2.00	C+
13369_238_24	XI	3	0.02	-0.32	0.36		0.16	-0.18	0.50	
13373_238_24	XI	4	1.15	0.77	1.52	B+	1.36	0.99	1.72	B+
13372_238_24	XI	5	0.73	0.22	1.24		0.09	-0.40	0.59	
13369_238_24	XI	6	1.62	1.26	1.99	C+	1.21	0.84	1.57	B+

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

Table 7.1.83b Gender-Based DIF Statistics for OR Items: OSSLT (English)

Item Code	Section	Sequence	Sample 1			Sample 2		
			Effect Size	<i>p</i> -value	DIF Level	Effect Size	<i>p</i> -value	DIF Level
13056_23	I	6	-0.06	0.0785		-0.05	0.2153	
13110_T	III	1	0.03	0.0000		0.04	0.0000	
13110_V	III	1	-0.08	0.0036		-0.11	0.0002	
18340_T	IV	1	-0.08	0.0507		-0.11	0.0002	
18340_V	IV	1	-0.15	0.0000		-0.15	0.0000	
11715_102	V	7	-0.07	0.0315		-0.08	0.0001	
18339_T	VI	1	-0.20	0.0000	B-	-0.19	0.0000	B-
18339_V	VI	1	-0.23	0.0000	B-	-0.28	0.0000	C-
10333_30	IX	6	-0.15	0.0000		-0.14	0.0000	
10334_30	IX	7	-0.07	0.0071		-0.07	0.0165	
13173_T	X	1	-0.11	0.0006		-0.07	0.0230	
13173_V	X	1	-0.08	0.0021		0.00	0.5418	

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

*Separate components are presented for short- and long-writing items; T=topic development and V=use of conventions.

Table 7.1.84a Gender-Based DIF Statistics for MC Items: OSSLT (French)

Item Code	Section	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
10491_42	I	1	0.97	0.51	1.44	
18336_42	I	2	0.03	-0.54	0.59	
10490_42	I	3	0.77	0.45	1.09	
10488_42	I	4	0.25	-0.09	0.59	
10492_42	I	5	0.39	-0.35	1.12	
10360	II	1	0.42	0.03	0.81	
10346	II	2	0.61	0.12	1.10	
10381	II	3	0.21	-0.17	0.60	
10480	II	4	-0.31	-0.77	0.15	
10447_40	V	1	-0.41	-0.81	-0.02	
10453_40	V	2	0.26	-0.22	0.73	
10450_40	V	3	-0.19	-0.51	0.12	
10452_40	V	4	0.44	0.06	0.81	
10451_40	V	5	-0.37	-0.92	0.19	
10449_40	V	6	-0.20	-0.66	0.25	
10482	VII	1	-0.39	-0.78	0.00	
10348	VII	2	0.16	-0.19	0.51	
10361	VII	3	-0.58	-1.05	-0.11	
10383	VII	4	-0.11	-0.43	0.22	
10371_33	VIII	1	0.98	0.62	1.35	

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

Table 7.1.84a Gender-Based DIF Statistics for MC Items: OSSLT (French) (continued)

Item Code	Section	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
10365 33	VIII	2	2.26	1.92	2.59	C+
10367 33	VIII	3	0.42	0.05	0.80	
10374 33	VIII	4	1.20	0.88	1.52	B+
10366 33	VIII	5	1.42	1.00	1.84	B+
10379 33	VIII	6	0.65	0.19	1.11	
10375 33	VIII	7	0.14	-0.40	0.68	
10376 33	VIII	8	0.16	-0.20	0.52	
10373 33	VIII	9	0.71	0.38	1.04	
10397 34	IX	1	0.10	-0.22	0.41	
10399 34	IX	2	0.16	-0.16	0.47	
10400 34	IX	3	-0.03	-0.65	0.59	
10394 34	IX	4	1.47	1.16	1.79	B+
10395 34	IX	5	-0.49	-0.83	-0.16	
10337 32 11	XI	1	1.33	0.89	1.76	B+
10342 32 11	XI	2	1.10	0.67	1.53	B+
10339 32 11	XI	3	-0.15	-0.66	0.35	
10336 32 11	XI	4	0.18	-0.14	0.49	
10338 32 11	XI	5	0.19	-0.11	0.50	
10344 32 11	XI	6	0.28	-0.09	0.65	

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

Table 7.1.84b Gender-Based DIF Statistics for OR Items: OSSLT (French)

ItemCode	Section	Sequence	Sample 1		
			Effect Size	p-value	DIF Level
10494 42	I	6	-0.13	0.0001	
10498 T	III	1	0.03	0.6965	
10498 V	III	1	-0.02	0.2835	
2É3 030T	IV	1	-0.16	0.0000	
2É3 030V	IV	1	-0.21	0.0000	B-
10460 40	V	7	-0.06	0.0288	
2É2 033T	VI	1	-0.20	0.0000	B-
2É2 033V	VI	1	-0.22	0.0000	B-
10402 34	IX	6	-0.08	0.0013	
10404 34	IX	7	-0.15	0.0000	
10499 T	X	1	-0.03	0.5630	
10499 V	X	1	-0.03	0.2624	

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

*Separate components are presented for short- and long-writing items; T=topic development and V=use of conventions.

