

EQAO's Technical Report

for the 2011–2012 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 2011–2012 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

ISBN 978-1-4606-2930-7, ISSN 1927-7105

© 2013 Queen's Printer for Ontario | Ctrc_report_ne_0813

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	5
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	18
Ongoing Daily Training	19
Daily Scoring-Centre Reports for Monitoring the Quality of Open-Response Item Scoring	19
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.....	23
CHAPTER 5: EQUATING	25
IRT MODELS	25
EQUATING DESIGN	26
EQUATING SAMPLE	26
CALIBRATION SAMPLE	27

IDENTIFICATION OF ITEMS TO BE EXCLUDED FROM EQUATING	27
THE ASSESSMENTS OF READING, WRITING AND MATHEMATICS: PRIMARY AND JUNIOR DIVISIONS.....	28
Description of the IRT Model	28
Equating Sample: Exclusion Rules.....	28
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.....	34
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	40
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	42
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).....	66
Classical Test Theory (CTT) Analysis	66
Item Response Theory (IRT) Analysis.....	66
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT)	69
DIFFERENTIAL ITEM FUNCTIONING (DIF)	69
The Primary- and Junior-Division Assessments.....	71
The Grade 9 Mathematics Assessment.....	73
The OSSLT.....	75
DECISION ACCURACY AND CONSISTENCY	76
Accuracy.....	77
Consistency	77
Estimation from One Test Form.....	78
The Primary and Junior Assessments	79
The Grade 9 Assessment of Mathematics	79
The OSSLT.....	80
REFERENCES	80
CHAPTER 8: VALIDITY EVIDENCE.....	81
INTRODUCTION.....	81
The Purposes of EQAO Assessments.....	81
Conceptual Framework for the Validity Argument.....	81
VALIDITY EVIDENCE BASED ON THE CONTENT OF THE ASSESSMENTS AND THE ASSESSMENT PROCESSES...	82
Test Specifications for EQAO Assessments.....	82

Appropriateness of Test Questions	82
Quality Assurance in Administration	83
Scoring of Open-Response Items	83
Equating	84
VALIDITY EVIDENCE BASED ON THE TEST CONSTRUCTS AND INTERNAL STRUCTURE.....	84
Test Dimensionality.....	84
Technical Quality of the Assessments	84
VALIDITY EVIDENCE BASED ON EXTERNAL ASSESSMENT DATA	85
Linkages to International Assessment Programs	85
VALIDITY EVIDENCE SUPPORTING APPROPRIATE INTERPRETATIONS OF RESULTS.....	86
Setting Standards	86
Reporting	86
REFERENCES	87
APPENDIX 4.1: SCORING VALIDITY AND INTERRATER RELIABILITY.....	88
APPENDIX 7.1: SCORE DISTRIBUTIONS AND ITEM STATISTICS	103

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 applied and academic versions of the Grade 9 Assessment of Mathematics measure 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);
- representation of rural and urban regions;
- current elementary and secondary experience (teachers, administrators, subject experts and consultants) and
- expertise in assessment, evaluation and large-scale assessment.

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 2011–2012 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.

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. Further item reviews are conducted by external experts prior to the final revisions by the education officers and prior to Assessment Development and Sensitivity Committee reviews.

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.

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 8–10 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 students 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).

Beginning in 2010, 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 for students with special education needs 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 (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 four Ontario Ministry of Education policy documents: *Individual Education Plans: Standards for Development, Program Planning, and Implementation* (2000); *English Language Learners / ESL and ELD Programs and Services: Policies and Procedures for Ontario Elementary and Secondary Schools, Kindergarten to Grade 12* (2007); *Growing Success: Assessment, Evaluation, and Reporting in Ontario Schools, First Edition, Covering Grades 1 to 12* (2010) and *Ontario Schools, Kindergarten to Grade 12: Policy and Program Requirements* (2011), available at www.edu.gov.on.ca.

Definition of “Accommodations”

Accommodations are defined in the accommodations guides (modified from *Ontario Schools, Kindergarten to Grade 12: Policy and Program Requirements* (2011)) 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 at 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. (Examples of unacceptable responses are also selected for training purposes.) 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 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, and 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 is scored. The number of French Grade 9 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. Qualifying tests are also developed for each open-response and short-writing item for scoring of field-test items. 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.

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 and
- 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 2011–2012, the EQAO target of 95% exact-plus-adjacent agreement for validity was met for all but one item in the mathematics component and for all items in the reading and writing components of the assessments. The aggregate validity estimates for exact-plus-adjacent agreement ranged from 98.8–100%.

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 2012. 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 99–100%.

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 2011–2012, the EQAO target of 95% exact-plus-adjacent agreement for validity was met for all but one item (in short writing). The aggregate validity estimates ranged from 97.1–99.7%.

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. Scoring reliability estimates for the primary-division, junior-division 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 2011–2012, the EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for all but five reading items, all but three rubrics for writing and all but two mathematics items. The aggregate reliability estimates ranged from 95.2–99.4%.

The Grade 9 Assessment of Mathematics: Academic and Applied

In 2011–2012, the EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for all but five items, and the aggregate reliability estimates ranged from 95.0–99.8%.

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 but one short-writing prompt, for all but one conventions rubric for the long-writing prompts and for none of the topic-development rubrics for the long-writing prompts. The aggregate reliability estimates ranged from 93.2–98.3%.

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 2010–2011 to 2011–2012, only equating procedures were used in 2011–2012.

The following sections describe the Item Response Theory (IRT) models, equating design, equating samples and calibration procedures used during the 2011–2012 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).

Identification of Items to be Excluded from Equating

A key assumption in the common-item non-equivalent groups design is that the common items should behave similarly from field testing (FT) to operational testing (OP). In order to determine which items did not behave similarly, and thus should be excluded from equating, a four-step process is followed to examine each assessment. This process relies on judgment, as Kolen and Brennan (2004) stated: “removal of items that appear to be outliers is clearly a judgmental process” (p. 188).

First, scatter plots are produced to compare the common-item parameter estimates (both discrimination estimates and difficulty estimates, including item-category difficulty estimates of OR items) from field testing to operational 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. For each open-response item, an individual plot is constructed of its OP- and FT-category difficulty estimates. If the category difficulty estimates are not monotonically increasing and/or they are not far off the best-fit line, then this open-response item is also flagged for further analysis.

Second, in order to determine which of the outlying items identified in the first step to focus on for further investigation, several additional factors are considered: whether an item is flagged by both the OP-FT difficulty and OP-FT discrimination plots, whether it has a large difference between OP and FT classical item statistics and whether there is a large change in its position in the booklets from FT to OP.

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. The resulting achievement levels of students are compared with their initial levels.

Finally, another factor that informs the final decision making concerns the slopes of the best-fit lines in the plots of the parameter estimates. Theoretically, the slope of the best-fit line in the plot of item-difficulty estimates should be the reciprocal of that in the plot of item-discrimination estimates (see Kolen and Brennan, 2004, for example), so these slopes are examined with and without excluding an outlier to see in which case the reciprocal relationship holds.

When it comes to excluding items from equating, the overarching principle is to be conservative. That is, a common item should not be excluded from equating unless there is strong evidence to support exclusion. A common item is part of an equating link, and generally, the larger the number of common items, the stronger the link.

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 2010–2011 and 2011–2012:

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 2010–2011 population who responded to both the 2011–2012 operational-test items and the field-test items that had been brought forward to form the 2011–2012 operational tests and who were not excluded by the rules stated above (calibration sample);
2. students from the 2010–2011 population who wrote the operational test and who were not excluded by the rules stated above (equating sample) and
3. students from the 2011–2012 population who wrote the operational test and who were not excluded by the rules stated above (calibration and equating samples).

Table 5.1 Number of Students in the Calibration and Equating Samples for the 2010–2011 and 2011–2012 Primary- and Junior-Division Assessments (English and French)

Assessment	Number of Students		
	2011–2012 Calibration and Equating Samples	2010–2011 Calibration Sample	2010–2011 Equating Sample
Primary Reading (English)	100 902	27 404	101 000
Junior Reading (English)	107 699	30 665	111 823
Primary Reading (French)	6 379	2 109	6 182
Junior Reading (French)	5 696	1 933	5 844
Primary Writing (English)	101 290	45 691	101 463
Junior Writing (English)	107 773	50 631	111 998
Primary Writing (French)	6 420	3 249	6 224
Junior Writing (French)	5 720	4 256	5 861
Primary Mathematics (English)	97 457	88 557	97 884
Junior Mathematics (English)	109 630	97 074	113 687
Primary Mathematics (French)	6 482	6 256	6 271
Junior Mathematics (French)	5 782	5 912	5 930

Equating Steps

In equating the 2010–2011 and 2011–2012 tests, the forward fixed common-item-parameter non-equivalent group design was implemented as follows:

1. The 2011–2012 operational items were calibrated independently to obtain item parameter estimates and student-ability scores for the 2011–2012 calibration and equating sample.
2. The 2010–2011 operational items were calibrated (using the calibration sample) together with the field-test items that were brought forward to the 2011–2012 operational assessments. In this calibration, the item parameter estimates of the field-test items were fixed at the values obtained from the 2011–2012 calibration runs (Step 1).
3. The 2010–2011 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 2010–2011 equating sample from the levels assigned in 2010–2011. 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 2011–2012.
6. The operational item parameter estimates of 2011–2012 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 2011–2012 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, five multiple-choice items and five open-response items across the 12 assessment components were excluded from equating. The four reading items were modified between field- and operational-test administrations. One multiple-choice writing item and one short-writing prompt were excluded due to large parameter shifts from field- to operational-test administrations. Long-writing prompts are not field-tested in the previous year. 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 (2011–2012)

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)	2	0	0
Junior Reading (French)	0	0	0
Primary Writing (English)	1	1	0
Junior Writing (English)	0	1	0
Primary Writing (French)	0	2	0
Junior Writing (French)	0	1	0
Primary Mathematics (English)	0	0	0
Junior Mathematics (English)	0	0	0
Primary Mathematics (French)	0	0	0
Junior Mathematics (French)	0	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 2010–2011

and 2011–2012 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 1.06 for Levels 3 and 4, –0.54 for Levels 2 and 3, –1.69 for Levels 1 and 2 and –2.75 for “not enough evidence for Level 1” (NE1) and Level 1.

Since the 2010–2011 and 2011–2012 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	
		2011	2012		2011	2012
Number of Students		101 000	100 902		6182	6379
Level 4		9.0%	11.5%		27.1%	34.8%
Level 3	1.06	63.3%	62.6%	0.41	48.8%	45.1%
Level 2	-0.54	22.5%	21.4%	-0.78	21.3%	17.9%
Level 1	-1.69	4.4%	4.0%	-1.99	2.8%	1.9%
NE1	-2.75	0.7%	0.5%	-3.28	0.1%	0.2%
% of Students at or Above the Provincial Standard		72.4%	74.1%		75.8%	79.9%

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	
		2011	2012		2011	2012
Number of Students		111 823	107 699		5844	5696
Level 4		13.3%	15.9%		23.9%	28.0%
Level 3	0.92	68.7%	67.6%	0.56	63.7%	63.0%
Level 2	-0.91	16.1%	14.9%	-1.27	12.0%	8.9%
Level 1	-2.13	1.8%	1.4%	-2.78	0.5%	0.1%
NE1	-3.39	0.1%	0.1%	-3.62	0.0%	0.0%
% of Students at or Above the Provincial Standard		81.9%	83.6%		87.5%	91.0%

Since the percentages of students in the 2010–2011 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 2011–2012 assessments. To determine the 0/1 cut point for the 2011–2012 assessments, regression equations that predicted the 1/2 and 2/3 2011–2012 cuts from the 1/2 and 2/3 2010–2011 cuts were determined. The constant and multiplier values obtained from these regression equations were used to predict the 2011–2012 0/1 cut from the 2010–2011 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	
		2011	2012		2011	2012
Number of Students		101 463	101 290		6224	6420
Level 4		5.3%	7.6%		27.3%	26.3%
Level 3	1.33	72.9%	73.9%	0.59	59.5%	59.6%
Level 2	-0.84	21.7%	18.5%	-1.03	13.2%	13.9%
Level 1	-3.34	0.1%	0.0%	-2.79	0.1%	0.2%
NE1	-3.50	0.1%	0.0%	-3.72	0.0%	0.0%
% of Students at or Above the Provincial Standard		78.2%	81.4%		86.7%	85.9%

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	
		2011	2012		2011	2012
Number of Students		111 998	107 773		5861	5720
Level 4		11.5%	13.8%		19.2%	22.3%
Level 3	1.00	69.2%	68.6%	0.69	68.3%	66.8%
Level 2	-0.84	19.1%	17.5%	-1.14	12.5%	10.8%
Level 1	-3.03	0.2%	0.2%	-3.37	0.1%	0.0%
NE1	-3.60	0.0%	0.0%	-3.73	0.0%	0.0%
% of Students at or Above the Provincial Standard		80.7%	82.4%		87.5%	89.2%

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	
		2011	2012		2011	2012
Number of Students		97 884	97 457		6271	6482
Level 4		13.3%	14.0%		20.0%	23.5%
Level 3	1.02	61.4%	60.9%	0.72	57.6%	55.5%
Level 2	-0.61	23.3%	22.8%	-0.78	20.4%	19.6%
Level 1	-2.03	2.0%	2.0%	-2.15	1.9%	1.2%
NE1	-2.76	0.1%	0.2%	-2.93	0.1%	0.1%
% of Students at or Above the Provincial Standard		74.7%	74.9%		77.6%	79.1%

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	
		2011	2012		2011	2012
Number of Students		113 687	109 630		5930	5782
Level 4		14.3%	15.5%		40.9%	44.8%
Level 3	0.99	50.7%	49.9%	0.14	43.2%	41.5%
Level 2	-0.37	28.8%	28.5%	-1.04	15.5%	13.3%
Level 1	-1.51	6.2%	6.0%	-2.42	0.3%	0.3%
NE1	-3.07	0.1%	0.1%	-2.82	0.1%	0.2%
% of Students at or Above the Provincial Standard		65.0%	65.4%		84.0%	86.3%

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 2010–2011 and 2011–2012 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 2010–2011 and the 2011–2012 assessments.

Table 5.9 Number of Grade 9 Students in the Equating Samples

Version	2010–2011				2011–2012	
	English		French		English	French
	Calibration Sample	Equating Sample	Calibration Sample	Equating Sample	Calibration and Equating Sample	Calibration and Equating Sample
Academic	80 978	96 703	3361	3985	95 309	3898
Applied	31 505	37 497	968	1167	35 718	1233

Equating Steps

The calibration and equating of the 2010–2011 and 2011–2012 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 2011–2012 winter and spring samples.
2. Calibration and equating were conducted for the 2010–2011 winter and spring samples (including the field-test items that were brought forward to the 2011–2012 operational test). In this calibration, the item parameter estimates of the field-test items that were brought forward to the 2011–2012 operational tests were fixed at the values obtained from the 2011–2012 calibration runs (Step 1). The parameter estimates of the 2010–2011 operational items that were repeated on the 2011–2012 tests were also fixed at the values obtained from the 2011–2012 calibration runs (Step 1). The 2010–2011 equating samples (operational items only) were scored using the scaled 2010–2011 operational-item parameter estimates.
3. The percentage of students in each achievement level was determined for the 2010–2011 equating sample using the levels assigned in 2010–2011. 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 2011–2012 assessments.
5. The parameter estimates for the 2011–2012 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 2011–2012 student theta values to classify students to achievement 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, two items were excluded from the equating process. These two items were multiple-choice matrix items from the 2007 assessment and were not re-field-tested in 2011 (see Table 5.10).

Table 5.10 Number of Items Excluded from the Equating Process and Dropped: Grade 9 (2011–2012)

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)	0	0	0
Academic, Winter (English)	0	0	0
Academic, Spring (English)	0	0	0
Applied, Winter (French)	1	0	0
Applied, Spring (French)	1	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 Assessment of Mathematics are summarized in Table 5.11. The results for the academic version are summarized in Table 5.12. The theta cut scores and number of students in 2010–2011 and 2011–2012 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 1.29 for Levels 3 and 4, 0.06 for Levels 2 and 3, –0.96 for Levels 1 and 2 and –1.71 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	
		2011	2012		2011	2012
Number of Students		37 497	35 718		1167	1233
Level 4		7.7%	8.2%		6.5%	9.2%
Level 3	1.29	37.9%	40.2%	1.29	33.8%	38.2%
Level 2	0.06	37.5%	35.6%	0.07	41.6%	38.9%
Level 1	-0.96	13.5%	13.0%	-1.02	14.6%	10.8%
Below Level 1	-1.71	3.4%	3.0%	-1.72	3.6%	2.8%
% of Students at or Above the Provincial Standard		45.6%	48.4%		40.3%	47.4%

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	
		2011	2012		2011	2012
Number of Students		96 703	95 309		3985	3898
Level 4		10.8%	13.0%		4.8%	6.9%
Level 3	1.12	73.6%	72.1%	1.45	66.9%	72.4%
Level 2	-0.10	11.6%	10.8%	-0.81	17.9%	14.7%
Level 1	-1.60	3.9%	4.1%	-1.45	9.9%	5.9%
Below Level 1	-2.66	0.1%	0.1%	-2.49	0.6%	0.1%
% of Students at or Above the Provincial Standard		84.4%	85.0%		71.6%	79.3%

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 2010–2011 and 2011–2012 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: 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 2010–2011 and 2011–2012 tests.

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

OSSLT	2011		2012	
	English	French	English	French
First-Time Eligible	139 338	5055	133 760	4875

Equating Steps

The following steps were implemented to calibrate and equate the 2011 and 2012 OSSLT:

1. The parameter estimates of the operational items administered in 2012 were calibrated using the 2012 equating sample.
2. The operational items that formed the 2011 test and the field-test items brought forward to the 2012 test were recalibrated using the 2011 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 2011 test, obtained in Step 2, were used to score the 2011 equating sample data.
4. The percentage of successful students for the 2011 equating sample was determined from the student results reported in 2011. The theta-value cut point that replicated this percentage was identified for the distribution of scores in the 2011 equating sample.
5. This theta value was applied to student scores for the 2012 assessment to determine which students would be successful. The results are presented in Table 5.14.

Scale Score

The reporting scale scores for the 2012 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

Five multiple-choice items in the English-language OSSLT and six multiple-choice items in the French-language OSSLT were excluded from the equating process due to modifications made to the items between the 2011 field-test and the 2012 operational-test administrations. An additional six multiple-choice items and one open-response item in the English-language OSSLT were excluded from the equating process because the percentage of successful students in the sample that responded to these items during the field test was much smaller than the percentage of successful students in the provincial population.

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 for 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 2011 and 2012 are reported for English-language and French-language students. For example, the equated cut score for the English-language test was -0.73 . The percentage of successful students in the equating samples was 85.8% in 2011 and 84.5% in 2012.

Table 5.14 Equating Results for the OSSLT

	English-Language			French-Language		
	Theta Cut Point 2011–2012	Equating Sample 2011	Equating Sample 2012	Theta Cut Point 2011–2012	Equating Sample 2011	Equating Sample 2012
No. of Students		139 338	133 760		5055	4875
% Successful	-0.73	85.8%	84.5%	-0.68	88.4%	88.0%
Unsuccessful		14.2%	15.5%		11.6%	12.0%

References

- Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: 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, NY: 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 2011–2012 English-language assessments are available on www.eqao.com:

- *EQAO's Provincial Elementary School Report: Results of the 2011–2012 Assessments of Reading, Writing and Mathematics, Primary Division (Grades 1–3) and Junior Division (Grades 4–6)* and
- *EQAO's Provincial Secondary School Report: Results of the Grade 9 Assessment of Mathematics and the Ontario Secondary School Literacy Test, 2011–2012.*

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 that wrote an assessment 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:

- 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. Elementary-school principals are provided with an online reporting application that enables them to examine achievement results in various ways, including globally for their school, in relation to demographic and student attitude and behaviour data, and also for groups of students with specified characteristics.

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. The ISR also includes a description of the student work typically provided by students at each achievement level and suggestions for assisting students to progress beyond their achieved level.

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 2012, 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 students 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;
- feedback for students on areas of strength and areas for improvement and
- the Grade 3 and Grade 6 reading and writing results for the student, if available.

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 2011–2012 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 56.8% (for primary English-language reading) to 67.8% (for primary English-language math).

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 0.89 for reading, 0.81 to 0.83 for writing and 0.88 to 0.90 for mathematics. The corresponding standard errors of measurement range from 4.3% to 5.0% of the possible maximum score for reading, 6.8% to 7.2% of the possible maximum score for writing and 6.1% to 6.4% 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	117 920	0	65	37.50	9.55	0.88	3.31
Junior Reading (English)	36	26	10	66	126 119	0	66	38.92	9.62	0.89	3.19
Primary Reading (French)	36	26	10	66	7 023	4	61	38.36	8.20	0.88	2.84
Junior Reading (French)	36	26	10	66	6 298	6	65	42.59	8.63	0.86	3.23
Primary Writing (English)	14	8	6*	29	118 315	0	29	18.50	5.08	0.83	2.09
Junior Writing (English)	14	8	6*	29	126 199	0	29	17.08	4.55	0.81	1.98
Primary Writing (French)	14	8	6*	29	7 036	1	29	16.74	4.55	0.81	1.98
Junior Writing (French)	14	8	6*	29	6 300	0	29	18.74	4.88	0.82	2.07
Primary Mathematics (English)	36	28	8	60	122 783	0	60	40.65	11.22	0.89	3.72
Junior Mathematics (English)	36	28	8	60	125 681	0	60	37.53	12.12	0.90	3.83
Primary Mathematics (French)	36	28	8	60	7 032	1	60	38.39	10.60	0.88	3.67
Junior Mathematics (French)	36	28	8	60	6 270	6	60	38.06	11.54	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)	117 920	-3.87	3.37	-0.14	1.00
Junior Reading (English)	126 119	-3.89	3.47	-0.16	1.02
Primary Reading (French)	7 023	-3.87	3.05	-0.08	0.97
Junior Reading (French)	6 298	-3.90	2.97	-0.10	0.99
Primary Writing (English)	118 315	-3.58	2.36	-0.09	0.95
Junior Writing (English)	126 199	-3.69	2.66	-0.15	0.98
Primary Writing (French)	7 036	-3.62	2.82	-0.04	0.93
Junior Writing (French)	6 300	-3.84	2.29	-0.09	0.97
Primary Mathematics (English)	122 783	-3.76	2.26	-0.12	1.00
Junior Mathematics (English)	125 681	-3.69	2.39	-0.15	1.02
Primary Mathematics (French)	7 032	-3.73	2.51	-0.06	0.97
Junior Mathematics (French)	6 270	-3.68	2.41	-0.08	0.99

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 -0.20 . 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)

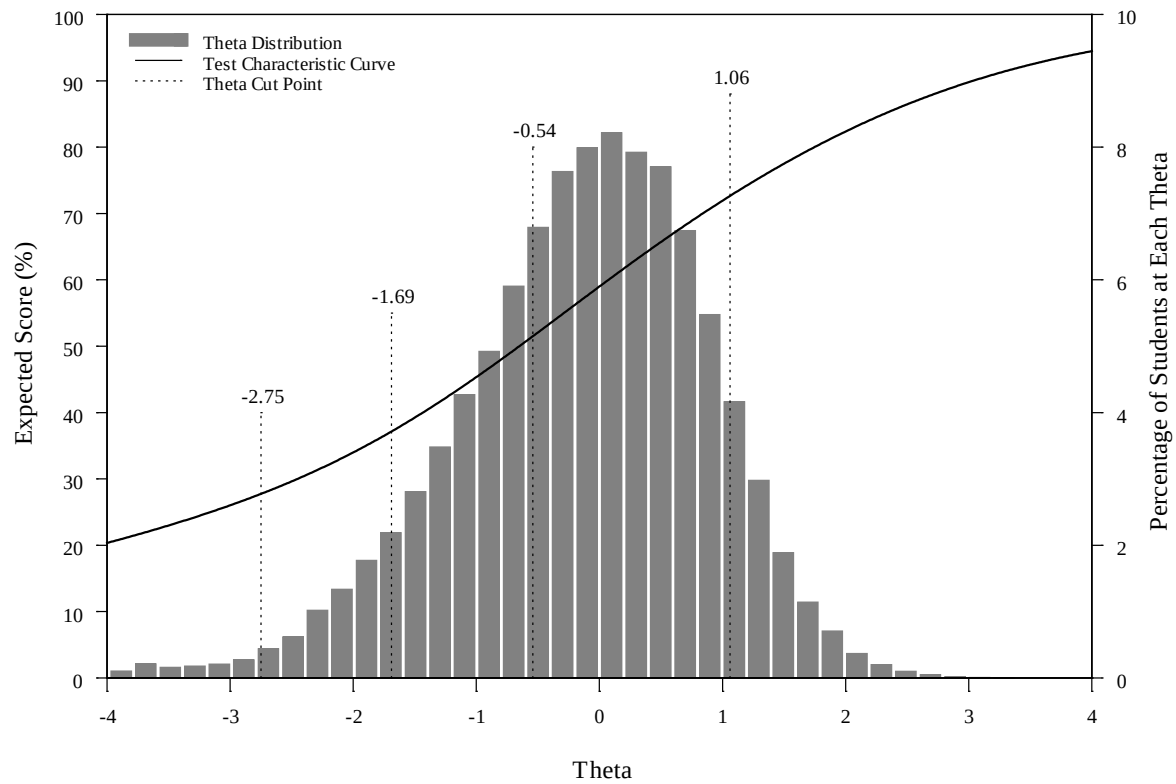


Figure 7.2 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (English)

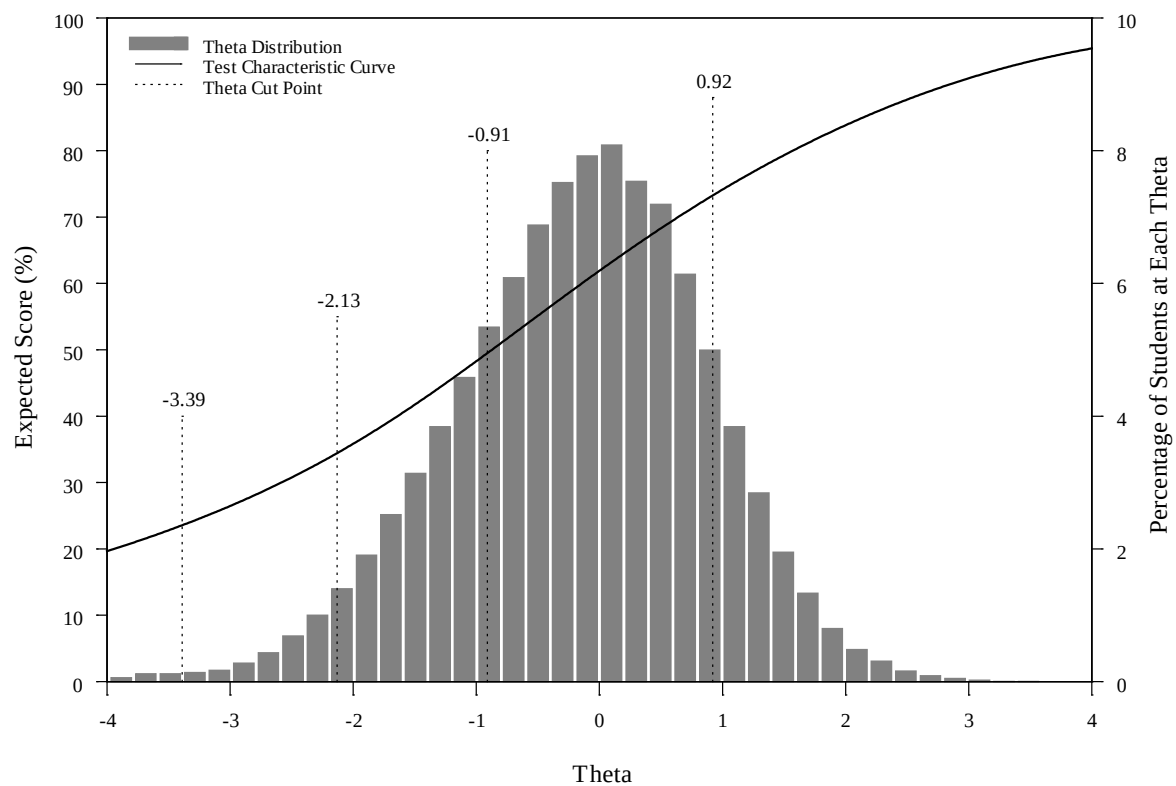


Figure 7.3 Test Characteristic Curve and Distribution of Thetas for Reading: Primary Division (French)

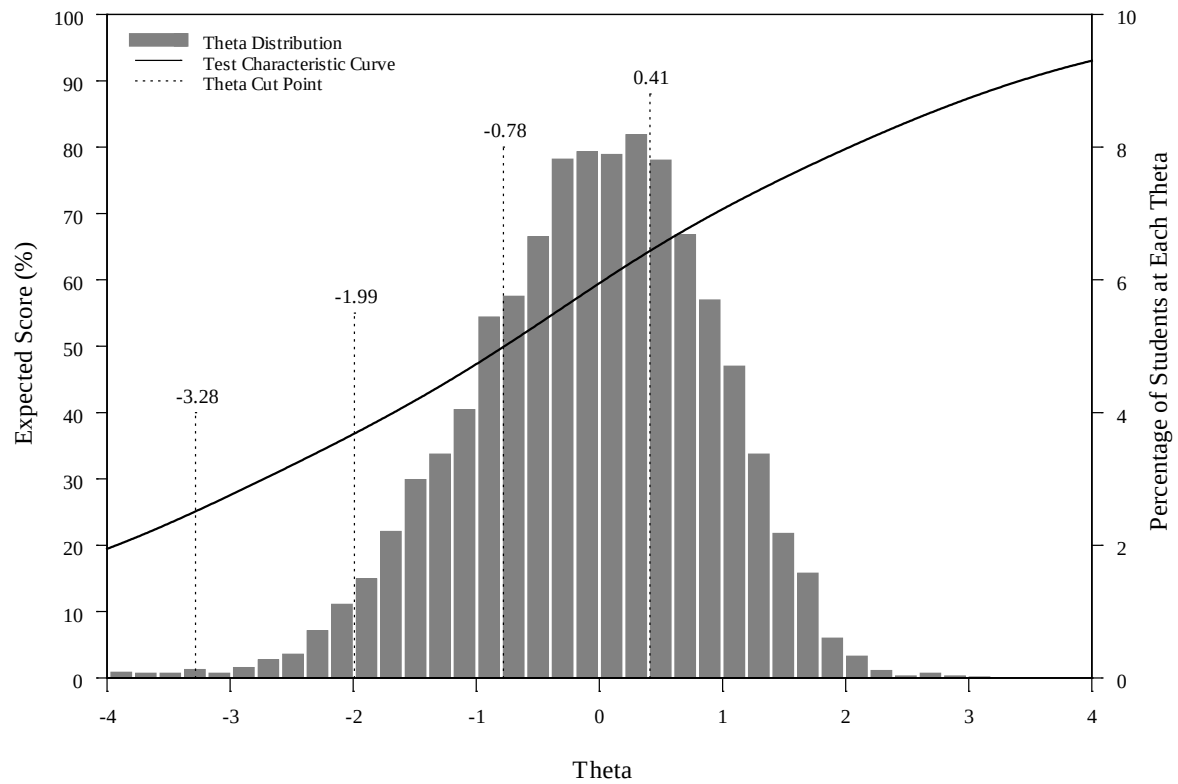


Figure 7.4 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (French)

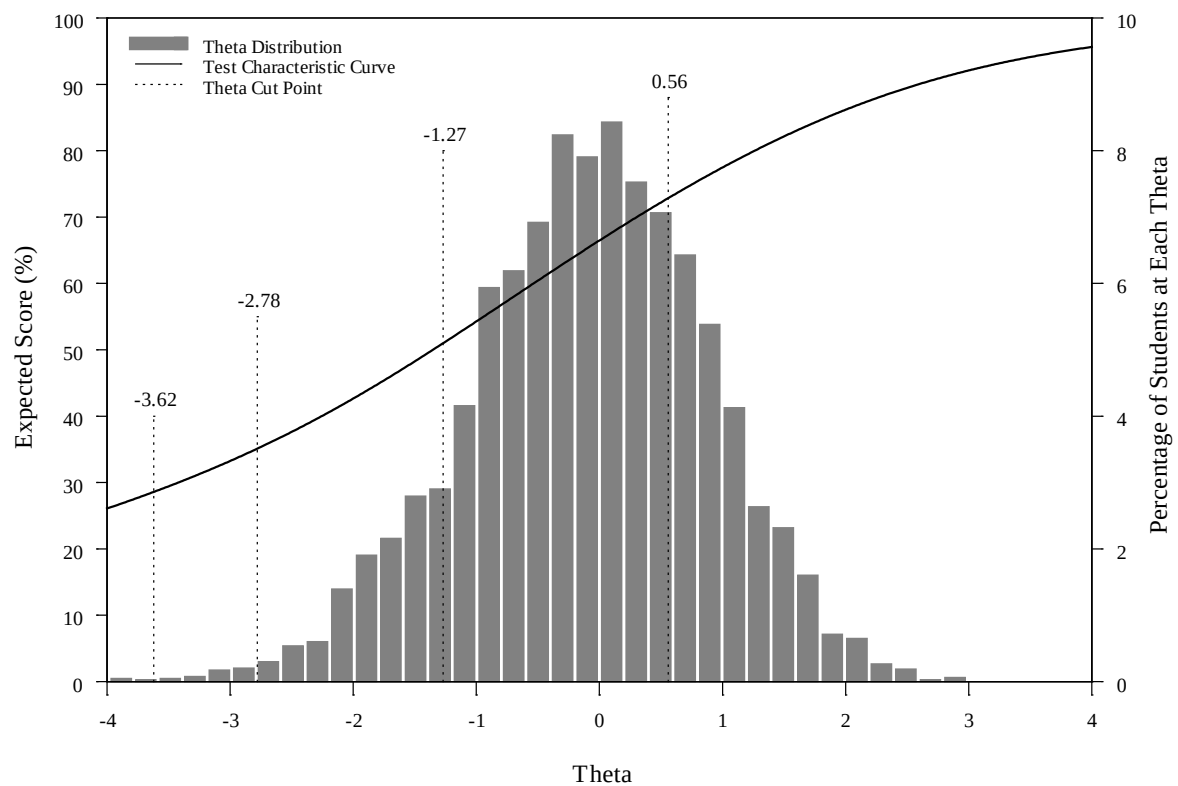


Figure 7.5 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (English)

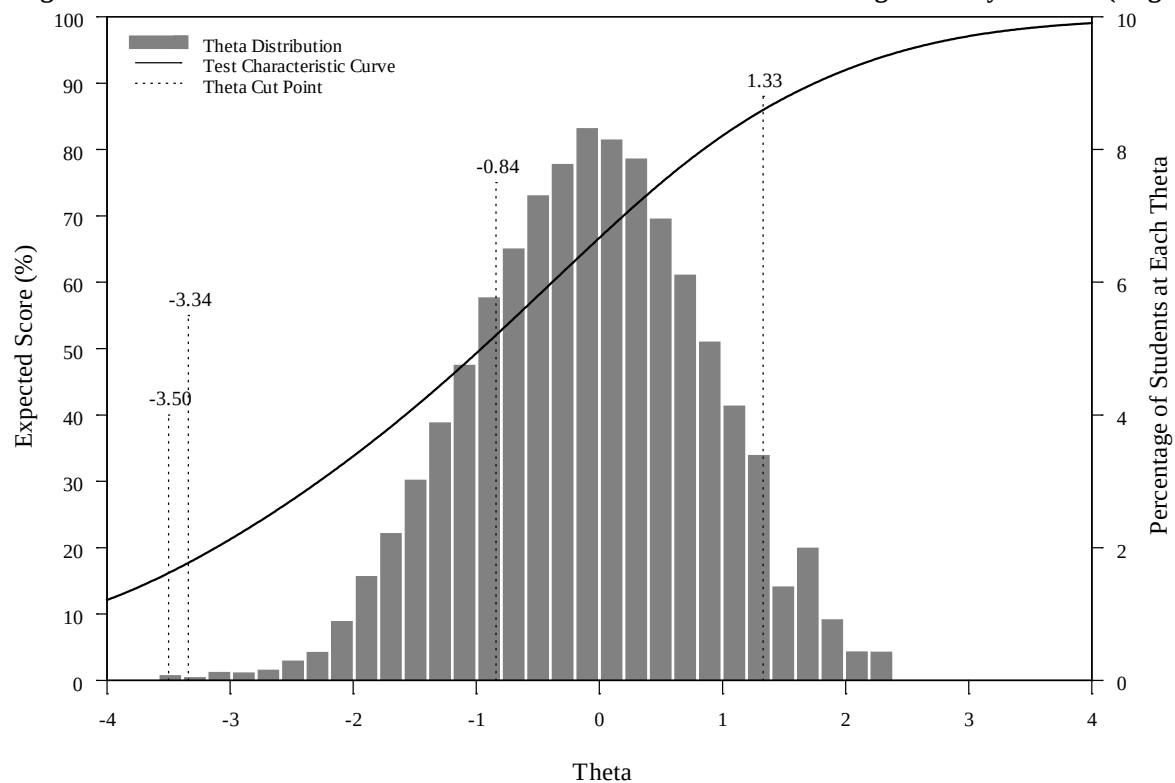


Figure 7.6 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (English)

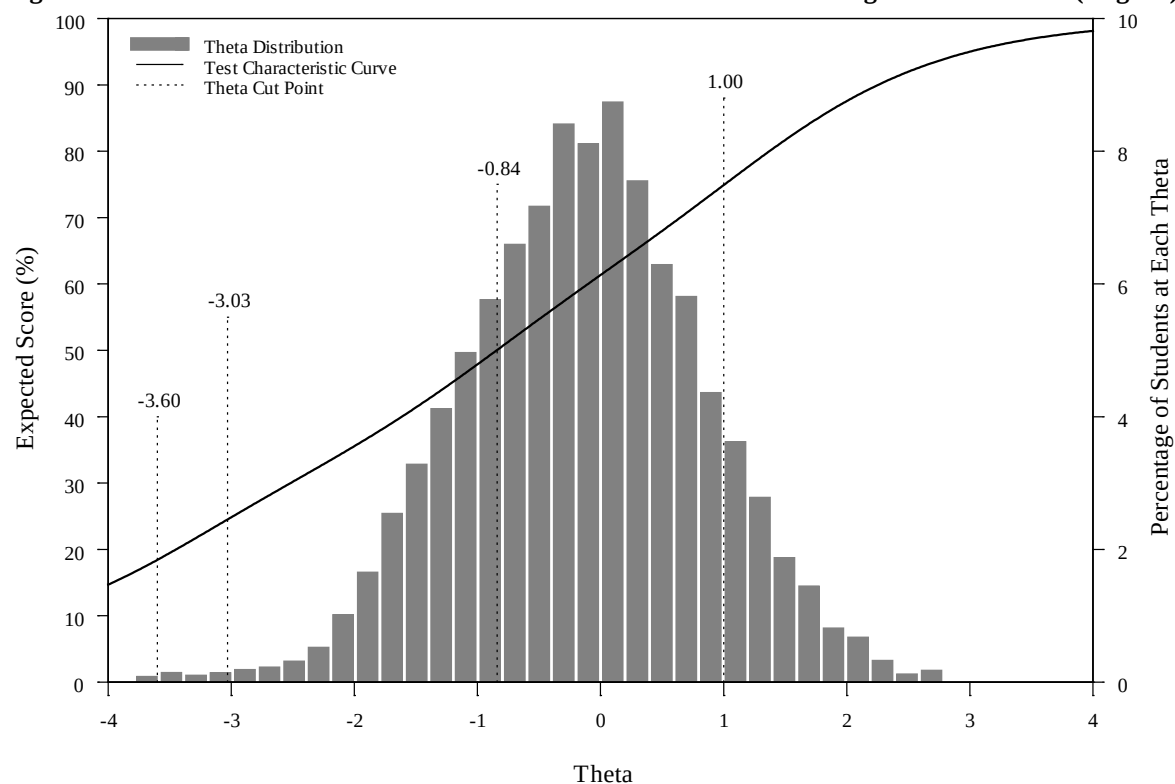


Figure 7.7 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (French)

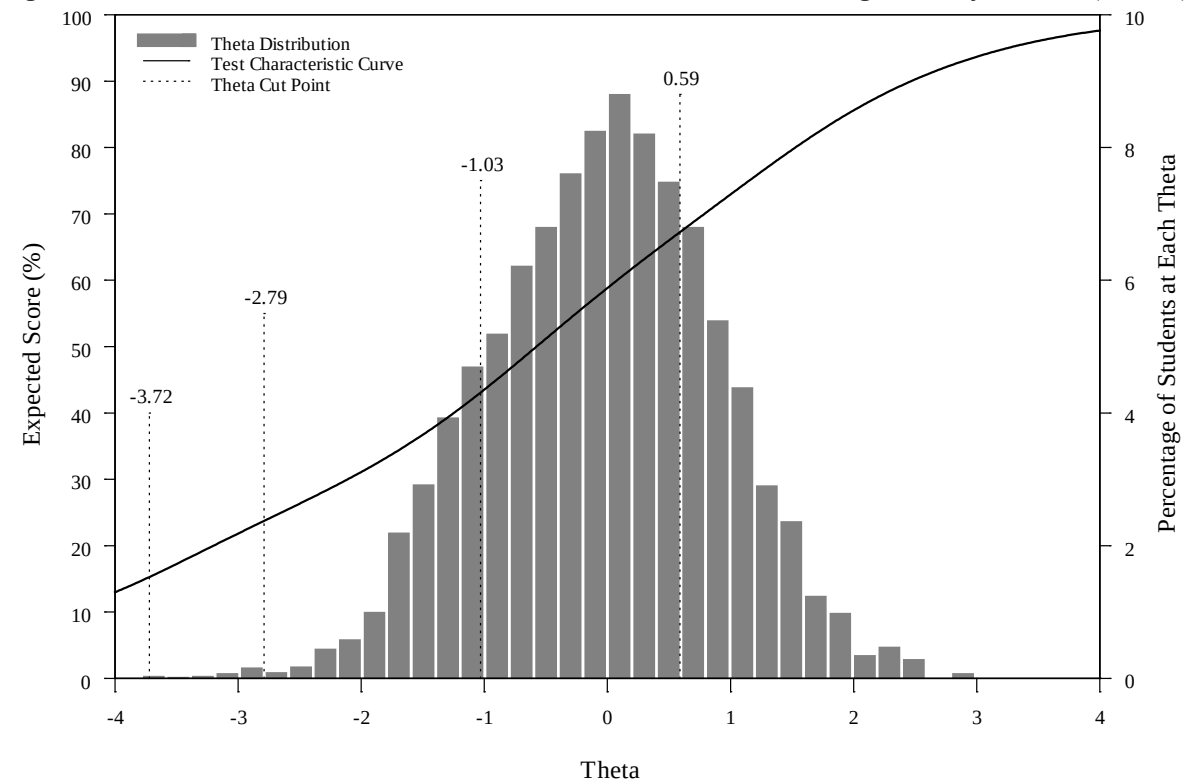


Figure 7.8 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (French)