Table 7.1.85a SLL-Based DIF Statistics for MC Items: OSSLT (English)

Item Code	Section	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
13052 23	I	1	-0.81	-1.16	-0.45		-1.25	-1.62	-0.88	B-
13057 23	I	2	1.86	1.27	2.45	C+	2.81	2.22	3.40	C+
13058 23	I	3	-0.95	-1.35	-0.55		-1.20	-1.61	-0.80	B-
13059 23	I	4	0.08	-0.31	0.46		0.84	0.47	1.21	
13061 23	I	5	0.31	-0.02	0.65		0.75	0.41	1.08	
13407	II	1	0.21	-0.17	0.58		-0.32	-0.70	0.06	
13423	II	2	0.15	-0.22	0.52		-0.27	-0.64	0.09	
13415	II	3	-1.59	-1.92	-1.26	C-	-2.01	-2.34	-1.67	C-
13400	II	4	-0.74	-1.09	-0.38		-1.46	-1.83	-1.10	B-
11709 102	V	1	0.07	-0.30	0.44		-0.05	-0.43	0.32	
11710 102	V	2	1.66	1.33	1.99	C+	1.65	1.31	1.98	C+
11711 102	V	3	1.84	1.27	2.42	C+	2.53	2.01	3.05	C+
11712 102	V	4	0.14	-0.18	0.45		-0.14	-0.45	0.18	
11713 102	V	5	0.61	0.18	1.04		0.55	0.12	0.97	
11714 102	V	6	-0.68	-0.99	-0.36		-1.27	-1.60	-0.95	B-
13424	VII	1	0.28	-0.10	0.67		0.38	0.00	0.75	
13425	VII	2	1.83	1.50	2.16	C+	1.73	1.41	2.06	C+
13401	VII	3	-0.12	-0.56	0.32		-0.24	-0.70	0.21	
13375	VII	4	-1.13	-1.46	-0.81	B-	-1.40	-1.73	-1.06	B-
10126 15	VIII	1	0.48	-0.13	1.10		1.24	0.64	1.83	B+
9743 15	VIII	2	1.33	0.97	1.70	B+	2.12	1.77	2.47	C+
13119 15	VIII	3	0.12	-0.23	0.46		-0.45	-0.79	-0.10	
13118 15	VIII	4	0.48	0.11	0.85		-0.05	-0.41	0.31	
13117 15	VIII	5	1.40	1.07	1.74	B+	1.68	1.35	2.01	C+
13116 15	VIII	6	2.08	1.75	2.41	C+	2.52	2.19	2.85	C+
13115 15	VIII	7	0.65	0.32	0.98		0.39	0.06	0.72	
13114 15	VIII	8	0.80	0.43	1.16		1.39	1.03	1.75	B+
9736 15	VIII	9	1.61	1.24	1.99	C+	1.72	1.36	2.07	C+
9957 30	IX	1	0.35	-0.03	0.74		-0.10	-0.47	0.28	
9955 30	IX	2	-0.95	-1.31	-0.59		-0.98	-1.34	-0.62	
9960 30	IX	3	1.11	0.71	1.52	B+	0.77	0.37	1.17	
9959 30	IX	4	0.05	-0.30	0.40		0.28	-0.07	0.64	
13141 30	IX	5	1.24	0.82	1.67	B+	2.15	1.75	2.56	C+
13370 238 24	XI	1	1.47	1.08	1.87	B+	1.30	0.92	1.68	B+
13374 238 24	XI	2	0.10	-0.27	0.48		-0.70	-1.10	-0.30	
13369 238 24	XI	3	0.15	-0.19	0.49		-0.05	-0.39	0.28	
13373 238 24	XI	4	-0.05	-0.40	0.31		-0.66	-1.02	-0.30	
13372 238 24	XI	5	0.74	0.28	1.20		1.11	0.68	1.53	B+
13369 238 24	XI	6	-0.27	-0.61	0.08		-1.21	-1.57	-0.86	B-

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

Table 7.1.85b SLL-Based DIF Statistics for OR Items: OSSLT (English)

ItemCode	Section	Sequence	Sample 1			Sample 2		
			Effect Size	<i>p</i> -value	DIF Level	Effect Size	<i>p</i> -value	DIF Level
13056_23	I	6	-0.21	0.0000	B-	-0.18	0.0000	B-
13110_T	III	1	-0.14	0.0008		-0.25	0.0000	B-
13110_V	III	1	0.00	0.0883		0.02	0.0320	
18340_T	IV	1	-0.10	0.0002		-0.13	0.0006	
18340_V	IV	1	0.11	0.0001		0.18	0.0000	B+
11715_102	V	7	-0.29	0.0000	C-	-0.35	0.0000	C-
18339_T	VI	1	-0.25	0.0000	C-	-0.21	0.0000	B-
18339_V	VI	1	0.02	0.0000		0.12	0.0000	
10333_30	IX	6	-0.05	0.0000		-0.03	0.2616	
10334_30	IX	7	-0.11	0.0013		-0.13	0.0001	
13173_T	X	1	-0.07	0.0083		0.04	0.0000	
13173_V	X	1	-0.02	0.4743		0.07	0.0177	

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

*Separate components are presented for short- and long-writing items; T=topic development and V=use of conventions.

**Education Quality and
Accountability Office**



2 Carlton Street, Suite 1200, Toronto ON M5B 2M9

Telephone: 1-888-327-7377 Web site: www.eqao.com

© 2012 Queen's Printer for Ontario