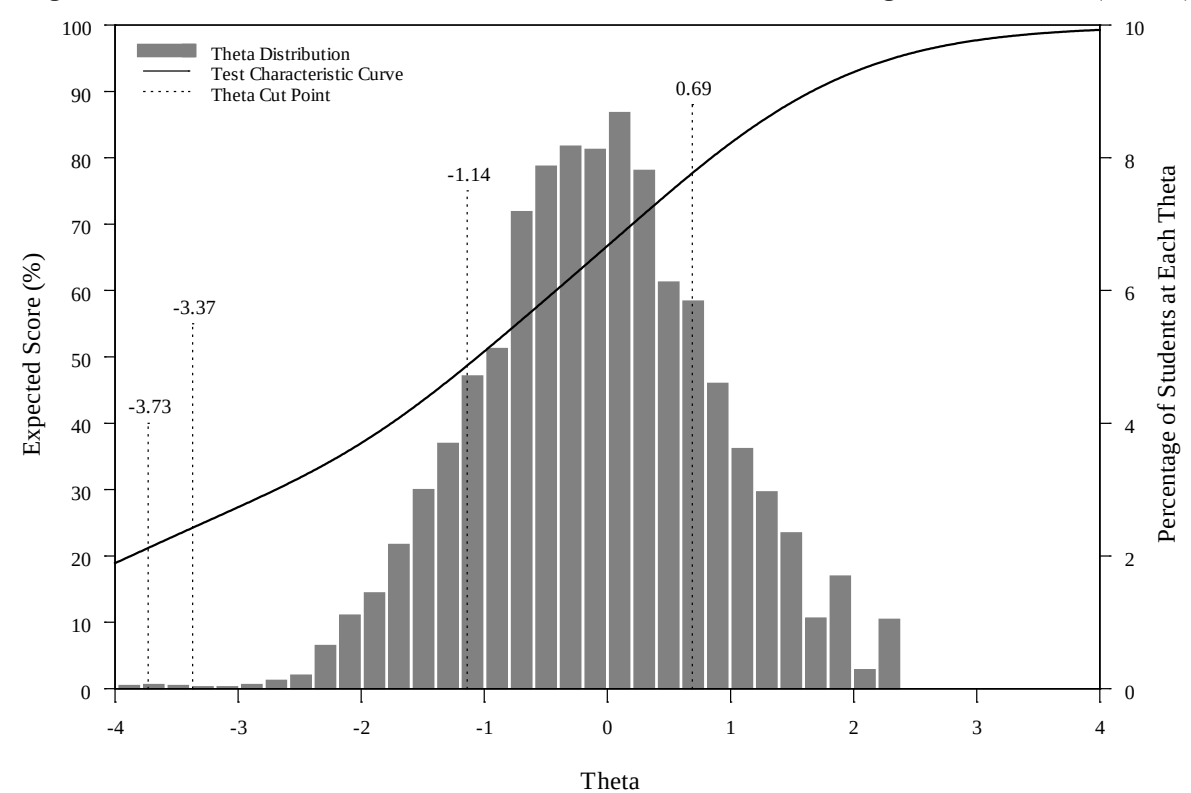


Figure 7.9 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (English)

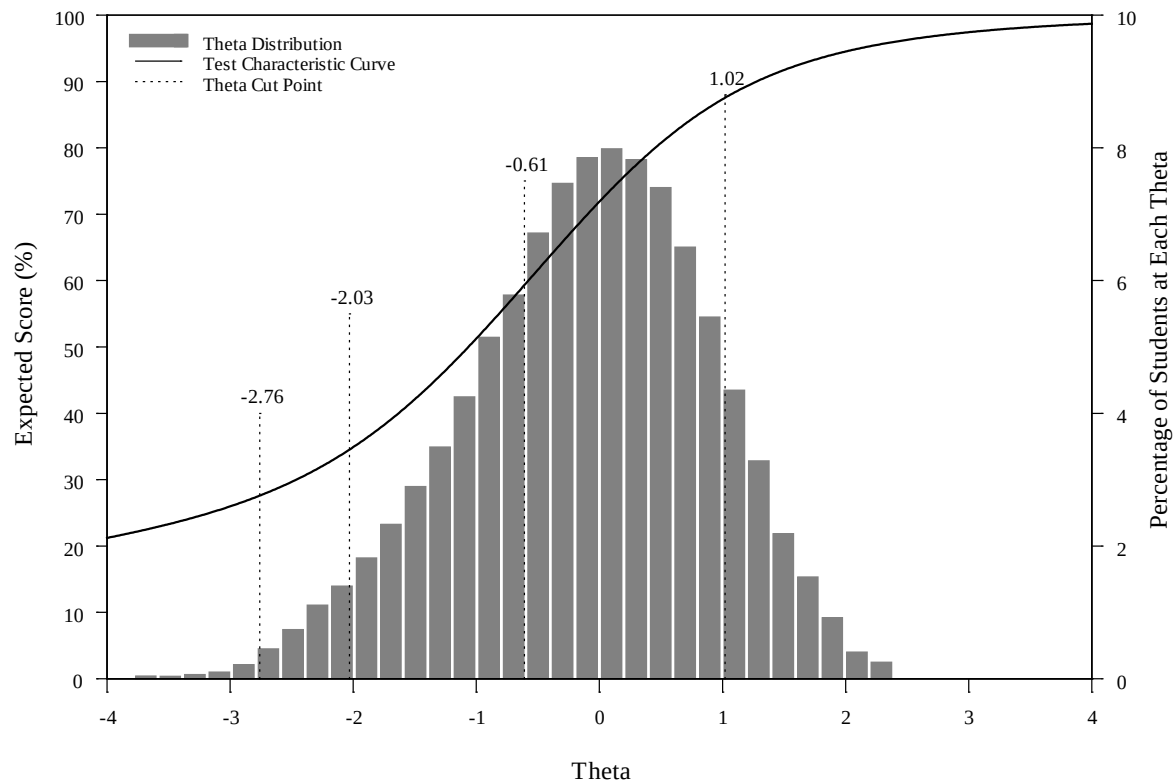


Figure 7.10 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (English)

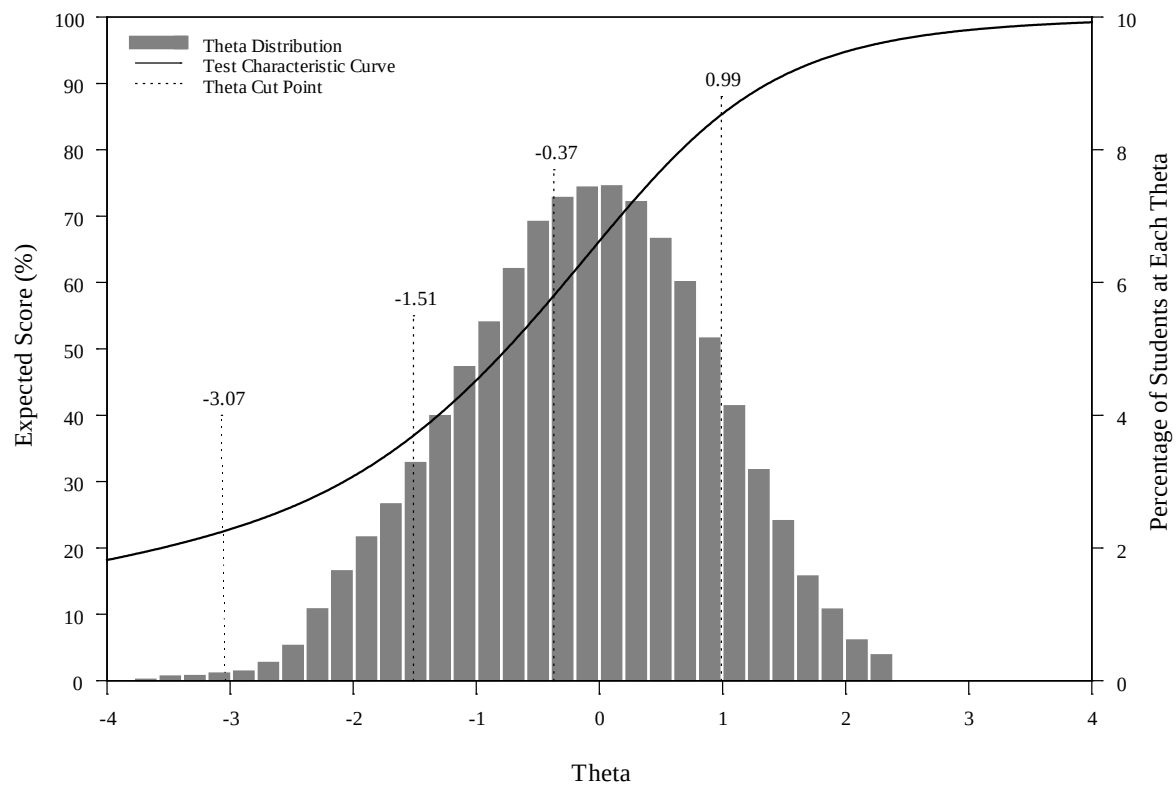


Figure 7.11 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (French)

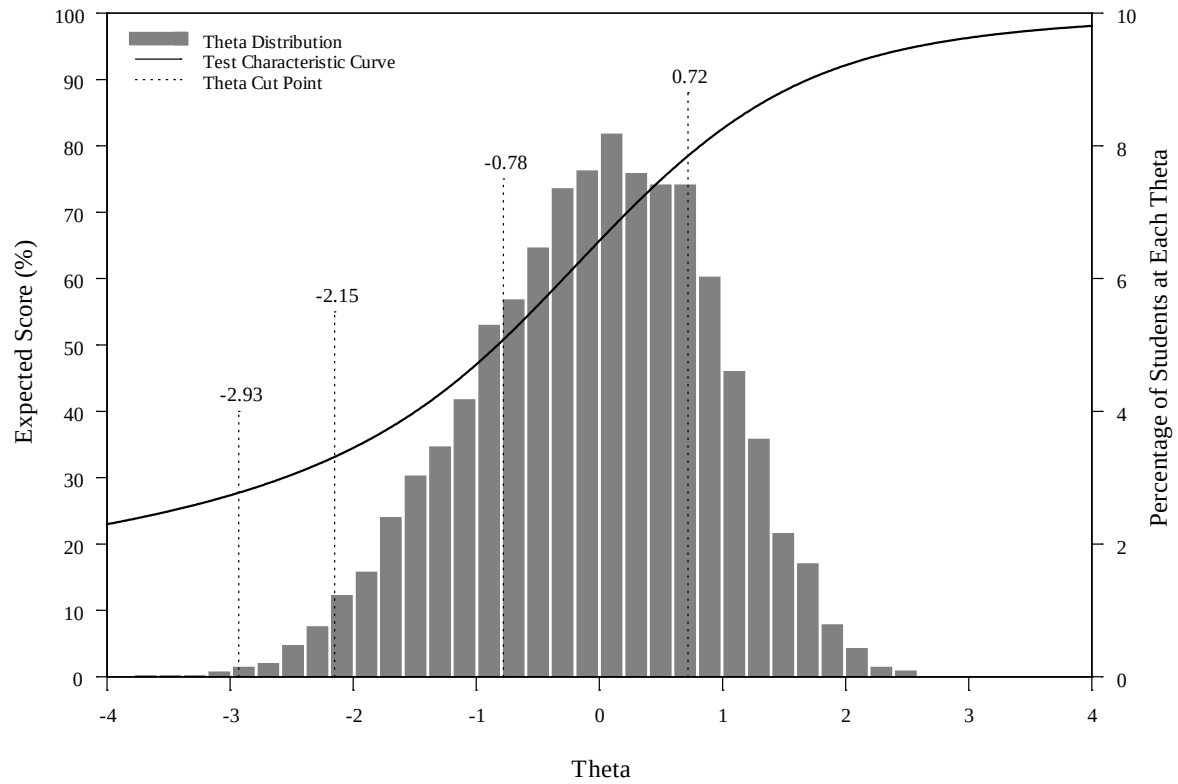


Figure 7.12 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (French)

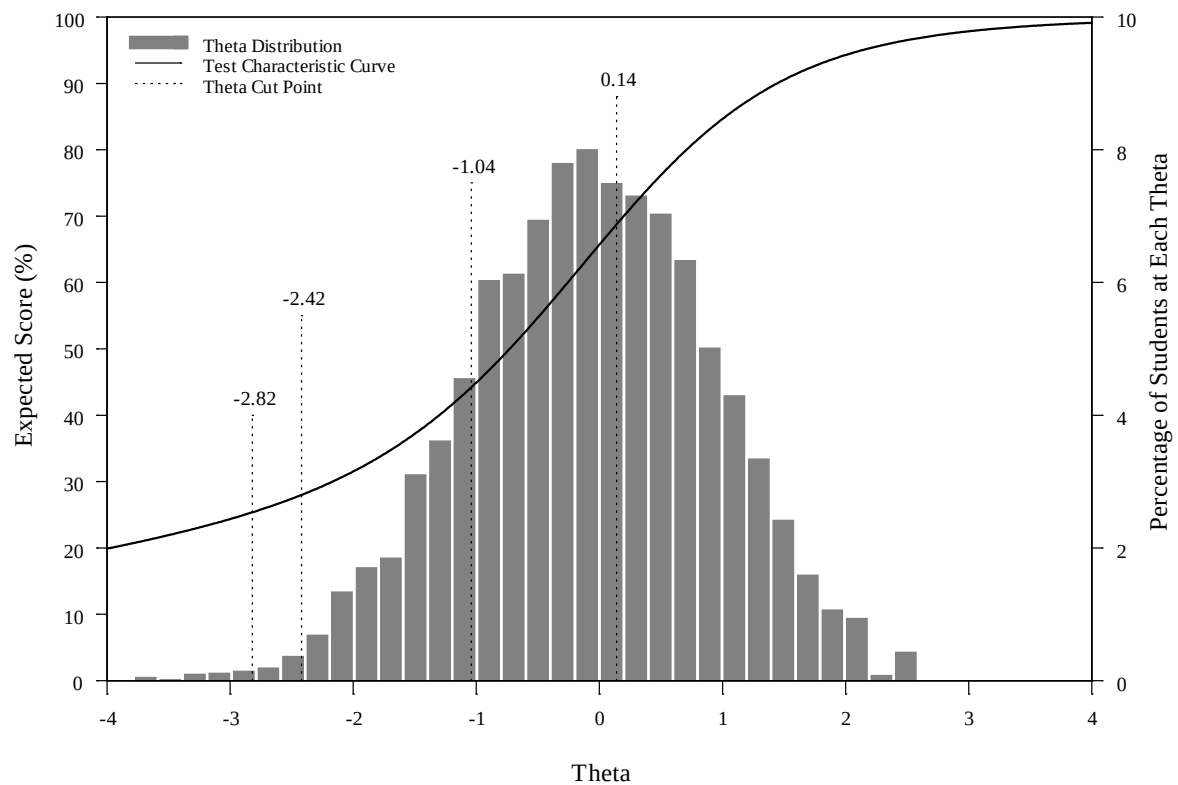


Figure 7.13 Test Information Function for Reading: Primary Division (English)

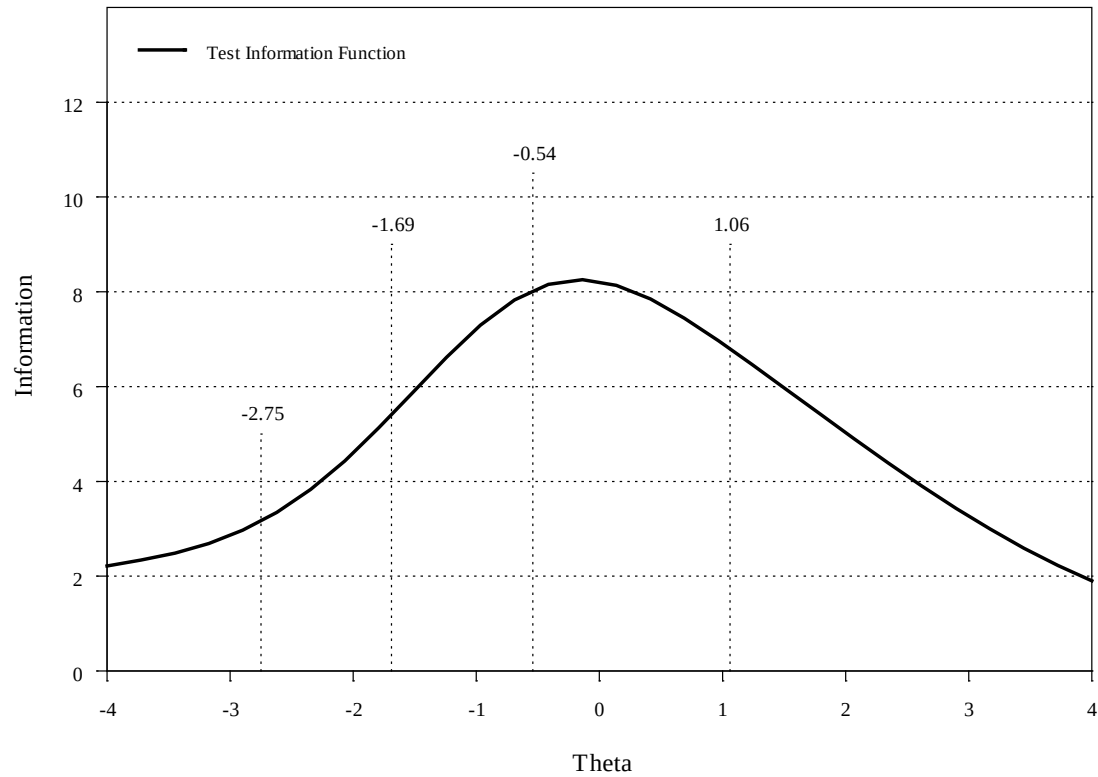


Figure 7.14 Test Information Function for Reading: Junior Division (English)

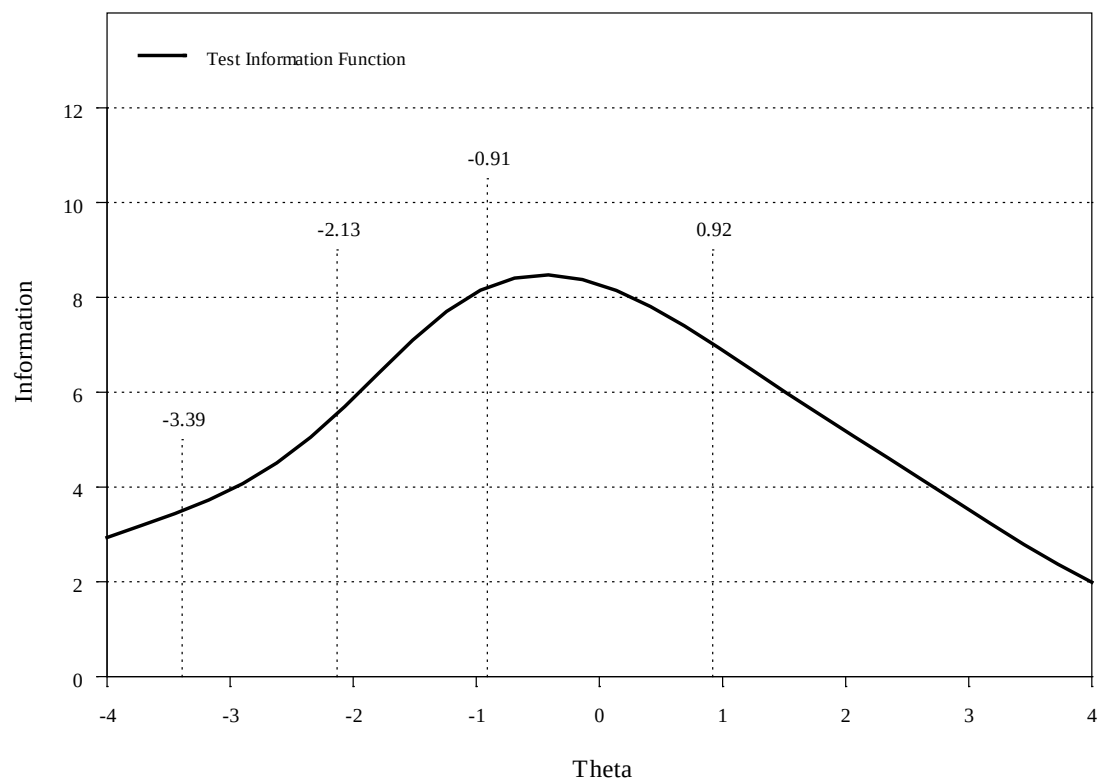


Figure 7.15 Test Information Function for Reading: Primary Division (French)

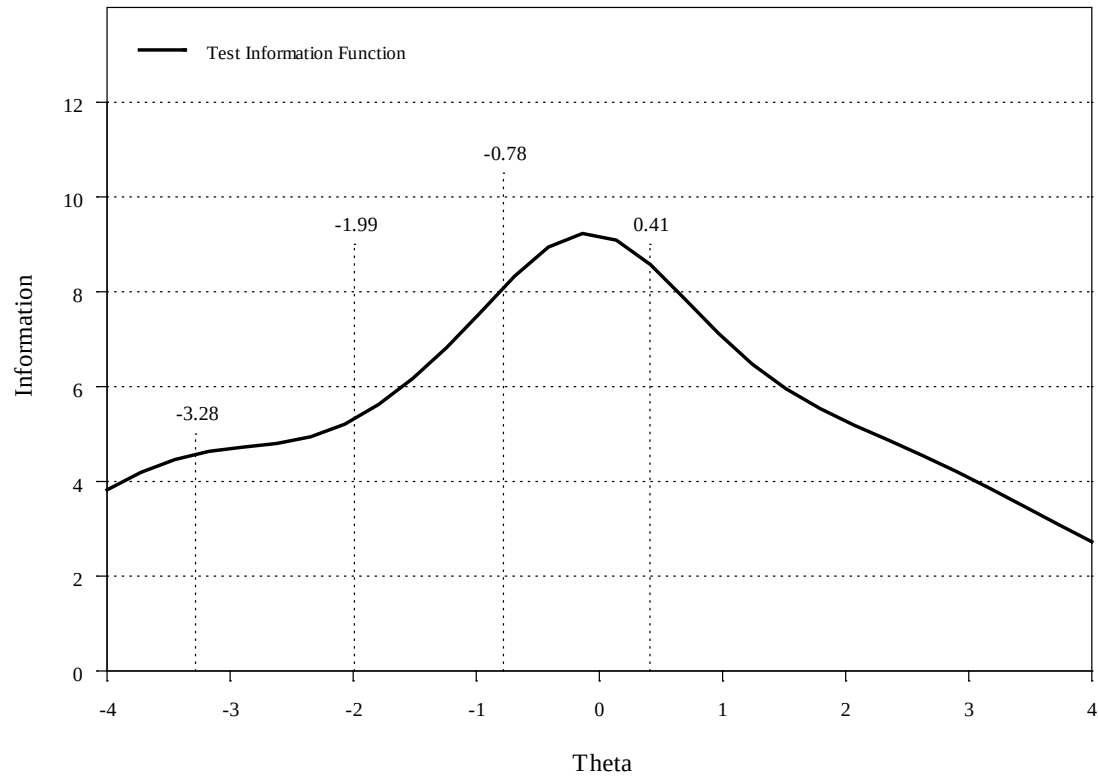


Figure 7.16 Test Information Function for Reading: Junior Division (French)

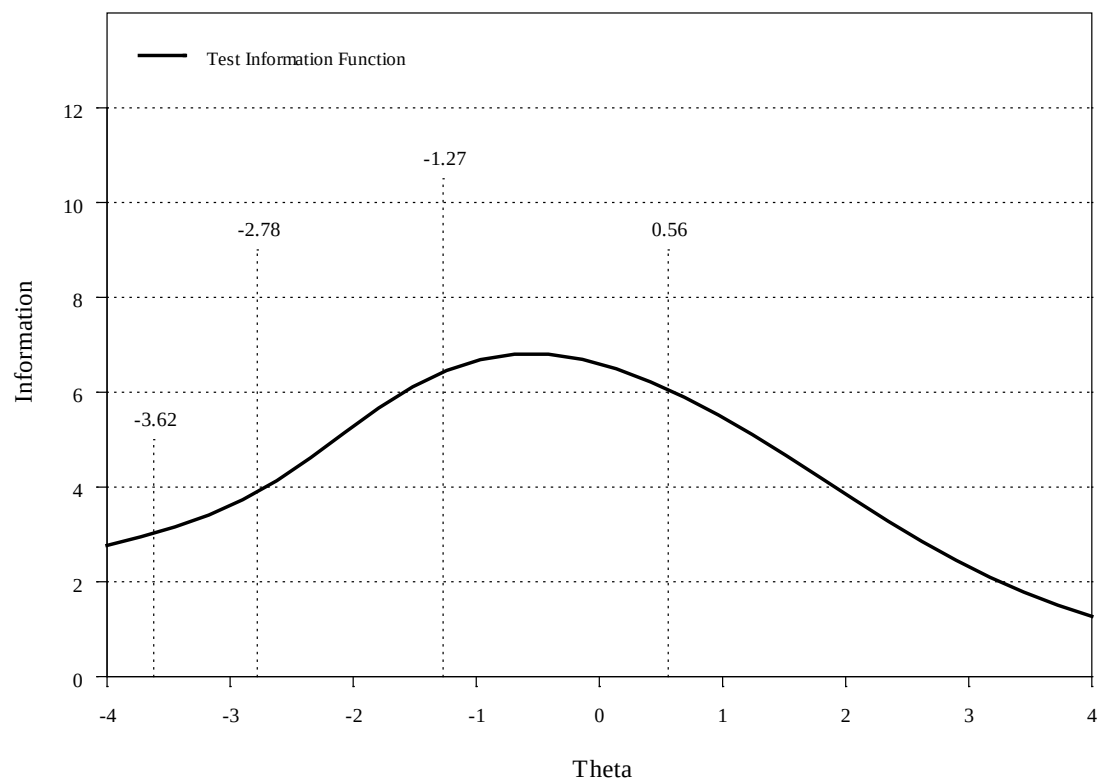


Figure 7.17 Test Information Function for Writing: Primary Division (English)

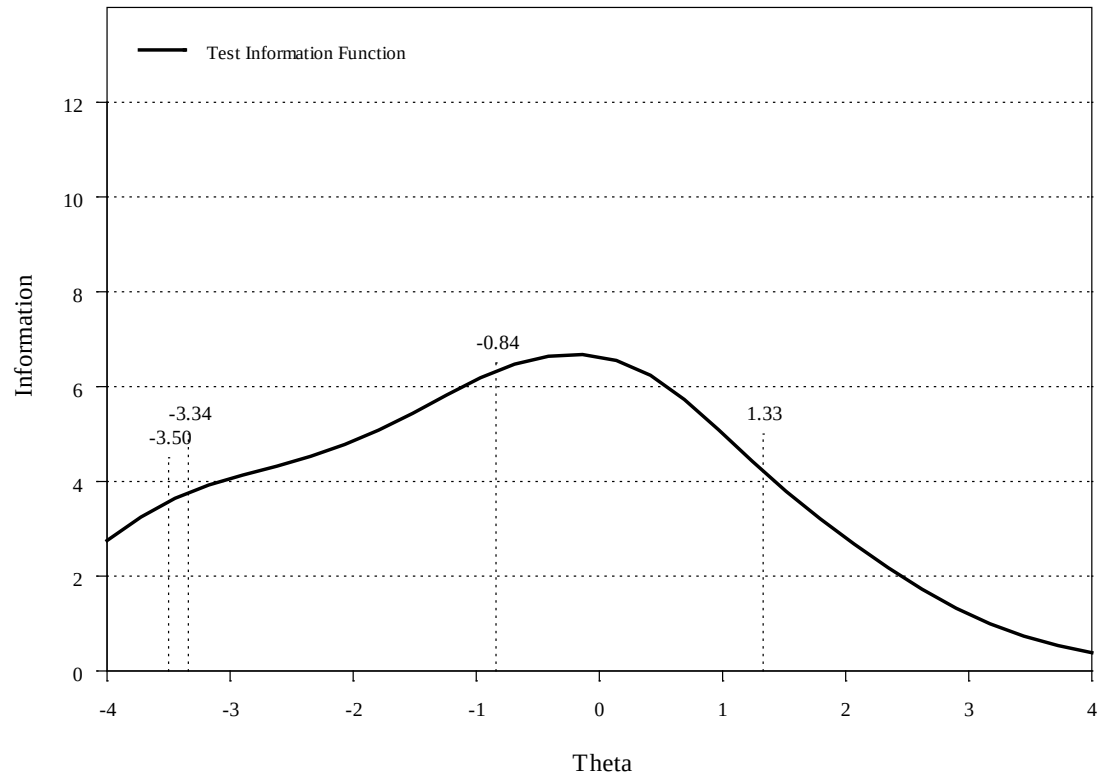


Figure 7.18 Test Information Function for Writing: Junior Division (English)

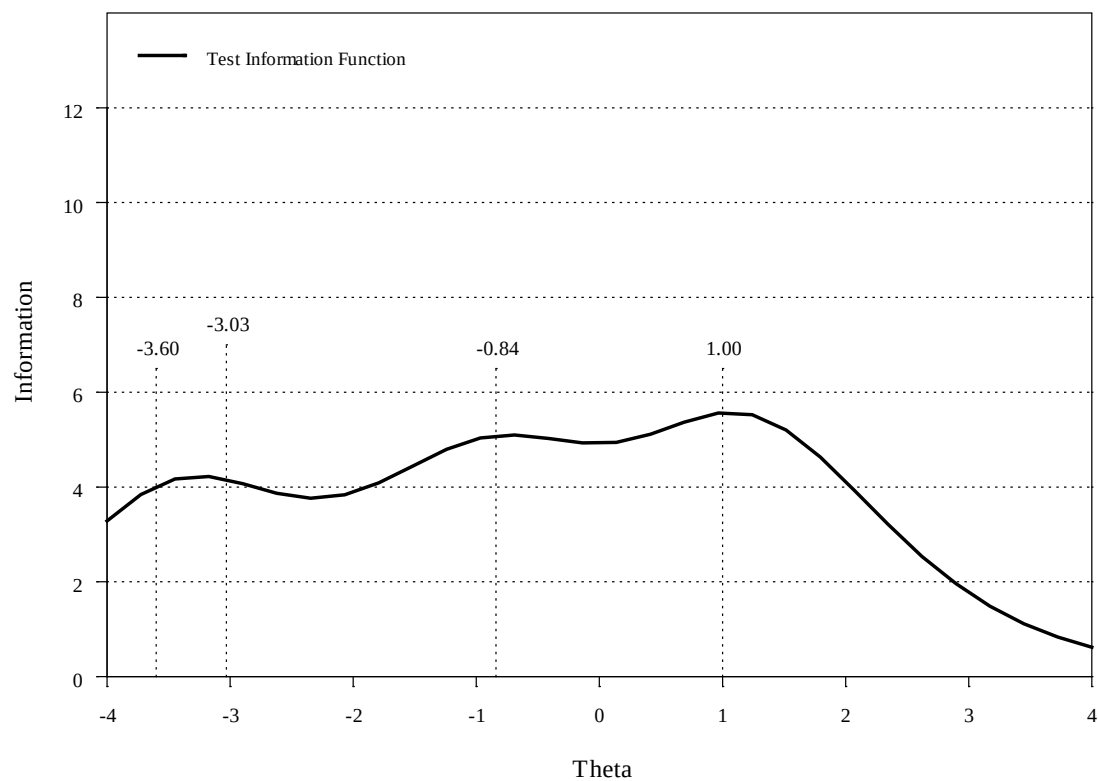


Figure 7.19 Test Information Function for Writing: Primary Division (French)

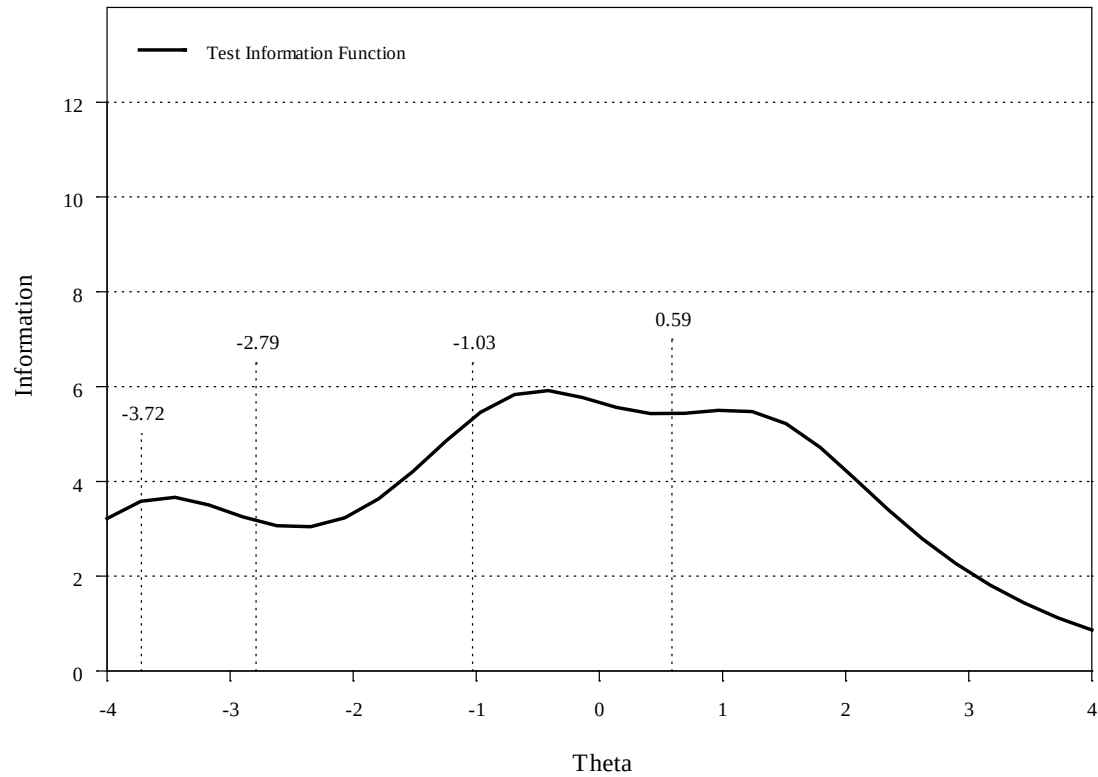


Figure 7.20 Test Information Function for Writing: Junior Division (French)

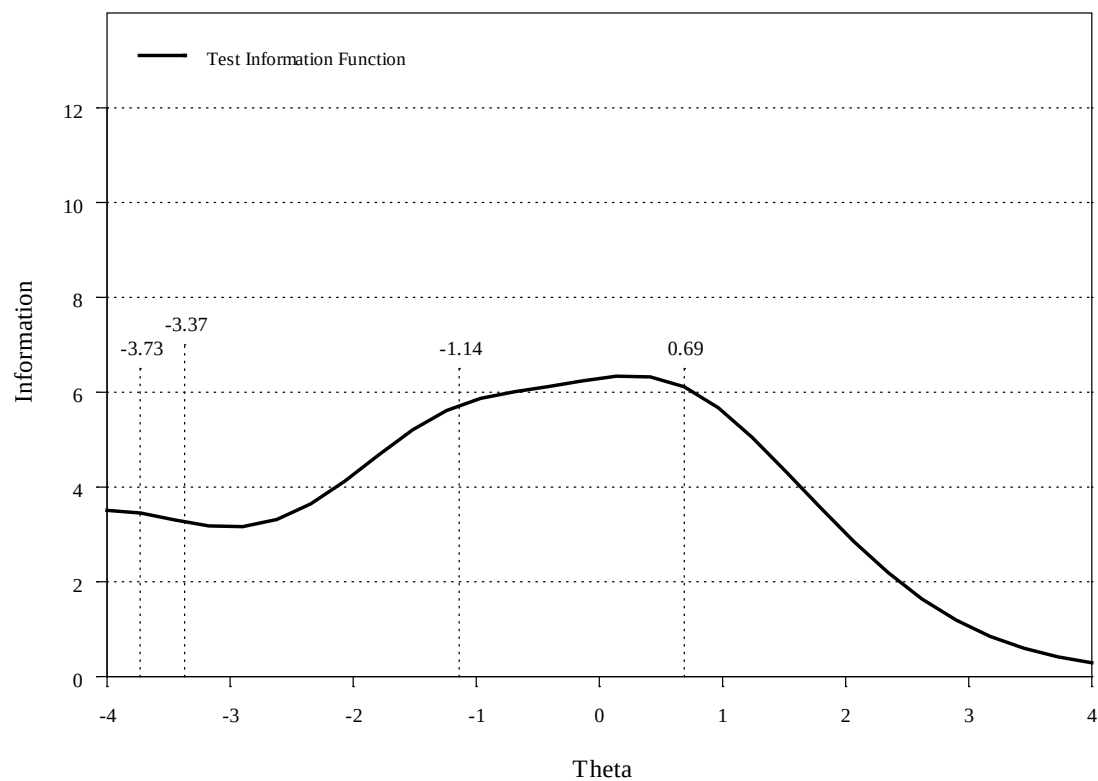


Figure 7.21 Test Information Function for Mathematics: Primary Division (English)

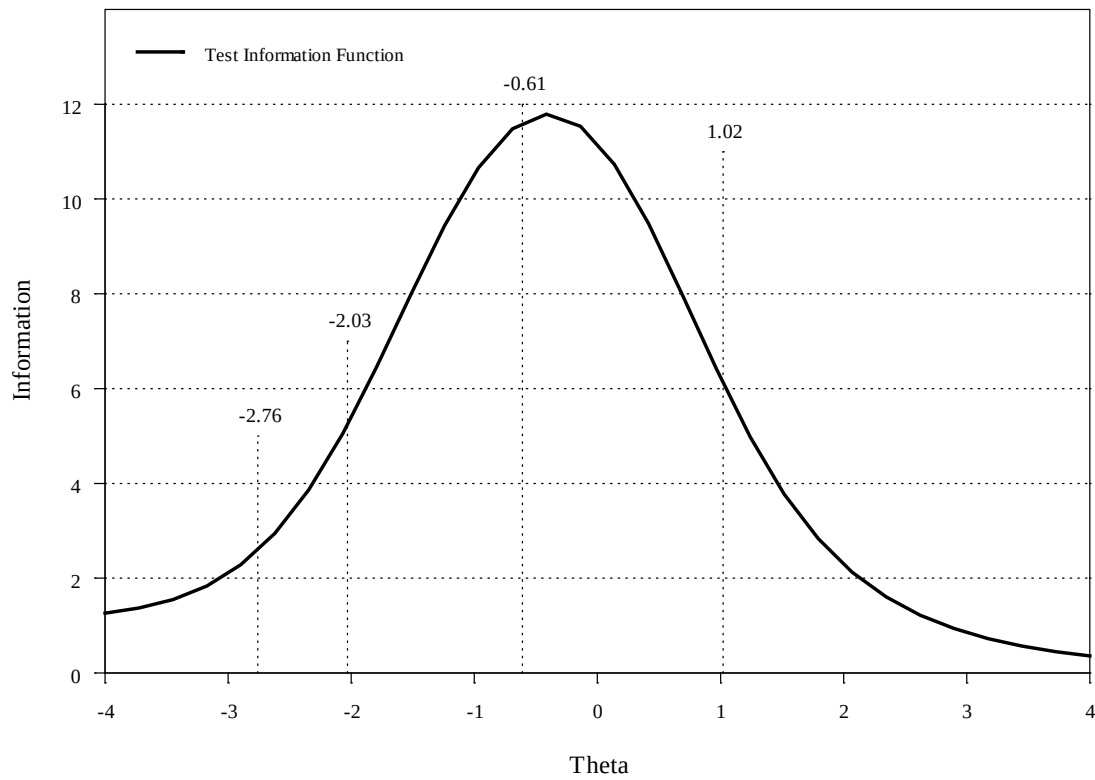


Figure 7.22 Test Information Function for Mathematics: Junior Division (English)

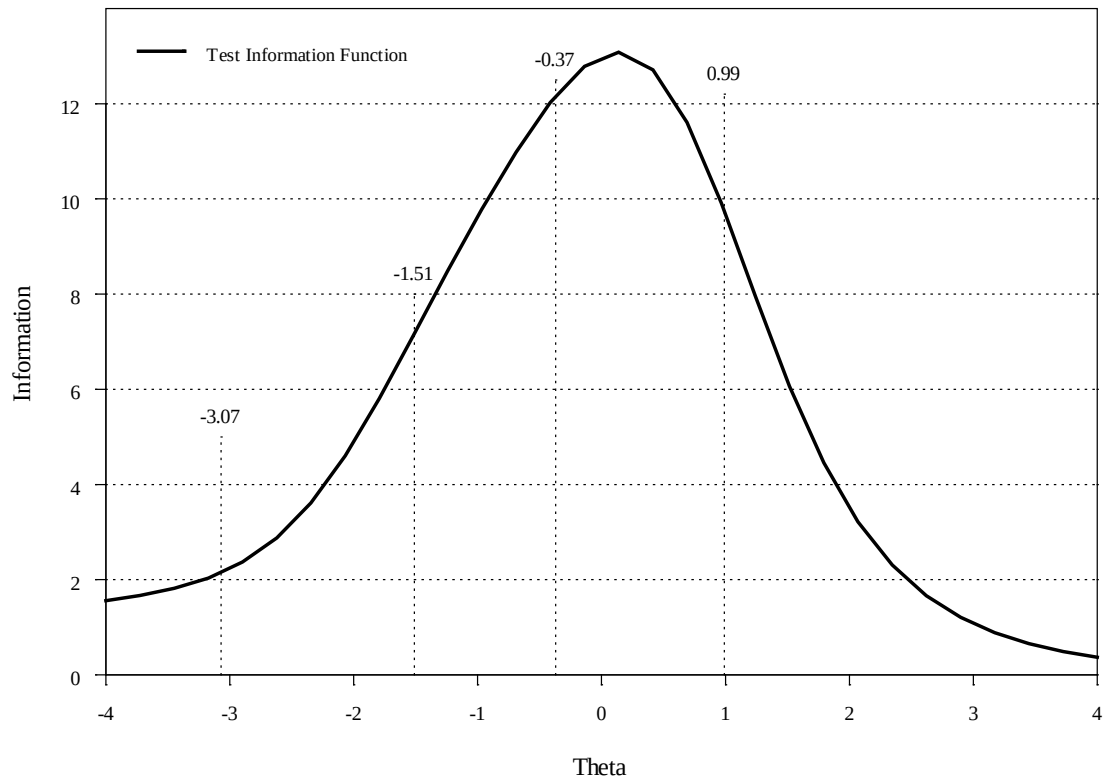


Figure 7.23 Test Information Function for Mathematics: Primary Division (French)

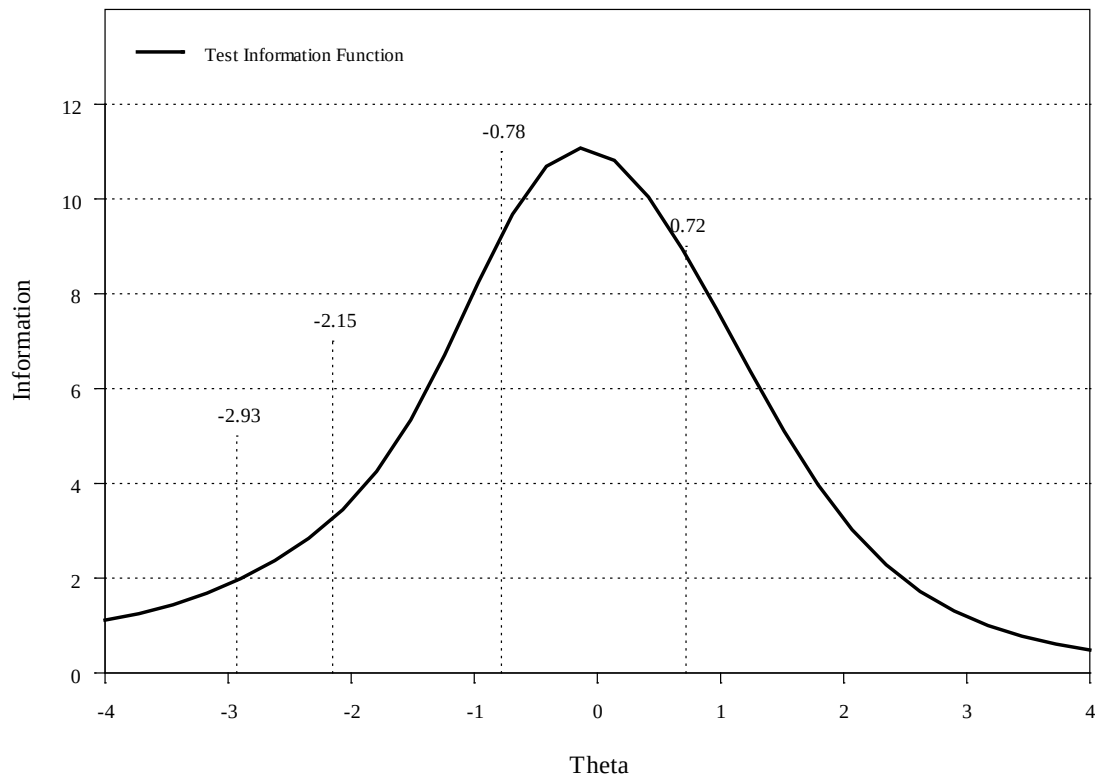
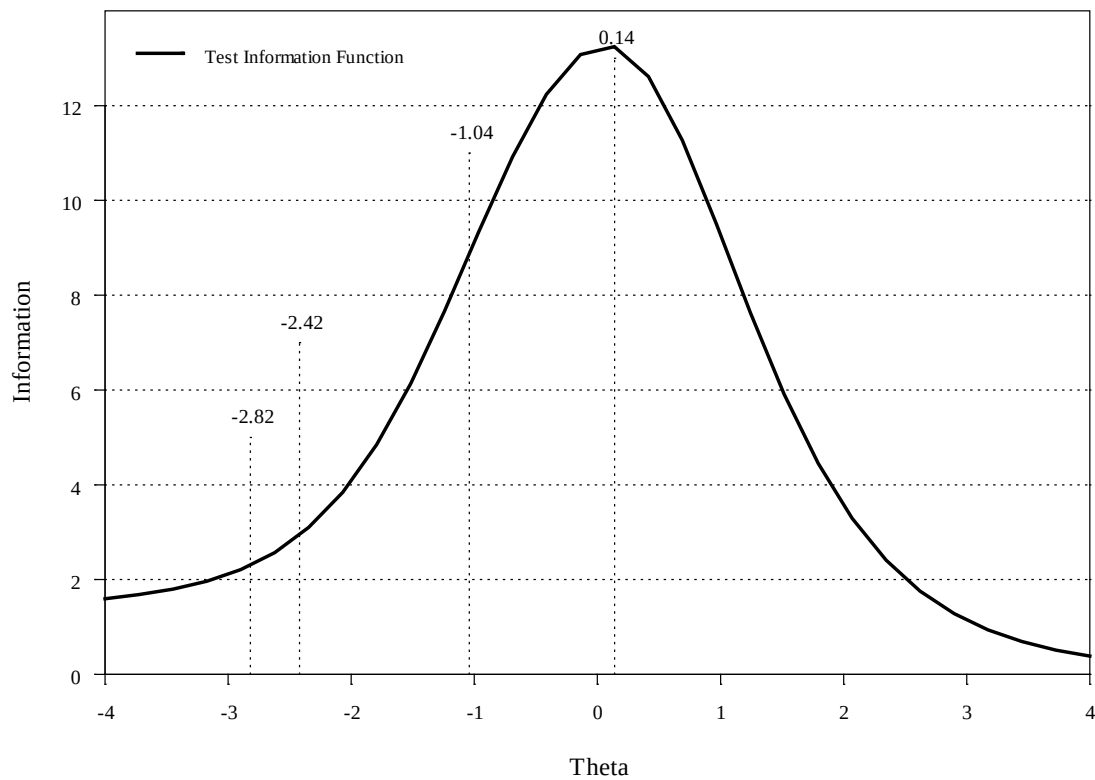


Figure 7.24 Test Information Function for Mathematics: Junior Division (French)



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.	0.39	0.21	0.36	-2.28
		Max.	0.89	0.51	0.96	1.21
		Mean	0.65	0.37**	0.61	-0.58
		SD	0.14	0.08	0.15	0.76
Junior Reading (English)	36	Min.	0.47	0.15	0.25	-2.14
		Max.	0.87	0.56	0.98	0.94
		Mean	0.68	0.37	0.58	-0.79
		SD	0.13	0.10	0.15	0.71
Primary Reading (French)	36	Min.	0.34	0.18	0.39	-2.20
		Max.	0.90	0.50	1.24	1.91
		Mean	0.64	0.38	0.73	-0.44
		SD	0.14	0.08	0.24	0.81
Junior Reading (French)	36	Min.	0.38	0.20	0.30	-2.75
		Max.	0.94	0.49	1.01	1.63
		Mean	0.68	0.33	0.58	-0.91
		SD	0.14	0.08	0.19	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.

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.	0.55	0.30	0.48	-1.48
		Max.	0.74	0.60	1.04	0.08
		Mean	0.66	0.47**	0.72	-0.79
		SD	0.06	0.12	0.18	0.42
Junior Writing (English)	14	Min.	0.42	0.29	0.46	-2.24
		Max.	0.85	0.62	1.19	0.87
		Mean	0.62	0.44	0.70	-0.76
		SD	0.13	0.13	0.24	0.84
Primary Writing (French)	14	Min.	0.45	0.28	0.49	-1.88
		Max.	0.86	0.59	1.12	0.39
		Mean	0.63	0.44	0.73	-0.62
		SD	0.13	0.12	0.19	0.65
Junior Writing (French)	14	Min.	0.51	0.24	0.34	-1.98
		Max.	0.76	0.62	1.04	0.35
		Mean	0.65	0.45	0.71	-0.91
		SD	0.07	0.15	0.21	0.67
Primary Mathematics (English)	36	Min.	0.44	0.25	0.29	-2.41
		Max.	0.87	0.61	1.17	0.55
		Mean	0.70	0.41	0.71	-0.97
		SD	0.11	0.11	0.24	0.73
Junior Mathematics (English)	36	Min.	0.33	0.27	0.38	-2.01
		Max.	0.87	0.66	1.28	1.03
		Mean	0.63	0.42	0.76	-0.58
		SD	0.13	0.14	0.25	0.76
Primary Mathematics (French)	36	Min.	0.36	0.13	0.21	-2.71
		Max.	0.93	0.57	1.25	1.05
		Mean	0.67	0.39	0.71	-0.91
		SD	0.14	0.11	0.27	1.01
Junior Mathematics (French)	36	Min.	0.40	0.29	0.26	-2.21
		Max.	0.88	0.63	1.32	0.94
		Mean	0.64	0.42	0.76	-0.58
		SD	0.11	0.12	0.24	0.80

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 2011–2012 the mean percentages ranged from 52.10% (French, winter) to 54.58% (English, spring) for applied mathematics. The mean percentages ranged from 65.09% (French, winter) to 71.90% (English, spring) for academic mathematics.

Cronbach's alpha estimates ranged from 0.82 to 0.85 for applied mathematics and from 0.86 to 0.87 for academic mathematics. The corresponding SEMs were likewise similar, ranging from 7.23% to 7.52% of the possible maximum score for applied mathematics and from 6.62% to 7.35% 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	18 505	1	52	27.88	9.33	0.84	3.76
Applied, Spring (English)	31	24	7	52	21 339	0	52	28.38	9.34	0.85	3.62
Academic, Winter (English)	31	24	7	52	42 784	1	52	36.19	9.57	0.86	3.56
Academic, Spring (English)	31	24	7	52	54 150	2	52	37.39	9.05	0.87	3.31
Applied, Winter (French)	31	24	7	52	441	1	50	27.09	8.84	0.82	3.75
Applied, Spring (French)	31	24	7	52	992	5	50	28.12	9.40	0.85	3.62
Academic, Winter (French)	31	24	7	52	1 442	5	52	33.85	10.25	0.87	3.68
Academic, Spring (French)	31	24	7	52	2 550	7	52	34.87	9.74	0.87	3.56

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 -0.10 to 0.03 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)	18 505	-3.17	2.82	-0.08	0.96
Applied, Spring (English)	21 339	-3.37	2.76	-0.04	0.98
Academic, Winter (English)	42 784	-3.54	2.08	-0.06	0.94
Academic, Spring (English)	54 150	-3.57	2.16	0.02	0.95
Applied, Winter (French)	441	-3.21	2.31	-0.14	0.93
Applied, Spring (French)	992	-2.81	2.65	-0.04	0.97
Academic, Winter (French)	1 442	-3.23	2.21	-0.10	0.95
Academic, Spring (French)	2 550	-3.02	2.29	0.03	0.95

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, whereas 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)

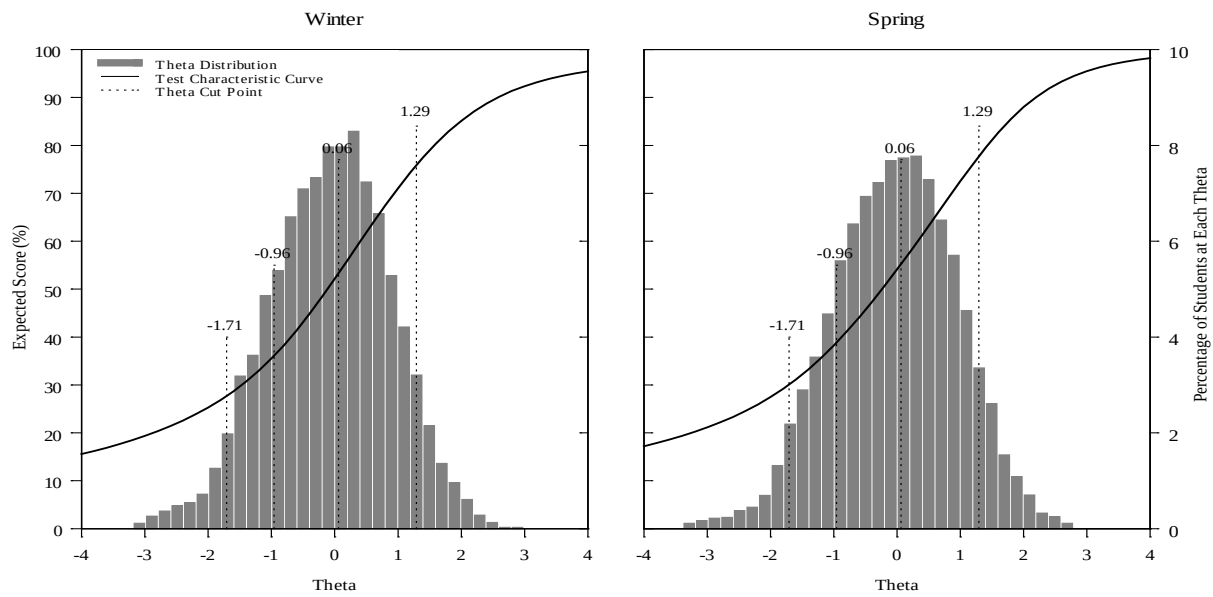


Figure 7.26 Test Characteristic Curves and Distributions of Thetas: Grade 9 Academic Math (English)

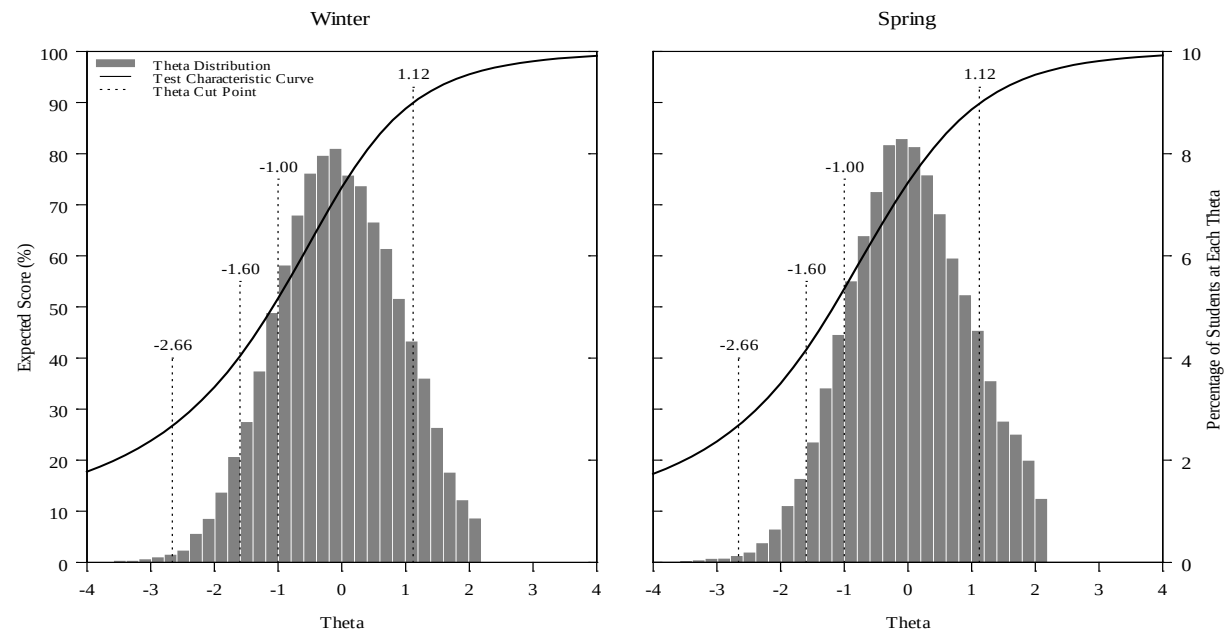


Figure 7.27 Test Characteristic Curves and Distributions of Thetas: Grade 9 Applied Math (French)

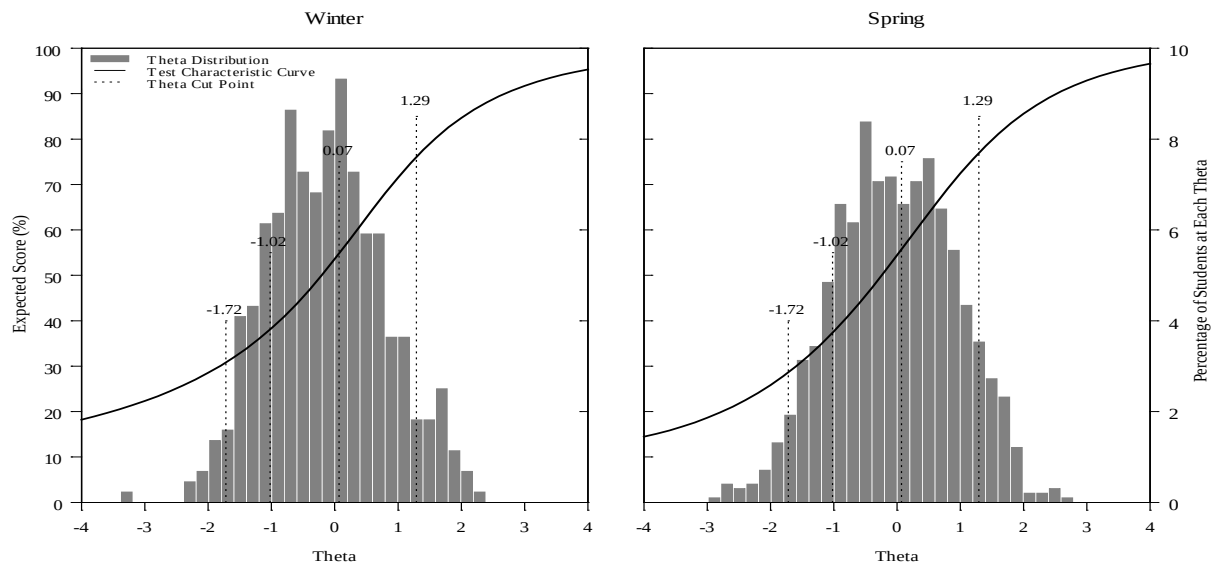


Figure 7.28 Test Characteristic Curves and Distributions of Thetas: Grade 9 Academic Math (French)

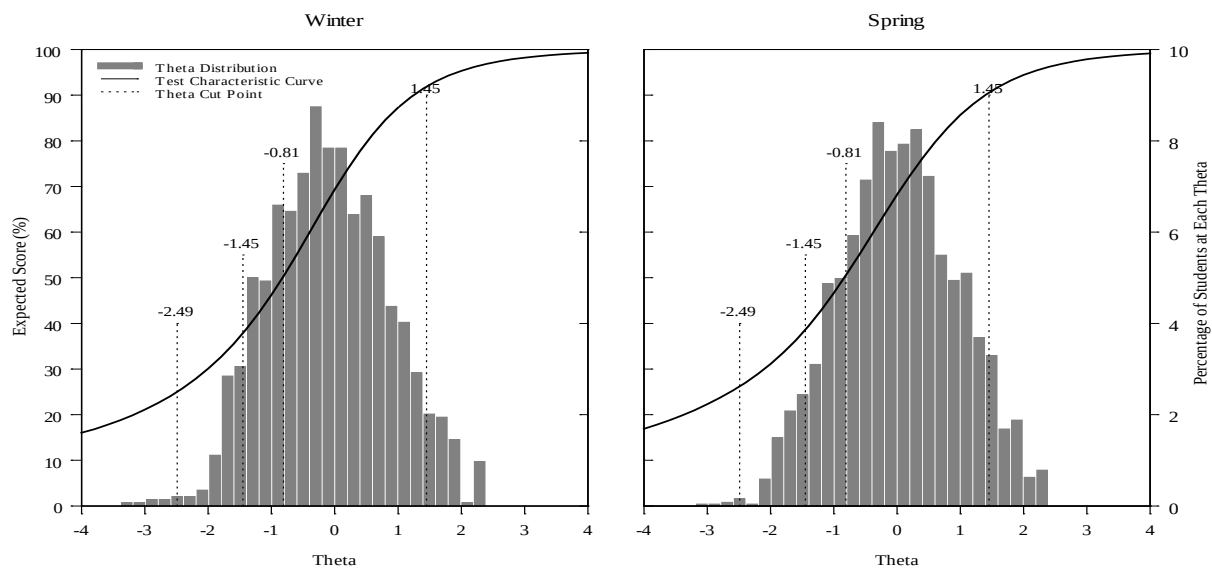


Figure 7.29 Test Information Functions: Grade 9 Applied Math (English)

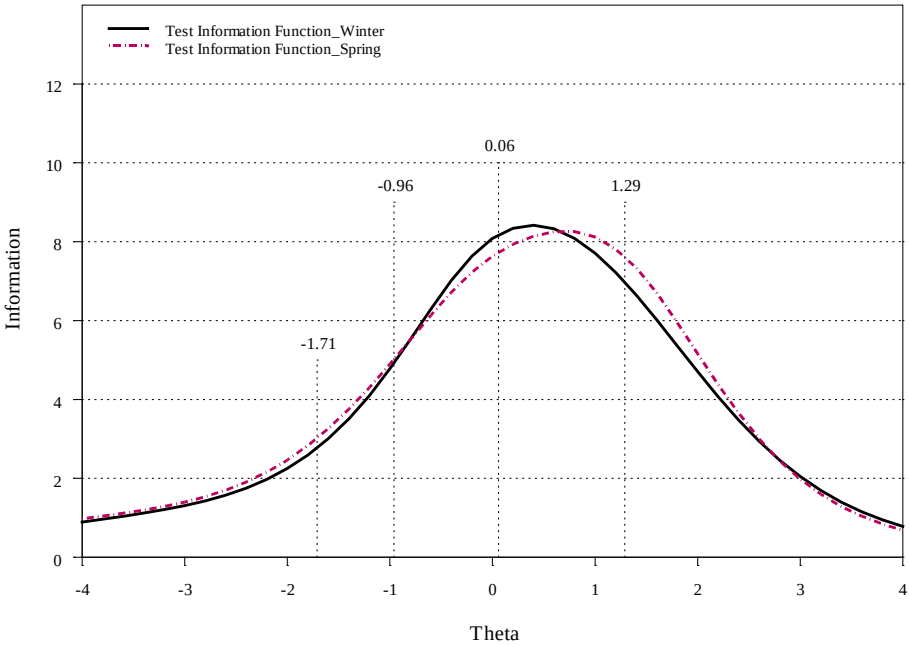


Figure 7.30 Test Information Functions: Grade 9 Academic Math (English)

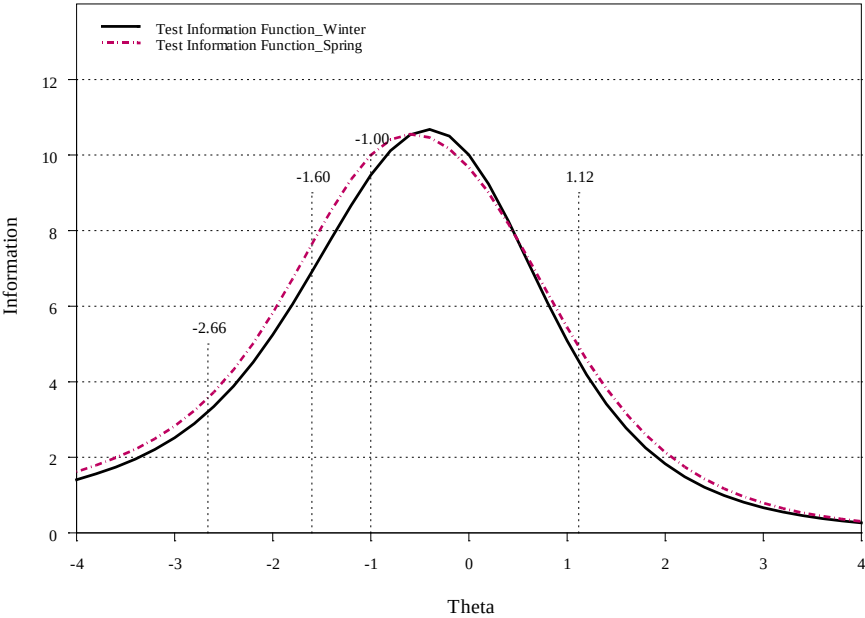


Figure 7.31 Test Information Functions: Grade 9 Applied Math (French)

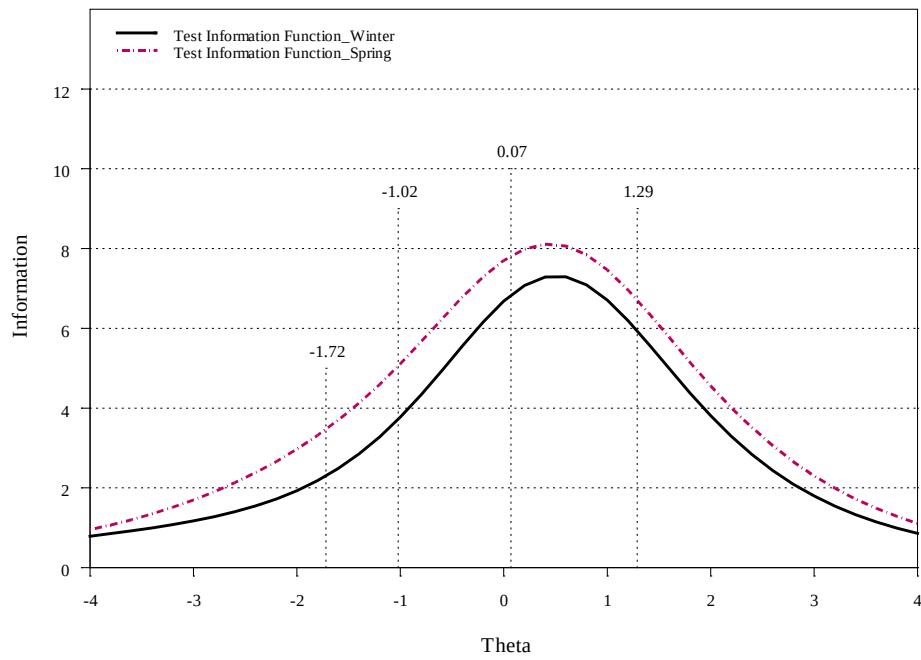
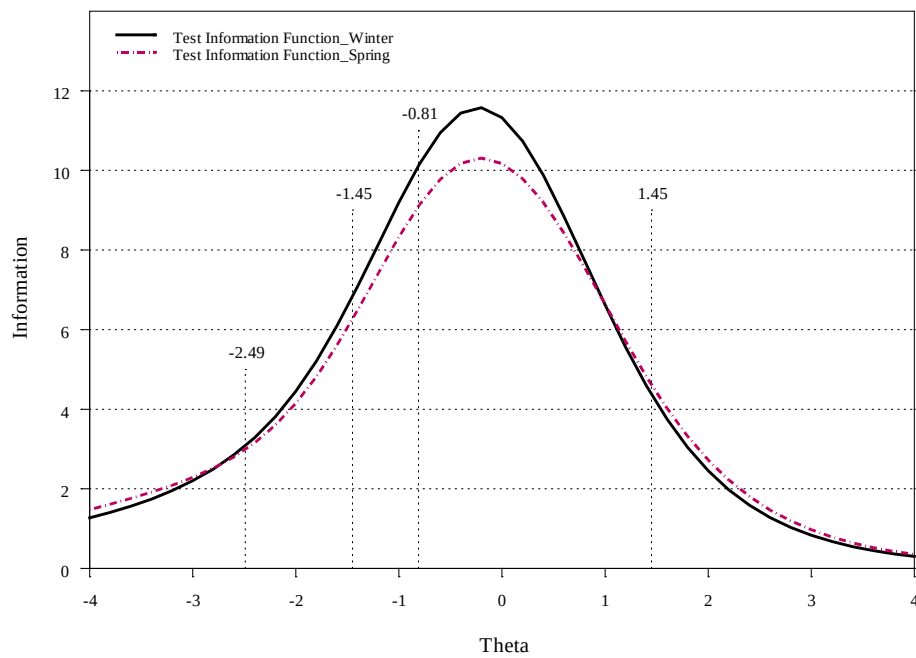


Figure 7.32 Test Information Functions: Grade 9 Academic Math (French)



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.	29.00	0.20	-1.96	0.14
		Max.	82.88	0.58	1.81	1.22
		Mean	55.07	0.35**	0.06	0.63
		SD	14.05	0.11	0.89	0.23
Applied, Spring (English)	31	Min.	31.41	0.21	-2.90	0.26
		Max.	77.16	0.54	1.60	1.22
		Mean	56.88	0.37	-0.08	0.63
		SD	13.99	0.09	0.95	0.21
Academic, Winter (English)	31	Min.	51.73	0.22	-2.25	0.33
		Max.	89.70	0.62	0.10	1.03
		Mean	72.47	0.40	-1.01	0.67
		SD	10.22	0.11	0.64	0.22
Academic, Spring (English)	31	Min.	52.74	0.22	-2.35	0.34
		Max.	90.76	0.59	0.20	1.06
		Mean	72.77	0.40	-0.91	0.70
		SD	10.51	0.10	0.69	0.23
Applied, Winter (French)	31	Min.	34.06	0.19	-2.63	0.15
		Max.	78.20	0.51	1.68	1.01
		Mean	54.38	0.32	-0.02	0.55
		SD	10.21	0.10	0.91	0.21
Applied, Spring (French)	31	Min.	20.44	0.11	-1.36	0.29
		Max.	72.40	0.57	3.13	1.10
		Mean	54.94	0.36	0.17	0.61
		SD	12.59	0.12	0.94	0.21
Academic, Winter (French)	31	Min.	40.29	0.23	-1.87	0.37
		Max.	85.71	0.63	0.68	1.19
		Mean	66.29	0.41	-0.67	0.68
		SD	11.44	0.12	0.69	0.21
Academic, Spring (French)	31	Min.	47.28	0.27	-1.99	0.34
		Max.	89.03	0.65	0.68	0.98
		Mean	67.79	0.40	-0.61	0.65
		SD	9.03	0.10	0.64	0.18

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 75.9% for English-language students and 77.4% for French-language students.

Cronbach's alpha estimates are 0.89 and 0.85 for English and French, respectively. The corresponding SEMs are 4.1% and 4.0% 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	137 002	4.0	81.0	61.5	10.16	0.89	3.36
French	47	39	4	2	2	81	5 318	20.0	80.5	62.7	8.26	0.85	3.20

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	137 002	-3.79	2.88	0.11	0.89
French	5 318	-3.47	2.89	0.19	0.84

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)

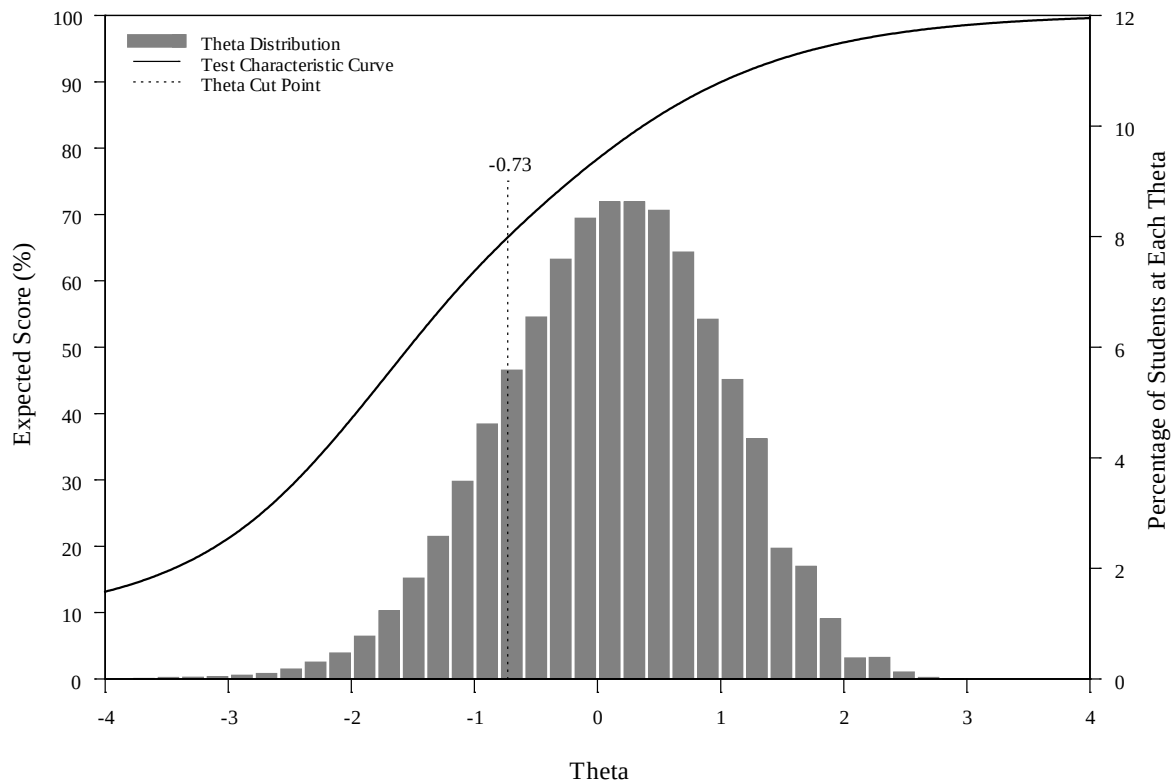


Figure 7.34 Test Characteristic Curve and Distribution of Theta: OSSLT (French)

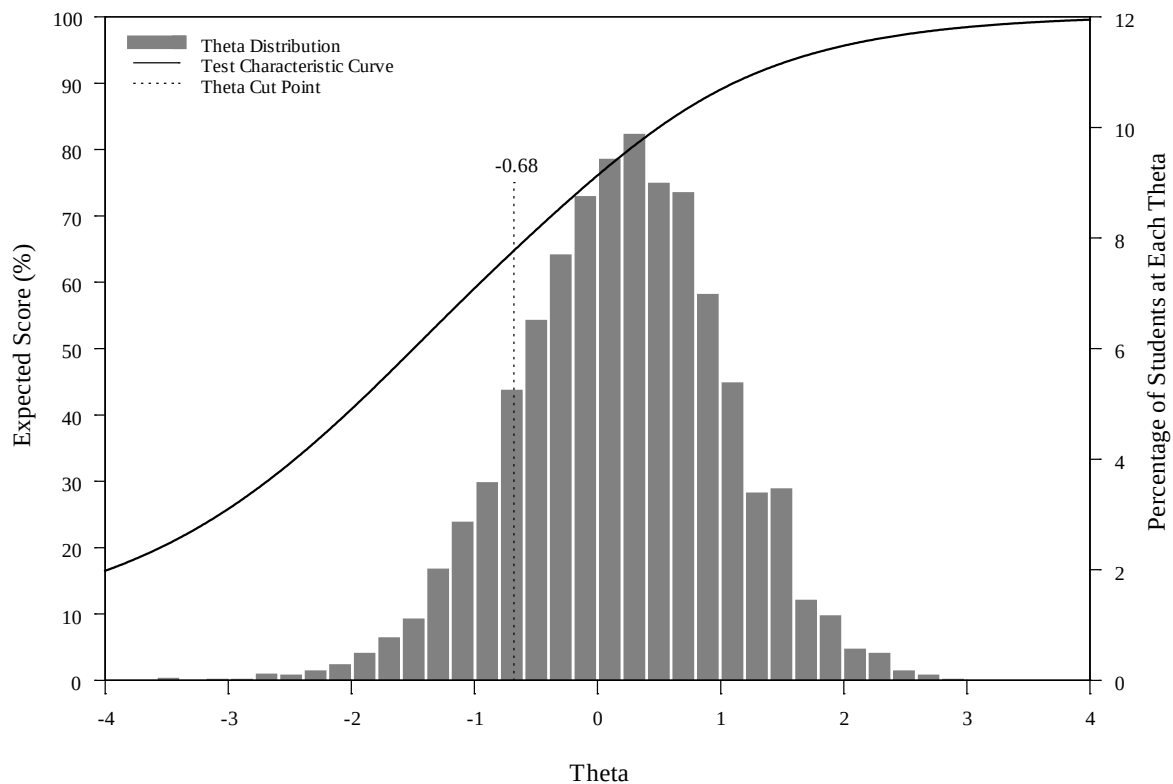


Figure 7.35 Test Information Function: OSSLT (English)

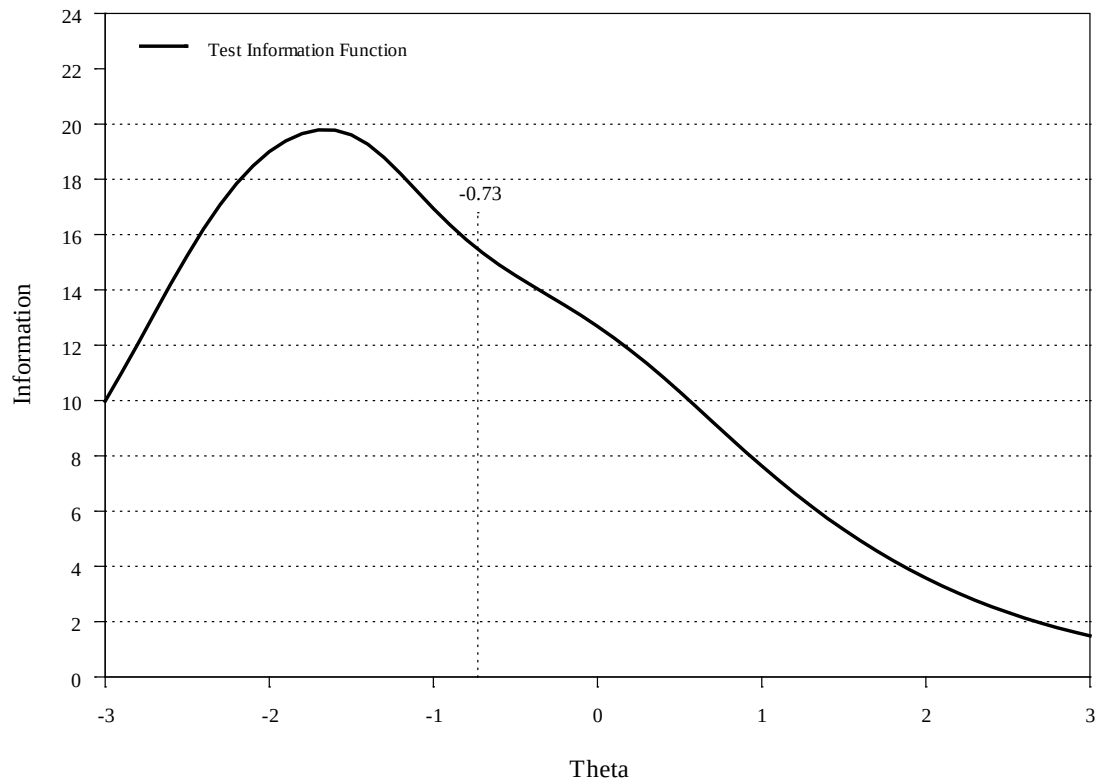
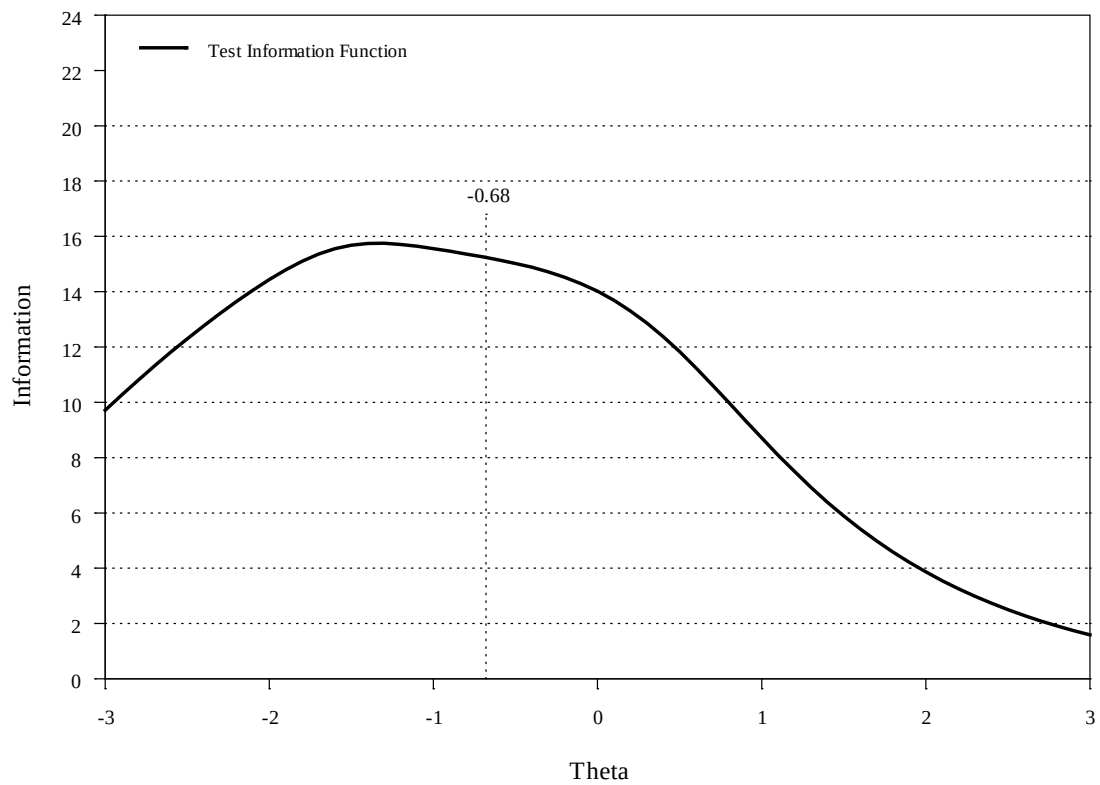


Figure 7.36 Test Information Function: OSSLT (French)



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 for 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 that for the primary, junior and Grade 9 assessments, the a -parameter was set at a constant value for all items on the OSSLT. 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.	0.48	0.19	-2.96
		Max.	0.97	0.56	0.77
		Mean	0.75**	0.35	-1.00
		SD	0.13	0.09	0.94
French	47	Min.	0.42	0.12	-3.51
		Max.	0.98	0.59	1.32
		Mean	0.79	0.30	-1.31
		SD	0.12	0.11	1.06

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 2011–2012 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 ($MH \chi^2$), 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	$MH \chi^2$ not significantly different from 0 or $ \text{Effect size} \leq .17$
B	Moderate DIF	$MH \chi^2$ significantly different from 0 and $.17 < \text{Effect size} \leq .25$
C	Strong DIF	$MH \chi^2$ 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 identified as having B-level or C-level DIF in both samples were considered DIF items. In addition, if an item was flagged with B-level DIF in one sample and C-level DIF in the other sample, then this item was considered to have B-level DIF.

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 2011–2012 primary- and junior-division

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.

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.

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)	3 (MC) 0 (OR)	0 (MC) 3 (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)	3 (MC) 0 (OR)	0 (MC) 0 (OR)	4 (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	1 (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> = 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, 19 items were found to have gender-based DIF. The majority of the DIF items had B-level DIF (17), and only two items had C-level DIF. Of the 16 multiple-choice items showing B- or C-level DIF, 15 favoured the boys, while one favoured the girls. The three open-response DIF items 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 (7).

The summaries for SLL-based DIF are reported in Tables 7.15 and 7.16. Five out of the 344 items across all the assessments showed SLL-based DIF, with all of them having B-level DIF. The three multiple-choice DIF items favoured SLL students, and the two open-response DIF items favoured non-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 k = 36	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 2 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Writing k = 14	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)
Math k = 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. SLL = second-language learner; MC = multiple choice; OR = open response.

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 k = 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)
Writing k = 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 k = 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. SLL = second language learner; MC = multiple choice; OR = open response.

All items identified as having B-level or C-level DIF on the primary- and junior-division assessments were reviewed by the assessment team. Since the reviewers did not identify any apparent 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)	4 (MC)	2 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Applied, Spring <i>k</i> = 31	1 (MC)	1 (MC)	2 (MC)	1 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Winter <i>k</i> = 31	0 (MC)	0 (MC)	2 (MC)	1 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Spring <i>k</i> = 31	1 (MC)	0 (MC)	3 (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	1 (MC)	0 (MC)	0 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Applied, Spring <i>k</i> = 31	0 (MC)	0 (MC)	0 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Winter <i>k</i> = 31	0 (MC)	0 (MC)	0 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Spring <i>k</i> = 31	0 (MC)	0 (MC)	1 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)

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

Of the 248 items across the eight Grade 9 math assessments, 19 multiple-choice items showed B-level gender-based DIF. Of the 19 B-level gender-based multiple-choice DIF items, 13 favoured boys and six favoured girls. No open-response items showed B-level gender-based DIF. Two multiple-choice items showed C-level gender-based DIF, and both favoured boys. Again, no open-response items showed C-level gender-based DIF. Overall, more multiple-choice items were found to favour boys.

Across the Grade 9 math assessments, six to 12 multiple-choice items were used in both versions of the winter and spring administrations. Altogether four common items showed gender-based DIF. Of the four items, two showed DIF in the spring administration, and two were found having 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)	3 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Spring k = 31	0 (MC)	1 (MC)	1 (MC)	3 (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	2 (MC)	0 (MC)	0 (MC)	0 (MC)
	1 (OR)	1 (OR)	0 (OR)	0 (OR)
Spring k = 31	1 (MC)	1 (MC)	0 (MC)	1 (MC)
	0 (OR)	1 (OR)	0 (OR)	0 (OR)

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

SLL-based DIF was conducted only for the English-language courses. There are 11 multiple-choice B-level SLL-based DIF items. Of these items, 9 items favoured SLL students and 2 favoured non-SLL students. No open-response items showed B-level SLL-based DIF. Eight items (five multiple-choice and three open-response) showed C-level SLL-based DIF. Of the five multiple-choice items, two favoured SLL students and three favoured non-SLL students. Of the three open-response items, two favoured SLL students and one favoured non-SLL students.

Seven common items showed SLL-based DIF, with three items showing DIF in both administrations. Overall, a larger number of SLL-based DIF was found to favour SLL students. More B-level SLL-based DIF items were found for the English-language academic course, while more C-level SLL-based DIF items were found for the English-language applied 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 (k = 51)		French (k = 51)	
Males	Females	Males	Females
2 (MC)	1(MC) 2(LW)	8 (MC)	2 (MC) 1(OR) 1(LW)

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

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

English (k = 51)		French (k = 51)	
Males	Females	Males	Females
6 (MC)	1(LW)	1 (MC)	1(OR)

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

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

There were 12 B-level DIF items on the French-language version of the OSSLT. Eight multiple-choice items favoured the males, and two multiple-choice items favoured the females. One open-response reading item and one long-writing item exhibited B-level DIF favouring females for topic development and use of conventions. One C-level DIF multiple-choice item favoured the males, and one open-response reading item favoured the females.

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 open-response reading item exhibited C-level DIF favouring SLLs. Five 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 (k = 51)			
B-Level DIF		C-Level DIF	
Non-SLLs	SLLs	Non-SLLs	SLLs
4 (MC)	1 (MC) 1 (OR) 1 (LW)	5 (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 2/3 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	0.91	0.87
Junior Reading (English)	0.83	0.76	0.92	0.88
Primary Reading (French)	0.79	0.70	0.92	0.88
Junior Reading (French)	0.83	0.76	0.94	0.90
Primary Writing (English)	0.86	0.80	0.89	0.84
Junior Writing (English)	0.82	0.75	0.88	0.83
Primary Writing (French)	0.80	0.73	0.90	0.86
Junior Writing (French)	0.82	0.75	0.90	0.86
Primary Mathematics (English)	0.83	0.76	0.91	0.87
Junior Mathematics (English)	0.80	0.72	0.91	0.87
Primary Mathematics (French)	0.82	0.74	0.91	0.87
Junior Mathematics (French)	0.83	0.76	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)	0.70	0.59	0.87	0.82
Applied, Spring (English)	0.71	0.60	0.88	0.83
Academic, Winter (English)	0.82	0.76	0.92	0.89
Academic, Spring (English)	0.83	0.77	0.93	0.90
Applied, Winter (French)	0.67	0.56	0.88	0.84
Applied, Spring (French)	0.68	0.58	0.88	0.84
Academic, Winter (French)	0.83	0.76	0.91	0.87
Academic, Spring (French)	0.84	0.78	0.92	0.89

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. In *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. In *Journal of Educational Measurement*, 32, 179–197.
- Mantel, N. (1963). Chi-square tests with one degree of freedom; Extensions of the Mantel-Haenszel procedure. In *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. In *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 mathematics components and for all items in the reading and writing components. The aggregate exact-plus-adjacent validity estimates for the items

in each component ranged from 98.8 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 99 to 100%. For the OSSLT, the EQAO target of 95% exact-plus-adjacent agreement was met for all but one item (in short writing). The aggregate validity estimates ranged from 97.1 to 99.7% (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 0.81 to 0.90 for the primary and junior assessments, 0.82 to 0.87 for Grade 9 mathematics, and 0.85 to 0.89 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 0.94, 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 the content and standards of the reading component of 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 conducted research to examine literacy skills by linking performance on the OSSLT with performance on the reading component of the 2009 Programme for International Student Assessment (PISA). Both assessments were administered to the same group of students between April and May 2009.

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 the 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 (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
17293	1 (A)	5	1 666	100.0	90.0	10.0	5.7	4.3	0.1
17294	1 (A)	6	1 511	99.2	81.2	18.0	12.2	5.8	0.8
17241	1 (A)	11	1 338	99.0	88.7	10.3	8.4	1.9	1.0
17240	1 (A)	12	1 337	99.6	79.8	19.8	6.8	13.0	0.4
17344	1 (B)	5	1 302	99.7	80.4	19.3	10.7	8.6	0.3
17343	1 (B)	6	1 343	99.8	96.7	3.1	1.6	1.5	0.2
10912	2 (C)	11	1 510	99.5	81.9	17.6	8.5	9.1	0.5
10899	2 (C)	12	1 421	99.6	82.6	17.0	9.1	7.9	0.4
17351	2 (D)	5	1 433	99.7	86.7	13.0	9.2	3.8	0.2
17350	2 (D)	6	1 478	99.5	80.9	18.6	10.1	8.5	0.5
Aggregate			14 339	99.6	84.9	14.7	8.2	6.4	0.4

Table 4.1.2 Validity Estimates for Reading: Junior Division (English)

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
17894	1 (A)	5	1 772	99.4	80.1	19.3	10.3	9.0	0.6
17895	1 (A)	6	1 634	99.8	79.9	19.9	16.2	3.7	0.2
17915	1 (A)	11	1 596	99.9	84.9	15.0	12.1	2.9	0.1
17914	1 (A)	12	1 585	99.7	86.9	12.8	5.3	7.5	0.3
17908	1 (B)	5	1 443	99.5	79.8	19.7	8.0	11.7	0.5
17909	1 (B)	6	1 444	99.4	74.1	25.3	17.3	8.0	0.6
12405	2 (C)	11	1 917	98.9	81.2	17.7	10.7	7.0	1.0
12404	2 (C)	12	1 758	99.5	78.7	20.8	11.4	9.4	0.5
17872	2 (D)	5	1 831	99.8	80.0	19.8	12.0	7.8	0.2
17871	2 (D)	6	1 661	100.0	84.2	15.8	11.2	4.6	0.1
Aggregate			16 641	99.6	81.0	18.6	11.4	7.1	0.4

Table 4.1.3 Validity Estimates for Reading: Primary Division (French)

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
17639	1 (A)	5	118	100.0	94.1	5.9	2.5	3.4	0.0
17640	1 (A)	6	183	99.4	98.9	0.5	0.5	0.0	0.5
17657	1 (A)	11	90	100.0	96.7	3.3	2.2	1.1	0.0
17658	1 (A)	12	57	100.0	96.5	3.5	3.5	0.0	0.0
17621	1 (B)	5	185	99.5	94.6	4.9	0.0	4.9	0.5
17622	1 (B)	6	126	98.4	64.3	34.1	34.1	0.0	1.6
17591	2 (C)	11	56	100.0	87.5	12.5	12.5	0.0	0.0
17592	2 (C)	12	82	100.0	97.6	2.4	0.0	2.4	0.0
17675	2 (D)	5	126	100.0	97.6	2.4	0.0	2.4	0.0
17676	2 (D)	6	86	100.0	100.0	0.0	0.0	0.0	0.0
Aggregate			1109	99.6	92.7	6.9	5.2	1.7	0.3

Table 4.1.4 Validity Estimates for Reading: Junior Division (French)

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
18229	1 (A)	5	93	100.0	79.6	20.4	0.0	20.4	0.0
18230	1 (A)	6	60	100.0	90.0	10.0	1.7	8.3	0.0
18248	1 (A)	11	179	100.0	95.5	4.5	4.5	0.0	0.0
18247	1 (A)	12	118	99.1	70.3	28.8	2.5	26.3	0.8
18211	1 (B)	5	62	100.0	100.0	0.0	0.0	0.0	0.0
18212	1 (B)	6	93	100.0	96.8	3.2	3.2	0.0	0.0
18205	2 (C)	11	139	100.0	78.4	21.6	11.5	10.1	0.0
18206	2 (C)	12	138	100.0	92.8	7.2	2.9	4.3	0.0
18217	2 (D)	5	110	100.0	97.3	2.7	1.8	0.9	0.0
18218	2 (D)	6	95	100.0	94.7	5.3	4.2	1.1	0.0
Aggregate			1087	99.9	89.1	10.9	3.8	7.1	0.1

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
16475_T	2(D)	7	1788	99.9	89.4	10.5	7.8	2.7	0.1
16475_V	2(D)	7	1678	99.5	91.2	8.3	2.7	5.5	0.5
Aggregate Long Writing			3466	99.7	90.3	9.4	5.3	4.1	0.3
17386_T	1(A)	13	1397	99.6	83.5	16.2	8.9	7.2	0.4
17386_V	1(A)	13	1523	99.9	79.3	20.6	7.9	12.7	0.1
18179_T	2(C)	13	1370	99.7	85.5	14.2	8.5	5.7	0.3
18179_V	2(C)	13	1652	100.0	79.5	20.5	7.3	13.1	0.0
Aggregate Short Writing			5942	99.8	81.8	18.1	8.1	9.9	0.2
Aggregate All Items			9408	99.8	84.9	14.9	7.1	7.8	0.2

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
18996_T	2(D)	7	2769	99.7	86.8	12.9	6.4	6.5	0.3
18996_V	2(D)	7	2522	99.8	90.2	9.7	3.8	5.8	0.2
Aggregate Long Writing			5291	99.8	88.4	11.3	5.2	6.2	0.2
17958_T	1(A)	13	2039	99.4	86.8	12.6	8.6	4.0	0.6
17958_V	1(A)	13	2153	100.0	92.0	8.0	2.7	5.3	0.0
17950_T	2(C)	13	1985	99.6	84.7	14.9	10.4	4.4	0.4
17950_V	2(C)	13	1991	99.9	87.1	12.8	6.2	6.6	0.1
Aggregate Short Writing			8168	99.7	87.7	12.0	6.9	5.1	0.3
Aggregate All Items			13459	99.7	88.0	11.7	6.2	5.5	0.2

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
15509_T	2(D)	7	192	99.5	81.3	18.2	9.9	8.3	0.5
15509_V	2(D)	7	179	98.3	91.1	7.3	4.5	2.8	1.7
Aggregate Long Writing			371	98.9	86.0	12.9	7.3	5.7	1.1
17708_T	1(A)	13	198	99.0	80.3	18.7	8.1	10.6	1.0
17708_V	1(A)	13	170	100.0	82.9	17.1	5.9	11.2	0.0
17692_T	2(C)	13	169	100.0	86.4	13.6	6.5	7.1	0.0
17692_V	2(C)	13	184	100.0	89.7	10.3	7.1	3.3	0.0
Aggregate Short Writing			721	99.7	84.7	15.0	6.9	8.0	0.3
Aggregate All Items			1092	99.5	85.2	14.3	7.1	7.2	0.5

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
16819_T	2(D)	7	185	100.0	75.1	24.9	10.8	14.1	0.0
16819_V	2(D)	7	161	100.0	82.0	18.0	3.1	14.9	0.0
Aggregate Long Writing			346	100.0	78.3	21.7	7.2	14.5	0.0
18303_T	1(A)	13	177	98.9	80.2	18.6	13.6	5.1	1.1
18303_V	1(A)	13	194	98.5	85.6	12.9	5.2	7.7	1.5
18302_T	2(C)	13	172	99.4	84.9	14.5	7.6	7.0	0.6
18302_V	2(C)	13	125	100.0	88.8	11.2	8.0	3.2	0.0
Aggregate Short Writing			668	99.1	84.6	14.5	8.5	6.0	0.9
Aggregate All Items			1014	99.4	82.4	17.0	8.1	8.9	0.6

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

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
15093	3(1)	7	631	99.8	96.4	3.5	2.1	1.4	0.2
10847	3(1)	8	720	100.0	90.4	9.6	8.8	0.8	0.0
15173	3(1)	9	888	100.0	95.3	4.7	2.0	2.7	0.0
15132	3(1)	10	888	99.7	93.7	6.0	1.7	4.3	0.3
10846	3(2)	8	989	100.0	97.2	2.8	0.6	2.2	0.0
10744	3(2)	9	894	99.9	94.0	5.9	4.4	1.6	0.1
15116	3(2)	10	1122	99.9	91.4	8.6	4.5	4.1	0.1
15095	3(2)	11	899	99.9	92.7	7.2	0.8	6.5	0.1
Aggregate			7031	99.9	93.8	6.1	0.1	3.1	0.1

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

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
17405	3(1)	7	993	99.0	89.1	9.9	5.7	4.1	1.0
15068	3(1)	8	818	99.9	93.2	6.7	1.6	5.1	0.1
11409	3(1)	9	726	99.4	93.8	5.6	2.8	2.9	0.6
15010	3(1)	10	634	99.8	91.5	8.4	7.6	0.8	0.2
18368	3(2)	8	858	99.3	95.0	4.3	3.3	1.0	0.7
11310	3(2)	9	851	99.6	91.3	8.3	4.5	3.9	0.4
11329	3(2)	10	1168	99.7	96.5	3.2	0.6	2.6	0.3
15034	3(2)	11	744	98.7	88.2	10.5	5.5	5.0	1.3
Aggregate			6792	99.4	92.5	6.9	3.7	3.2	0.6

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

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
17769	3(1)	7	49	100.0	98.0	2.0	0.0	2.0	0.0
14631	3(1)	8	42	100.0	92.9	7.1	0.0	7.1	0.0
14667	3(1)	9	49	98.0	93.9	4.1	0.0	4.1	2.0
14599	3(1)	10	35	100.0	91.4	8.6	8.6	0.0	0.0
17773	3(2)	8	93	98.9	92.5	6.5	5.4	1.1	1.1
11059	3(2)	9	89	100.0	96.6	3.4	3.4	0.0	0.0
14668	3(2)	10	83	95.2	83.1	12.0	6.0	6.0	4.8
17766	3(2)	11	94	100.0	96.8	3.2	3.2	0.0	0.0
Aggregate			534	98.9	93.1	5.8	3.6	2.3	1.1

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

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
18317	3(1)	7	63	98.4	88.9	9.5	4.8	4.8	1.6
18316	3(1)	8	62	100.0	77.4	22.6	11.3	11.3	0.0
18324	3(1)	9	63	100.0	98.4	1.6	0.0	1.6	0.0
14688	3(1)	10	56	100.0	96.4	3.6	1.8	1.8	0.0
11533	3(2)	8	95	98.9	72.6	26.3	6.3	20.0	1.1
18004	3(2)	9	110	94.5	81.8	12.7	6.4	6.4	5.5
11508	3(2)	10	97	100.0	82.5	17.5	7.2	10.3	0.0
14744	3(2)	11	109	100.0	78.9	21.1	15.6	5.5	0.0
Aggregate			655	98.8	83.2	15.6	7.3	8.2	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	14864	NR	225	100.0	96.4	3.6	0.9	2.7	0.0
	14843	NR	232	99.6	78.9	20.7	14.2	6.5	0.4
	14791	8	219	99.5	92.7	6.8	2.3	4.6	0.5
	14870	31	247	99.6	91.1	8.5	1.6	6.9	0.4
	14844	NR	250	99.2	90.0	9.2	4.0	5.2	0.4
	14797	NR	245	98.0	89.4	8.6	4.5	4.1	2.0
	10164	21	232	97.8	89.7	8.2	6.0	2.2	2.2
Aggregate			1650	99.1	89.7	9.4	4.8	4.6	0.8
Spring	14788	NR	207	99.5	90.8	8.7	5.3	3.4	0.5
	14863	30	218	95.9	87.6	8.3	4.1	4.1	4.1
	14847	23	224	100.0	88.8	11.2	4.0	7.1	0.0
	14871	NR	201	100.0	90.0	10.0	6.5	3.5	0.0
	14837	NR	336	99.4	93.8	5.7	3.9	1.8	0.6
	14793	9	336	99.7	95.2	4.5	2.7	1.8	0.3
	14840	22	350	100.0	98.3	1.7	1.1	0.6	0.0
Aggregate			1872	99.3	92.8	6.5	3.6	2.8	0.7
Aggregate Across Administrations			3522	99.2	91.4	7.8	4.2	3.7	0.8

Note. NR = not released.

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	14939	NR	456	99.3	90.6	8.8	6.8	2.0	0.7
	10289	6	454	99.8	95.6	4.2	3.5	0.7	0.2
	14969	NR	398	99.5	94.2	5.3	4.3	1.0	0.5
	14909	NR	460	100.0	95.2	4.8	3.3	1.5	0.0
	14945	23	471	99.8	94.7	5.1	1.1	4.0	0.2
	14963	30	421	100.0	93.6	6.4	3.1	3.3	0.0
	14910	14	466	98.5	93.3	5.2	2.6	2.6	1.5
Aggregate			3126	99.6	93.9	5.7	3.5	2.2	0.4
Spring	14947	NR	594	100.0	86.7	13.3	9.6	3.7	0.0
	14965	NR	597	99.0	89.1	9.9	7.9	2.0	0.8
	10322	22	557	99.6	83.7	16.0	11.1	4.8	0.4
	14911	NR	572	98.4	89.0	9.4	0.5	8.9	1.6
	14908	13	757	99.5	94.6	4.9	3.0	1.8	0.5
	14971	31	659	99.5	87.7	11.8	1.8	10.0	0.5
	14884	NR	788	99.5	90.7	8.8	4.4	4.3	0.5
Aggregate			4524	99.4	89.1	10.3	5.3	5.0	0.6
Aggregate Across Administrations			7650	99.5	91.1	8.4	4.5	3.8	0.5

Note. NR = not released.

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	16910	NR	15	100.0	60.0	40.0	20.0	20.0	0.0
	16933	NR	15	100.0	93.3	6.7	6.7	0.0	0.0
	16517	9	12	100.0	91.7	8.3	8.3	0.0	0.0
	CMS9759	19	6	100.0	50.0	50.0	33.3	16.7	0.0
	OQMF31839	28	22	95.5	90.9	4.5	4.5	0.0	4.5
	16546	NR	14	100.0	92.9	7.1	7.1	0.0	0.0
	16519	NR	19	100.0	94.7	5.3	5.3	0.0	0.0
Aggregate			103	99.0	85.4	13.6	9.7	3.9	1.0
Spring	14418	NR	25	100.0	100.0	0.0	0.0	0.0	0.0
	CMS9626	8	22	100.0	95.5	4.5	4.5	0.0	0.0
	14456	NR	21	100.0	100.0	0.0	0.0	0.0	0.0
	OQMF31791	18	16	100.0	81.3	18.8	12.5	6.3	0.0
	9927	NR	33	100.0	97.0	3.0	0.0	3.0	0.0
	12839	NR	25	100.0	96.0	4.0	4.0	0.0	0.0
	OQMF21395	27	23	100.0	100.0	0.0	0.0	0.0	0.0
Aggregate			165	100.0	96.4	3.6	2.4	1.2	0.0
Aggregate Across Administrations			268	99.6	92.2	7.5	5.2	2.2	0.4

Note. NR = not released.

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	39	100.0	84.6	15.4	12.8	2.6	0.0
	10064	40	42	100.0	97.6	2.4	2.4	0.0	0.0
	17064	NR	33	100.0	87.9	12.1	0.0	12.1	0.0
	14493	8	39	100.0	100.0	0.0	0.0	0.0	0.0
	9967	NR	56	96.4	89.3	7.1	7.1	0.0	3.6
	CMS10066	31	52	100.0	98.1	1.9	1.9	0.0	0.0
	14500	16	46	100.0	84.8	15.2	2.2	13.0	0.0
Aggregate			307	99.5	94.3	5.2	1.9	3.3	0.5
Spring	CMS9968	15	66	98.5	95.5	3.0	1.5	1.5	1.5
	14564	30	53	100.0	94.3	5.7	1.9	3.8	0.0
	OQMF31900	NR	67	100.0	100.0	0.0	0.0	0.0	0.0
	17065	NR	75	100.0	97.3	2.7	0.0	2.7	0.0
	OQMF20902	7	73	100.0	90.4	9.6	1.4	8.2	0.0
	14524	22	104	100.0	94.2	5.8	3.8	1.9	0.0
	14566	NR	49	100.0	93.9	6.1	2.0	4.1	0.0
Aggregate			487	99.8	95.1	4.7	1.6	3.1	0.2
Aggregate Across Administrations			794	99.6	93.8	5.8	2.5	3.3	0.4

Note. NR = not released.

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
18163_269	n/a	6	9 270	98.0	80.4	17.7	12.8	4.9	2.0
17237_265	n/a	7	8 255	98.2	86.0	12.1	4.4	7.7	1.8
13498_262	IV	6	9 827	98.5	67.5	31.0	11.2	19.8	1.5
13499_262	IV	7	8 747	97.8	73.5	24.3	8.7	15.6	2.2
Aggregate			36 099	98.1	76.5	21.6	9.5	12.2	1.9

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
19840_T	n/a	1	19 825	95.3	64.2	31.1	17.5	13.6	4.7
19840_V	n/a	1	19 825	98.8	72.8	26.0	12.8	13.1	1.2
18378_T	I	1	23 201	98.4	69.8	28.6	11.3	17.3	1.6
18378_V	I	1	23 201	99.0	70.0	29.0	14.4	14.6	1.0
Aggregate Long Writing			86 052	97.9	69.3	28.7	13.9	14.8	2.1
18022_T&V	n/a	1	9 013	99.0	86.8	12.3	7.2	5.1	1.0
18021_T&V	V	1	8 165	97.8	69.8	28.0	13.0	15.0	2.2
Aggregate Short Writing			17 178	98.5	78.7	19.7	9.9	9.8	1.5
Aggregate All Items			103 230	98.0	70.8	27.2	13.2	13.9	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
18325_232	s. o.	6	335	100.0	91.0	9.0	0.6	8.4	0.0
18181_227	s. o.	7	247	99.2	95.1	4.0	0.8	3.2	0.8
13242_236	IV	6	342	99.7	90.9	8.8	6.7	2.0	0.3
18328_236	IV	7	252	100.0	99.6	0.4	0.0	0.4	0.0
Aggregate			1176	99.7	93.7	6.0	2.3	3.7	0.3

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
19077_T	s. o.	1	692	95.7	80.5	15.2	8.4	6.8	4.3
19077_V	s. o.	1	692	98.6	67.8	30.8	5.9	24.9	1.4
19076_T	I	1	1017	97.6	69.8	27.8	14.4	13.5	2.4
19076_V	I	1	1017	98.2	64.6	33.6	11.4	22.2	1.8
Aggregate Long Writing			3418	97.6	70.0	27.6	10.6	17.0	2.4
13278_T&V	s. o.	1	412	94.9	69.2	25.7	15.3	10.4	5.1
13272_T&V	V	1	394	99.5	81.2	18.3	5.8	12.4	0.5
Aggregate Short Writing			806	97.1	75.1	22.1	10.7	11.4	2.9
Aggregate All Items			4224	97.5	71.0	26.5	10.6	16.0	2.5

Interrater Reliability: The Primary and Junior Assessments

Table 4.1.21 Interrater Reliability Estimates for Reading: Primary Division (English)

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
17293	1 (A)	5	1 046	93.6	59.9	33.7	6.4
17294	1 (A)	6	1 119	97.1	66.8	30.2	2.9
17241	1 (A)	11	1 715	95.2	71.9	23.3	4.8
17240	1 (A)	12	1 716	96.3	64.7	31.6	3.7
17344	1 (B)	5	991	94.2	65.1	29.2	5.8
17343	1 (B)	6	970	99.2	75.2	24.0	0.8
10912	2 (C)	11	859	97.8	66.8	31.0	2.2
10899	2 (C)	12	832	97.6	64.5	33.1	2.4
17351	2 (D)	5	939	95.1	60.9	34.2	4.9
17350	2 (D)	6	916	97.6	57.0	40.6	2.4
Aggregate			11 103	96.2	65.7	30.5	3.8

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
17894	1 (A)	5	1 195	95.9	62.1	33.8	4.1
17895	1 (A)	6	1 263	94.9	62.5	32.3	5.1
17915	1 (A)	11	1 018	89.0	61.3	27.7	11.0
17914	1 (A)	12	1 020	95.8	58.9	36.9	4.2
17908	1 (B)	5	879	96.2	61.2	35.0	3.8
17909	1 (B)	6	856	96.0	57.5	38.6	4.0
12405	2 (C)	11	1 285	95.7	58.8	37.0	4.3
12404	2 (C)	12	1 365	98.0	65.0	33.0	2.0
17872	2 (D)	5	879	98.1	68.6	29.5	1.9
17871	2 (D)	6	964	98.9	66.2	32.7	1.1
Aggregate			10 724	95.8	62.2	33.6	4.2

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
17639	1 (A)	5	160	98.8	79.4	19.4	1.3
17640	1 (A)	6	128	98.4	82.0	16.4	1.6
17657	1 (A)	11	80	98.8	78.8	20.0	1.3
17658	1 (A)	12	97	100.0	70.1	29.9	0.0
17621	1 (B)	5	85	100.0	87.1	12.9	0.0
17622	1 (B)	6	114	98.2	73.7	24.6	1.8
17591	2 (C)	11	42	100.0	83.3	16.7	0.0
17592	2 (C)	12	29	100.0	96.6	3.4	0.0
17675	2 (D)	5	52	100.0	82.7	17.3	0.0
17676	2 (D)	6	74	100.0	85.1	14.9	0.0
Aggregate			861	99.2	80.1	19.0	0.8

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
18229	1 (A)	5	80	96.3	87.5	8.8	3.8
18230	1 (A)	6	96	97.9	72.9	25.0	2.1
18248	1 (A)	11	191	99.5	85.3	14.1	0.5
18247	1 (A)	12	221	95.5	73.3	22.2	4.5
18211	1 (B)	5	89	100.0	86.5	13.5	0.0
18212	1 (B)	6	74	100.0	79.7	20.3	0.0
18205	2 (C)	11	77	94.8	63.6	31.2	5.2
18206	2 (C)	12	78	100.0	57.7	42.3	0.0
18217	2 (D)	5	80	100.0	96.3	3.8	0.0
18218	2 (D)	6	88	98.9	76.1	22.7	1.1
Aggregate			1074	98.0	78.1	19.9	2.0

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
16475_T	2(D)	7	970	97.8	70.3	27.5	2.2
16475_V	2(D)	7	1026	98.9	72.3	26.6	1.1
Aggregate Long Writing			1996	98.4	71.3	27.1	1.6
17386_T	1(A)	13	907	94.7	64.9	29.8	5.3
17386_V	1(A)	13	845	96.4	72.2	24.3	3.6
18179_T	2(C)	13	952	98.5	68.8	29.7	1.5
18179_V	2(C)	13	804	98.4	77.5	20.9	1.6
Aggregate Short Writing			3508	97.0	70.6	26.4	3.0
Aggregate All Items			5504	97.5	70.9	26.6	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
18996_T	2(D)	7	1541	97.7	65.9	31.8	2.3
18996_V	2(D)	7	1625	99.5	66.6	32.9	0.5
Aggregate Long Writing			3166	98.6	66.3	32.3	1.4
17958_T	1(A)	13	1170	98.7	68.6	30.1	1.3
17958_V	1(A)	13	1112	98.7	68.2	30.5	1.3
17950_T	2(C)	13	1538	96.8	65.0	31.9	3.2
17950_V	2(C)	13	1533	99.4	75.0	24.4	0.6
Aggregate Short Writing			5353	98.4	69.3	29.0	1.6
Aggregate All Items			8519	98.5	68.2	30.3	1.5

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
15509_T	2(D)	7	78	98.7	60.3	38.5	1.3
15509_V	2(D)	7	84	100.0	75.0	25.0	0.0
Aggregate Long Writing			162	99.4	67.9	31.5	0.6
17708_T	1(A)	13	264	94.7	67.4	27.3	5.3
17708_V	1(A)	13	275	98.2	74.9	23.3	1.8
17692_T	2(C)	13	76	100.0	68.4	31.6	0.0
17692_V	2(C)	13	64	100.0	76.6	23.4	0.0
Aggregate Short Writing			679	97.2	71.4	25.8	2.8
Aggregate All Items			841	97.6	70.7	26.9	2.4

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
16819_T	2(D)	7	84	96.4	67.9	28.6	3.6
16819_V	2(D)	7	94	98.9	68.1	30.9	1.1
Aggregate Long Writing			178	97.8	68.0	29.8	2.8
18303_T	1(A)	13	94	94.7	54.3	40.4	5.3
18303_V	1(A)	13	86	97.7	60.5	37.2	2.3
18302_T	2(C)	13	81	100.0	75.3	24.7	0.0
18302_V	2(C)	13	101	98.0	68.3	29.7	2.0
Aggregate Short Writing			362	97.5	64.4	33.1	2.5
Aggregate All Items			540	97.6	65.6	32.0	2.4

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
15093	3(1)	7	1 381	99.0	85.7	13.3	1.0
10847	3(1)	8	1 336	96.3	68.9	27.5	3.7
15173	3(1)	9	1 252	98.5	79.9	18.6	1.5
15132	3(1)	10	1 251	97.2	81.4	15.8	2.8
10846	3(2)	8	1 692	96.0	77.9	18.1	4.0
10744	3(2)	9	1 739	98.1	87.6	10.5	1.9
15116	3(2)	10	1 628	97.2	72.2	24.9	2.8
15095	3(2)	11	1 737	96.0	76.1	19.9	4.0
Aggregate			12 016	97.2	78.7	18.5	2.8

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
17405	3(1)	7	1 617	96.2	73.0	23.1	3.8
15068	3(1)	8	1 701	97.5	74.8	22.7	2.5
11409	3(1)	9	1 747	97.8	77.4	20.4	2.2
15010	3(1)	10	1 793	98.2	85.0	13.2	1.8
18368	3(2)	8	2 131	96.5	82.4	14.1	3.5
11310	3(2)	9	2 134	97.9	84.8	13.2	2.1
11329	3(2)	10	1 976	97.2	91.2	6.0	2.8
15034	3(2)	11	2 189	90.6	57.8	32.8	9.4
Aggregate			15 288	96.4	78.3	18.1	3.6

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
17769	3(1)	7	182	97.3	91.2	6.0	2.7
14631	3(1)	8	185	97.8	94.1	3.8	2.2
14667	3(1)	9	182	96.7	80.2	16.5	3.3
14599	3(1)	10	188	98.9	97.9	1.1	1.1
17773	3(2)	8	268	98.5	68.7	29.9	1.5
11059	3(2)	9	268	95.9	84.7	11.2	4.1
14668	3(2)	10	268	94.4	74.6	19.8	5.6
17766	3(2)	11	263	95.4	84.0	11.4	4.6
Aggregate			1804	96.7	83.3	13.5	3.3

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
18317	3(1)	7	95	98.9	89.5	9.5	1.1
18316	3(1)	8	95	98.9	76.8	22.1	1.1
18324	3(1)	9	95	97.9	86.3	11.6	2.1
14688	3(1)	10	98	99.0	90.8	8.2	1.0
11533	3(2)	8	170	95.9	83.5	12.4	4.1
18004	3(2)	9	162	98.1	88.9	9.3	1.9
11508	3(2)	10	168	98.2	81.5	16.7	1.8
14744	3(2)	11	162	98.1	85.2	13.0	1.9
Aggregate			1045	98.0	85.2	12.8	2.0

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	14864	NR	291	97.9	79.4	18.6	2.1
	14843	NR	298	97.0	76.5	20.5	3.0
	14791	8	293	95.9	71.7	24.2	4.1
	14870	31	286	97.9	82.9	15.0	2.1
	14844	NR	294	94.6	72.8	21.8	5.4
	14797	NR	299	95.7	81.6	14.0	4.3
	10164	21	303	94.1	82.2	11.9	5.9
Aggregate			2064	96.1	78.2	18.0	3.9
Spring	14788	NR	294	98.3	90.5	7.8	1.7
	14863	30	290	99.0	95.5	3.4	1.0
	14847	23	288	99.0	84.0	14.9	1.0
	14871	NR	293	99.3	87.7	11.6	0.7
	14837	NR	284	97.5	82.4	15.1	2.5
	14793	9	284	97.5	88.0	9.5	2.5
	14840	22	277	99.3	81.2	18.1	0.7
Aggregate			2010	98.6	87.1	11.4	1.4
Aggregate Across Administrations			4074	97.3	870.7	14.7	2.7

Note. NR = not released.

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	14939	NR	708	97.2	78.2	18.9	2.8
	10289	6	704	99.1	91.2	8.0	0.9
	14969	NR	730	99.9	93.6	6.3	0.1
	14909	NR	703	97.4	75.8	21.6	2.6
	14945	23	497	95.2	83.9	11.3	4.8
	14963	30	515	97.3	89.3	8.0	2.7
	14910	14	502	95.8	72.9	22.9	4.2
Aggregate			4359	97.6	83.8	13.8	2.4
Spring	14947	NR	868	97.2	73.0	24.2	2.8
	14965	NR	868	99.2	81.3	17.9	0.8
	10322	22	894	96.4	73.0	23.4	3.6
	14911	NR	882	98.4	86.1	12.4	1.6
	14908	13	687	97.2	76.0	21.3	2.8
	14971	31	724	98.5	94.6	3.9	1.5
	14884	NR	687	97.5	87.6	9.9	2.5
Aggregate			5610	97.8	81.3	16.5	2.2
Aggregate Across Administrations			9969	97.7	82.6	15.1	2.3

Note. NR = not released.

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	16910	NR	19	94.7	89.5	5.3	5.3
	16933	NR	20	95.0	90.0	5.0	5.0
	16517	9	17	88.2	70.6	17.6	11.8
	CMS9759	19	25	96.0	84.0	12.0	4.0
	OQMF31839	28	20	95.0	80.0	15.0	5.0
	16546	NR	19	100.0	94.7	5.3	0.0
	16519	NR	19	94.7	84.2	10.5	5.3
Aggregate			139	95.0	84.9	10.1	5.0
Spring	14418	NR	31	100.0	93.5	6.5	0.0
	CMS9626	8	32	100.0	81.3	18.8	0.0
	14456	NR	32	100.0	93.8	6.3	0.0
	OQMF31791	18	34	97.1	97.1	0.0	2.9
	9927	NR	20	100.0	65.0	35.0	0.0
	12839	NR	21	100.0	90.5	9.5	0.0
	OQMF21395	27	21	100.0	90.5	9.5	0.0
Aggregate			191	99.5	88.5	11.0	0.5
Aggregate Across Administrations			330	97.2	86.7	10.5	2.8

Note. NR = not released.

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	42	97.6	88.1	9.5	2.4
	10064	40	43	100.0	88.4	11.6	0.0
	17064	NR	42	95.2	78.6	16.7	4.8
	14493	8	41	95.1	90.2	4.9	4.9
	9967	NR	44	97.7	75.0	22.7	2.3
	CMS10066	31	46	95.7	91.3	4.3	4.3
	14500	16	47	100.0	89.4	10.6	0.0
Aggregate			305	97.4	85.9	11.5	2.6
Spring	CMS9968	15	67	100.0	97.0	3.0	0.0
	14564	30	71	100.0	85.9	14.1	0.0
	OQMF31900	NR	71	100.0	91.5	8.5	0.0
	17065	NR	70	100.0	92.9	7.1	0.0
	OQMF20902	7	56	98.2	76.8	21.4	1.8
	14524	22	46	100.0	82.6	17.4	0.0
	14566	NR	64	98.4	90.6	7.8	1.6
Aggregate			445	96.7	85.1	11.5	3.3
Aggregate Across Administrations			750	97.0	85.5	11.5	3.0

Note. NR = not released.

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
18163_269	n/a	6	181 209	97.7	60.4	37.4	2.3
17237_265	n/a	7	181 209	96.6	70.5	26.1	3.4
13498_262	IV	6	181 207	95.1	54.1	41.0	4.9
13499_262	IV	7	181 207	96.3	59.8	36.5	3.7
Aggregate			724 832	96.4	61.2	35.2	3.6

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
19840_T	n/a	1	181 208	86.7	42.5	44.2	13.3
19840_V	n/a	1	181 208	96.8	55.8	41.0	3.2
18378_T	I	1	181 207	92.7	47.7	45.0	7.3
18378_V	I	1	181 207	97.4	56.1	41.4	2.6
Aggregate Long Writing			724 830	93.4	50.5	42.9	6.6
18022_T&V	n/a	1	181 209	96.3	72.4	23.9	3.7
18021_T&V	V	1	181 208	97.3	61.0	36.3	2.7
Aggregate Short Writing			362 417	96.8	66.7	30.1	3.2
Aggregate All Items			1 087 247	94.5	55.9	38.6	5.5

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
18325_232	s. o.	6	6 262	99.8	77.4	22.4	0.2
18181_227	s. o.	7	6 262	95.3	69.9	25.4	4.7
13242_236	IV	6	6 262	99.1	71.5	27.6	0.9
18328_236	IV	7	6 262	99.0	66.7	32.3	1.0
Aggregate			25 048	98.3	71.3	26.9	1.7

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
19077_T	s. o.	1	6 262	89.2	44.8	44.4	10.8
19077_V	s. o.	1	6 262	97.2	54.3	42.9	2.8
19076_T	I	1	6 262	92.0	49.0	43.1	8.0
19076_V	I	1	6 262	94.3	50.0	44.3	5.7
Aggregate Long Writing			25 048	93.2	49.5	43.6	6.8
13278_T&V	s. o.	1	6 262	94.5	56.6	37.8	5.5
13272_T&V	V	1	6 262	97.0	64.9	32.1	3.0
Aggregate Short Writing			12 524	95.8	60.8	35.0	4.2
Aggregate All Items			37 572	94.0	53.3	40.8	6.0

APPENDIX 7.1: SCORE DISTRIBUTIONS AND ITEM STATISTICS

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 7.1.77–7.1.82 for the OSSLT. The distribution of score points and the item-category difficulty estimates are also provided for open-response items. Note that the IRT model that was fit to EQAO open-response item data is the generalized partial credit model, so the step parameter estimates from PARSCALE calibration are intersection points of adjacent item-category response curves; for students with a theta value smaller than the intersection point of category 1 and category 2, for example, it is more likely that they will achieve score category 1, and vice versa. In order to convey the difficulties of various item categories (as in the graded response model), the step parameter estimates were transformed, by first obtaining the cumulative item-category response functions and then, through each of these functions, locating the value on the theta scale that corresponds to a probability of 0.5. In this document, the resulting estimates are called item-category-difficulty parameter estimates.

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 (Section)	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
17297	1 (A)	1	R3.0	I	2	0.73	0.31	-1.13	0.50
17296	1 (A)	2	R2.0	I	3	0.50	0.30	0.46	0.64
17298	1 (A)	3	R3.0	C	4	0.63	0.40	-0.38	0.70
17295	1 (A)	4	R1.0	C	4	0.51	0.26	0.53	0.47
17293	1 (A)	5	R1.0	C	4*	0.57 (2.27)	0.49	-1.59	0.36
17294	1 (A)	6	R1.0	C	4*	0.51 (2.04)	0.47	-0.36	0.44
17244	1 (A)	7	R3.0	E	4	0.68	0.41	-0.53	0.77
17243	1 (A)	8	R2.0	C	1	0.82	0.38	-1.61	0.70
17245	1 (A)	9	R3.0	I	1	0.70	0.34	-0.91	0.54
17242	1 (A)	10	R1.0	I	3	0.60	0.34	-0.10	0.64
17241	1 (A)	11	R2.0	I	4*	0.47 (1.87)	0.49	-0.39	0.45
17240	1 (A)	12	R1.0	C	4*	0.47 (1.87)	0.48	-0.50	0.44
17346	1 (B)	1	R1.0	I	3	0.57	0.42	-0.06	0.79
17348	1 (B)	2	R3.0	E	4	0.80	0.42	-1.24	0.85
17347	1 (B)	3	R2.0	C	3	0.71	0.39	-0.78	0.71
17349	1 (B)	4	R3.0	C	1	0.38	0.23	1.21	0.70
17344	1 (B)	5	R1.0	C	4*	0.57 (2.28)	0.45	-1.04	0.42
17343	1 (B)	6	R1.0	E	4*	0.60 (2.41)	0.54	-1.08	0.51
10910	2 (C)	1	R2.0	C	3	0.82	0.39	-1.57	0.70
10913	2 (C)	2	R3.0	C	1	0.65	0.33	-0.43	0.60
10894	2 (C)	3	R1.0	E	4	0.75	0.34	-1.05	0.58
10901	2 (C)	4	R3.0	I	1	0.87	0.35	-1.95	0.70
10902	2 (C)	5	R2.0	I	4	0.67	0.47	-0.50	0.96
10900	2 (C)	6	R2.0	C	3	0.81	0.42	-1.31	0.83
10897	2 (C)	7	R1.0	I	4	0.79	0.34	-1.40	0.58
10904	2 (C)	8	R3.0	I	3	0.52	0.33	0.35	0.56
10909	2 (C)	9	R1.0	C	2	0.58	0.33	0.05	0.67
10908	2 (C)	10	R1.0	I	1	0.67	0.42	-0.52	0.81
10912	2 (C)	11	R2.0	I	4*	0.44 (1.76)	0.45	0.13	0.46
10899	2 (C)	12	R1.0	C	4*	0.47 (1.86)	0.48	-0.12	0.47
17354	2 (D)	1	R2.0	I	1	0.83	0.28	-2.28	0.42
17353	2 (D)	2	R1.0	I	1	0.67	0.38	-0.53	0.69
17352	2 (D)	3	R1.0	I	4	0.55	0.39	0.08	0.78
17355	2 (D)	4	R3.0	I	4	0.57	0.22	0.18	0.37
17351	2 (D)	5	R2.0	C	4*	0.47 (1.86)	0.48	-0.16	0.51
17350	2 (D)	6	R1.0	C	4*	0.49 (1.97)	0.52	-0.29	0.58

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

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

() = mean score for OR items.

Table 7.1.2 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Primary Reading (English)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
17293	1 (A)	5	% of Students	0.43	0.57	28.44	31.14	21.43	18.00
			Parameters			-7.09	-1.24	0.37	1.63
17294	1 (A)	6	% of Students	0.65	0.51	24.38	45.34	27.71	1.41
			Parameters			-5.68	-1.49	0.99	4.72
17241	1 (A)	11	% of Students	1.28	1.57	33.12	43.82	14.55	5.67
			Parameters			-4.67	-0.86	1.38	2.61
17240	1 (A)	12	% of Students	1.20	0.81	38.22	36.14	17.62	6.00
			Parameters			-5.27	-0.58	1.18	2.68
17344	1 (B)	5	% of Students	0.57	0.73	12.21	51.49	26.89	8.11
			Parameters			-5.02	-2.55	0.73	2.69
17343	1 (B)	6	% of Students	0.85	0.68	13.38	41.01	30.55	13.53
			Parameters			-4.27	-2.02	0.17	1.81
10912	2 (C)	11	% of Students	1.28	1.60	34.73	47.68	13.00	1.71
			Parameters			-4.45	-0.73	1.88	3.81
10899	2 (C)	12	% of Students	1.29	1.41	36.94	34.95	22.10	3.31
			Parameters			-4.43	-0.51	1.08	3.37
17351	2 (D)	5	% of Students	0.86	0.98	30.40	50.14	15.53	2.08
			Parameters			-4.76	-1.00	1.61	3.51
17350	2 (D)	6	% of Students	1.12	0.70	26.28	48.53	20.28	3.09
			Parameters			-4.20	-1.08	1.14	2.98

Note. The total number of students is 117 920.

Table 7.1.3 Item Statistics: Junior Reading (English)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
17898	1 (A)	1	R2.0	C	4	0.68	0.28	-0.91	0.41
17896	1 (A)	2	R1.0	E	3	0.83	0.28	-1.94	0.48
17897	1 (A)	3	R1.0	I	2	0.54	0.40	0.12	0.64
17899	1 (A)	4	R3.0	I	2	0.83	0.39	-1.68	0.68
17894	1 (A)	5	R1.0	C	4*	0.59 (2.34)	0.43	-1.50	0.40
17895	1 (A)	6	R2.0	I	4*	0.49 (1.94)	0.49	-0.48	0.44
17916	1 (A)	7	R1.0	E	3	0.71	0.31	-1.03	0.48
17917	1 (A)	8	R1.0	I	4	0.66	0.40	-0.62	0.60
17920	1 (A)	9	R3.0	C	1	0.61	0.36	-0.32	0.57
17919	1 (A)	10	R2.0	C	4	0.77	0.30	-1.62	0.44
17915	1 (A)	11	R1.0	C	4*	0.55 (2.21)	0.55	-0.54	0.55
17914	1 (A)	12	R1.0	C	4*	0.56 (2.24)	0.53	-0.75	0.54
17910	1 (B)	1	R1.0	I	2	0.67	0.25	-0.72	0.37
17913	1 (B)	2	R3.0	I	1	0.45	0.27	0.94	0.46
17912	1 (B)	3	R2.0	I	2	0.73	0.40	-0.97	0.69
17911	1 (B)	4	R1.0	C	3	0.62	0.40	-0.23	0.74
17908	1 (B)	5	R1.0	C	4*	0.53 (2.12)	0.51	-0.64	0.55
17909	1 (B)	6	R1.0	C	4*	0.52 (2.07)	0.55	-0.57	0.62
12394	2 (C)	1	R1.0	E	2	0.75	0.19	-2.14	0.25
12403	2 (C)	2	R2.0	I	4	0.54	0.41	0.12	0.84
12395	2 (C)	3	R1.0	E	4	0.81	0.41	-1.31	0.83
12396	2 (C)	4	R3.0	C	2	0.65	0.34	-0.44	0.59
12397	2 (C)	5	R3.0	E	3	0.74	0.48	-0.89	0.98
12398	2 (C)	6	R1.0	I	3	0.76	0.34	-1.23	0.57
12400	2 (C)	7	R1.0	I	2	0.84	0.43	-1.44	0.92
12399	2 (C)	8	R1.0	I	4	0.80	0.36	-1.58	0.57
12401	2 (C)	9	R1.0	I	4	0.78	0.35	-1.40	0.58
12402	2 (C)	10	R2.0	I	1	0.79	0.36	-1.58	0.58
12405	2 (C)	11	R2.0	I	4*	0.51 (2.05)	0.55	-0.33	0.59
12404	2 (C)	12	R1.0	I	4*	0.53 (2.11)	0.59	-0.47	0.70
17876	2 (D)	1	R2.0	I	4	0.64	0.42	-0.46	0.70
17873	2 (D)	2	R1.0	E	3	0.79	0.33	-1.48	0.57
17875	2 (D)	3	R1.0	I	2	0.69	0.32	-0.87	0.48
17878	2 (D)	4	R3.0	C	4	0.47	0.25	0.86	0.48
17872	2 (D)	5	R2.0	I	4*	0.46 (1.84)	0.48	-0.25	0.46
17871	2 (D)	6	R1.0	C	4*	0.49 (1.95)	0.56	-0.22	0.63

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

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

() = mean score for OR items.

Table 7.1.4 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Junior Reading (English)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
17894	1 (A)	5	% of Students	0.23	0.17	11.87	47.14	34.97	5.62
			Parameters			-6.92	-2.83	0.36	3.38
17895	1 (A)	6	% of Students	1.27	0.70	35.52	33.81	24.17	4.52
			Parameters			-5.25	-0.71	0.85	3.17
17915	1 (A)	11	% of Students	1.35	1.63	19.45	37.93	32.51	7.13
			Parameters			-3.52	-1.38	0.30	2.43
17914	1 (A)	12	% of Students	1.04	0.58	16.64	43.66	31.89	6.19
			Parameters			-4.24	-1.77	0.39	2.62
17908	1 (B)	5	% of Students	0.56	0.43	18.64	51.71	24.89	3.77
			Parameters			-4.68	-1.76	0.88	3.00
17909	1 (B)	6	% of Students	0.63	0.57	22.54	48.78	23.36	4.12
			Parameters			-4.42	-1.37	0.85	2.68
12405	2 (C)	11	% of Students	1.19	1.17	23.76	44.17	25.97	3.74
			Parameters			-3.72	-1.23	0.74	2.90
12404	2 (C)	12	% of Students	1.03	0.77	20.72	46.54	26.53	4.41
			Parameters			-3.64	-1.36	0.63	2.51
17872	2 (D)	5	% of Students	1.41	0.64	40.34	32.38	21.83	3.41
			Parameters			-4.98	-0.45	1.04	3.38
17871	2 (D)	6	% of Students	1.62	0.46	30.64	40.71	23.56	3.02
			Parameters			-3.83	-0.85	0.85	2.96

Note. The total number of students is 126 119.

Table 7.1.5 Item Statistics: Primary Reading (French)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
17635	1 (A)	1	A	L	3	0.48	0.39	0.48	0.96
17638	1 (A)	2	C	L	2	0.74	0.47	-0.78	0.98
17636	1 (A)	3	B	I	4	0.71	0.39	-0.68	0.74
17637	1 (A)	4	C	I	3	0.68	0.33	-0.63	0.54
17639	1 (A)	5	A	L	4*	0.50 (1.99)	0.48	-0.33	0.52
17640	1 (A)	6	A	L	4*	0.51 (2.05)	0.43	-0.51	0.43
17654	1 (A)	7	B	L	3	0.83	0.38	-1.47	0.75
17656	1 (A)	8	C	L	2	0.55	0.22	0.35	0.39
17653	1 (A)	9	A	I	3	0.34	0.18	1.91	0.54
17655	1 (A)	10	C	E	4	0.57	0.41	0.10	0.95
17657	1 (A)	11	A	I	4*	0.54 (2.15)	0.42	-0.62	0.52
17658	1 (A)	12	A	L	4*	0.48 (1.90)	0.50	0.00	0.64
17618	1 (B)	1	B	L	2	0.89	0.30	-2.20	0.64
17620	1 (B)	2	C	I	1	0.54	0.34	0.23	0.60
17619	1 (B)	3	C	E	3	0.86	0.39	-1.50	0.92
17617	1 (B)	4	A	I	2	0.49	0.35	0.49	0.81
17621	1 (B)	5	A	I	4*	0.58 (2.32)	0.47	-0.98	0.56
17622	1 (B)	6	A	L	4*	0.54 (2.14)	0.36	-0.73	0.44
17582	2 (C)	1	A	I	4	0.88	0.41	-1.65	1.01
17581	2 (C)	2	A	E	2	0.76	0.31	-1.14	0.55
17583	2 (C)	3	B	I	2	0.67	0.38	-0.46	0.72
17590	2 (C)	4	C	L	3	0.73	0.38	-0.79	0.70
17584	2 (C)	5	B	L	3	0.51	0.33	0.42	0.67
17585	2 (C)	6	C	I	2	0.71	0.48	-0.53	1.24
17589	2 (C)	7	C	L	3	0.70	0.43	-0.57	0.90
17587	2 (C)	8	C	I	4	0.67	0.45	-0.31	1.13
17586	2 (C)	9	C	I	3	0.59	0.37	-0.02	0.77
17588	2 (C)	10	C	L	2	0.83	0.29	-1.70	0.56
17591	2 (C)	11	A	I	4*	0.58 (2.30)	0.48	-0.84	0.51
17592	2 (C)	12	A	L	4*	0.49 (1.96)	0.37	0.50	0.64
17672	2 (D)	1	A	I	2	0.49	0.26	0.66	0.52
17673	2 (D)	2	B	I	4	0.84	0.45	-1.31	1.04
17671	2 (D)	3	A	I	2	0.58	0.48	0.00	1.19
17674	2 (D)	4	C	I	2	0.57	0.47	0.05	1.11
17675	2 (D)	5	A	I	4*	0.57 (2.28)	0.41	-0.84	0.54
17676	2 (D)	6	B	L	4*	0.52 (2.08)	0.39	-0.28	0.47

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.

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

() = mean score for OR items.

Table 7.1.6 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Primary Reading (French)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
17639	1 (A)	5	% of Students	0.27	0.44	26.22	47.76	23.92	1.39
			Parameters			-5.44	-1.21	1.15	4.18
17640	1 (A)	6	% of Students	0.50	0.54	21.20	51.58	23.67	2.50
			Parameters			-5.55	-1.73	1.25	4.00
17657	1 (A)	11	% of Students	0.56	0.27	8.92	67.88	19.00	3.37
			Parameters			-4.37	-2.60	1.40	3.08
17658	1 (A)	12	% of Students	1.14	0.39	23.33	60.28	13.43	1.43
			Parameters			-3.80	-1.19	1.67	3.33
17621	1 (B)	5	% of Students	0.49	0.27	6.56	59.34	26.93	6.42
			Parameters			-4.33	-2.80	0.74	2.47
17622	1 (B)	6	% of Students	0.46	0.23	6.93	72.05	17.87	2.46
			Parameters			-5.05	-3.34	1.74	3.74
17591	2 (C)	11	% of Students	0.37	0.70	9.69	54.22	28.28	6.74
			Parameters			-4.23	-2.44	0.69	2.62
17592	2 (C)	12	% of Students	0.64	0.81	6.66	86.80	4.97	0.11
			Parameters			-3.31	-2.45	2.94	4.81
17675	2 (D)	5	% of Students	0.34	0.36	2.37	68.91	24.19	3.83
			Parameters			-3.98	-3.46	1.08	3.00
17676	2 (D)	6	% of Students	1.23	1.04	6.13	75.77	12.82	3.01
			Parameters			-3.52	-2.74	1.92	3.21

Note. The total number of students is 7023.

Table 7.1.7 Item Statistics: Junior Reading (French)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
18225	1 (A)	1	A	I	4	0.66	0.30	-0.59	0.46
18226	1 (A)	2	B	L	1	0.72	0.37	-0.78	0.70
18228	1 (A)	3	C	I	3	0.54	0.37	0.19	0.77
18227	1 (A)	4	C	I	4	0.66	0.28	-0.51	0.48
18229	1 (A)	5	A	I	4*	0.65 (2.59)	0.47	-2.21	0.42
18230	1 (A)	6	A	L	4*	0.57 (2.29)	0.44	-1.54	0.41
18246	1 (A)	7	C	L	2	0.54	0.21	0.48	0.37
18243	1 (A)	8	A	E	2	0.89	0.33	-2.09	0.69
18244	1 (A)	9	A	I	3	0.65	0.35	-0.43	0.61
18245	1 (A)	10	B	L	1	0.50	0.25	0.61	0.46
18248	1 (A)	11	A	L	4*	0.52 (2.06)	0.41	-1.66	0.32
18247	1 (A)	12	A	L	4*	0.65 (2.58)	0.42	-1.77	0.30
18208	1 (B)	1	A	I	2	0.63	0.45	-0.25	0.99
18209	1 (B)	2	C	I	1	0.43	0.26	1.05	0.50
18207	1 (B)	3	A	E	2	0.84	0.41	-1.41	0.91
18210	1 (B)	4	B	I	2	0.61	0.28	-0.19	0.44
18211	1 (B)	5	A	I	4*	0.73 (2.92)	0.50	-2.14	0.49
18212	1 (B)	6	A	L	4*	0.72 (2.86)	0.49	-1.99	0.50
18198	2 (C)	1	A	I	2	0.81	0.28	-1.80	0.46
18196	2 (C)	2	A	E	3	0.92	0.27	-2.75	0.62
18201	2 (C)	3	B	I	4	0.91	0.37	-1.88	1.01
18202	2 (C)	4	C	I	3	0.77	0.39	-1.00	0.80
18204	2 (C)	5	C	L	2	0.42	0.32	0.80	0.88
18203	2 (C)	6	C	I	1	0.49	0.22	0.76	0.44
18195	2 (C)	7	A	E	2	0.76	0.32	-1.24	0.55
18197	2 (C)	8	A	I	3	0.37	0.20	1.63	0.53
18200	2 (C)	9	A	I	3	0.80	0.38	-1.31	0.74
18199	2 (C)	10	A	I	2	0.75	0.44	-0.87	0.89
18205	2 (C)	11	A	I	4*	0.64 (2.54)	0.51	-1.25	0.53
18206	2 (C)	12	A	L	4*	0.63 (2.52)	0.46	-1.52	0.47
18215	2 (D)	1	B	I	1	0.83	0.30	-1.56	0.63
18216	2 (D)	2	C	L	1	0.72	0.30	-1.00	0.52
18214	2 (D)	3	A	L	3	0.61	0.28	-0.19	0.49
18213	2 (D)	4	A	I	1	0.79	0.31	-1.58	0.50
18217	2 (D)	5	A	I	4*	0.57 (2.28)	0.47	-1.25	0.54
18218	2 (D)	6	B	L	4*	0.59 (2.37)	0.36	-1.53	0.33

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.

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

() = mean score for OR items.

Table 7.1.8 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Junior Reading (French)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
18229	1 (A)	5	% of Students	0.14	0.05	11.07	34.96	36.67	17.11
			Parameters			-7.79	-2.57	-0.27	1.79
18230	1 (A)	6	% of Students	0.19	0.11	15.17	46.54	30.81	7.18
			Parameters			-7.30	-2.33	0.53	2.95
18248	1 (A)	11	% of Students	0.16	0.13	34.80	30.78	27.24	6.89
			Parameters			-9.93	-0.94	0.78	3.43
18247	1 (A)	12	% of Students	0.85	0.64	12.26	34.26	30.33	21.66
			Parameters			-5.87	-2.70	-0.23	1.74
18211	1 (B)	5	% of Students	0.27	0.00	4.24	26.46	41.20	27.83
			Parameters			-5.30	-3.30	-0.99	1.02
18212	1 (B)	6	% of Students	0.21	0.10	4.53	23.06	53.25	18.85
			Parameters			-5.32	-3.09	-1.21	1.66
18205	2 (C)	11	% of Students	0.54	0.19	10.06	35.52	42.20	11.48
			Parameters			-4.59	-2.28	-0.25	2.12
18206	2 (C)	12	% of Students	0.27	0.10	6.92	42.98	39.54	10.19
			Parameters			-5.39	-3.03	-0.07	2.42
18217	2 (D)	5	% of Students	0.33	0.00	6.83	65.23	19.38	8.23
			Parameters			-5.06	-3.06	0.97	2.15
18218	2 (D)	6	% of Students	0.45	0.05	8.18	51.48	33.56	6.28
			Parameters			-6.78	-3.74	0.58	3.80

Note. The total number of students is 6298.

Table 7.1.9 Item Statistics: Primary Writing (English)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
17386_T	1(A)	13	W2.0	T	4*	0.55 (2.20)	0.59	-0.41	0.73
17386_V	1(A)	13	W3.0	V	3*	0.65 (1.96)	0.57	-1.26	0.83
18179_T	2(C)	13	W2.0	T	4*	0.58 (2.30)	0.58	-0.61	0.96
18179_V	2(C)	13	W3.0	V	3*	0.68 (2.04)	0.60	-1.28	1.04
17383	2(C)	14	W2.0	T	4	0.69	0.36	-0.73	0.61
17358	2(C)	15	W1.0	T	3	0.72	0.40	-0.93	0.67
17356	2(C)	16	W2.0	T	3	0.61	0.31	-0.27	0.48
17363	2(C)	17	W1.0	T	1	0.74	0.37	-1.07	0.66
16475_T	2(D)	7	W2.0	T	4*	0.63 (2.51)	0.56	-0.78	0.65
16475_V	2(D)	7	W3.0	V	3*	0.71 (2.12)	0.60	-1.48	0.93
17367	2(D)	8	W3.0	V	4	0.72	0.45	-0.79	0.90
17392	2(D)	9	W1.0	T	3	0.66	0.40	-0.55	0.67
17388	2(D)	10	W3.0	V	2	0.56	0.33	0.08	0.56
17988	2(D)	11	W1.0	T	3	0.71	0.30	-0.95	0.48

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

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

() = mean score for OR items.

Table 7.1.10 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Primary Writing (English)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
17386_T	1(A)	13	% of Students	0.61	2.04	17.91	42.56	30.14	6.74
			Parameters			-2.15	-1.81	-1.02	0.71
17386_V	1(A)	13	% of Students	0.61	0.37	33.21	34.86	30.95	
			Parameters			-2.81	-2.44	-0.37	
18179_T	2(C)	13	% of Students	0.51	0.58	10.30	52.67	29.81	6.13
			Parameters			-2.23	-2.04	-1.42	0.58
18179_V	2(C)	13	% of Students	0.51	0.21	25.73	42.23	31.31	
			Parameters			-2.61	-2.36	-0.75	
16475_T	2(D)	7	% of Students	0.52	2.21	8.27	36.73	40.10	12.17
			Parameters			-2.44	-2.00	-1.45	0.15
16475_V	2(D)	7	% of Students	0.52	0.21	21.31	42.96	34.99	
			Parameters			-2.87	-2.57	-0.97	

Note. The total number of students is 117 955.

Table 7.1.11 Item Statistics: Junior Writing (English)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
17958_T	1(A)	13	W2.0	T	4*	0.55 (2.21)	0.55	-0.60	0.71
17958_V	1(A)	13	W3.0	V	3*	0.64 (1.91)	0.56	-1.23	0.91
17950_T	2(C)	13	W2.0	T	4*	0.51 (2.04)	0.50	-0.41	0.52
17950_V	2(C)	13	W3.0	V	3*	0.63 (1.89)	0.58	-0.98	1.05
17926	2(C)	14	W1.0	T	3	0.85	0.30	-2.24	0.52
17728	2(C)	15	W3.0	V	2	0.76	0.31	-1.52	0.46
17929	2(C)	16	W2.0	T	4	0.64	0.42	-0.41	0.76
17925	2(C)	17	W1.0	T	2	0.70	0.31	-1.01	0.46
18996_T	2(D)	7	W2.0	T	4*	0.49 (1.94)	0.61	-0.38	0.92
18996_V	2(D)	7	W3.0	V	3*	0.62 (1.85)	0.62	-1.05	1.19
17954	2(D)	8	W3.0	V	4	0.42	0.30	0.87	0.70
17944	2(D)	9	W1.0	T	2	0.80	0.32	-1.71	0.52
17957	2(D)	10	W2.0	T	4	0.65	0.29	-0.60	0.46
17729	2(D)	11	W3.0	V	1	0.46	0.32	0.64	0.66

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

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

() = mean score for OR items.

Table 7.1.12 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Junior Writing (English)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
17958_T	1(A)	13	% of Students	0.59	0.75	13.15	53.53	27.51	4.47
			Parameters			-2.60	-2.30	-1.49	0.83
17958_V	1(A)	13	% of Students	0.59	0.2	23.40	59.82	15.99	
			Parameters			-2.79	-2.52	-1.04	
17950_T	2(C)	13	% of Students	0.95	1.53	24.91	43.93	23.57	5.12
			Parameters			-2.84	-2.32	-0.97	1.37
17950_V	2(C)	13	% of Students	0.95	0.33	19.61	67.73	11.38	
			Parameters			-2.34	-2.12	-1.12	
18996_T	2(D)	7	% of Students	0.67	0.53	32.79	41.49	19.82	4.69
			Parameters			-2.26	-2.08	-0.74	1.04
18996_V	2(D)	7	% of Students	0.67	0.21	30.82	51.04	17.25	
			Parameters			-2.40	-2.22	-0.72	

Note. The total number of students is 125 732.

Table 7.1.13 Item Statistics: Primary Writing (French)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
17708_T	1(A)	13	A	T	4*	0.45 (1.79)	0.53	0.31	0.65
17708_V	1(A)	13	C	V	3*	0.62 (1.87)	0.58	-1.00	1.02
17692_T	2(C)	13	A	T	4*	0.49 (1.96)	0.54	-0.13	0.73
17692_V	2(C)	13	C	V	3*	0.62 (1.85)	0.59	-1.06	1.12
17702	2(C)	14	A	T	3	0.77	0.36	-1.08	0.69
17786	2(C)	15	C	V	1	0.63	0.28	-0.29	0.49
17691	2(C)	16	C	V	2	0.53	0.29	0.39	0.54
17699	2(C)	17	C	V	3	0.67	0.36	-0.50	0.67
15509_T	2(D)	7	A	T	4*	0.47 (1.89)	0.55	-0.12	0.71
15509_V	2(D)	7	C	V	3*	0.59 (1.76)	0.58	-0.87	1.07
17690	2(D)	8	B	T	4	0.79	0.37	-1.25	0.67
17689	2(D)	9	A	T	1	0.61	0.35	-0.18	0.62
17706	2(D)	10	A	T	4	0.75	0.35	-0.97	0.61
17695	2(D)	11	C	V	3	0.86	0.31	-1.88	0.62

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.

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

() = mean score for OR items.

Table 7.1.14 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Primary Writing (French)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
17708_T	1(A)	13	% of Students	0.64	5.97	28.11	47.27	15.50	2.51
			Parameters			-1.61	-1.21	-0.33	1.76
17708_V	1(A)	13	% of Students	0.64	0.17	29.79	50.99	18.41	
			Parameters			-2.44	-2.21	-0.62	
17692_T	2(C)	13	% of Students	0.33	1.37	27.21	46.90	21.70	2.49
			Parameters			-2.23	-1.95	-0.77	1.28
17692_V	2(C)	13	% of Students	0.33	0.11	28.32	56.72	14.52	
			Parameters			-2.57	-2.40	-0.72	
15509_T	2(D)	7	% of Students	0.51	1.50	32.24	43.62	18.71	3.42
			Parameters			-2.20	-1.89	-0.60	1.42
15509_V	2(D)	7	% of Students	0.51	0.23	36.17	49.58	13.51	
			Parameters			-2.41	-2.23	-0.40	

Note. The total number of students is 7019.

Table 7.1.15 Item Statistics: Junior Writing (French)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
18303_T	1(A)	13	A	T	4*	0.57 (2.28)	0.61	-0.67	0.80
18303_V	1(A)	13	C	V	3*	0.68 (2.03)	0.58	-1.46	0.93
18302_T	2(C)	13	A	T	4*	0.58 (2.32)	0.62	-0.83	0.97
18302_V	2(C)	13	C	V	3*	0.69 (2.07)	0.60	-1.58	1.04
18294	2(C)	14	A	T	1	0.76	0.32	-1.13	0.58
18295	2(C)	15	B	T	3	0.63	0.37	-0.29	0.62
18293	2(C)	16	A	T	4	0.77	0.28	-1.44	0.45
18300	2(C)	17	C	V	3	0.61	0.35	-0.21	0.65
16819_T	2(D)	7	A	T	4*	0.68 (2.71)	0.62	-1.53	0.86
16819_V	2(D)	7	C	V	3*	0.70 (2.10)	0.59	-1.98	0.90
18311	2(D)	8	C	V	4	0.52	0.36	0.35	0.70
18296	2(D)	9	A	T	1	0.70	0.24	-1.12	0.34
18304	2(D)	10	B	T	3	0.61	0.32	-0.18	0.49
18297	2(D)	11	C	V	4	0.67	0.33	-0.65	0.56

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.

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

() = mean score for OR items.

Table 7.1.16 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Junior Writing (French)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
18303_T	1(A)	13	% of Students	0.27	1.18	16.49	43.75	28.89	9.42
			Parameters			-2.40	-2.12	-1.23	0.61
18303_V	1(A)	13	% of Students	0.27	0.24	21.03	52.98	25.48	
			Parameters			-2.99	-2.72	-1.08	
18302_T	2(C)	13	% of Students	0.13	0.35	15.08	43.61	33.13	7.70
			Parameters			-2.62	-2.45	-1.43	0.39
18302_V	2(C)	13	% of Students	0.13	0.03	20.12	52.34	27.38	
			Parameters			-3.08	-2.87	-1.16	
16819_T	2(D)	7	% of Students	0.16	0.10	7.43	33.51	38.87	19.93
			Parameters			-3.35	-3.12	-2.12	-0.18
16819_V	2(D)	7	% of Students	0.16	0.02	21.19	47.54	31.09	
			Parameters			-3.81	-3.58	-1.16	

Note. The total number of students is 6272.

Table 7.1.17 Item Statistics: Primary Mathematics (English)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
							Difficulty	Item-Total Correlation	Location	Slope
18359	3(1)	1	NV2	AP	N	4	82.86	0.40	-1.35	0.94
10967	3(1)	2	NV3	AP	N	2	77.44	0.37	-1.19	0.71
10815	3(1)	3	NV3	KU	N	2	71.22	0.44	-0.64	1.01
10783	3(1)	4	PV2	AP	P	1	64.67	0.43	-0.30	1.06
18348	3(1)	5	PV2	AP	P	2	59.84	0.38	-0.11	0.81
15140	3(1)	6	PV1	PS	P	3	67.82	0.45	-0.50	1.00
15093	3(1)	7	NV1	AP	N	4*	61.75(2.47)	0.56	-1.68	0.37
10847	3(1)	8	PV1	AP	P	4*	70.50(2.82)	0.51	-2.37	0.35
15173	3(1)	9	DV2	PS	D	4*	51.00(2.04)	0.46	-0.82	0.33
15132	3(1)	10	GV1	PS	G	4*	66.50(2.66)	0.48	-1.53	0.32
10687	3(1)	11	NV1	KU	N	3	76.91	0.36	-1.16	0.67
15097	3(1)	12	MV1	KU	M	4	74.16	0.32	-0.98	0.63
18360	3(1)	13	GV2	KU	G	4	72.27	0.36	-0.85	0.67
15146	3(1)	14	PV2	KU	P	3	67.77	0.31	-0.58	0.59
10973	3(1)	15	MV2	AP	M	4	43.88	0.40	0.55	1.17
15106	3(1)	16	MV2	PS	M	2	62.00	0.44	-0.18	1.10
15142	3(1)	17	PV1	KU	P	2	80.05	0.45	-1.09	1.12
15111	3(1)	18	MV2	KU	M	3	85.72	0.42	-1.57	0.94
10844	3(2)	1	DV1	KU	D	4	87.23	0.41	-1.66	0.94
15085	3(2)	2	NV3	AP	N	3	83.01	0.38	-1.48	0.77
15098	3(2)	3	MV1	KU	M	1	76.58	0.31	-1.39	0.51
12011	3(2)	4	NV3	PS	N	1	54.62	0.25	0.28	0.61
18350	3(2)	5	PV1	KU	P	2	80.68	0.32	-1.85	0.47
10974	3(2)	6	GV1	KU	G	3	67.77	0.36	-0.51	0.73
10980	3(2)	7	DV3	AP	D	2	83.44	0.37	-1.58	0.73
10846	3(2)	8	GV3	AP	G	4*	68.25(2.73)	0.47	-2.41	0.29
10744	3(2)	9	DV1	AP	D	4*	79.75(3.19)	0.61	-1.97	0.50
15116	3(2)	10	MV2	PS	M	4*	56.00(2.24)	0.61	-1.08	0.53
15095	3(2)	11	NV3	PS	N	4*	64.75(2.59)	0.61	-1.32	0.50
10696	3(2)	12	MV1	AP	M	2	85.35	0.37	-1.69	0.76
18355	3(2)	13	GV3	AP	G	1	56.86	0.28	0.13	0.60
10785	3(2)	14	DV1	AP	D	3	63.32	0.30	-0.30	0.56
15124	3(2)	15	GV2	AP	G	4	72.15	0.35	-0.79	0.68
15080	3(2)	16	NV1	PS	N	2	59.97	0.39	-0.12	0.88
12556	3(2)	17	MV1	AP	M	3	62.83	0.40	-0.25	0.84
15103	3(2)	18	MV2	KU	M	1	70.63	0.38	-0.71	0.74

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.

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

() = mean score for OR items.

**Table 7.1.18 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items:
Primary Mathematics (English)**

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
15093	3(1)	7	% of Students	0.93	0.55	37.73	10.96	11.78	38.06
			Parameters			-6.36	-0.51	-0.12	0.30
10847	3(1)	8	% of Students	0.35	0.11	20.24	12.45	30.44	36.42
			Parameters			-7.66	-1.56	-0.88	0.52
15173	3(1)	9	% of Students	1.87	0.72	34.14	29.83	23.16	10.27
			Parameters			-5.64	-0.86	0.68	2.53
15132	3(1)	10	% of Students	0.93	18.42	21.51	23.62	33.53	0.93
			Parameters			-4.73	-1.58	-0.44	0.68
10846	3(2)	8	% of Students	0.67	0.18	23.82	15.10	21.65	38.57
			Parameters			-8.17	-1.38	-0.56	0.46
10744	3(2)	9	% of Students	1.01	0.24	13.54	12.59	10.41	62.21
			Parameters			-4.62	-1.61	-0.93	-0.61
15116	3(2)	10	% of Students	0.85	0.34	36.65	22.54	16.64	22.97
			Parameters			-5.02	-0.51	0.29	0.96
15095	3(2)	11	% of Students	1.36	0.29	24.51	24.30	12.69	36.85
			Parameters			-4.51	-1.01	-0.04	0.37

Note. The total number of students is 122 760.

Table 7.1.19 Item Statistics: Junior Mathematics (English)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
							Difficulty	Item-Total Correlation	Location	Slope
11174	3(1)	1	NV3	KU	N	1	67.44	0.42	-0.52	0.95
15058	3(1)	2	DV2	AP	D	4	64.61	0.32	-0.47	0.58
18369	3(1)	3	PV2	KU	P	1	74.36	0.30	-1.23	0.52
11413	3(1)	4	PV1	KU	P	3	56.33	0.33	0.06	0.68
11305	3(1)	5	DV3	AP	D	3	62.75	0.35	-0.32	0.70
15018	3(1)	6	GV2	KU	G	1	77.39	0.35	-1.15	0.74
17405	3(1)	7	NV1	PS	N	4*	66.50(2.66)	0.60	-1.59	0.45
15068	3(1)	8	DV2	PS	D	4*	60.00(2.40)	0.54	-1.48	0.40
11409	3(1)	9	DV3	AP	D	4*	69.25(2.77)	0.60	-1.47	0.50
15010	3(1)	10	MV2	AP	M	4*	55.50(2.22)	0.65	-0.89	0.57
11319	3(1)	11	GV3	AP	G	1	64.55	0.36	-0.40	0.72
11318	3(1)	12	GV1	PS	G	2	50.87	0.27	0.44	0.65
18371	3(1)	13	NV1	PS	N	1	39.97	0.43	0.63	1.27
15006	3(1)	14	MV2	KU	M	4	79.43	0.39	-1.21	0.85
15026	3(1)	15	GV3	PS	G	1	56.50	0.38	0.03	0.86
15000	3(1)	16	MV2	AP	M	1	44.30	0.41	0.50	1.28
11315	3(1)	17	MV2	AP	M	2	69.11	0.34	-0.61	0.73
13042	3(1)	18	NV2	AP	N	3	69.44	0.36	-0.65	0.75
15048	3(2)	1	PV2	AP	P	4	83.48	0.40	-1.42	0.94
11321	3(2)	2	PV1	AP	P	1	84.19	0.43	-1.42	1.09
14984	3(2)	3	NV2	AP	N	2	50.69	0.42	0.24	1.14
11326	3(2)	4	DV3	PS	D	2	53.28	0.32	0.24	0.73
15055	3(2)	5	DV1	AP	D	4	73.92	0.38	-0.94	0.75
18375	3(2)	6	NV3	PS	N	4	61.27	0.35	-0.17	0.80
15002	3(2)	7	MV2	AP	M	1	43.03	0.31	0.70	1.02
18368	3(2)	8	NV3	PS	N	4*	63.00(2.52)	0.66	-1.04	0.50
11310	3(2)	9	PV1	PS	P	4*	58.50(2.34)	0.60	-1.01	0.38
11329	3(2)	10	GV1	AP	G	4*	71.75(2.87)	0.54	-2.01	0.42
15034	3(2)	11	GV3	AP	G	4*	49.50(1.98)	0.59	-0.47	0.48
11292	3(2)	12	NV1	KU	N	1	87.49	0.34	-1.77	0.86
11176	3(2)	13	MV2	PS	M	3	65.00	0.31	-0.41	0.63
14999	3(2)	14	MV2	PS	M	2	73.56	0.32	-0.90	0.67
15061	3(2)	15	DV2	KU	D	3	55.31	0.41	0.06	1.05
15037	3(2)	16	PV1	AP	P	4	32.95	0.33	1.03	1.24
14997	3(2)	17	MV2	KU	M	1	67.40	0.37	-0.58	0.71
15045	3(2)	18	PV1	PS	P	3	69.80	0.38	-0.67	0.80

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.

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

() = mean score for OR items.

**Table 7.1.20 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items:
Junior Mathematics (English)**

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
17405	3(1)	7	% of Students	0.79	0.37	23.79	17.08	23.80	34.17
			Parameters			-5.15	-1.25	-0.46	0.49
15068	3(1)	8	% of Students	0.94	0.21	20.95	39.60	13.52	24.79
			Parameters			-5.39	-1.80	0.38	0.99
11409	3(1)	9	% of Students	1.25	0.3	15.89	17.85	33.77	30.95
			Parameters			-4.20	-1.63	-0.75	0.68
15010	3(1)	10	% of Students	1.47	0.92	40.61	18.18	10.20	28.63
			Parameters			-4.01	-0.32	0.25	0.59
18368	3(2)	8	% of Students	3.47	1.17	29.83	11.60	16.27	37.67
			Parameters			-3.43	-0.68	-0.28	0.26
11310	3(2)	9	% of Students	4.08	1.44	38.51	10.33	7.65	37.98
			Parameters			-3.93	-0.32	0.02	0.29
11329	3(2)	10	% of Students	0.67	0.11	15.32	10.75	42.10	31.06
			Parameters			-5.91	-1.81	-1.22	0.82
15034	3(2)	11	% of Students	3.61	1.32	39.19	23.17	18.76	13.96
			Parameters			-3.65	-0.35	0.54	1.57

Note. The total number of students in Junior Mathematics (English) is 125 625.

Table 7.1.21 Item Statistics: Primary Mathematics (French)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
							Difficulty	Item-Total Correlation	Location	Slope
11214	3(1)	1	NA1	CC	N	2	88.17	0.25	-2.38	0.53
14596	3(1)	2	NA4	CC	N	3	77.77	0.35	-1.06	0.74
14595	3(1)	3	NA4	MA	N	1	67.23	0.42	-0.40	0.93
17779	3(1)	4	AA2	CC	A	2	39.72	0.36	0.85	0.99
14583	3(1)	5	NA3	MA	N	3	68.55	0.31	-0.62	0.55
14572	3(1)	6	NA2	MA	N	4	93.03	0.29	-2.32	0.83
17769	3(1)	7	MA1	HP	M	4*	69.00(2.76)	0.57	-2.06	0.38
14631	3(1)	8	GA1	HP	G	4*	54.00(2.16)	0.44	-1.28	0.33
14667	3(1)	9	TA1	MA	T	4*	53.50(2.14)	0.47	-1.31	0.32
14599	3(1)	10	NA2	MA	N	4*	52.00(2.08)	0.55	-0.77	0.42
17785	3(1)	11	TA2	MA	T	2	76.82	0.13	-2.64	0.21
14658	3(1)	12	TA1	HP	T	3	55.67	0.40	0.14	0.96
14649	3(1)	13	MA4	CC	M	4	73.72	0.46	-0.61	1.25
14608	3(1)	14	AA2	HP	A	3	58.55	0.39	0.01	0.84
17781	3(1)	15	AA3	CC	A	1	89.49	0.27	-2.33	0.58
10998	3(1)	16	GA1	HP	G	3	43.61	0.27	0.91	0.66
17767	3(1)	17	MA1	MA	M	4	63.45	0.42	-0.20	1.02
14630	3(1)	18	GA2	MA	G	4	73.96	0.35	-0.83	0.69
17721	3(2)	1	NA3	CC	N	4	91.54	0.28	-2.29	0.73
14628	3(2)	2	GA1	CC	G	2	70.81	0.35	-0.64	0.75
14615	3(2)	3	AA3	MA	A	2	74.20	0.39	-0.75	0.89
17768	3(2)	4	MA1	HP	M	4	60.67	0.35	-0.04	0.82
11240	3(2)	5	AA1	CC	A	2	88.82	0.21	-2.71	0.46
10997	3(2)	6	GA1	MA	G	1	73.82	0.40	-0.77	0.86
14661	3(2)	7	TA1	CC	T	2	57.96	0.47	0.01	1.20
17773	3(2)	8	GA1	MA	G	4*	71.50(2.86)	0.44	-2.70	0.33
11059	3(2)	9	AA2	HP	A	4*	61.00(2.44)	0.55	-1.28	0.39
14668	3(2)	10	TA1	HP	T	4*	50.25(2.01)	0.52	-1.00	0.39
17766	3(2)	11	NA4	HP	N	4*	64.00(2.56)	0.48	-1.61	0.31
11098	3(2)	12	MA3	CC	M	4	74.38	0.40	-0.82	0.80
11223	3(2)	13	NA2	HP	N	1	70.54	0.40	-0.55	0.88
14646	3(2)	14	MA2	CC	M	3	67.91	0.37	-0.46	0.81
17771	3(2)	15	MA2	MA	M	2	35.66	0.30	1.05	1.10
12616	3(2)	16	NA1	MA	N	3	63.70	0.40	-0.22	0.90
14589	3(2)	17	NA3	HP	N	2	61.54	0.33	-0.15	0.65
17784	3(2)	18	TA1	MA	T	1	75.74	0.44	-0.78	1.04

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.

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

() = mean score for OR items.

**Table 7.1.22 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items:
Primary Mathematics (French)**

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
17769	3(1)	7	% of Students	0.48	0.18	29.10	9.88	13.92	46.43
			Parameters			-7.01	-0.83	-0.44	0.04
14631	3(1)	8	% of Students	0.87	0.06	34.63	22.88	30.77	10.79
			Parameters			-7.45	-0.81	0.34	2.75
14667	3(1)	9	% of Students	0.23	36.73	27.25	15.26	19.25	0.23
			Parameters			-6.74	-0.62	0.66	1.54
14599	3(1)	10	% of Students	0.30	44.50	16.01	14.86	21.60	0.30
			Parameters			-4.60	-0.09	0.47	1.14
17773	3(2)	8	% of Students	0.18	0.07	10.59	25.03	30.77	33.35
			Parameters			-8.09	-2.74	-0.79	0.86
11059	3(2)	9	% of Students	1.71	0.71	26.68	27.72	10.38	32.81
			Parameters			-4.80	-0.99	0.21	0.63
14668	3(2)	10	% of Students	1.22	0.36	44.10	23.02	13.87	17.43
			Parameters			-5.93	-0.20	0.71	1.44
17766	3(2)	11	% of Students	1.44	0.26	27.25	12.23	31.20	27.63
			Parameters			-6.27	-1.05	-0.45	1.24

Note. The total number of students is 7031.

Table 7.1.23 Item Statistics: Junior Mathematics (French)

Item Code	Booklet (Section)	Sequence	Expectation	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
							Difficulty	Item-Total Correlation	Location	Slope
18323	3(1)	1	NA2	MA	N	3	72.44	0.33	-0.82	0.66
11648	3(1)	2	MA3	MA	M	2	71.34	0.37	-0.65	0.82
14734	3(1)	3	MA2	HP	M	4	51.81	0.47	0.21	1.32
14704	3(1)	4	AA2	HP	A	3	56.87	0.40	0.06	0.97
11457	3(1)	5	NA3	MA	N	1	72.63	0.43	-0.67	1.02
14715	3(1)	6	GA2	CC	G	4	71.37	0.36	-0.75	0.66
18317	3(1)	7	TA1	HP	T	4*	54.25(2.17)	0.53	-0.90	0.36
18316	3(1)	8	GA2	HP	G	4*	66.75(2.67)	0.57	-1.55	0.52
18324	3(1)	9	AA1	HP	A	4*	58.25(2.33)	0.44	-2.21	0.26
14688	3(1)	10	NA3	MA	N	4*	71.75(2.87)	0.53	-1.97	0.46
14750	3(1)	11	TA1	MA	T	3	78.11	0.36	-1.16	0.72
18007	3(1)	12	NA2	CC	N	4	83.89	0.31	-1.74	0.62
14758	3(1)	13	TA2	CC	T	2	50.60	0.35	0.37	0.84
14711	3(1)	14	GA1	CC	G	4	64.57	0.44	-0.29	1.00
11507	3(1)	15	TA1	HP	T	2	49.05	0.37	0.41	1.19
14703	3(1)	16	AA2	MA	A	4	55.46	0.40	0.13	0.89
18014	3(1)	17	GA2	CC	G	2	63.74	0.37	-0.26	0.78
14731	3(1)	18	MA2	CC	M	2	68.88	0.40	-0.49	0.90
18015	3(2)	1	MA3	CC	M	2	69.88	0.34	-0.62	0.71
14767	3(2)	2	TA2	HP	T	3	59.80	0.40	-0.06	0.95
14728	3(2)	3	MA1	HP	M	2	40.07	0.29	0.94	0.91
18318	3(2)	4	NA1	MA	N	4	73.86	0.39	-0.74	0.90
11484	3(2)	5	AA1	MA	A	2	63.87	0.30	-0.35	0.53
18321	3(2)	6	GA1	HP	G	4	68.77	0.39	-0.56	0.79
18315	3(2)	7	NA1	CC	N	2	52.83	0.34	0.29	0.74
11533	3(2)	8	GA1	MA	G	4*	67.00(2.68)	0.60	-1.60	0.44
18004	3(2)	9	AA2	MA	A	4*	65.00(2.60)	0.59	-1.40	0.40
11508	3(2)	10	NA1	HP	N	4*	56.75(2.27)	0.63	-0.79	0.59
14744	3(2)	11	MA2	MA	M	4*	64.50(2.58)	0.59	-1.39	0.51
11565	3(2)	12	GA2	HP	G	2	67.09	0.31	-0.49	0.61
18319	3(2)	13	NA3	HP	N	3	57.55	0.35	0.07	0.74
14763	3(2)	14	TA2	MA	T	4	73.14	0.42	-0.71	0.94
14696	3(2)	15	AA1	CC	A	2	44.87	0.33	0.67	1.00
18017	3(2)	16	MA1	CC	M	3	87.83	0.35	-1.70	0.87
14726	3(2)	17	MA1	MA	M	2	40.77	0.29	0.90	0.96
11520	3(2)	18	GA1	MA	G	2	78.21	0.41	-0.96	0.97

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.

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

() = Mean score for OR items.

Table 7.1.24 Distribution of Score Points and Category Difficulty Estimates for Open-Response Items: Junior Mathematics (French)

Item Code	Booklet (Section)	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
18317	3(1)	7	% of Students	3.29	0.72	36.93	22.08	12.24	24.74
			Parameters			-4.54	-0.44	0.46	1.02
18316	3(1)	8	% of Students	0.54	0.08	12.43	27.06	38.70	21.19
			Parameters			-5.00	-1.97	-0.51	1.30
18324	3(1)	9	% of Students	0.73	0.16	36.98	20.87	10.63	30.63
			Parameters			-9.27	-0.69	0.35	0.90
14688	3(1)	10	% of Students	0.45	0.05	7.16	29.13	31.37	31.84
			Parameters			-5.13	-2.83	-0.67	0.75
11533	3(2)	8	% of Students	0.89	0.40	27.90	16.46	10.40	43.94
			Parameters			-5.32	-0.86	-0.23	0.09
18004	3(2)	9	% of Students	2.09	0.46	33.07	9.81	10.50	44.07
			Parameters			-4.83	-0.58	-0.24	0.10
11508	3(2)	10	% of Students	2.01	0.37	26.63	33.42	16.70	20.87
			Parameters			-3.63	-0.92	0.38	1.05
14744	3(2)	11	% of Students	0.97	0.16	17.52	30.79	23.64	26.91
			Parameters			-4.73	-1.56	-0.10	0.87

Note. The total number of students is 6269.

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 Multiple-Choice Items: Primary Reading (English)

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17297	1 (A)	1	0.33	-0.03	0.69		0.65	0.30	1.01	
17296	1 (A)	2	-0.68	-1.00	-0.37		-0.63	-0.94	-0.31	
17298	1 (A)	3	-0.01	-0.35	0.33		0.24	-0.10	0.58	
17295	1 (A)	4	0.75	0.44	1.06		0.65	0.34	0.96	
17244	1 (A)	7	0.27	-0.08	0.63		0.33	-0.02	0.69	
17243	1 (A)	8	-0.26	-0.69	0.16		-0.04	-0.48	0.39	
17245	1 (A)	9	-0.38	-0.73	-0.02		-0.42	-0.77	-0.07	
17242	1 (A)	10	0.22	-0.11	0.55		0.08	-0.25	0.41	
17346	1 (B)	1	-0.62	-0.96	-0.28		-0.52	-0.85	-0.19	
17348	1 (B)	2	0.48	0.05	0.90		0.43	0.03	0.84	
17347	1 (B)	3	-0.40	-0.76	-0.04		-0.01	-0.37	0.35	
17349	1 (B)	4	0.19	-0.13	0.51		0.22	-0.09	0.54	
10910	2 (C)	1	0.07	-0.36	0.51		0.46	0.03	0.90	
10913	2 (C)	2	0.53	0.19	0.86		0.79	0.45	1.13	
10894	2 (C)	3	0.09	-0.27	0.46		0.17	-0.19	0.54	
10901	2 (C)	4	-0.57	-1.05	-0.09		-0.55	-1.01	-0.09	
10902	2 (C)	5	0.01	-0.36	0.37		0.13	-0.23	0.49	
10900	2 (C)	6	-0.24	-0.66	0.19		-0.05	-0.47	0.37	
10897	2 (C)	7	-0.10	-0.49	0.29		-0.29	-0.67	0.09	
10904	2 (C)	8	0.29	-0.03	0.61		0.37	0.05	0.69	
10909	2 (C)	9	0.03	-0.29	0.35		0.17	-0.15	0.50	
10908	2 (C)	10	-0.13	-0.48	0.23		-0.02	-0.38	0.33	
17354	2 (D)	1	0.95	0.53	1.37		0.50	0.08	0.91	
17353	2 (D)	2	1.07	0.73	1.42	B+	0.69	0.34	1.04	
17352	2 (D)	3	0.08	-0.24	0.41		0.10	-0.23	0.43	
17355	2 (D)	4	0.40	0.09	0.71		0.38	0.07	0.68	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17293	1 (A)	5	-0.03	0.59		0.02	0.12	
17294	1 (A)	6	0.05	0.09		0.06	0.07	
17241	1 (A)	11	0.02	0.43		0.04	0.05	
17240	1 (A)	12	0.07	0.09		0.08	0.00	
17344	1 (B)	5	0.04	0.01		0.02	0.54	
17343	1 (B)	6	0.08	0.00		0.04	0.07	
10912	2 (C)	11	0.10	0.00		0.07	0.11	
10899	2 (C)	12	0.05	0.60		0.00	0.07	
17351	2 (D)	5	0.10	0.01		0.10	0.00	
17350	2 (D)	6	-0.02	0.38		-0.03	0.15	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17898	1 (A)	1	0.17	-0.16	0.51		0.18	-0.15	0.51	
17896	1 (A)	2	0.43	0.01	0.84		0.39	-0.01	0.79	
17897	1 (A)	3	1.00	0.66	1.34	B+	0.57	0.24	0.90	
17899	1 (A)	4	0.58	0.14	1.02		0.64	0.20	1.07	
17916	1 (A)	7	0.18	-0.16	0.53		-0.02	-0.37	0.32	
17917	1 (A)	8	0.37	0.02	0.72		-0.32	-0.67	0.02	
17920	1 (A)	9	0.28	-0.05	0.60		0.23	-0.10	0.57	
17919	1 (A)	10	-0.58	-0.95	-0.20		-0.07	-0.43	0.30	
17910	1 (B)	1	0.90	0.57	1.23		0.93	0.60	1.25	
17913	1 (B)	2	0.32	0.01	0.63		0.27	-0.05	0.58	
17912	1 (B)	3	0.02	-0.34	0.38		-0.13	-0.50	0.24	
17911	1 (B)	4	0.91	0.57	1.25		0.74	0.41	1.07	
12394	2 (C)	1	0.51	0.16	0.85		0.46	0.11	0.80	
12403	2 (C)	2	0.18	-0.15	0.51		0.20	-0.13	0.53	
12395	2 (C)	3	0.72	0.30	1.14		0.25	-0.17	0.68	
12396	2 (C)	4	0.30	-0.04	0.63		0.21	-0.12	0.55	
12397	2 (C)	5	-0.12	-0.52	0.28		-0.11	-0.51	0.28	
12398	2 (C)	6	-0.15	-0.52	0.22		0.25	-0.12	0.62	
12400	2 (C)	7	1.16	0.70	1.61	B+	1.15	0.70	1.61	B+
12399	2 (C)	8	0.67	0.26	1.08		0.35	-0.04	0.75	
12401	2 (C)	9	-0.22	-0.61	0.17		-0.44	-0.82	-0.05	
12402	2 (C)	10	-0.83	-1.23	-0.43		-0.31	-0.70	0.09	
17876	2 (D)	1	0.53	0.18	0.88		0.76	0.41	1.11	
17873	2 (D)	2	1.27	0.88	1.66	B+	1.21	0.81	1.60	B+
17875	2 (D)	3	1.27	0.92	1.61	B+	1.35	1.00	1.69	B+
17878	2 (D)	4	1.81	1.49	2.13	C+	1.88	1.56	2.20	C+

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17894	1 (A)	5	0.03	0.09		0.11	0.00	
17895	1 (A)	6	0.11	0.00		0.13	0.00	
17915	1 (A)	11	0.10	0.00		0.09	0.01	
17914	1 (A)	12	0.17	0.00	B-	0.19	0.00	B-
17908	1 (B)	5	0.07	0.05		0.16	0.00	
17909	1 (B)	6	0.00	0.33		0.04	0.13	
12405	2 (C)	11	0.06	0.06		0.03	0.01	
12404	2 (C)	12	0.11	0.00		0.11	0.00	
17872	2 (D)	5	0.19	0.00	B-	0.20	0.00	B-
17871	2 (D)	6	0.19	0.00	B-	0.19	0.00	B-

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17635	1 (A)	1	0.10	-0.23	0.43		-0.18	-0.51	0.15	
17638	1 (A)	2	0.29	-0.11	0.68		0.35	-0.04	0.75	
17636	1 (A)	3	0.01	-0.35	0.37		0.03	-0.33	0.39	
17637	1 (A)	4	0.16	-0.18	0.50		0.34	0.00	0.69	
17654	1 (A)	7	-0.18	-0.62	0.25		-0.06	-0.49	0.36	
17656	1 (A)	8	0.06	-0.25	0.37		-0.02	-0.33	0.28	
17653	1 (A)	9	-0.24	-0.56	0.08		-0.10	-0.43	0.22	
17655	1 (A)	10	0.15	-0.19	0.48		0.23	-0.11	0.57	
17618	1 (B)	1	0.18	-0.32	0.68		0.25	-0.25	0.75	
17620	1 (B)	2	0.45	0.13	0.78		0.35	0.03	0.67	
17619	1 (B)	3	0.15	-0.33	0.62		0.31	-0.16	0.78	
17617	1 (B)	4	0.38	0.06	0.71		0.48	0.15	0.80	
17582	2 (C)	1	-0.55	-1.08	-0.01		-0.74	-1.27	-0.21	
17581	2 (C)	2	-0.64	-1.01	-0.27		-0.68	-1.05	-0.31	
17583	2 (C)	3	-0.01	-0.35	0.34		-0.17	-0.52	0.17	
17590	2 (C)	4	0.19	-0.17	0.55		0.17	-0.20	0.54	
17584	2 (C)	5	0.34	0.02	0.66		0.26	-0.06	0.58	
17585	2 (C)	6	-0.68	-1.06	-0.29		-0.83	-1.22	-0.45	
17589	2 (C)	7	-0.35	-0.71	0.01		-0.39	-0.75	-0.02	
17587	2 (C)	8	0.32	-0.04	0.68		0.31	-0.05	0.67	
17586	2 (C)	9	0.16	-0.17	0.49		0.17	-0.16	0.50	
17588	2 (C)	10	-0.02	-0.44	0.39		0.02	-0.40	0.43	
17672	2 (D)	1	0.50	0.19	0.80		0.55	0.24	0.86	
17673	2 (D)	2	0.18	-0.28	0.65		-0.01	-0.48	0.46	
17671	2 (D)	3	0.36	0.01	0.72		0.44	0.09	0.80	
17674	2 (D)	4	0.24	-0.11	0.59		0.34	-0.01	0.69	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17639	1 (A)	5	0.05	0.03		0.05	0.23	
17640	1 (A)	6	0.10	0.00		0.15	0.00	
17657	1 (A)	11	-0.02	0.49		-0.02	0.61	
17658	1 (A)	12	0.01	0.80		0.00	0.71	
17621	1 (B)	5	0.13	0.00		0.11	0.00	
17622	1 (B)	6	0.07	0.02		0.03	0.08	
17591	2 (C)	11	0.07	0.03		0.11	0.00	
17592	2 (C)	12	-0.01	0.69		0.03	0.25	
17675	2 (D)	5	0.07	0.01		0.08	0.00	
17676	2 (D)	6	0.10	0.01		0.08	0.02	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
18225	1 (A)	1	0.08	-0.25	0.42		0.00	-0.34	0.33	
18226	1 (A)	2	0.24	-0.12	0.60		0.27	-0.09	0.64	
18228	1 (A)	3	0.37	0.04	0.71		0.41	0.09	0.74	
18227	1 (A)	4	0.16	-0.17	0.50		0.30	-0.03	0.64	
18246	1 (A)	7	0.33	0.02	0.63		0.24	-0.07	0.55	
18243	1 (A)	8	-0.05	-0.56	0.45		-0.08	-0.59	0.43	
18244	1 (A)	9	0.60	0.25	0.94		0.26	-0.08	0.60	
18245	1 (A)	10	0.06	-0.26	0.37		0.01	-0.30	0.32	
18208	1 (B)	1	0.43	0.08	0.79		0.41	0.05	0.77	
18209	1 (B)	2	0.26	-0.06	0.57		0.14	-0.17	0.46	
18207	1 (B)	3	0.32	-0.13	0.78		0.13	-0.33	0.60	
18210	1 (B)	4	-0.16	-0.48	0.17		-0.38	-0.70	-0.06	
18198	2 (C)	1	-0.28	-0.67	0.12		-0.30	-0.70	0.11	
18196	2 (C)	2	0.26	-0.34	0.86		-0.32	-0.93	0.29	
18201	2 (C)	3	-0.31	-0.89	0.26		-0.40	-0.99	0.19	
18202	2 (C)	4	-0.14	-0.53	0.25		-0.29	-0.68	0.10	
18204	2 (C)	5	-0.07	-0.39	0.26		0.05	-0.27	0.38	
18203	2 (C)	6	0.26	-0.05	0.57		0.43	0.12	0.74	
18195	2 (C)	7	0.13	-0.24	0.51		-0.16	-0.54	0.21	
18197	2 (C)	8	-0.39	-0.71	-0.07		-0.32	-0.64	0.00	
18200	2 (C)	9	-0.10	-0.51	0.31		-0.03	-0.44	0.38	
18199	2 (C)	10	0.35	-0.05	0.74		0.05	-0.34	0.45	
18215	2 (D)	1	0.31	-0.10	0.73		0.30	-0.12	0.72	
18216	2 (D)	2	-0.09	-0.44	0.27		-0.04	-0.39	0.32	
18214	2 (D)	3	0.33	0.01	0.66		0.35	0.03	0.67	
18213	2 (D)	4	-0.25	-0.64	0.13		-0.15	-0.55	0.24	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
18229	1 (A)	5	0.04	0.24		0.07	0.07	
18230	1 (A)	6	0.03	0.33		0.05	0.04	
18248	1 (A)	11	-0.11	0.00		-0.06	0.32	
18247	1 (A)	12	0.04	0.08		0.06	0.02	
18211	1 (B)	5	0.11	0.00		0.09	0.02	
18212	1 (B)	6	0.08	0.00		0.11	0.00	
18205	2 (C)	11	0.03	0.12		0.05	0.22	
18206	2 (C)	12	0.00	0.00		0.00	0.00	
18217	2 (D)	5	0.11	0.00		0.08	0.00	
18218	2 (D)	6	0.13	0.00		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 Multiple-Choice Items: Primary Reading (English)

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17297	1 (A)	1	0.68	0.34	1.02		0.59	0.25	0.94	
17296	1 (A)	2	0.13	-0.19	0.44		0.19	-0.13	0.50	
17298	1 (A)	3	0.44	0.11	0.77		0.35	0.01	0.69	
17295	1 (A)	4	1.31	1.00	1.62	B+	0.93	0.62	1.24	
17244	1 (A)	7	1.69	1.35	2.03	C+	1.63	1.30	1.97	C+
17243	1 (A)	8	-0.11	-0.52	0.31		0.17	-0.25	0.58	
17245	1 (A)	9	-0.53	-0.88	-0.18		-0.21	-0.56	0.13	
17242	1 (A)	10	0.10	-0.22	0.43		0.10	-0.23	0.42	
17346	1 (B)	1	-0.67	-1.00	-0.33		-0.73	-1.07	-0.39	
17348	1 (B)	2	0.87	0.47	1.27		0.85	0.45	1.25	
17347	1 (B)	3	0.41	0.06	0.76		0.41	0.06	0.76	
17349	1 (B)	4	0.13	-0.18	0.45		-0.18	-0.50	0.14	
10910	2 (C)	1	-0.27	-0.68	0.15		0.03	-0.39	0.45	
10913	2 (C)	2	0.67	0.35	0.99		0.63	0.31	0.96	
10894	2 (C)	3	0.42	0.06	0.77		0.18	-0.17	0.53	
10901	2 (C)	4	0.04	-0.41	0.50		-0.29	-0.74	0.17	
10902	2 (C)	5	-0.07	-0.43	0.29		0.09	-0.26	0.45	
10900	2 (C)	6	0.14	-0.26	0.54		0.47	0.07	0.87	
10897	2 (C)	7	0.52	0.16	0.89		0.74	0.37	1.11	
10904	2 (C)	8	0.34	0.02	0.65		0.13	-0.19	0.44	
10909	2 (C)	9	0.60	0.29	0.92		0.24	-0.08	0.55	
10908	2 (C)	10	-0.29	-0.64	0.05		-0.33	-0.68	0.02	
17354	2 (D)	1	0.60	0.21	0.99		0.00	-0.40	0.40	
17353	2 (D)	2	0.63	0.30	0.97		0.47	0.14	0.81	
17352	2 (D)	3	0.60	0.28	0.92		0.07	-0.26	0.40	
17355	2 (D)	4	0.95	0.64	1.25		0.74	0.43	1.04	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17293	1 (A)	5	0.10	0.00		0.15	0.00	
17294	1 (A)	6	0.08	0.03		0.06	0.19	
17241	1 (A)	11	0.07	0.11		0.07	0.10	
17240	1 (A)	12	0.09	0.00		0.10	0.00	
17344	1 (B)	5	-0.01	0.96		0.01	0.10	
17343	1 (B)	6	0.08	0.01		0.07	0.01	
10912	2 (C)	11	0.10	0.00		0.10	0.01	
10899	2 (C)	12	0.01	0.87		0.01	0.05	
17351	2 (D)	5	0.12	0.00		0.07	0.11	
17350	2 (D)	6	0.09	0.00		0.06	0.04	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17898	1 (A)	1	0.14	-0.19	0.47		0.41	0.09	0.74	
17896	1 (A)	2	1.05	0.66	1.44	B+	0.55	0.16	0.93	
17897	1 (A)	3	0.36	0.03	0.69		0.32	-0.01	0.65	
17899	1 (A)	4	0.38	-0.02	0.79		0.15	-0.26	0.56	
17916	1 (A)	7	0.70	0.36	1.03		0.95	0.62	1.27	
17917	1 (A)	8	1.42	1.08	1.76	B+	1.58	1.24	1.91	C+
17920	1 (A)	9	0.15	-0.18	0.47		0.16	-0.17	0.49	
17919	1 (A)	10	0.44	0.08	0.80		0.45	0.09	0.81	
17910	1 (B)	1	0.70	0.38	1.02		0.89	0.57	1.22	
17913	1 (B)	2	-0.19	-0.51	0.13		-0.09	-0.41	0.24	
17912	1 (B)	3	-0.26	-0.63	0.11		0.15	-0.21	0.51	
17911	1 (B)	4	1.00	0.67	1.33	B+	1.18	0.85	1.51	B+
12394	2 (C)	1	-0.28	-0.63	0.07		-0.30	-0.65	0.05	
12403	2 (C)	2	0.53	0.19	0.87		0.75	0.40	1.09	
12395	2 (C)	3	1.04	0.65	1.42	B+	0.72	0.34	1.11	
12396	2 (C)	4	0.25	-0.07	0.58		0.51	0.18	0.84	
12397	2 (C)	5	0.45	0.08	0.83		0.39	0.02	0.76	
12398	2 (C)	6	0.04	-0.32	0.40		-0.13	-0.50	0.24	
12400	2 (C)	7	0.96	0.54	1.38		1.03	0.61	1.46	B+
12399	2 (C)	8	0.87	0.49	1.26		0.63	0.25	1.01	
12401	2 (C)	9	0.09	-0.29	0.48		-0.14	-0.52	0.24	
12402	2 (C)	10	0.07	-0.32	0.46		-0.12	-0.51	0.27	
17876	2 (D)	1	0.42	0.08	0.76		0.59	0.25	0.93	
17873	2 (D)	2	0.18	-0.20	0.56		-0.23	-0.61	0.15	
17875	2 (D)	3	0.55	0.22	0.88		0.11	-0.22	0.45	
17878	2 (D)	4	1.09	0.77	1.41	B+	0.92	0.60	1.25	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17894	1 (A)	5	0.06	0.03		0.07	0.00	
17895	1 (A)	6	0.13	0.00		0.12	0.00	
17915	1 (A)	11	0.00	0.46		0.02	0.01	
17914	1 (A)	12	0.08	0.00		0.08	0.00	
17908	1 (B)	5	0.06	0.11		0.03	0.33	
17909	1 (B)	6	0.05	0.12		0.01	0.70	
12405	2 (C)	11	0.15	0.00		0.13	0.00	
12404	2 (C)	12	0.10	0.00		0.12	0.00	
17872	2 (D)	5	0.09	0.00		0.11	0.00	
17871	2 (D)	6	0.10	0.00		0.04	0.00	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17635	1 (A)	1	0.25	-0.14	0.64		0.11	-0.27	0.50	
17638	1 (A)	2	0.57	0.13	1.02		0.52	0.07	0.97	
17636	1 (A)	3	-0.20	-0.62	0.22		-0.06	-0.48	0.36	
17637	1 (A)	4	0.64	0.25	1.03		0.55	0.16	0.94	
17654	1 (A)	7	-0.41	-0.90	0.08		-0.18	-0.68	0.32	
17656	1 (A)	8	-0.13	-0.49	0.23		-0.08	-0.45	0.28	
17653	1 (A)	9	-0.53	-0.91	-0.14		-0.46	-0.84	-0.08	
17655	1 (A)	10	0.18	-0.21	0.56		-0.03	-0.41	0.36	
17618	1 (B)	1	0.74	0.16	1.32		0.23	-0.32	0.78	
17620	1 (B)	2	0.09	-0.29	0.46		-0.30	-0.68	0.08	
17619	1 (B)	3	0.74	0.19	1.28		0.37	-0.16	0.90	
17617	1 (B)	4	-0.74	-1.13	-0.35		-0.06	-0.44	0.32	
17582	2 (C)	1	0.42	-0.18	1.02		-0.11	-0.70	0.48	
17581	2 (C)	2	-0.40	-0.83	0.03		-0.53	-0.96	-0.10	
17583	2 (C)	3	-0.01	-0.41	0.39		-0.51	-0.91	-0.11	
17590	2 (C)	4	0.39	-0.02	0.81		0.15	-0.27	0.57	
17584	2 (C)	5	0.21	-0.17	0.58		0.36	-0.02	0.73	
17585	2 (C)	6	0.12	-0.31	0.56		0.23	-0.21	0.67	
17589	2 (C)	7	0.55	0.14	0.97		0.32	-0.10	0.73	
17587	2 (C)	8	-0.48	-0.89	-0.06		-0.38	-0.80	0.03	
17586	2 (C)	9	0.33	-0.06	0.71		0.20	-0.18	0.59	
17588	2 (C)	10	-0.52	-1.00	-0.04		-0.46	-0.94	0.02	
17672	2 (D)	1	-0.17	-0.54	0.19		-0.01	-0.38	0.36	
17673	2 (D)	2	0.32	-0.21	0.85		-0.05	-0.57	0.47	
17671	2 (D)	3	-0.54	-0.95	-0.13		-0.67	-1.09	-0.25	
17674	2 (D)	4	0.38	-0.02	0.78		0.31	-0.09	0.70	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17639	1 (A)	5	-0.01	0.25		-0.01	0.95	
17640	1 (A)	6	0.01	0.42		0.02	0.52	
17657	1 (A)	11	0.00	0.70		-0.01	0.76	
17658	1 (A)	12	-0.02	0.81		-0.03	0.76	
17621	1 (B)	5	-0.03	0.37		-0.01	0.61	
17622	1 (B)	6	-0.09	0.03		-0.08	0.15	
17591	2 (C)	11	-0.07	0.32		0.00	0.79	
17592	2 (C)	12	-0.07	0.32		-0.04	0.66	
17675	2 (D)	5	-0.04	0.08		0.05	0.26	
17676	2 (D)	6	-0.10	0.13		-0.02	0.87	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
18225	1 (A)	1	-0.46	-1.01	0.08		-0.53	-1.07	0.01	
18226	1 (A)	2	-0.11	-0.70	0.47		-0.13	-0.70	0.44	
18228	1 (A)	3	0.27	-0.26	0.81		0.08	-0.45	0.61	
18227	1 (A)	4	1.18	0.66	1.70	B+	0.91	0.39	1.43	
18246	1 (A)	7	-0.25	-0.74	0.25		-0.24	-0.73	0.25	
18243	1 (A)	8	0.25	-0.53	1.03		0.32	-0.45	1.08	
18244	1 (A)	9	-0.51	-1.04	0.03		0.04	-0.50	0.57	
18245	1 (A)	10	-0.26	-0.76	0.24		-0.31	-0.81	0.19	
18208	1 (B)	1	-0.37	-0.93	0.20		-0.41	-0.97	0.15	
18209	1 (B)	2	-0.18	-0.70	0.34		-0.15	-0.66	0.37	
18207	1 (B)	3	0.67	-0.04	1.37		0.37	-0.32	1.05	
18210	1 (B)	4	0.84	0.33	1.35		0.52	0.01	1.03	
18198	2 (C)	1	-0.46	-1.10	0.18		-0.23	-0.88	0.41	
18196	2 (C)	2	-0.19	-1.13	0.75		-0.48	-1.38	0.43	
18201	2 (C)	3	-0.18	-1.05	0.68		-0.07	-0.94	0.80	
18202	2 (C)	4	-0.59	-1.20	0.02		-0.11	-0.72	0.51	
18204	2 (C)	5	-0.58	-1.11	-0.05		-0.17	-0.69	0.35	
18203	2 (C)	6	0.44	-0.06	0.93		0.69	0.19	1.18	
18195	2 (C)	7	0.14	-0.46	0.74		-0.06	-0.66	0.53	
18197	2 (C)	8	-0.67	-1.19	-0.15		-0.37	-0.88	0.15	
18200	2 (C)	9	-0.12	-0.76	0.52		0.40	-0.25	1.05	
18199	2 (C)	10	-0.32	-0.95	0.31		0.07	-0.56	0.69	
18215	2 (D)	1	-0.45	-1.10	0.20		-0.16	-0.80	0.49	
18216	2 (D)	2	-0.22	-0.77	0.34		-0.20	-0.75	0.36	
18214	2 (D)	3	0.16	-0.36	0.68		-0.02	-0.53	0.50	
18213	2 (D)	4	-0.17	-0.78	0.43		0.13	-0.48	0.73	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
18229	1 (A)	5	-0.11	0.02		-0.11	0.07	
18230	1 (A)	6	-0.13	0.04		-0.08	0.29	
18248	1 (A)	11	-0.05	0.32		0.00	0.84	
18247	1 (A)	12	0.04	0.33		-0.01	0.37	
18211	1 (B)	5	-0.04	0.76		-0.05	0.76	
18212	1 (B)	6	-0.05	0.70		-0.07	0.30	
18205	2 (C)	11	0.00	0.82		-0.07	0.29	
18206	2 (C)	12	0.07	0.61		0.11	0.04	
18217	2 (D)	5	0.00	0.27		-0.03	0.52	
18218	2 (D)	6	-0.06	0.08		-0.01	0.29	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17383	2(C)	14	-0.04	0.32	-0.40		0.16	0.52	-0.19	
17358	2(C)	15	0.00	0.37	-0.37		0.00	0.38	-0.38	
17356	2(C)	16	-0.20	0.13	-0.54		-0.20	0.14	-0.53	
17363	2(C)	17	-0.16	0.22	-0.54		-0.11	0.28	-0.49	
17367	2(D)	8	-0.25	0.13	-0.64		-0.19	0.20	-0.58	
17392	2(D)	9	-0.49	-0.14	-0.84		-0.57	-0.20	-0.93	
17388	2(D)	10	-0.32	0.01	-0.66		-0.03	0.30	-0.36	
17988	2(D)	11	-0.78	-0.42	-1.13		-0.33	0.02	-0.68	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17386_T	1(A)	13	-0.01	0.51		0.03	0.31	
17386_V	1(A)	13	0.04	0.37		0.01	0.39	
18179_T	2(C)	13	-0.05	0.21		-0.06	0.07	
18179_V	2(C)	13	-0.06	0.02		-0.10	0.00	
16475_T	2(D)	7	-0.06	0.07		0.05	0.07	
16475_V	2(D)	7	-0.03	0.07		0.00	0.07	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17926	2(C)	14	-0.34	0.13	-0.80		-0.76	-0.29	-1.23	
17728	2(C)	15	-0.58	-0.19	-0.96		-0.11	0.28	-0.50	
17929	2(C)	16	-0.64	-0.27	-1.01		-0.58	-0.21	-0.96	
17925	2(C)	17	0.25	0.62	-0.11		0.06	0.42	-0.30	
17954	2(D)	8	-0.03	0.30	-0.37		-0.15	0.18	-0.48	
17944	2(D)	9	-0.38	0.04	-0.79		-0.51	-0.09	-0.92	
17957	2(D)	10	-0.14	0.20	-0.48		-0.09	0.26	-0.44	
17729	2(D)	11	-0.88	-0.54	-1.22		-0.79	-0.46	-1.13	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17958_T	1(A)	13	-0.03	0.45		-0.04	0.26	
17958_V	1(A)	13	-0.13	0.00		-0.14	0.00	
17950_T	2(C)	13	0.02	0.35		0.04	0.29	
17950_V	2(C)	13	-0.09	0.00		-0.10	0.00	
18996_T	2(D)	7	0.05	0.09		0.00	0.98	
18996_V	2(D)	7	-0.09	0.00		-0.08	0.00	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17702	2(C)	14	-0.49	0.10	-1.07		-0.94	-0.38	-1.50	
17786	2(C)	15	-0.67	-0.19	-1.14		-0.32	0.16	-0.80	
17691	2(C)	16	-0.05	0.42	-0.51		-0.58	-0.12	-1.04	
17699	2(C)	17	-0.41	0.09	-0.92		-0.37	0.15	-0.89	
17690	2(D)	8	-0.81	-0.21	-1.42		-0.32	0.27	-0.90	
17689	2(D)	9	-0.29	0.20	-0.79		0.11	0.59	-0.37	
17706	2(D)	10	-0.57	-0.01	-1.12		0.17	0.70	-0.37	
17695	2(D)	11	-0.86	-0.14	-1.57		-0.30	0.32	-0.92	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17708_T	1(A)	13	0.00	0.25		-0.05	0.30	
17708_V	1(A)	13	-0.02	0.37		0.00	0.13	
17692_T	2(C)	13	-0.11	0.02		0.02	0.05	
17692_V	2(C)	13	0.00	0.58		0.01	0.52	
15509_T	2(D)	7	-0.11	0.01		-0.15	0.00	
15509_V	2(D)	7	-0.07	0.07		-0.08	0.01	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
18294	2(C)	14	-0.20	0.35	-0.76		0.03	0.58	-0.51	
18295	2(C)	15	-0.30	0.20	-0.81		-0.14	0.35	-0.63	
18293	2(C)	16	-0.12	0.42	-0.66		-0.66	-0.12	-1.21	
18300	2(C)	17	-0.28	0.21	-0.77		-0.22	0.27	-0.71	
18311	2(D)	8	-1.01	-0.53	-1.50	B+	-0.86	-0.37	-1.35	
18296	2(D)	9	-0.06	0.45	-0.57		0.15	0.63	-0.34	
18304	2(D)	10	-1.27	-0.76	-1.77	B+	-1.01	-0.53	-1.50	B+
18297	2(D)	11	0.21	0.71	-0.30		0.40	0.90	-0.09	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
18303_T	1(A)	13	0.00	0.30		-0.07	0.11	
18303_V	1(A)	13	-0.02	0.09		0.00	0.36	
18302_T	2(C)	13	-0.05	0.01		-0.08	0.06	
18302_V	2(C)	13	-0.02	0.04		-0.02	0.41	
16819_T	2(D)	7	-0.07	0.03		-0.06	0.41	
16819_V	2(D)	7	-0.10	0.00		0.00	0.81	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17383	2(C)	14	0.34	0.69	0.00		0.19	0.54	-0.16	
17358	2(C)	15	-0.06	0.31	-0.43		-0.12	0.25	-0.48	
17356	2(C)	16	0.27	0.60	-0.05		0.39	0.72	0.07	
17363	2(C)	17	0.34	0.71	-0.02		0.43	0.79	0.06	
17367	2(D)	8	0.43	0.80	0.06		0.04	0.43	-0.34	
17392	2(D)	9	0.22	0.56	-0.13		0.14	0.49	-0.22	
17388	2(D)	10	0.15	0.47	-0.17		-0.29	0.04	-0.62	
17988	2(D)	11	0.73	1.07	0.39		0.54	0.88	0.19	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17386_T	1(A)	13	-0.02	0.43		-0.02	0.37	
17386_V	1(A)	13	-0.02	0.83		-0.05	0.03	
18179_T	2(C)	13	0.01	0.20		0.08	0.00	
18179_V	2(C)	13	-0.08	0.01		-0.04	0.02	
16475_T	2(D)	7	-0.02	0.86		0.02	0.03	
16475_V	2(D)	7	-0.07	0.01		-0.04	0.10	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17926	2(C)	14	0.56	0.99	0.14		1.16	1.58	0.73	B-
17728	2(C)	15	0.16	0.53	-0.20		0.62	0.99	0.25	
17929	2(C)	16	1.00	1.36	0.65	B-	1.32	1.68	0.97	B-
17925	2(C)	17	0.26	0.61	-0.08		0.44	0.78	0.09	
17954	2(D)	8	0.21	0.55	-0.13		0.10	0.43	-0.23	
17944	2(D)	9	0.41	0.80	0.02		0.79	1.18	0.40	
17957	2(D)	10	-0.25	0.08	-0.58		-0.40	-0.05	-0.74	
17729	2(D)	11	0.05	0.39	-0.28		0.28	0.62	-0.05	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17958_T	1(A)	13	-0.02	0.53		-0.01	0.00	
17958_V	1(A)	13	-0.03	0.09		0.00	0.00	
17950_T	2(C)	13	-0.05	0.19		-0.07	0.00	
17950_V	2(C)	13	-0.07	0.01		-0.06	0.00	
18996_T	2(D)	7	0.01	0.32		-0.02	0.44	
18996_V	2(D)	7	-0.06	0.03		-0.08	0.00	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
17702	2(C)	14	0.51	1.02	0.00		0.44	0.93	-0.06	
17786	2(C)	15	0.31	0.74	-0.12		0.24	0.66	-0.18	
17691	2(C)	16	1.16	1.59	0.74	B-	1.19	1.61	0.77	B-
17699	2(C)	17	-0.58	-0.12	-1.05		-0.49	-0.03	-0.95	
17690	2(D)	8	0.26	0.81	-0.28		-0.12	0.40	-0.64	
17689	2(D)	9	0.37	0.82	-0.07		0.13	0.56	-0.29	
17706	2(D)	10	0.09	0.58	-0.41		-0.42	0.05	-0.90	
17695	2(D)	11	1.06	1.68	0.45	B-	0.01	0.55	-0.52	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17708_T	1(A)	13	-0.06	0.20		-0.08	0.09	
17708_V	1(A)	13	-0.05	0.22		0.00	0.61	
17692_T	2(C)	13	0.01	0.48		-0.02	0.90	
17692_V	2(C)	13	-0.02	0.73		0.01	0.98	
15509_T	2(D)	7	-0.07	0.13		-0.04	0.17	
15509_V	2(D)	7	-0.10	0.01		0.01	0.33	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
18294	2(C)	14	-0.56	0.01	-1.14		-0.51	0.05	-1.07	
18295	2(C)	15	1.15	1.65	0.64	B-	1.58	2.09	1.07	C-
18293	2(C)	16	0.18	0.74	-0.38		0.38	0.94	-0.18	
18300	2(C)	17	-0.43	0.08	-0.95		-0.30	0.22	-0.81	
18311	2(D)	8	-0.15	0.35	-0.65		0.45	0.95	-0.04	
18296	2(D)	9	0.50	1.02	-0.01		0.14	0.64	-0.37	
18304	2(D)	10	-0.29	0.21	-0.79		-0.63	-0.13	-1.13	
18297	2(D)	11	-0.72	-0.17	-1.26		-0.49	0.03	-1.02	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
18303_T	1(A)	13	0.02	0.98		0.05	0.41	
18303_V	1(A)	13	0.04	0.29		0.04	0.46	
18302_T	2(C)	13	0.00	0.91		-0.02	0.16	
18302_V	2(C)	13	0.02	0.37		-0.01	0.99	
16819_T	2(D)	7	0.01	0.25		-0.05	0.31	
16819_V	2(D)	7	0.02	0.53		-0.05	0.28	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
18359	3(1)	1	0.46	0.90	0.02		0.73	1.17	0.29	
10967	3(1)	2	-0.01	0.38	-0.40		-0.14	0.24	-0.53	
10815	3(1)	3	0.18	0.56	-0.19		0.10	0.48	-0.27	
10783	3(1)	4	0.41	0.77	0.06		0.48	0.84	0.12	
18348	3(1)	5	-0.51	-0.18	-0.85		-0.46	-0.13	-0.80	
15140	3(1)	6	-0.04	0.33	-0.42		0.35	0.71	-0.02	
10687	3(1)	11	-0.04	0.35	-0.43		0.37	0.76	-0.01	
15097	3(1)	12	0.04	0.40	-0.32		0.06	0.42	-0.31	
18360	3(1)	13	-0.37	-0.01	-0.72		-0.12	0.24	-0.48	
15146	3(1)	14	-0.26	0.08	-0.60		-0.35	-0.01	-0.69	
10973	3(1)	15	1.04	1.38	0.70	B+	1.18	1.52	0.83	B+
15106	3(1)	16	-0.16	0.19	-0.52		0.01	0.37	-0.35	
15142	3(1)	17	0.59	1.02	0.16		0.85	1.28	0.41	
15111	3(1)	18	1.11	1.60	0.61	B+	0.86	1.34	0.39	
10844	3(2)	1	-0.60	-0.11	-1.09		-0.68	-0.18	-1.19	
15085	3(2)	2	-0.08	0.35	-0.51		-0.51	-0.07	-0.94	
15098	3(2)	3	0.64	1.02	0.27		0.61	0.97	0.24	
12011	3(2)	4	-0.03	0.28	-0.34		0.10	0.42	-0.21	
18350	3(2)	5	0.33	0.73	-0.07		0.26	0.66	-0.14	
10974	3(2)	6	0.21	0.55	-0.14		0.18	0.53	-0.16	
10980	3(2)	7	-0.95	-0.51	-1.40		-0.13	0.30	-0.56	
10696	3(2)	12	-0.54	-0.09	-1.00		-0.22	0.23	-0.67	
18355	3(2)	13	-0.07	0.25	-0.39		0.03	0.35	-0.29	
10785	3(2)	14	0.25	0.57	-0.08		0.32	0.64	-0.01	
15124	3(2)	15	-0.24	0.12	-0.60		-0.29	0.07	-0.64	
15080	3(2)	16	0.71	1.05	0.37		0.68	1.02	0.34	
12556	3(2)	17	-0.01	0.33	-0.35		-0.04	0.30	-0.38	
15103	3(2)	18	0.53	0.89	0.18		0.69	1.04	0.33	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
15093	3(1)	7	0.05	0.04		0.07	0.00	
10847	3(1)	8	0.08	0.00		0.11	0.00	
15173	3(1)	9	0.05	0.30		0.10	0.00	
15132	3(1)	10	0.09	0.01		0.08	0.01	
10846	3(2)	8	0.06	0.03		0.02	0.69	
10744	3(2)	9	-0.08	0.00		-0.04	0.00	
15116	3(2)	10	0.02	0.08		0.06	0.11	
15095	3(2)	11	-0.03	0.00		-0.02	0.01	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
11174	3(1)	1	1.12	1.48	0.75	B+	0.39	0.76	0.03	
15058	3(1)	2	0.24	0.57	-0.10		0.13	0.46	-0.20	
18369	3(1)	3	-0.49	-0.12	-0.85		-0.26	0.10	-0.62	
11413	3(1)	4	0.32	0.64	0.00		0.30	0.63	-0.02	
11305	3(1)	5	0.36	0.70	0.02		0.21	0.55	-0.13	
15018	3(1)	6	0.68	1.08	0.29		0.94	1.33	0.56	
11319	3(1)	11	-0.34	-0.01	-0.68		-0.30	0.04	-0.63	
11318	3(1)	12	0.06	0.37	-0.25		0.18	0.50	-0.14	
18371	3(1)	13	0.13	0.48	-0.22		0.55	0.90	0.20	
15006	3(1)	14	-0.35	0.06	-0.76		-0.19	0.21	-0.60	
15026	3(1)	15	0.03	0.36	-0.31		0.15	0.48	-0.18	
15000	3(1)	16	-0.16	0.19	-0.51		-0.42	-0.08	-0.76	
11315	3(1)	17	-0.60	-0.25	-0.95		0.02	0.36	-0.33	
13042	3(1)	18	0.11	0.46	-0.24		0.23	0.59	-0.12	
15048	3(2)	1	-0.50	-0.05	-0.96		-0.27	0.18	-0.72	
11321	3(2)	2	-0.27	0.21	-0.75		-0.25	0.21	-0.71	
14984	3(2)	3	1.01	1.35	0.66	B+	0.83	1.18	0.49	
11326	3(2)	4	0.90	1.22	0.57		0.99	1.31	0.67	
15055	3(2)	5	-0.71	-0.33	-1.09		-0.33	0.05	-0.70	
18375	3(2)	6	-0.15	0.18	-0.48		0.00	0.33	-0.34	
15002	3(2)	7	0.17	0.49	-0.15		0.39	0.71	0.07	
11292	3(2)	12	1.03	1.53	0.54	B+	1.50	1.99	1.00	B+
11176	3(2)	13	-0.25	0.08	-0.58		0.12	0.45	-0.21	
14999	3(2)	14	1.11	1.47	0.75	B+	1.69	2.06	1.32	C+
15061	3(2)	15	-0.43	-0.10	-0.77		-0.40	-0.06	-0.73	
15037	3(2)	16	0.84	1.19	0.49		0.59	0.94	0.25	
14997	3(2)	17	1.30	1.66	0.95	B+	1.06	1.40	0.71	B+
15045	3(2)	18	0.25	0.61	-0.11		-0.06	0.29	-0.41	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17405	3(1)	7	-0.04	0.01		0.01	0.00	
15068	3(1)	8	0.11	0.00		0.17	0.00	
11409	3(1)	9	0.04	0.28		0.04	0.25	
15010	3(1)	10	0.08	0.00		0.07	0.00	
18368	3(2)	8	-0.01	0.04		0.03	0.00	
11310	3(2)	9	0.05	0.11		0.05	0.10	
11329	3(2)	10	0.09	0.00		0.05	0.20	
15034	3(2)	11	0.10	0.00		0.18	0.00	B-

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
11214	3(1)	1	0.85	1.33	0.37		0.61	1.09	0.13	
14596	3(1)	2	-1.03	-0.64	-1.43	B-	-1.22	-0.83	-1.62	B-
14595	3(1)	3	-0.11	0.25	-0.46		0.13	0.48	-0.22	
17779	3(1)	4	0.08	0.41	-0.26		0.01	0.34	-0.33	
14583	3(1)	5	0.40	0.74	0.06		0.25	0.59	-0.09	
14572	3(1)	6	1.02	1.67	0.38	B+	1.09	1.73	0.45	B+
17785	3(1)	11	0.11	0.46	-0.24		-0.01	0.34	-0.36	
14658	3(1)	12	-0.04	0.29	-0.38		0.01	0.34	-0.33	
14649	3(1)	13	0.35	0.75	-0.04		0.30	0.70	-0.10	
14608	3(1)	14	-0.09	0.24	-0.42		-0.32	0.01	-0.65	
17781	3(1)	15	-0.11	0.38	-0.60		-0.03	0.46	-0.52	
10998	3(1)	16	-0.49	-0.18	-0.80		-0.39	-0.08	-0.71	
17767	3(1)	17	0.26	0.60	-0.09		0.09	0.43	-0.26	
14630	3(1)	18	0.04	0.40	-0.33		0.18	0.54	-0.19	
17721	3(2)	1	0.19	0.75	-0.36		0.11	0.67	-0.45	
14628	3(2)	2	-0.86	-0.50	-1.22		-0.85	-0.49	-1.20	
14615	3(2)	3	0.36	0.73	-0.02		0.15	0.53	-0.22	
17768	3(2)	4	-0.02	0.31	-0.36		-0.10	0.23	-0.43	
11240	3(2)	5	-0.91	-0.42	-1.41		-0.80	-0.31	-1.28	
10997	3(2)	6	-0.61	-0.24	-0.98		-0.59	-0.22	-0.97	
14661	3(2)	7	1.31	1.66	0.95	B+	1.01	1.36	0.65	B+
11098	3(2)	12	0.09	0.47	-0.29		0.11	0.49	-0.26	
11223	3(2)	13	0.61	0.98	0.25		0.60	0.96	0.24	
14646	3(2)	14	1.23	1.59	0.88	B+	1.29	1.64	0.93	B+
17771	3(2)	15	1.14	1.47	0.80	B+	1.19	1.53	0.86	B+
12616	3(2)	16	0.29	0.63	-0.06		0.03	0.37	-0.31	
14589	3(2)	17	0.37	0.70	0.05		0.43	0.76	0.11	
17784	3(2)	18	0.22	0.61	-0.18		0.34	0.74	-0.05	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17769	3(1)	7	-0.05	0.05		-0.04	0.17	
14631	3(1)	8	0.02	0.19		0.00	0.54	
14667	3(1)	9	0.12	0.00		0.16	0.00	
14599	3(1)	10	-0.06	0.00		-0.07	0.00	
17773	3(2)	8	0.10	0.00		0.05	0.29	
11059	3(2)	9	-0.08	0.02		-0.10	0.00	
14668	3(2)	10	0.16	0.00		0.17	0.00	B-
17766	3(2)	11	0.07	0.01		0.09	0.00	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
18323	3(1)	1	0.08	0.43	-0.28		-0.09	0.27	-0.44	
11648	3(1)	2	1.15	1.51	0.79	B+	1.32	1.69	0.95	B+
14734	3(1)	3	0.08	0.43	-0.27		0.30	0.65	-0.05	
14704	3(1)	4	-0.01	0.32	-0.35		-0.16	0.17	-0.49	
11457	3(1)	5	-0.05	0.33	-0.44		0.01	0.39	-0.36	
14715	3(1)	6	-0.58	-0.22	-0.94		-0.43	-0.07	-0.79	
14750	3(1)	11	0.52	0.91	0.13		0.55	0.94	0.16	
18007	3(1)	12	-0.75	-0.31	-1.18		-1.25	-0.82	-1.69	B-
14758	3(1)	13	0.63	0.96	0.31		0.44	0.77	0.12	
14711	3(1)	14	-0.20	0.15	-0.55		0.00	0.35	-0.35	
11507	3(1)	15	0.64	0.98	0.30		0.56	0.89	0.22	
14703	3(1)	16	-0.01	0.33	-0.35		-0.21	0.13	-0.54	
18014	3(1)	17	0.24	0.58	-0.10		0.04	0.38	-0.30	
14731	3(1)	18	0.18	0.54	-0.18		0.37	0.73	0.02	
18015	3(2)	1	-0.28	0.07	-0.62		-0.15	0.20	-0.50	
14767	3(2)	2	0.11	0.45	-0.22		0.11	0.45	-0.23	
14728	3(2)	3	0.17	0.49	-0.15		0.32	0.65	0.00	
18318	3(2)	4	0.56	0.94	0.19		0.48	0.85	0.10	
11484	3(2)	5	-0.37	-0.04	-0.69		-0.60	-0.27	-0.93	
18321	3(2)	6	-0.36	-0.01	-0.72		-0.19	0.16	-0.54	
18315	3(2)	7	0.36	0.68	0.03		0.40	0.72	0.08	
11565	3(2)	12	0.20	0.53	-0.14		0.16	0.50	-0.17	
18319	3(2)	13	0.31	0.64	-0.01		0.32	0.65	0.00	
14763	3(2)	14	-0.16	0.21	-0.54		-0.62	-0.25	-1.00	
14696	3(2)	15	0.36	0.68	0.04		0.20	0.53	-0.13	
18017	3(2)	16	-0.27	0.22	-0.76		-0.45	0.05	-0.94	
14726	3(2)	17	0.75	1.08	0.43		0.52	0.84	0.19	
11520	3(2)	18	0.23	0.62	-0.17		0.11	0.52	-0.29	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
18317	3(1)	7	0.15	0.00		0.10	0.00	
18316	3(1)	8	0.12	0.00		0.07	0.00	
18324	3(1)	9	0.00	0.27		0.00	0.02	
14688	3(1)	10	0.04	0.20		0.06	0.02	
11533	3(2)	8	0.04	0.02		0.04	0.00	
18004	3(2)	9	0.03	0.20		0.03	0.03	
11508	3(2)	10	-0.03	0.02		-0.04	0.00	
14744	3(2)	11	-0.12	0.00		-0.12	0.00	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
18359	3(1)	1	-0.41	0.03	-0.85		-0.33	0.10	-0.77	
10967	3(1)	2	-0.38	0.00	-0.77		-0.12	0.27	-0.51	
10815	3(1)	3	-0.55	-0.18	-0.93		-0.43	-0.06	-0.80	
10783	3(1)	4	0.30	0.65	-0.05		0.19	0.55	-0.16	
18348	3(1)	5	-0.69	-0.34	-1.03		-0.55	-0.21	-0.89	
15140	3(1)	6	-0.39	-0.02	-0.77		-0.76	-0.39	-1.13	
10687	3(1)	11	0.00	0.38	-0.37		0.03	0.41	-0.35	
15097	3(1)	12	0.48	0.83	0.13		0.36	0.71	0.01	
18360	3(1)	13	0.34	0.70	-0.01		0.06	0.41	-0.28	
15146	3(1)	14	0.24	0.58	-0.09		0.03	0.36	-0.31	
10973	3(1)	15	0.16	0.50	-0.18		0.06	0.40	-0.27	
15106	3(1)	16	0.29	0.64	-0.06		0.36	0.71	0.02	
15142	3(1)	17	0.14	0.56	-0.28		0.07	0.48	-0.34	
15111	3(1)	18	0.14	0.61	-0.32		0.32	0.79	-0.16	
10844	3(2)	1	-0.33	0.17	-0.83		-0.92	-0.41	-1.42	
15085	3(2)	2	-0.42	0.02	-0.85		-0.41	0.03	-0.85	
15098	3(2)	3	-0.60	-0.22	-0.97		-0.50	-0.13	-0.87	
12011	3(2)	4	-0.01	0.30	-0.32		0.30	0.61	-0.01	
18350	3(2)	5	-0.34	0.05	-0.74		-0.37	0.03	-0.77	
10974	3(2)	6	-0.50	-0.16	-0.85		-0.40	-0.05	-0.75	
10980	3(2)	7	0.08	0.51	-0.34		0.24	0.67	-0.19	
10696	3(2)	12	-0.06	0.40	-0.52		-0.28	0.17	-0.73	
18355	3(2)	13	0.14	0.46	-0.17		0.01	0.32	-0.30	
10785	3(2)	14	0.34	0.67	0.02		0.31	0.63	-0.01	
15124	3(2)	15	0.12	0.48	-0.23		0.09	0.45	-0.26	
15080	3(2)	16	0.31	0.65	-0.03		0.02	0.36	-0.31	
12556	3(2)	17	-0.33	0.01	-0.67		-0.28	0.06	-0.62	
15103	3(2)	18	0.22	0.58	-0.13		0.04	0.40	-0.31	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
15093	3(1)	7	0.06	0.05		0.01	0.28	
10847	3(1)	8	0.06	0.09		0.09	0.00	
15173	3(1)	9	0.00	0.33		0.03	0.47	
15132	3(1)	10	-0.01	0.61		0.05	0.24	
10846	3(2)	8	-0.03	0.21		-0.02	0.26	
10744	3(2)	9	0.00	0.23		0.05	0.13	
15116	3(2)	10	-0.01	0.02		-0.02	0.22	
15095	3(2)	11	0.01	0.06		0.03	0.33	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
11174	3(1)	1	-0.58	-0.21	-0.94		-0.29	0.06	-0.65	
15058	3(1)	2	0.25	0.58	-0.09		-0.10	0.23	-0.44	
18369	3(1)	3	-0.19	0.17	-0.55		-0.01	0.35	-0.37	
11413	3(1)	4	0.22	0.54	-0.11		0.41	0.74	0.09	
11305	3(1)	5	0.44	0.78	0.11		0.72	1.05	0.39	
15018	3(1)	6	0.49	0.88	0.11		0.53	0.92	0.15	
11319	3(1)	11	-0.13	0.21	-0.47		0.08	0.42	-0.26	
11318	3(1)	12	-0.14	0.18	-0.46		-0.42	-0.10	-0.74	
18371	3(1)	13	-0.66	-0.30	-1.02		-0.11	0.25	-0.46	
15006	3(1)	14	-0.11	0.30	-0.52		-0.08	0.32	-0.48	
15026	3(1)	15	0.09	0.42	-0.25		0.20	0.54	-0.13	
15000	3(1)	16	-0.43	-0.08	-0.78		-0.80	-0.44	-1.15	
11315	3(1)	17	-0.56	-0.21	-0.92		-0.47	-0.11	-0.82	
13042	3(1)	18	-0.44	-0.09	-0.79		-0.37	-0.01	-0.72	
15048	3(2)	1	-0.51	-0.07	-0.96		-0.24	0.20	-0.68	
11321	3(2)	2	-0.78	-0.32	-1.25		-0.62	-0.15	-1.10	
14984	3(2)	3	-0.08	0.26	-0.42		-0.07	0.27	-0.41	
11326	3(2)	4	0.02	0.35	-0.30		-0.07	0.25	-0.40	
15055	3(2)	5	-0.17	0.21	-0.55		-0.27	0.10	-0.65	
18375	3(2)	6	-0.57	-0.24	-0.91		-0.20	0.14	-0.53	
15002	3(2)	7	-0.14	0.18	-0.47		0.00	0.33	-0.33	
11292	3(2)	12	-0.01	0.46	-0.48		0.46	0.94	-0.01	
11176	3(2)	13	-0.14	0.19	-0.48		-0.34	-0.01	-0.67	
14999	3(2)	14	0.08	0.44	-0.28		-0.10	0.25	-0.46	
15061	3(2)	15	-0.91	-0.56	-1.25		-0.60	-0.26	-0.95	
15037	3(2)	16	0.47	0.83	0.11		0.64	1.00	0.28	
14997	3(2)	17	-0.17	0.18	-0.53		-0.15	0.20	-0.50	
15045	3(2)	18	-0.30	0.06	-0.66		-0.26	0.09	-0.62	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17405	3(1)	7	-0.02	0.80		0.03	0.38	
15068	3(1)	8	0.01	0.95		0.08	0.02	
11409	3(1)	9	0.06	0.02		0.11	0.00	
15010	3(1)	10	-0.06	0.07		-0.10	0.00	
18368	3(2)	8	-0.05	0.05		-0.08	0.00	
11310	3(2)	9	0.03	0.44		0.00	0.19	
11329	3(2)	10	0.13	0.00		0.16	0.00	
15034	3(2)	11	0.10	0.00		0.10	0.00	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
11214	3(1)	1	-0.04	0.50	-0.59		0.01	0.56	-0.54	
14596	3(1)	2	0.08	0.52	-0.37		0.04	0.49	-0.40	
14595	3(1)	3	-0.22	0.19	-0.63		-0.25	0.16	-0.66	
17779	3(1)	4	0.38	0.77	-0.02		0.14	0.54	-0.25	
14583	3(1)	5	-0.19	0.21	-0.59		-0.03	0.38	-0.43	
14572	3(1)	6	-0.03	0.65	-0.70		0.04	0.73	-0.65	
17785	3(1)	11	-0.36	0.05	-0.76		-0.27	0.14	-0.69	
14658	3(1)	12	0.24	0.62	-0.15		0.19	0.58	-0.20	
14649	3(1)	13	0.00	0.45	-0.45		-0.19	0.27	-0.65	
14608	3(1)	14	0.14	0.52	-0.25		0.07	0.46	-0.32	
17781	3(1)	15	-0.42	0.15	-0.99		-0.46	0.12	-1.04	
10998	3(1)	16	0.03	0.40	-0.34		0.05	0.42	-0.33	
17767	3(1)	17	0.29	0.70	-0.11		0.49	0.89	0.09	
14630	3(1)	18	-0.12	0.30	-0.54		-0.24	0.20	-0.67	
17721	3(2)	1	0.30	0.90	-0.31		0.47	1.09	-0.14	
14628	3(2)	2	0.25	0.66	-0.16		0.09	0.50	-0.33	
14615	3(2)	3	-0.36	0.09	-0.80		-0.28	0.16	-0.72	
17768	3(2)	4	-0.02	0.36	-0.41		0.08	0.46	-0.31	
11240	3(2)	5	0.44	0.99	-0.11		0.17	0.71	-0.38	
10997	3(2)	6	0.04	0.47	-0.39		-0.04	0.39	-0.48	
14661	3(2)	7	-0.06	0.36	-0.47		-0.10	0.32	-0.51	
11098	3(2)	12	0.63	1.06	0.21		0.67	1.11	0.23	
11223	3(2)	13	0.29	0.71	-0.13		0.02	0.44	-0.40	
14646	3(2)	14	-0.65	-0.24	-1.07		-0.64	-0.22	-1.05	
17771	3(2)	15	0.06	0.45	-0.33		-0.13	0.26	-0.53	
12616	3(2)	16	0.40	0.80	0.01		0.30	0.70	-0.09	
14589	3(2)	17	-0.01	0.36	-0.39		0.15	0.53	-0.24	
17784	3(2)	18	0.11	0.56	-0.34		-0.20	0.26	-0.66	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
17769	3(1)	7	0.00	0.90		0.00	0.61	
14631	3(1)	8	-0.01	0.47		0.02	0.78	
14667	3(1)	9	-0.07	0.17		-0.02	0.80	
14599	3(1)	10	0.08	0.05		0.07	0.07	
17773	3(2)	8	-0.03	0.61		0.00	0.69	
11059	3(2)	9	0.02	0.03		0.00	0.01	
14668	3(2)	10	-0.02	0.37		-0.03	0.68	
17766	3(2)	11	-0.02	0.33		0.00	0.02	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Upper Limit	Lower Limit	DIF Level	Δ	Upper Limit	Lower Limit	DIF Level
18323	3(1)	1	0.25	0.82	-0.31		0.37	0.93	-0.20	
11648	3(1)	2	-0.04	0.54	-0.61		-0.39	0.18	-0.96	
14734	3(1)	3	-0.07	0.50	-0.64		-0.12	0.45	-0.69	
14704	3(1)	4	-0.33	0.22	-0.88		-0.62	-0.06	-1.17	
11457	3(1)	5	0.11	0.72	-0.50		0.16	0.76	-0.43	
14715	3(1)	6	-0.16	0.40	-0.73		-0.32	0.26	-0.90	
14750	3(1)	11	0.50	1.12	-0.12		0.68	1.30	0.07	
18007	3(1)	12	0.63	1.30	-0.04		0.28	0.94	-0.38	
14758	3(1)	13	0.03	0.56	-0.49		-0.32	0.21	-0.85	
14711	3(1)	14	0.36	0.92	-0.21		-0.05	0.52	-0.61	
11507	3(1)	15	-0.03	0.51	-0.57		-0.29	0.26	-0.84	
14703	3(1)	16	-0.16	0.39	-0.70		-0.19	0.35	-0.74	
18014	3(1)	17	-0.39	0.15	-0.93		-0.45	0.09	-1.00	
14731	3(1)	18	-0.10	0.48	-0.68		-0.42	0.16	-1.00	
18015	3(2)	1	-0.14	0.43	-0.71		-0.58	-0.02	-1.14	
14767	3(2)	2	-0.28	0.28	-0.83		-0.06	0.50	-0.61	
14728	3(2)	3	0.07	0.59	-0.46		0.22	0.75	-0.31	
18318	3(2)	4	-0.72	-0.10	-1.33		-0.30	0.32	-0.91	
11484	3(2)	5	0.12	0.65	-0.41		-0.13	0.38	-0.65	
18321	3(2)	6	-0.24	0.33	-0.82		-0.47	0.10	-1.05	
18315	3(2)	7	0.23	0.74	-0.28		-0.07	0.45	-0.59	
11565	3(2)	12	-0.04	0.51	-0.59		-0.42	0.11	-0.96	
18319	3(2)	13	-0.08	0.46	-0.61		0.29	0.82	-0.23	
14763	3(2)	14	-0.10	0.50	-0.70		0.11	0.71	-0.50	
14696	3(2)	15	-0.53	-0.01	-1.04		-0.42	0.10	-0.94	
18017	3(2)	16	0.56	1.34	-0.21		0.65	1.40	-0.10	
14726	3(2)	17	-0.08	0.45	-0.61		-0.15	0.38	-0.67	
11520	3(2)	18	0.39	1.04	-0.27		-0.15	0.49	-0.78	

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

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

Item Code	Booklet (Section)	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
18317	3(1)	7	0.02	0.46		0.06	0.57	
18316	3(1)	8	0.09	0.33		0.06	0.27	
18324	3(1)	9	-0.01	0.64		-0.02	0.24	
14688	3(1)	10	0.06	0.05		0.07	0.02	
11533	3(2)	8	0.04	0.53		0.02	0.09	
18004	3(2)	9	0.00	0.84		0.03	0.36	
11508	3(2)	10	0.00	0.06		0.00	0.25	
14744	3(2)	11	-0.04	0.16		-0.09	0.04	

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
10236	1	NA1.03	KU	N	3	74.30	0.32	-0.96	0.63
10150	2	NA1.04	PS	N	3	49.35	0.30	0.43	0.67
10127	3	NA1.05	PS	N	4	60.34	0.33	-0.03	0.64
10240	4	NA1.06	AP	N	1	52.97	0.33	0.41	0.66
14774	5	NA2.02	KU	N	3	44.62	0.32	0.79	0.78
14784	7	NA2.07	AP	N	2	33.24	0.20	1.74	0.66
14791	8	NA1.05	PS	N	4*	48.35(1.90)	0.50	-0.26	0.39
10196	10	LR1.01	AP	R	4	52.99	0.37	0.25	0.80
EQME31814	11	LR2.01	KU	R	1	82.88	0.22	-1.96	0.44
14802	12	LR2.01	AP	R	4	49.99	0.35	0.51	0.74
EQME10508	14	LR3.01	AP	R	2	50.89	0.31	0.47	0.73
EQME31796	15	LR3.03	KU	R	4	46.07	0.29	0.78	0.68
EQME21490	16	LR3.04	PS	R	3	42.69	0.34	0.75	0.96
10176	17	LR4.01	PS	R	3	29.00	0.24	1.81	0.81
10103	18	LR4.03	AP	R	4	55.35	0.34	0.10	0.72
10164	21	LR1.01	AP	R	4*	67.78(2.65)	0.52	-1.35	0.36
10135	24	MG2.03	AP	M	2	70.74	0.48	-0.22	1.22
10104	25	MG2.05	KU	M	4	70.03	0.41	-0.57	0.85
10158	26	MG2.05	PS	M	2	31.39	0.21	1.60	0.89
EQME21446	28	MG3.01	KU	M	3	73.98	0.34	-0.79	0.69
14870	31	MG3.02	AP	M	4*	45.51(1.79)	0.49	-0.07	0.36
14785	NR	NA2.08	KU	N	3	49.38	0.37	0.50	0.86
14797	NR	NA2.07	AP	N	4*	39.18(1.54)	0.46	0.22	0.32
14811	NR	LR2.03	AP	R	1	49.33	0.20	0.89	0.39
14827	NR	LR4.04	PS	R	3	58.86	0.21	0.09	0.35
14831	NR	LR4.06	KU	R	3	78.47	0.32	-1.13	0.65
14843	NR	LR3.05	AP	R	4*	43.30(1.70)	0.52	0.14	0.42
14844	NR	LR4.01	PS	R	4*	69.16(2.73)	0.25	-0.97	0.14
14851	NR	MG1.03	PS	M	4	67.85	0.25	-0.59	0.45
14857	NR	MG3.01	AP	M	1	53.93	0.30	0.35	0.65
14864	NR	MG2.03	PS	M	4*	65.12(2.50)	0.58	-1.04	0.59

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.

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

() = mean score for OR items.

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
10164	21	% of Students	3.39	0.22	26.70	10.23	13.87	45.59
		Parameters			-4.28	-0.77	-0.37	0.11
14791	8	% of Students	4.82	1.82	40.76	18.95	19.83	13.81
		Parameters			-3.49	-0.07	0.65	1.85
14797	NR	% of Students	14.14	2.36	46.13	15.95	7.00	14.42
		Parameters			-2.48	0.60	1.24	1.68
14843	NR	% of Students	11.43	1.26	39.21	21.95	14.48	11.66
		Parameters			-2.36	0.11	0.96	1.88
14844	NR	% of Students	0.38	0.03	20.99	21.36	16.01	41.23
		Parameters				-2.83	-0.60	0.72
14864	NR	% of Students	1.55	0.38	20.58	22.96	24.17	30.37
		Parameters			-3.63	-1.05	-0.16	0.69
14870	31	% of Students	10.53	1.20	39.81	19.32	12.97	16.16
		Parameters			-2.75	0.10	0.85	1.59

Note. The total number of students is 16 494. 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
10236	1	NA1.03	KU	N	3	77.16	0.31	-0.96	0.63
10150	2	NA1.04	PS	N	3	55.08	0.34	0.43	0.67
10127	3	NA1.05	PS	N	4	60.55	0.34	-0.03	0.64
10240	4	NA1.06	AP	N	1	52.44	0.33	0.41	0.66
14779	6	NA2.05	KU	N	1	41.34	0.26	1.19	0.62
14793	9	NA2.02	AP	N	4*	39.85(1.56)	0.49	0.22	0.34
10196	10	LR1.01	AP	R	4	56.84	0.39	0.25	0.80
14825	13	LR2.01	AP	R	1	68.10	0.28	-0.54	0.50
EQME10508	14	LR3.01	AP	R	2	52.22	0.34	0.47	0.73
EQME21490	16	LR3.04	PS	R	3	45.51	0.39	0.75	0.96
10103	18	LR4.03	AP	R	4	61.25	0.35	0.10	0.72
14826	19	LR4.04	KU	R	2	74.39	0.38	-0.72	0.80
14833	20	LR4.06	PS	R	4	37.69	0.33	1.17	0.81
14840	22	LR3.02	AP	R	4*	40.58(1.60)	0.50	-0.09	0.51
14847	23	LR4.04	PS	R	4*	69.54(2.74)	0.54	-1.26	0.47
10135	24	MG2.03	AP	M	2	60.87	0.49	-0.22	1.22
10104	25	MG2.05	KU	M	4	72.85	0.40	-0.57	0.85
10158	26	MG2.05	PS	M	2	31.41	0.26	1.60	0.89
14856	27	MG2.05	PS	M	3	50.56	0.38	0.48	0.83
EQME21446	28	MG3.01	KU	M	3	74.06	0.35	-0.79	0.69
14858	29	MG3.01	AP	M	2	46.87	0.22	1.07	0.42
14863	30	MG2.02	PS	M	4*	37.25(1.47)	0.48	-0.06	0.39
EQME21477	NR	NA2.02	AP	N	1	73.74	0.24	-1.10	0.42
EQME31752	NR	NA2.07	KU	N	1	46.09	0.25	0.98	0.52
14788	NR	NA1.04	PS	N	4*	76.13(3.00)	0.49	-1.51	0.37
14803	NR	LR2.01	AP	R	4	71.89	0.36	-0.64	0.71
14808	NR	LR2.03	KU	R	3	64.54	0.34	-0.24	0.64
14816	NR	LR3.03	KU	R	3	70.59	0.21	-0.96	0.35
14822	NR	LR4.02	PS	R	2	42.88	0.31	0.97	0.72
14837	NR	LR2.02	AP	R	4*	67.95(2.68)	0.40	-2.90	0.26
14871	NR	MG3.02	AP	M	4*	43.06(1.70)	0.52	0.04	0.39

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.

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

() = mean score for OR items.

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
14788	NR	% of Students	2.55	1.37	14.43	10.90	14.74	56.01
		Parameters			-3.42	-1.34	-0.82	-0.31
14793	9	% of Students	12.03	3.20	50.13	8.78	11.71	14.15
		Parameters			-2.57	0.70	1.02	1.75
14837	NR	% of Students	0.32	0.06	20.30	26.84	12.09	40.39
		Parameters			-9.90	-1.86	-0.11	0.49
14840	22	% of Students	2.27	0.24	55.54	26.67	7.68	7.60
		Parameters			-4.26	0.44	1.48	2.03
14847	23	% of Students	1.88	0.36	12.44	25.77	24.02	35.52
		Parameters			-3.58	-1.66	-0.36	0.58
14863	30	% of Students	4.97	1.57	71.07	4.32	2.98	15.08
		Parameters			-4.07	1.15	1.28	1.43
14871	NR	% of Students	11.14	1.87	40.96	23.78	5.25	16.99
		Parameters			-2.46	0.25	1.15	1.42

Note. The total number of students is 19 224. 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
10278	1	NA1.01	KU	N	4	80.52	0.22	-1.80	0.42
EQME10477	3	NA2.02	AP	N	4	75.39	0.35	-0.95	0.71
EQME21249	4	NA2.06	KU	N	1	67.60	0.28	-0.58	0.49
EQME31889	5	NA2.09	PS	N	2	82.10	0.35	-1.32	0.74
10289	6	NA2.02	AP	N	4*	74.24(2.96)	0.47	-1.78	0.35
EQME21223	7	LR1.04	PS	R	3	68.34	0.26	-0.36	0.46
10283	8	LR2.02	PS	R	2	59.71	0.30	0.01	0.58
10179	9	LR2.03	AP	R	1	71.09	0.35	-0.77	0.68
14896	11	LR3.02	KU	R	4	89.51	0.25	-2.25	0.59
14910	14	LR3.01	PS	R	4*	67.60(2.69)	0.47	-1.66	0.33
EQME10321	15	AG1.02	KU	G	3	85.48	0.26	-2.00	0.54
EQME10437	16	AG1.03	AP	G	1	84.46	0.36	-1.40	0.83
14920	17	AG2.01	PS	G	3	56.68	0.42	0.10	0.97
10180	19	AG3.03	PS	G	1	55.92	0.43	0.09	1.03
14932	20	AG3.04	KU	G	2	88.10	0.26	-2.09	0.59
14945	23	AG3.04	PS	G	4*	70.50(2.80)	0.53	-1.37	0.38
10312	24	MG2.01	KU	M	4	74.87	0.32	-0.91	0.68
EQME21204	25	MG2.02	PS	M	3	62.36	0.46	-0.21	1.00
14953	26	MG2.06	AP	M	1	78.32	0.43	-0.92	0.97
10313	27	MG2.06	PS	M	4	64.28	0.44	-0.30	0.95
14961	29	MG3.02	AP	M	2	73.87	0.32	-0.87	0.63
14963	30	MG2.02	AP	M	4*	72.64(2.89)	0.61	-1.18	0.50
14875	NR	NA1.03	AP	N	2	64.30	0.39	-0.28	0.77
EQME31994	NR	LR2.01	KU	R	4	89.70	0.36	-1.70	0.95
14895	NR	LR3.01	AP	R	1	73.51	0.41	-0.73	0.84
14909	NR	LR2.05	AP	R	4*	51.73(2.06)	0.52	-0.62	0.43
14919	NR	AG2.02	AP	G	2	78.64	0.43	-0.94	0.98
EQME31884	NR	AG3.05	AP	G	2	72.73	0.28	-0.94	0.50
14939	NR	AG2.04	AP	G	4*	60.44(2.40)	0.61	-1.00	0.53
14957	NR	MG3.01	KU	M	3	84.58	0.35	-1.40	0.83
14969	NR	MG3.01	PS	M	4*	67.31(2.68)	0.62	-1.04	0.53

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.

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

() = mean score for OR items.

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
10289	6	% of Students	1.34	0.45	12.84	14.63	28.08	42.66
		Parameters			-4.56	-1.82	-0.99	0.25
14909	NR	% of Students	2.46	0.31	36.43	25.56	21.61	13.63
		Parameters			-4.37	-0.47	0.55	1.81
14910	14	% of Students	1.59	0.42	19.34	24.37	14.83	39.46
		Parameters			-5.15	-1.51	-0.24	0.39
14939	NR	% of Students	1.82	0.48	31.69	20.37	13.21	32.43
		Parameters			-3.97	-0.58	0.09	0.53
14945	23	% of Students	3.12	0.58	14.39	22.65	14.76	44.50
		Parameters			-3.51	-1.53	-0.43	0.12
14963	30	% of Students	4.11	0.81	17.98	11.73	12.39	52.99
		Parameters			-2.81	-1.01	-0.58	-0.23
14969	NR	% of Students	3.71	0.54	15.81	29.40	7.50	43.03
		Parameters			-2.84	-1.26	-0.08	0.15

Note. The total number of students is 41 861. 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
10278	1	NA1.01	KU	N	4	81.44	0.22	-1.80	0.42
14877	2	NA1.03	AP	N	1	66.83	0.35	-0.35	0.68
EQME10477	3	NA2.02	AP	N	4	78.19	0.35	-0.95	0.71
EQME21249	4	NA2.06	KU	N	1	68.80	0.28	-0.58	0.49
EQME21223	7	LR1.04	PS	R	3	62.27	0.27	-0.36	0.46
10283	8	LR2.02	PS	R	2	59.71	0.33	0.01	0.58
10179	9	LR2.03	AP	R	1	75.35	0.34	-0.77	0.68
EQME31904	10	LR2.05	KU	R	4	79.74	0.43	-0.90	1.04
14900	12	LR3.03	AP	R	4	55.50	0.44	0.20	1.01
14908	13	LR2.04	AP	R	4*	65.41(2.61)	0.59	-1.19	0.56
EQME10321	15	AG1.02	KU	G	3	87.23	0.24	-2.00	0.54
EQME10437	16	AG1.03	AP	G	1	85.88	0.35	-1.40	0.83
14918	18	AG2.02	AP	G	1	66.31	0.37	-0.31	0.73
10180	19	AG3.03	PS	G	1	58.76	0.45	0.09	1.03
14933	21	AG3.04	AP	G	4	66.17	0.37	-0.31	0.71
10322	22	AG2.04	AP	G	4*	52.74(2.10)	0.58	-0.30	0.58
10312	24	MG2.01	KU	M	4	76.47	0.35	-0.91	0.68
EQME21204	25	MG2.02	PS	M	3	65.90	0.46	-0.21	1.00
10313	27	MG2.06	PS	M	4	67.63	0.44	-0.30	0.95
EQME31931	28	MG3.01	KU	M	2	90.76	0.32	-1.82	0.86
14971	31	MG3.02	PS	M	4*	79.16(3.15)	0.58	-1.62	0.51
14881	NR	NA2.07	PS	N	2	80.14	0.29	-1.36	0.55
14884	NR	NA2.02	AP	N	4*	85.56(3.41)	0.43	-2.35	0.34
EQME10310	NR	LR3.04	KU	R	4	85.31	0.33	-1.50	0.73
14911	NR	LR3.02	PS	R	4*	72.62(2.90)	0.50	-1.32	0.41
10154	NR	AG2.01	PS	G	3	59.41	0.42	0.05	0.94
14927	NR	AG3.01	KU	G	2	86.17	0.41	-1.28	1.06
14947	NR	AG3.05	PS	G	4*	70.28(2.80)	0.53	-1.46	0.41
14948	NR	MG1.01	AP	M	3	67.04	0.29	-0.43	0.53
EQME31966	NR	MG3.01	AP	M	4	86.27	0.41	-1.29	1.05
14965	NR	MG2.03	PS	M	4*	72.73(2.90)	0.54	-1.36	0.47

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.

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

() = mean score for OR items.

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
10322	22	% of Students	4.73	0.27	29.35	29.75	21.50	14.40
		Parameters			-2.77	-0.55	0.54	1.60
14884	NR	% of Students	0.95	0.19	7.54	8.79	13.03	69.51
		Parameters			-4.65	-2.13	-1.48	-0.97
14908	13	% of Students	0.86	0.18	17.79	29.22	22.40	29.55
		Parameters			-4.18	-1.31	-0.05	0.78
14911	NR	% of Students	2.83	0.09	8.98	20.48	29.94	37.69
		Parameters			-3.24	-1.86	-0.75	0.56
14947	NR	% of Students	1.48	0.18	17.17	17.49	25.76	37.93
		Parameters			-4.45	-1.36	-0.53	0.50
14965	NR	% of Students	1.70	0.17	8.75	21.84	31.64	35.89
		Parameters			-3.43	-1.90	-0.70	0.61
14971	31	% of Students	1.24	0.26	16.89	8.01	10.67	62.93
		Parameters			-4.03	-1.12	-0.80	-0.50

Note. The total number of students is 53 448. 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
CMS9700	2	PN02	MA	N	2	34.06	0.19	1.68	0.60
9723	3	PN06	CC	N	1	51.50	0.23	0.53	0.50
OQMF31752	4	PN07	CC	N	1	51.77	0.36	0.52	0.63
18450	5	PN08	MA	N	1	49.32	0.35	0.45	0.79
CMS10021	6	PN09	HP	N	2	62.13	0.25	-0.18	0.45
16517	9	PN13	HP	N	4*	43.32(1.70)	0.34	0.71	0.27
OQMF31770	10	PR03	MA	R	1	50.14	0.37	0.45	1.01
CMS9887	11	PR04	CC	R	2	58.31	0.32	-0.29	0.70
OQMF10298	12	PR04	HP	R	3	41.42	0.19	1.58	0.39
16512	13	PR09	CC	R	2	42.78	0.28	0.99	0.58
CMS10031	15	PR11	CC	R	2	71.39	0.34	-0.74	0.59
OQMF31783	16	PR13	HP	R	4	60.22	0.33	-0.01	0.63
CMS9759	19	PR09	MA	R	4*	50.48(1.95)	0.49	-0.52	0.39
OQMF21399	21	PM05	MA	M	1	57.22	0.35	0.12	0.89
OQMF31807	22	PM06	CC	M	2	49.86	0.21	0.56	0.44
OQMF31806	23	PM06	HP	M	3	46.32	0.46	0.46	0.82
OQMF31826	24	PM14	MA	M	3	55.31	0.28	0.01	0.82
18455	25	PM14	HP	M	1	51.23	0.27	0.52	0.51
14449	26	PM17	HP	M	2	68.39	0.19	-0.80	0.33
OQMF31839	28	PM08	HP	M	4*	60.69(2.35)	0.48	-0.91	0.39
16465	NR	PN02	CC	N	4	41.96	0.30	0.84	0.89
9725	NR	PN11	MA	N	3	78.20	0.19	-1.99	0.32
16519	NR	PN01	MA	N	4*	45.78(1.80)	0.51	0.10	0.39
14407	NR	PR10	MA	R	3	43.60	0.25	0.93	0.63
14413	NR	PR14	MA	R	4	59.95	0.37	-0.08	0.75
16546	NR	PR04	MA	R	4*	48.02(1.88)	0.45	-0.68	0.29
16910	NR	PR18	HP	R	4*	61.31(2.43)	0.25	-2.63	0.15
14423	NR	PM03	HP	M	3	68.66	0.33	-0.56	0.66
OQMF31850	NR	PM17	CC	M	3	63.76	0.31	-0.38	0.62
16467	NR	PM17	MA	M	2	66.49	0.23	-0.59	0.42
16933	NR	PM03	MA	M	4*	52.32(2.08)	0.46	-0.68	0.32

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.

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

() = mean score for OR items.

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
16517	9	% of Students	4.90	6.27	25.07	46.32	14.17	3.27
		Parameters			-3.21	-0.97	2.33	4.68
16519	NR	% of Students	10.35	3.54	34.06	17.71	23.71	10.63
		Parameters			-2.25	-0.13	0.55	2.24
16546	NR	% of Students	7.36	1.36	54.22	3.81	2.72	30.52
		Parameters			-4.31	0.46	0.59	0.70
16910	NR	% of Students	1.63	1.09	19.07	40.60	5.45	32.15
		Parameters			-9.09	-3.12	0.90	1.36
16933	NR	% of Students	4.36	2.18	30.25	33.51	6.81	22.89
		Parameters			-3.80	-0.70	0.82	1.17
CMS9759	19	% of Students	4.63	1.91	35.69	27.52	9.81	20.44
		Parameters			-3.47	-0.36	0.72	1.19
OQMF31839	28	% of Students	3.00	1.09	16.35	34.33	23.16	22.07
		Parameters			-3.49	-1.56	0.13	1.31

Note. The total number of students is 367. 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
18454	1	PN02	CC	N	2	57.27	0.29	0.51	0.21
CMS9700	2	PN02	MA	N	2	36.26	0.22	0.60	1.68
OQMF31752	4	PN07	CC	N	1	50.69	0.30	0.63	0.52
CMS10021	6	PN09	HP	N	2	62.70	0.26	0.45	-0.18
14382	7	PN12	MA	N	2	63.86	0.42	0.95	-0.13
CMS9626	8	PN02	MA	N	4*	54.21(2.11)	0.49	0.35	-0.73
OQMF31770	10	PR03	MA	R	1	49.65	0.44	1.01	0.45
CMS9887	11	PR04	CC	R	2	68.59	0.36	0.70	-0.29
9989	14	PR09	MA	R	1	28.41	0.20	0.85	1.95
CMS10031	15	PR11	CC	R	2	72.06	0.31	0.59	-0.74
OQMF31783	16	PR13	HP	R	4	59.93	0.35	0.63	-0.01
14412	17	PR14	MA	R	2	39.03	0.21	0.59	1.47
OQMF31791	18	PR01	MA	R	4*	43.97(1.72)	0.52	0.44	0.08
18460	20	PM04	HP	M	2	39.15	0.38	1.10	0.96
OQMF21399	21	PM05	MA	M	1	58.20	0.41	0.89	0.12
OQMF31807	22	PM06	CC	M	2	53.58	0.25	0.44	0.56
OQMF31806	23	PM06	HP	M	3	52.08	0.36	0.82	0.46
OQMF31826	24	PM14	MA	M	3	61.89	0.43	0.82	0.01
14449	26	PM17	HP	M	2	68.24	0.21	0.33	-0.80
OQMF21395	27	PM04	MA	M	4*	61.14(2.38)	0.53	0.38	-0.78
14372	NR	PN07	CC	N	1	59.93	0.35	0.68	0.04
14369	NR	PN05	MA	N	1	20.44	0.11	0.57	3.13
12839	NR	PN09	HP	N	4*	52.19(2.04)	0.49	0.46	0.11
OQMF10299	NR	PR04	HP	R	3	47.00	0.26	0.52	0.91
14397	NR	PR06	CC	R	3	72.40	0.17	0.29	-1.36
9927	NR	PR09	MA	R	4*	67.49(2.62)	0.52	0.44	-1.00
14418	NR	PR13	HP	R	4*	55.34(2.18)	0.53	0.40	-0.46
9907	NR	PM14	HP	M	3	53.00	0.36	0.77	0.39
OQMF31850	NR	PM17	CC	M	3	67.90	0.33	0.62	-0.38
18467	NR	PM17	MA	M	3	67.78	0.28	0.47	-0.53
14456	NR	PM05	HP	M	4*	58.83(2.31)	0.57	0.43	-0.36

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.

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

() = mean score for OR items.

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
12839	NR	% of Students	5.08	1.73	16.51	42.26	29.91	4.50
		Parameters			-2.50	-1.21	0.76	3.39
14418	NR	% of Students	4.73	2.19	23.67	30.60	18.71	20.09
		Parameters			-2.90	-0.82	0.50	1.44
14456	NR	% of Students	11.32	1.39	12.47	26.44	23.56	24.83
		Parameters			-1.74	-0.96	0.07	1.15
9927	NR	% of Students	2.77	0.81	11.43	23.67	34.06	27.25
		Parameters			-3.06	-1.62	-0.46	1.12
CMS9626	8	% of Students	2.31	1.50	33.72	24.13	18.48	19.86
		Parameters			-4.44	-0.51	0.54	1.53
OQMF21395	27	% of Students	5.43	1.39	19.98	29.33	9.58	34.30
		Parameters			-2.88	-0.99	0.26	0.64
OQMF31791	18	% of Students	7.51	1.15	40.88	27.25	12.36	10.85
		Parameters			-2.83	0.07	1.17	1.97

Note. The total number of students is 866. 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
CMS9657	1	DN01	MA	N	2	64.79	0.33	-0.50	0.57
17060	2	DN06	CC	N	4	72.21	0.41	-0.67	0.88
OQMF10391	4	DN11	MA	N	2	66.14	0.41	-0.43	0.79
OQMF20950	5	DN13	HP	N	4	40.29	0.40	0.68	0.98
14487	6	DN19	MA	N	4	78.93	0.30	-1.28	0.60
14493	8	DN18	MA	N	4*	48.91(1.93)	0.61	-0.25	0.53
CMS9660	10	DR04	HP	R	3	75.00	0.42	-0.79	0.93
CMS9947	11	DR09	CC	R	2	69.86	0.29	-0.50	0.59
9789	12	DR09	MA	R	1	49.64	0.41	0.39	0.92
CMS9806	14	DR14	HP	R	2	47.71	0.29	0.55	0.57
14500	16	DR11	HP	R	4*	64.20(2.55)	0.56	-1.40	0.49
OQMF10331	17	DG04	MA	G	3	65.21	0.45	-0.27	0.97
CMS10054	18	DG07	CC	G	1	69.21	0.38	-0.65	0.82
16949	19	DG11	CC	G	2	70.21	0.39	-0.57	0.87
17064	21	DG05	MA	G	4*	61.18(2.43)	0.57	-0.81	0.47
CMS9669	23	DM03	HP	M	3	72.07	0.31	-0.78	0.59
OQMF21008	24	DM05	CC	M	4	71.21	0.30	-1.07	0.52
14528	25	DM05	MA	M	2	62.07	0.36	-0.21	0.63
14539	26	DM15	CC	M	2	76.57	0.24	-1.39	0.44
CMS9694	28	DM20	CC	M	3	82.79	0.24	-1.71	0.56
CMS10066	31	DM20	HP	M	4*	64.57(2.56)	0.63	-1.00	0.53
14470	NR	DN09	CC	N	3	85.71	0.29	-1.71	0.67
9967	NR	DN03	MA	N	4*	67.96(2.69)	0.59	-1.26	0.43
14495	NR	DR04	MA	R	4	82.86	0.41	-1.16	1.05
17062	NR	DR15	MA	R	3	73.71	0.32	-0.94	0.59
10064	NR	DR04	MA	R	4*	68.54(2.71)	0.61	-1.03	0.51
14509	NR	DG07	HP	G	1	55.86	0.48	0.06	1.19
9668	NR	DG12	HP	G	2	54.00	0.32	0.27	0.69
18462	NR	DM18	HP	M	3	45.93	0.39	0.59	0.84
9672	NR	DM20	MA	M	3	70.64	0.23	-0.97	0.39
OQMF10361	NR	DM18	HP	M	4*	77.14(3.06)	0.48	-1.87	0.37

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.

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

() = mean score for OR items.

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
10064	NR	% of Students	4.07	1.57	17.07	19.00	14.07	44.21
		Parameters			-2.60	-1.08	-0.38	0.05
14493	8	% of Students	6.71	3.07	36.29	22.64	11.07	20.21
		Parameters			-2.36	-0.17	0.56	1.03
14500	16	% of Students	0.50	0.21	23.50	17.07	35.71	23.00
		Parameters			-5.29	-1.10	-0.42	1.17
17064	21	% of Students	4.86	1.00	20.43	27.57	15.43	30.71
		Parameters			-2.80	-1.03	0.05	0.63
9967	NR	% of Students	2.07	1.64	24.86	14.71	9.29	47.43
		Parameters			-3.68	-0.85	-0.33	-0.05
CMS10066	31	% of Students	2.50	1.57	29.21	12.21	13.36	41.14
		Parameters			-3.25	-0.61	-0.24	0.15
OQMF10361	NR	% of Students	1.50	0.21	10.71	14.79	22.86	49.93
		Parameters			-4.27	-1.93	-1.07	-0.15

Note. The total number of students is 1400. 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
CMS9657	1	DN01	MA	N	2	69.90	0.29	-0.50	0.57
10048	3	DN11	CC	N	4	68.49	0.30	-0.50	0.56
OQMF10391	4	DN11	MA	N	2	69.58	0.37	-0.43	0.79
OQMF20950	5	DN13	HP	N	4	47.28	0.39	0.68	0.98
OQMF20902	7	DN11	MA	N	4*	65.15(1.85)	0.51	-1.50	0.48
9658	9	DR03	MA	R	4	67.37	0.34	-0.41	0.60
CMS9660	10	DR04	HP	R	3	76.98	0.43	-0.79	0.93
CMS9947	11	DR09	CC	R	2	67.05	0.34	-0.50	0.59
9948	13	DR12	MA	R	4	72.70	0.37	-0.66	0.72
CMS9806	14	DR14	HP	R	2	53.00	0.29	0.55	0.57
CMS9968	15	DR03	MA	R	4*	67.25(2.68)	0.57	-1.26	0.40
OQMF10331	17	DG04	MA	G	3	65.81	0.44	-0.27	0.97
CMS10054	18	DG07	CC	G	1	74.54	0.40	-0.65	0.82
14520	20	DG12	HP	G	3	58.61	0.28	0.14	0.53
14524	22	DG06	MA	G	4*	59.08(2.35)	0.65	-0.66	0.71
CMS9669	23	DM03	HP	M	3	72.66	0.33	-0.78	0.59
OQMF21008	24	DM05	CC	M	4	77.58	0.29	-1.07	0.52
14548	27	DM18	HP	M	3	60.21	0.35	0.04	0.73
CMS9694	28	DM20	CC	M	3	84.55	0.30	-1.71	0.56
9842	29	DM20	MA	M	2	65.73	0.34	-0.31	0.62
14564	30	DM18	HP	M	4*	63.19(2.52)	0.47	-1.38	0.34
18448	NR	DN06	CC	N	2	70.14	0.40	-0.47	0.83
14480	NR	DN16	MA	N	2	89.03	0.27	-1.99	0.65
17065	NR	DN18	MA	N	4*	75.42(3.01)	0.54	-1.27	0.39
9662	NR	DR09	MA	R	2	61.73	0.35	-0.07	0.66
OQMF31900	NR	DR15	HP	R	4*	68.27(2.73)	0.58	-0.92	0.44
14511	NR	DG08	HP	G	4	54.84	0.40	0.28	0.98
18466	NR	DG11	CC	G	2	57.21	0.35	0.19	0.71
16939	NR	DM09	MA	M	3	79.62	0.40	-0.97	0.88
17063	NR	DM14	CC	M	1	71.82	0.35	-0.63	0.68
14566	NR	DM20	HP	M	4*	66.69(2.66)	0.58	-1.15	0.47

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.

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

() = mean score for OR items.

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

Item Code	Sequence		Score Points					
			Missing	Illegible	10	20	30	40
14524	22	% of Students	1.88	0.24	22.98	32.47	21.34	21.10
		Parameters			-3.14	-0.89	0.29	1.09
14564	30	% of Students	1.20	0.56	16.89	36.27	16.97	28.10
		Parameters			-4.97	-1.80	0.26	1.09
14566	NR	% of Students	1.80	0.32	20.66	18.33	26.10	32.79
		Parameters			-3.92	-1.06	-0.31	0.70
17065	NR	% of Students	4.28	1.68	11.17	16.73	7.49	58.65
		Parameters			-2.56	-1.35	-0.60	-0.36
CMS9968	15	% of Students	2.04	1.00	30.46	7.85	11.73	46.92
		Parameters			-4.26	-0.54	-0.28	0.08
OQMF20902	7	% of Students	0.40	0.12	7.17	49.00	17.81	25.50
		Parameters			-4.64	-2.72	0.33	1.06
OQMF31900	NR	% of Students	4.72	1.76	16.77	14.45	21.74	40.55
		Parameters			-2.53	-1.00	-0.44	0.34

Note. The total number of students is 2498. 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 assessment are the results from two random samples of 2000 examinees. For the French-language assessment, gender-based DIF analysis was conducted for one sample of students, but 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 assessment. 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 and upper limits of the confidence band, and 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 Multiple-Choice 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
10236	11101	1	0.10	-0.41	0.61		-0.40	-0.91	0.10	
10150	11101	2	0.03	-0.40	0.47		0.48	0.04	0.92	
10127	11101	3	0.97	0.51	1.43		0.74	0.28	1.20	
10240	11101	4	0.98	0.53	1.42		1.40	0.95	1.85	B+
14774	11101	5	-0.41	-0.87	0.04		-0.02	-0.47	0.44	
14784	11101	7	-0.55	-1.00	-0.09		-0.43	-0.89	0.03	
10196	11101	10	-0.17	-0.62	0.28		0.03	-0.41	0.48	
EQME31814	11101	11	-0.31	-0.89	0.27		0.35	-0.22	0.92	
14802	11101	12	0.24	-0.21	0.70		0.44	-0.01	0.88	
EQME10508	11101	14	1.16	0.71	1.60	B+	0.69	0.24	1.13	
EQME31796	11101	15	-0.12	-0.57	0.33		0.62	0.17	1.07	
EQME21490	11101	16	0.17	-0.29	0.63		0.48	0.02	0.93	
10176	11101	17	1.75	1.25	2.24	C+	1.84	1.35	2.34	C+
10103	11101	18	0.55	0.10	1.00		0.46	0.01	0.90	
10135	11101	24	-0.07	-0.61	0.47		0.24	-0.29	0.76	
10104	11101	25	-1.12	-1.65	-0.60	B-	-0.70	-1.21	-0.19	
10158	11101	26	0.34	-0.12	0.79		0.00	-0.46	0.46	
EQME21446	11101	28	0.74	0.23	1.25		0.47	-0.06	0.99	
14785	11101	NR	-0.54	-0.99	-0.08		-0.94	-1.40	-0.47	
14811	11101	NR	-0.13	-0.55	0.30		-0.04	-0.46	0.39	
14827	11101	NR	0.14	-0.30	0.58		0.01	-0.43	0.45	
14831	11101	NR	0.22	-0.32	0.76		0.17	-0.37	0.72	
14851	11101	NR	-0.12	-0.59	0.34		-0.48	-0.94	-0.02	
14857	11101	NR	-0.01	-0.46	0.44		-0.12	-0.57	0.32	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

Table 7.1.65b Gender-Based DIF Statistics for Open-Response 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
14791	11101	8	-0.18	0.00	B-	-0.14	0.00	
10164	11101	21	-0.01	0.27		-0.13	0.00	
14870	11101	31	0.09	0.05		0.00	0.07	
14797	11101	NR	0.14	0.00		0.20	0.00	B+
14843	11101	NR	-0.13	0.00		-0.10	0.00	
14844	11101	NR	0.10	0.04		0.12	0.00	
14864	11101	NR	0.10	0.01		0.15	0.00	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

Table 7.1.66a Gender-Based DIF Statistics for Multiple-Choice 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
10236	11201	1	-0.13	-0.67	0.40		-0.42	-0.96	0.12	
10150	11201	2	0.65	0.20	1.10		0.31	-0.14	0.77	
10127	11201	3	1.21	0.75	1.67	B+	0.89	0.43	1.35	
10240	11201	4	1.29	0.84	1.75	B+	1.19	0.74	1.65	B+
14779	11201	6	0.00	-0.44	0.44		-0.65	-1.10	-0.21	
10196	11201	10	-0.03	-0.50	0.44		-0.12	-0.58	0.34	
14825	11201	13	0.65	0.17	1.13		0.67	0.19	1.14	
EQME10508	11201	14	0.98	0.52	1.43		0.43	-0.02	0.88	
EQME21490	11201	16	0.38	-0.09	0.84		0.63	0.17	1.10	
10103	11201	18	0.39	-0.07	0.86		0.60	0.14	1.05	
14826	11201	19	0.61	0.08	1.13		1.43	0.87	1.99	B+
14833	11201	20	-0.28	-0.74	0.19		0.30	-0.16	0.76	
10135	11201	24	0.41	-0.10	0.93		-0.13	-0.64	0.38	
10104	11201	25	-1.07	-1.59	-0.54	B-	-1.11	-1.65	-0.57	B-
10158	11201	26	-0.03	-0.50	0.44		0.01	-0.46	0.48	
14856	11201	27	0.04	-0.43	0.50		-0.28	-0.74	0.18	
EQME21446	11201	28	0.58	0.07	1.09		0.37	-0.13	0.87	
14858	11201	29	0.71	0.27	1.15		0.64	0.21	1.08	
EQME21477	11201	NR	-0.42	-0.90	0.07		-0.14	-0.62	0.35	
EQME31752	11201	NR	0.15	-0.28	0.59		-0.09	-0.53	0.34	
14803	11201	NR	0.68	0.17	1.19		0.97	0.47	1.48	
14808	11201	NR	-0.23	-0.70	0.24		-0.21	-0.69	0.26	
14816	11201	NR	-0.45	-0.92	0.03		-0.86	-1.33	-0.40	
14822	11201	NR	0.67	0.22	1.11		0.43	-0.02	0.88	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

Table 7.1.66b Gender-Based DIF Statistics for Open-Response 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
14793	11201	9	0.02	0.93		0.07	0.27	
14840	11201	22	0.06	0.04		0.02	0.07	
14847	11201	23	0.23	0.00	B+	0.13	0.00	
14863	11201	30	0.06	0.01		-0.02	0.49	
14788	11201	NR	-0.02	0.81		0.06	0.04	
14837	11201	NR	-0.03	0.19		0.06	0.06	
14871	11201	NR	-0.06	0.00		-0.01	0.84	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

Table 7.1.67a Gender-Based DIF Statistics for Multiple-Choice 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
10278	12101	1	-0.07	-0.61	0.46		-0.45	-0.99	0.09	
EQME10477	12101	3	1.44	0.91	1.98	B+	0.93	0.40	1.46	
EQME21249	12101	4	-0.42	-0.88	0.05		0.22	-0.25	0.69	
EQME31889	12101	5	0.99	0.41	1.58		0.80	0.22	1.37	
EQME21223	12101	7	0.30	-0.16	0.76		0.69	0.23	1.14	
10283	12101	8	0.09	-0.36	0.53		0.02	-0.42	0.47	
10179	12101	9	-0.32	-0.81	0.17		0.13	-0.35	0.61	
14896	12101	11	-0.19	-0.90	0.53		0.04	-0.66	0.75	
EQME10321	12101	15	-1.01	-1.60	-0.42	B-	-0.69	-1.31	-0.08	
EQME10437	12101	16	-0.22	-0.84	0.40		-0.38	-1.01	0.26	
14920	12101	17	-0.55	-1.03	-0.08		-0.85	-1.32	-0.38	
10180	12101	19	0.04	-0.43	0.51		0.14	-0.33	0.62	
14932	12101	20	-0.78	-1.45	-0.10		-0.20	-0.87	0.47	
10312	12101	24	0.56	0.05	1.07		0.32	-0.18	0.82	
EQME21204	12101	25	0.39	-0.11	0.89		0.10	-0.40	0.59	
14953	12101	26	0.24	-0.32	0.81		0.00	-0.56	0.56	
10313	12101	27	0.19	-0.29	0.67		0.74	0.25	1.23	
14961	12101	29	0.46	-0.05	0.97		0.73	0.23	1.23	
14875	12101	NR	0.27	-0.22	0.75		0.52	0.05	1.00	
EQME31994	12101	NR	-0.32	-1.04	0.39		-0.01	-0.73	0.71	
14895	12101	NR	-0.81	-1.33	-0.29		-0.59	-1.11	-0.06	
14919	12101	NR	0.00	-0.57	0.57		0.12	-0.45	0.68	
EQME31884	12101	NR	0.75	0.27	1.24		0.86	0.37	1.35	
14957	12101	NR	0.46	-0.15	1.08		0.26	-0.34	0.86	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

Table 7.1.67b Gender-Based DIF Statistics for Open-Response 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
10289	12101	6	-0.12	0.00		-0.11	0.00	
14910	12101	14	-0.13	0.00		-0.08	0.06	
14945	12101	23	0.08	0.00		0.07	0.01	
14963	12101	30	0.03	0.08		0.02	0.06	
14909	12101	NR	0.17	0.00		0.17	0.00	
14939	12101	NR	0.05	0.13		0.03	0.50	
14969	12101	NR	0.00	0.96		-0.02	0.13	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

Table 7.1.68a Gender-Based DIF Statistics for Multiple-Choice 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
10278	12201	1	-0.38	-0.94	0.19		-0.08	-0.64	0.47	
14877	12201	2	-0.07	-0.56	0.42		-0.29	-0.77	0.18	
EQME10477	12201	3	0.94	0.39	1.49		0.81	0.26	1.37	
EQME21249	12201	4	-0.60	-1.08	-0.12		0.05	-0.42	0.52	
EQME21223	12201	7	0.62	0.17	1.08		0.24	-0.22	0.69	
10283	12201	8	0.06	-0.40	0.52		0.00	-0.45	0.45	
10179	12201	9	-0.07	-0.60	0.46		-0.39	-0.90	0.13	
EQME31904	12201	10	-0.09	-0.66	0.49		-0.55	-1.13	0.03	
14900	12201	12	0.17	-0.32	0.66		0.09	-0.39	0.58	
EQME10321	12201	15	-0.74	-1.37	-0.10		-0.74	-1.41	-0.06	
EQME10437	12201	16	-0.77	-1.44	-0.10		-0.99	-1.67	-0.32	
14918	12201	18	-0.53	-1.02	-0.05		-0.27	-0.75	0.21	
10180	12201	19	0.47	-0.02	0.96		0.24	-0.26	0.73	
14933	12201	21	-0.20	-0.68	0.28		-0.50	-0.98	-0.02	
10312	12201	24	0.28	-0.26	0.82		0.98	0.45	1.51	
EQME21204	12201	25	0.33	-0.17	0.83		0.24	-0.27	0.75	
10313	12201	27	0.31	-0.21	0.82		0.89	0.37	1.41	
EQME31931	12201	28	-0.06	-0.82	0.71		0.61	-0.18	1.41	
14881	12201	NR	-0.80	-1.37	-0.23		-0.48	-1.03	0.06	
EQME10310	12201	NR	1.40	0.75	2.06	B+	1.35	0.72	1.99	B+
10154	12201	NR	0.24	-0.24	0.72		-0.15	-0.63	0.33	
14927	12201	NR	-0.24	-0.91	0.43		-0.57	-1.24	0.10	
14948	12201	NR	0.63	0.16	1.09		0.62	0.15	1.08	
EQME31966	12201	NR	0.37	-0.31	1.11		0.39	-0.29	1.08	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

Table 7.1.68b Gender-Based DIF Statistics for Open-Response 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
14908	12201	13	0.05	0.30		0.05	0.39	
10322	12201	22	0.13	0.00		0.11	0.00	
14971	12201	31	0.02	0.92		0.02	0.27	
14884	12201	NR	0.01	0.50		0.00	0.19	
14911	12201	NR	-0.17	0.00		-0.20	0.00	B-
14947	12201	NR	-0.04	0.75		-0.07	0.11	
14965	12201	NR	0.08	0.00		0.06	0.03	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

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

Item Code	Booklet	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
CMS9700	21101	2	0.22	-0.71	1.16	
9723	21101	3	-0.55	-1.43	0.32	
OQMF31752	21101	4	-1.33	-2.24	-0.41	B-
18450	21101	5	-1.38	-2.31	-0.44	B-
CMS10021	21101	6	0.63	-0.27	1.53	
OQMF31770	21101	10	0.19	-0.73	1.11	
CMS9887	21101	11	0.06	-0.86	0.98	
OQMF10298	21101	12	1.30	0.40	2.19	B+
16512	21101	13	0.28	-0.63	1.19	
CMS10031	21101	15	0.14	-0.82	1.11	
OQMF31783	21101	16	1.08	0.18	1.97	B+
OQMF21399	21101	21	0.71	-0.22	1.64	
OQMF31807	21101	22	-0.24	-1.12	0.63	
OQMF31806	21101	23	0.91	-0.04	1.87	
OQMF31826	21101	24	0.68	-0.22	1.57	
18455	21101	25	0.93	0.05	1.81	
14449	21101	26	0.14	-0.76	1.03	
16465	21101	NR	1.31	0.38	2.23	B+
9725	21101	NR	0.93	-0.07	1.93	
14407	21101	NR	1.18	0.27	2.10	B+
14413	21101	NR	-0.28	-1.21	0.64	
14423	21101	NR	0.54	-0.40	1.47	
OQMF31850	21101	NR	0.63	-0.28	1.54	
16467	21101	NR	0.55	-0.36	1.47	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

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

Item Code	Booklet	Sequence	Sample 1		
			Effect Size	p-Value	DIF Level
16517	21101	9	0.07	0.51	
CMS9759	21101	19	0.08	0.06	
OQMF31839	21101	28	-0.01	0.56	
16519	21101	NR	0.13	0.06	
16546	21101	NR	0.02	0.18	
16910	21101	NR	-0.01	0.55	
16933	21101	NR	0.09	0.67	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

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

Item Code	Booklet	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
18454	21201	1	0.42	-0.22	1.06	
CMS9700	21201	2	0.30	-0.35	0.95	
OQMF31752	21201	4	-0.37	-1.00	0.27	
CMS10021	21201	6	0.93	0.28	1.58	
14382	21201	7	-1.03	-1.74	-0.33	B-
OQMF31770	21201	10	-0.72	-1.42	-0.03	
CMS9887	21201	11	1.02	0.33	1.71	B+
9989	21201	14	-0.15	-0.84	0.53	
CMS10031	21201	15	0.28	-0.42	0.97	
OQMF31783	21201	16	1.17	0.51	1.83	B+
14412	21201	17	0.06	-0.57	0.69	
18460	21201	20	0.23	-0.45	0.91	
OQMF21399	21201	21	-0.30	-0.98	0.38	
OQMF31807	21201	22	-0.35	-0.98	0.28	
OQMF31806	21201	23	-0.51	-1.18	0.15	
OQMF31826	21201	24	-0.22	-0.93	0.49	
14449	21201	26	-0.14	-0.79	0.52	
14372	21201	NR	-0.97	-1.64	-0.29	
14369	21201	NR	-0.54	-1.31	0.23	
OQMF10299	21201	NR	0.57	-0.06	1.20	
14397	21201	NR	-0.68	-1.36	0.00	
9907	21201	NR	-0.22	-0.88	0.44	
OQMF31850	21201	NR	0.12	-0.57	0.81	
18467	21201	NR	0.60	-0.06	1.27	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

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

Item Code	Booklet	Sequence	Sample 1		
			Effect Size	p-Value	DIF Level
CMS9626	21201	8	-0.15	0.03	
OQMF31791	21201	18	-0.03	0.34	B+
OQMF21395	21201	27	0.03	0.38	
12839	21201	NR	0.12	0.02	
9927	21201	NR	-0.07	0.00	
14418	21201	NR	-0.16	0.00	
14456	21201	NR	0.01	0.76	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

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

Item Code	Booklet	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
CMS9657	22101	1	-0.11	-0.64	0.43	
17060	22101	2	-0.71	-1.29	-0.13	
OQMF10391	22101	4	0.13	-0.42	0.69	
OQMF20950	22101	5	0.93	0.39	1.47	
14487	22101	6	0.11	-0.49	0.71	
CMS9660	22101	10	1.35	0.73	1.97	B+
CMS9947	22101	11	0.72	0.17	1.27	
9789	22101	12	0.48	-0.05	1.01	
CMS9806	22101	14	1.17	0.67	1.67	B+
OQMF10331	22101	17	0.20	-0.36	0.75	
CMS10054	22101	18	0.06	-0.50	0.62	
16949	22101	19	-0.47	-1.03	0.10	
CMS9669	22101	23	0.19	-0.37	0.75	
OQMF21008	22101	24	-0.18	-0.73	0.37	
14528	22101	25	0.04	-0.49	0.58	
14539	22101	26	-0.82	-1.40	-0.24	
CMS9694	22101	28	0.30	-0.33	0.94	
14470	22101	NR	-0.43	-1.11	0.26	
14495	22101	NR	-1.04	-1.73	-0.35	B-
17062	22101	NR	0.43	-0.14	1.01	
14509	22101	NR	-0.64	-1.19	-0.08	
9668	22101	NR	0.20	-0.31	0.70	
18462	22101	NR	0.39	-0.14	0.92	
9672	22101	NR	0.15	-0.39	0.69	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

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

Item Code	Booklet	Sequence	Sample 1		
			Effect Size	p-Value	DIF Level
14493	22101	8	0.04	0.63	
14500	22101	16	-0.21	0.00	B-
17064	22101	21	0.17	0.00	
CMS10066	22101	31	0.15	0.00	
9967	22101	NR	-0.07	0.31	
10064	22101	NR	0.09	0.00	
OQMF10361	22101	NR	-0.07	0.31	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

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

Item Code	Booklet	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
CMS9657	22201	1	0.07	-0.35	0.50	
10048	22201	3	-0.55	-0.98	-0.13	
OQMF10391	22201	4	-0.77	-1.21	-0.33	
OQMF20950	22201	5	0.91	0.50	1.33	
9658	22201	9	-0.19	-0.61	0.24	
CMS9660	22201	10	1.60	1.10	2.10	C+
CMS9947	22201	11	0.07	-0.35	0.50	
9948	22201	13	0.07	-0.39	0.52	
CMS9806	22201	14	1.43	1.04	1.83	B+
OQMF10331	22201	17	0.10	-0.34	0.54	
CMS10054	22201	18	-0.05	-0.51	0.42	
14520	22201	20	0.77	0.37	1.16	
CMS9669	22201	23	0.22	-0.22	0.67	
OQMF21008	22201	24	0.24	-0.22	0.71	
14548	22201	27	0.04	-0.37	0.45	
CMS9694	22201	28	0.25	-0.29	0.79	
9842	22201	29	0.06	-0.36	0.48	
18448	22201	NR	-0.40	-0.85	0.05	
14480	22201	NR	0.13	-0.49	0.75	
9662	22201	NR	1.32	0.90	1.73	B+
14511	22201	NR	-0.42	-0.83	-0.01	
18466	22201	NR	-0.85	-1.26	-0.45	
16939	22201	NR	1.37	0.85	1.89	B+
17063	22201	NR	-0.19	-0.63	0.25	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

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

Item Code	Booklet	Sequence	Sample 1		
			Effect Size	p-Value	DIF Level
OQMF20902	22201	7	0.08	0.00	
CMS9968	22201	15	-0.05	0.10	
14524	22201	22	0.04	0.33	
14564	22201	30	0.12	0.00	
17065	22201	NR	0.13	0.00	
OQMF31900	22201	NR	-0.09	0.04	
14566	22201	NR	0.06	0.00	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students. NR = not released.

Table 7.1.73a SLL-Based DIF Statistics for Multiple-Choice 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
10236	11101	1	-0.7	-1.2	-0.3		-0.96	-1.44	-0.47	
10150	11101	2	0.5	0.0	0.9		0.44	0.03	0.85	
10127	11101	3	1.6	1.1	2.0	C+	1.58	1.16	1.99	C+
10240	11101	4	0.4	0.0	0.8		0.14	-0.28	0.56	
14774	11101	5	-1.2	-1.6	-0.8	B-	-1.35	-1.78	-0.91	B-
14784	11101	7	-0.3	-0.7	0.2		-0.09	-0.51	0.34	
10196	11101	10	-0.8	-1.3	-0.4		-1.01	-1.44	-0.59	B-
EQME31814	11101	11	0.6	0.1	1.1		0.92	0.41	1.42	
14802	11101	12	0.3	-0.1	0.7		0.26	-0.16	0.69	
EQME10508	11101	14	0.2	-0.2	0.7		0.30	-0.12	0.71	
EQME31796	11101	15	0.0	-0.5	0.4		-0.06	-0.48	0.37	
EQME21490	11101	16	-0.2	-0.6	0.2		-0.06	-0.49	0.36	
10176	11101	17	-0.4	-0.8	0.1		-0.09	-0.54	0.35	
10103	11101	18	0.0	-0.4	0.4		-0.22	-0.64	0.20	
10135	11101	24	0.8	0.3	1.2		0.85	0.38	1.32	
10104	11101	25	-1.3	-1.8	-0.9	B-	-1.06	-1.55	-0.57	B-
10158	11101	26	0.4	-0.1	0.8		0.46	0.02	0.90	
EQME21446	11101	28	0.1	-0.3	0.6		0.14	-0.32	0.60	
14785	11101	NR	-0.3	-0.7	0.2		-0.28	-0.71	0.15	
14811	11101	NR	-1.0	-1.5	-0.6	B-	-0.87	-1.27	-0.47	
14827	11101	NR	0.3	-0.1	0.7		0.27	-0.14	0.68	
14831	11101	NR	1.7	1.2	2.2	C+	1.61	1.15	2.08	C+
14851	11101	NR	0.5	0.1	1.0		0.29	-0.14	0.71	
14857	11101	NR	-0.8	-1.2	-0.4		-0.76	-1.18	-0.34	

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

Table 7.1.73b SLL-Based DIF Statistics for Open-Response 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
14791	11101	8	0.06	0.00		0.07	0.07	
10164	11101	21	-0.12	0.00		-0.16	0.00	
14870	11101	31	0.13	0.00		0.10	0.00	
14797	11101	NR	0.23	0.00	B+	0.21	0.00	B+
14843	11101	NR	-0.18	0.00	B-	-0.17	0.00	B-
14844	11101	NR	-0.10	0.07		-0.10	0.02	
14864	11101	NR	-0.04	0.13		-0.06	0.13	

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

Table 7.1.74a SLL-Based DIF Statistics for Multiple-Choice 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
10236	11201	1	-0.89	-1.29	-0.48		-0.91	-1.31	-0.50	
10150	11201	2	0.89	0.55	1.23		0.48	0.14	0.82	
10127	11201	3	1.46	1.12	1.80	B+	1.66	1.32	2.00	C+
10240	11201	4	0.27	-0.07	0.61		-0.03	-0.37	0.31	
14779	11201	6	-1.92	-2.26	-1.58	C-	-2.09	-2.44	-1.75	C-
10196	11201	10	-0.51	-0.87	-0.16		-0.56	-0.92	-0.21	
14825	11201	13	0.70	0.35	1.04		0.77	0.42	1.12	
EQME10508	11201	14	0.09	-0.25	0.44		0.39	0.04	0.73	
EQME21490	11201	16	-0.25	-0.60	0.10		-0.50	-0.85	-0.14	
10103	11201	18	-0.16	-0.51	0.19		-0.19	-0.54	0.16	
14826	11201	19	0.96	0.59	1.34		1.20	0.82	1.57	B+
14833	11201	20	0.15	-0.21	0.51		0.35	-0.01	0.71	
10135	11201	24	0.90	0.52	1.27		1.08	0.70	1.45	B+
10104	11201	25	-1.35	-1.75	-0.95	B-	-1.25	-1.65	-0.85	B-
10158	11201	26	0.15	-0.21	0.51		0.20	-0.16	0.57	
14856	11201	27	-0.20	-0.55	0.15		0.02	-0.33	0.37	
EQME21446	11201	28	-0.12	-0.50	0.26		-0.13	-0.51	0.25	
14858	11201	29	0.45	0.12	0.79		0.44	0.10	0.77	
EQME21477	11201	NR	0.05	-0.31	0.42		0.07	-0.29	0.44	
EQME31752	11201	NR	-0.77	-1.11	-0.44		-0.66	-0.99	-0.32	
14803	11201	NR	1.21	0.84	1.57	B+	0.97	0.61	1.34	
14808	11201	NR	0.28	-0.06	0.62		0.59	0.25	0.93	
14816	11201	NR	0.45	0.11	0.80		0.43	0.09	0.78	
14822	11201	NR	-0.21	-0.55	0.14		-0.14	-0.49	0.21	

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

Table 7.1.74b SLL-Based DIF Statistics for Open-Response 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
14793	11201	9	0.05	0.18		0.08	0.00	
14840	11201	22	0.05	0.02		0.04	0.00	
14847	11201	23	0.04	0.01		0.02	0.72	
14863	11201	30	0.13	0.00		0.15	0.00	
14788	11201	NR	-0.04	0.00		-0.08	0.00	
14837	11201	NR	-0.09	0.01		-0.07	0.11	
14871	11201	NR	0.06	0.00		0.05	0.00	

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

Table 7.1.75a SLL-Based DIF Statistics for Multiple-Choice 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
10278	12101	1	-0.97	-1.39	-0.54		-0.89	-1.32	-0.47	
EQME10477	12101	3	-0.01	-0.39	0.36		-0.12	-0.49	0.25	
EQME21249	12101	4	-1.48	-1.84	-1.13	B-	-1.47	-1.83	-1.11	B-
EQME31889	12101	5	0.87	0.46	1.27		1.22	0.81	1.63	B+
EQME21223	12101	7	0.72	0.39	1.04		0.92	0.59	1.24	
10283	12101	8	1.10	0.78	1.42	B+	1.14	0.82	1.47	B+
10179	12101	9	-0.34	-0.69	0.02		-0.04	-0.40	0.32	
14896	12101	11	0.42	-0.07	0.91		0.12	-0.36	0.61	
EQME10321	12101	15	-0.47	-0.92	-0.03		-0.54	-0.98	-0.10	
EQME10437	12101	16	-1.07	-1.55	-0.59	B-	-1.06	-1.54	-0.58	B-
14920	12101	17	-0.38	-0.73	-0.04		-0.49	-0.83	-0.15	
10180	12101	19	0.19	-0.15	0.54		0.26	-0.08	0.60	
14932	12101	20	0.80	0.34	1.25		0.89	0.43	1.34	
10312	12101	24	-0.37	-0.74	0.01		-0.41	-0.79	-0.03	
EQME21204	12101	25	-0.32	-0.68	0.04		-0.44	-0.80	-0.09	
14953	12101	26	0.60	0.20	0.99		0.54	0.15	0.94	
10313	12101	27	-0.20	-0.56	0.16		0.14	-0.22	0.49	
14961	12101	29	0.06	-0.31	0.42		-0.09	-0.45	0.28	
14875	12101	NR	-1.19	-1.54	-0.84	B-	-1.36	-1.72	-1.01	B-
EQME31994	12101	NR	0.18	-0.34	0.70		0.10	-0.41	0.62	
14895	12101	NR	-0.90	-1.29	-0.51		-1.18	-1.57	-0.79	B-
14919	12101	NR	-0.43	-0.84	-0.01		-0.38	-0.79	0.03	
EQME31884	12101	NR	0.73	0.39	1.07		0.83	0.49	1.18	
14957	12101	NR	0.06	-0.38	0.50		0.16	-0.29	0.60	

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

Table 7.1.75b SLL-Based DIF Statistics for Open-Response 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
10289	12101	6	-0.05	0.34		-0.02	0.81	
14910	12101	14	0.07	0.00		0.02	0.19	
14945	12101	23	-0.25	0.00	B-	-0.25	0.00	C-
14963	12101	30	0.05	0.06		0.10	0.00	
14909	12101	NR	0.05	0.12		0.05	0.14	
14939	12101	NR	-0.05	0.01		-0.03	0.00	
14969	12101	NR	0.00	0.13		-0.02	0.21	

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

Table 7.1.76a SLL-Based DIF Statistics for Multiple-Choice 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
10278	12201	1	-1.35	-1.68	-1.03	B-	-1.20	-1.53	-0.87	B-
14877	12201	2	-1.19	-1.46	-0.91	B-	-1.16	-1.44	-0.89	B-
EQME10477	12201	3	-0.28	-0.58	0.02		-0.38	-0.68	-0.08	
EQME21249	12201	4	-1.20	-1.48	-0.93	B-	-1.16	-1.43	-0.89	B-
EQME21223	12201	7	0.49	0.25	0.74		0.22	-0.02	0.47	
10283	12201	8	0.80	0.55	1.05		0.73	0.48	0.98	
10179	12201	9	-0.01	-0.29	0.27		0.23	-0.06	0.51	
EQME31904	12201	10	-1.00	-1.33	-0.67		-0.89	-1.22	-0.56	
14900	12201	12	-0.62	-0.89	-0.36		-0.65	-0.92	-0.39	
EQME10321	12201	15	-0.54	-0.91	-0.17		-0.50	-0.87	-0.13	
EQME10437	12201	16	-1.96	-2.35	-1.56	C-	-1.80	-2.19	-1.40	C-
14918	12201	18	-0.72	-0.98	-0.45		-0.78	-1.05	-0.51	
10180	12201	19	0.21	-0.06	0.48		0.09	-0.18	0.35	
14933	12201	21	-0.32	-0.58	-0.05		-0.28	-0.55	-0.02	
10312	12201	24	-0.36	-0.66	-0.06		-0.64	-0.93	-0.34	
EQME21204	12201	25	-0.34	-0.62	-0.07		-0.20	-0.48	0.07	
10313	12201	27	-0.23	-0.51	0.05		-0.22	-0.50	0.06	
EQME31931	12201	28	0.18	-0.22	0.59		0.50	0.09	0.91	
14881	12201	NR	-0.46	-0.77	-0.15		-0.41	-0.72	-0.10	
EQME10310	12201	NR	1.67	1.35	1.99	C+	1.47	1.16	1.79	B+
10154	12201	NR	-0.47	-0.73	-0.21		-0.63	-0.89	-0.37	
14927	12201	NR	-0.67	-1.05	-0.29		-0.70	-1.07	-0.33	
14948	12201	NR	0.67	0.42	0.93		0.77	0.51	1.03	
EQME31966	12201	NR	-0.16	-0.53	0.21		-0.05	-0.42	0.31	

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

Table 7.1.76b SLL-Based DIF Statistics for Open-Response 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
14908	12201	13	-0.10	0.00		-0.12	0.00	
10322	12201	22	-0.04	0.01		-0.06	0.00	
14971	12201	31	0.09	0.00		0.07	0.00	
14884	12201	NR	-0.05	0.04		-0.02	0.24	
14911	12201	NR	-0.17	0.00		-0.17	0.00	B-
14947	12201	NR	-0.12	0.00		-0.12	0.00	
14965	12201	NR	0.01	0.24		0.03	0.01	

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
18158_269	n/a	1	R2	2	0.89	0.25	-2.00
18159_269	n/a	2	R1	3	0.95	0.30	-2.83
18160_269	n/a	3	R1	2	0.79	0.40	-1.10
18161_269	n/a	4	R3	1	0.95	0.28	-2.96
18162_269	n/a	5	R2	3	0.49	0.27	0.75
18163_269	n/a	6	R3	3*	0.83 (2.50)	0.36	-2.22
17970	n/a	7	W1	3	0.77	0.23	-0.94
17980	n/a	8	W2	3	0.54	0.25	0.44
17996	n/a	9	W3	2	0.93	0.27	-2.67
17995	n/a	10	W3	4	0.48	0.31	0.77
18022_T	n/a	11	W4	3*	0.74 (2.23)	0.36	-0.91
18022_V	n/a	11	W3	2*	0.92 (1.85)	0.32	-1.72
19840_T	n/a	12	W4	6*	0.71 (4.26)	0.54	-0.93
19840_V	n/a	12	W3	4*	0.85 (3.39)	0.55	-1.86
17236_265	n/a	13	R2	2	0.82	0.37	-1.37
17246_265	n/a	14	R1	1	0.70	0.38	-0.54
17357_265	n/a	15	R3	3	0.57	0.39	0.25
18065_265	n/a	16	R2	1	0.69	0.26	-0.38
18072_265	n/a	17	R1	2	0.69	0.41	-0.50
18075_265	n/a	18	R2	1	0.51	0.23	0.65
17237_265	n/a	19	R2	3*	0.62 (1.87)	0.32	-0.63
18378_T	I	1	W4	6*	0.76 (4.54)	0.56	-1.29
18378_V	I	1	W3	4*	0.82 (3.26)	0.52	-1.89
17968	II	1	W1	3	0.75	0.28	-0.82
17977	II	2	W2	3	0.85	0.32	-1.55
17992	II	3	W3	3	0.56	0.32	0.31
18002	II	4	W3	2	0.88	0.34	-2.04
18068_255	III	1	R2	2	0.72	0.40	-0.65
18069_255	III	2	R1	3	0.63	0.39	-0.12
18070_255	III	3	R3	3	0.54	0.33	0.40
18071_255	III	4	R2	3	0.70	0.32	-0.48
18073_255	III	5	R3	2	0.67	0.38	-0.34

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.

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

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

() = mean score for open-response items.

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
18074_255	III	6	R2	2	0.83	0.32	-1.41
18076_255	III	7	R1	3	0.90	0.32	-2.09
18077_255	III	8	R2	2	0.75	0.35	-0.81
18078_255	III	9	R2	2	0.72	0.34	-0.70
13495_262	IV	1	R2	3	0.79	0.29	-1.02
13494_262	IV	2	R2	4	0.74	0.37	-0.74
13508_262	IV	3	R2	1	0.90	0.22	-2.11
13502_262	IV	4	R2	3	0.80	0.43	-1.19
13492_262	IV	5	R3	1	0.61	0.36	0.01
13498_262	IV	6	R2	3*	0.75 (2.25)	0.51	-1.10
13499_262	IV	7	R3	3*	0.73 (2.19)	0.53	-0.97
18021_T	V	1	W4	3*	0.87 (2.60)	0.42	-1.81
18021_V	V	1	W3	2*	0.97 (1.95)	0.26	-2.61
18052_254_21	VI	1	R1	2	0.81	0.33	-1.26
18054_254_21	VI	2	R3	4	0.73	0.36	-0.74
18056_254_21	VI	3	R2	1	0.73	0.35	-0.66
19524_254	VI	4	R2	2	0.87	0.34	-1.77
18059_254_21	VI	5	R2	1	0.80	0.26	-1.13
18061_254_21	VI	6	R2	3	0.63	0.19	0.02

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.

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

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

() = mean score for open-response items.

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

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1	Score 2	Score 3
18163_269	n/a	6	% of Students	0.23	0.18	0.01	3.95	40.65	54.98
			Parameters				-3.73	-3.03	-1.98
18022_T	n/a	11	% of Students	4.74	2.09	0.08	15.19	26.10	51.80
			Parameters				-2.12	-1.21	0.03
18022_V	n/a	11	% of Students	4.74	2.09	0.08	1.36	91.74	
			Parameters				-2.01	1.16	
17237_265	n/a	19	% of Students	1.60	3.55	0.09	31.31	34.76	28.70
			Parameters				-2.18	-1.44	0.29
13498_262	IV	6	% of Students	2.03	1.86	0.06	10.70	41.97	43.38
			Parameters				-2.42	-1.61	-0.64
13499_262	IV	7	% of Students	2.72	2.24	0.05	11.20	43.88	39.91
			Parameters				-2.28	-1.47	-0.55
18021_T	V	1	% of Students	0.27	1.45	0.04	1.62	31.25	65.37
			Parameters				-2.88	-1.99	-1.53
18021_V	V	1	% of Students	0.27	1.45	0.04	1.77	96.47	
			Parameters				-3.06	-0.59	

Note. The total number of students is 136 949.

Table 7.1.79 Distribution of Score Points and Category Difficulty Estimates 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
19840_T	n/a	12	% of Students	0.03	0.52	0.03	0.52	0.67	1.61	2.46
			Parameters				-2.81*		-2.38	-1.90
19840_V	n/a	12	% of Students	0.00	0.52	0.03	0.26	0.40	2.42	7.45
			Parameters				-3.21*		-2.64	-2.35
18378_T	I	1	% of Students	0.02	0.36	0.03	0.18	0.18	0.84	1.63
			Parameters				-3.00*		-2.58	-2.30
18378_V	I	1	% of Students	0.00	0.36	0.03	0.23	0.33	2.69	9.09
			Parameters				-3.42*		-2.90	-2.58
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
19840_T	n/a	12	% of Students	4.82	12.45	25.04	23.62	13.09	10.52	4.62
			Parameters	-1.64	-1.44	-1.22	-0.90	-0.37	0.30	1.20
19840_V	n/a	12	% of Students	23.94	34.38	30.60				
			Parameters	-2.05	-1.63	-0.55				
18378_T	I	1	% of Students	4.61	8.43	16.86	23.14	22.11	15.55	6.07
			Parameters	-2.07	-1.87	-1.60	-1.30	-0.87	-0.26	0.84
18378_V	I	1	% of Students	34.86	33.34	19.06				
			Parameters	-2.22	-1.72	-0.30				

Note. The total number of students is 136 949.

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

Table 7.1.80 Item Statistics: TPCL (French)

Item Code	Section	Sequence	Cognitive Skill	Answer Key/Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
13196_232	s. o.	1	R1	3	0.97	0.20	-3.24
13198_232	s. o.	2	R2	4	0.82	0.38	-1.30
13201_232	s. o.	3	R1	1	0.81	0.27	-1.11
13202_232	s. o.	4	R2	1	0.89	0.22	-1.83
13200_232	s. o.	5	R3	3	0.97	0.22	-3.42
18325_232	s. o.	6	R3	3*	0.80 (2.39)	0.16	-2.52
13294	s. o.	7	W1	4	0.75	0.24	-0.70
13305	s. o.	8	W3	4	0.92	0.18	-2.33
13289	s. o.	9	W2	4	0.79	0.26	-0.97
13300	s. o.	10	W3	3	0.84	0.23	-1.36
13278_T	s. o.	11	W4	3*	0.89 (2.66)	0.42	-1.91
13278_V	s. o.	11	W3	2*	0.97 (1.95)	0.23	-2.75
19077_T	s. o.	12	W4	6*	0.67 (4.01)	0.54	-0.77
19077_V	s. o.	12	W3	4*	0.72 (2.86)	0.55	-1.30
13158_227	s. o.	13	R3	2	0.92	0.31	-2.32
13167_227	s. o.	14	R2	3	0.42	0.23	1.32
13161_227	s. o.	15	R1	2	0.65	0.26	-0.04
13157_227	s. o.	16	R1	1	0.97	0.12	-3.37
13165_227	s. o.	17	R2	2	0.59	0.17	0.34
13168_227	s. o.	18	R2	4	0.92	0.34	-2.29
18181_227	s. o.	19	R2	3*	0.64 (1.90)	0.37	-0.51
19076_T	I	1	W4	6*	0.69 (4.13)	0.59	-0.80
19076_V	I	1	W3	4*	0.68 (2.73)	0.59	-1.10
13293	II	1	W1	4	0.82	0.37	-1.28
13287	II	2	W2	1	0.67	0.24	-0.18
10387	II	3	W3	3	0.97	0.21	-3.51
13306	II	4	W3	4	0.71	0.23	-0.43
13223_235	III	1	R2	4	0.62	0.29	0.08
13218_235	III	2	R1	4	0.96	0.26	-3.08
13228_235	III	3	R2	1	0.79	0.22	-0.97
13230_235	III	4	R3	3	0.73	0.27	-0.60
13217_235	III	5	R1	3	0.70	0.35	-0.43
13196_232	s. o.	1	R1	3	0.97	0.20	-3.24
13198_232	s. o.	2	R2	4	0.82	0.38	-1.30
13201_232	s. o.	3	R1	1	0.81	0.27	-1.11
13202_232	s. o.	4	R2	1	0.89	0.22	-1.83

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.

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

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

() = mean score for open-response items.

Table 7.1.80 Item Statistics: TPCL (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
13221_235	III	6	R2	3	0.86	0.27	-1.52
13225_235	III	7	R3	2	0.76	0.44	-0.85
13229_235	III	8	R2	2	0.77	0.27	-0.81
13227_235	III	9	R2	3	0.79	0.29	-1.01
13233_236	IV	1	R2	2	0.83	0.16	-1.21
13234_236	IV	2	R3	3	0.84	0.18	-1.33
13240_236	IV	3	R2	4	0.86	0.32	-1.69
13237_236	IV	4	R2	1	0.90	0.32	-2.08
13238_236	IV	5	R2	2	0.56	0.33	0.43
13242_236	IV	6	R2	3*	0.74 (2.21)	0.29	-1.50
18328_236	IV	7	R3	3*	0.76 (2.28)	0.44	-1.33
13272_T	V	1	W4	3*	0.91 (2.72)	0.40	-2.03
13272_V	V	1	W3	2*	0.98 (1.95)	0.25	-2.83
13149_225_12	VI	1	R2	3	0.81	0.29	-1.11
13471_225_12	VI	2	R2	2	0.76	0.26	-0.73

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.

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

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

() = mean score for open-response items.

Table 7.1.81 Distribution of Score Points and Category Difficulty Estimates for Open-Response Reading and Short-Writing Tasks: TPCL (French)

Item Code	Section	Sequence		Off Topic	Missing	Illegible	Score 1.0	Score 2.0	Score 3.0
18325_232	s. o.	6	% of Students	0.00	0.08	0.00	0.87	58.68	40.38
			Parameters				-4.00	-3.59	-3.01
13278_T	s. o.	11	% of Students	0.17	0.77	0.08	1.94	26.85	70.19
			Parameters				-3.03	-2.14	-1.52
13278_V	s. o.	11	% of Students	0.17	0.77	0.08	3.05	95.94	
			Parameters				-3.33	-1.22	
18181_227	s. o.	19	% of Students	1.81	2.94	0.08	25.27	44.51	25.40
			Parameters				-2.04	-1.31	0.10
13242_236	IV	6	% of Students	0.41	0.34	0.00	15.13	46.27	37.84
			Parameters				-3.27	-2.64	-0.83
18328_236	IV	7	% of Students	0.72	0.73	0.00	7.58	52.28	38.69
			Parameters				-2.80	-2.07	-1.12
13272_T	V	1	% of Students	0.08	0.73	0.04	1.04	23.65	74.46
			Parameters				-3.07	-2.15	-1.67
13272_V	V	1	% of Students	0.08	0.73	0.04	3.20	95.95	
			Parameters				-3.44	-1.39	

Note. The total number of students is 5314.

Table 7.1.82 Distribution of Score Points and Category Difficulty Estimates for Long-Writing Tasks: TPCL (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	0.17	0.02	0.28	0.51	1.66	2.99
			Parameters				-2.94*		-2.58	-2.05
2É3_030V	IV	1	% of Students	0.00	0.17	0.02	0.98	1.69	12.50	24.78
			Parameters				-3.34*		-2.92	-2.37
2É2_033T	VI	1	% of Students	0.02	0.28	0.09	0.28	0.24	2.16	3.52
			Parameters				-2.64*		-2.25	-1.97
2É2_033V	VI	1	% of Students	0.00	0.28	0.09	2.82	6.72	18.22	22.68
			Parameters				-3.25*		-2.80	-1.76
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	12.38	18.31	25.59	20.66	8.90	6.45	2.07
			Parameters	-1.70	-1.44	-0.99	-0.53	0.03	0.69	1.53
2É3_030V	IV	1	% of Students	34.76	18.10	7.00				
			Parameters	-1.55	-0.65	0.86				
2É2_033T	VI	1	% of Students	10.54	15.32	22.71	19.63	12.12	8.62	4.46
			Parameters	-1.63	-1.36	-0.99	-0.61	-0.12	0.47	1.37
2É2_033V	VI	1	% of Students	20.91	17.99	10.29				
			Parameters	-0.96	-0.24	0.81				

Note. The total number of students is 5314.

*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 assessment are the results from two random samples of 2000 students. For the French-language assessment, gender-based DIF analysis was conducted for one sample of students, but 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 Multiple-Choice 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
18158_269	n/a	1	0.39	-0.09	0.87		0.65	0.16	1.14	
18159_269	n/a	2	1.83	1.04	2.62	C+	2.10	1.32	2.87	C+
18160_269	n/a	3	0.41	0.01	0.81		-0.10	-0.50	0.31	
18161_269	n/a	4	1.09	0.38	1.79	B+	1.11	0.38	1.84	B+
18162_269	n/a	5	0.95	0.63	1.27		0.76	0.44	1.07	
17970	n/a	7	-0.23	-0.59	0.14		-0.32	-0.68	0.05	
17980	n/a	8	-0.22	-0.53	0.09		-0.36	-0.67	-0.05	
17996	n/a	9	0.41	-0.22	1.04		-0.30	-0.91	0.32	
17995	n/a	10	-0.26	-0.57	0.06		-0.69	-1.00	-0.37	
17236_265	n/a	13	2.74	2.30	3.18	C+	2.42	1.97	2.87	C+
17246_265	n/a	14	-0.18	-0.53	0.18		-0.56	-0.92	-0.21	
17357_265	n/a	15	0.66	0.32	1.00		0.45	0.12	0.78	
18065_265	n/a	16	-0.31	-0.64	0.03		-0.44	-0.78	-0.10	
18072_265	n/a	17	0.51	0.16	0.87		0.25	-0.11	0.62	
18075_265	n/a	18	0.21	-0.10	0.52		0.74	0.43	1.05	
17968	II	1	0.51	0.15	0.87		0.50	0.14	0.87	
17977	II	2	-0.28	-0.71	0.16		0.00	-0.44	0.43	
17992	II	3	-1.04	-1.36	-0.72	B-	-1.04	-1.36	-0.72	B-
18002	II	4	-0.41	-0.92	0.09		-0.16	-0.66	0.34	
18068_255	III	1	0.69	0.32	1.05		0.80	0.43	1.17	
18069_255	III	2	1.52	1.16	1.87	C+	1.18	0.83	1.53	B+
18070_255	III	3	0.81	0.49	1.14		0.80	0.48	1.12	
18071_255	III	4	0.63	0.28	0.98		1.12	0.76	1.48	B+
18073_255	III	5	0.10	-0.25	0.44		0.01	-0.34	0.35	
18074_255	III	6	1.59	1.17	2.02	C+	1.73	1.29	2.17	C+
18076_255	III	7	0.10	-0.44	0.63		-0.61	-1.15	-0.07	
18077_255	III	8	0.52	0.14	0.90		0.47	0.10	0.84	
18078_255	III	9	-0.76	-1.12	-0.40		-0.36	-0.72	0.00	
13495_262	IV	1	0.82	0.43	1.20		0.03	-0.35	0.41	
13494_262	IV	2	0.50	0.13	0.87		0.21	-0.16	0.58	
13508_262	IV	3	-0.29	-0.76	0.19		-0.03	-0.54	0.49	
13502_262	IV	4	0.22	-0.19	0.64		-0.17	-0.59	0.24	
13492_262	IV	5	0.64	0.30	0.97		0.75	0.41	1.09	
18052_254_21	VI	1	1.96	1.55	2.37	C+	1.67	1.25	2.09	C+
18054_254_21	VI	2	-0.06	-0.42	0.30		-0.67	-1.04	-0.30	
18056_254_21	VI	3	2.08	1.70	2.45	C+	2.13	1.75	2.50	C+
19524_254	VI	4	0.24	-0.24	0.71		-0.10	-0.57	0.37	
18059_254_21	VI	5	0.80	0.41	1.19		0.82	0.43	1.21	
18061_254_21	VI	6	1.56	1.24	1.88	C+	1.54	1.22	1.87	C+

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

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

Item Code	Section	Sequence	Sample 1			Sample 2		
			Effect Size	p-Value	DIF Level	Effect Size	p-Value	DIF Level
18163_269	n/a	6	-0.06	0.0231		-0.01	0.3266	
17237_265	n/a	19	-0.09	0.0017		-0.09	0.0055	
18022_T	n/a	11	-0.12	0.0000		0.03	0.4082	
18022_V	n/a	11	-0.01	0.7887		0.00	0.3771	
19840_T	n/a	12	-0.14	0.0000		-0.20	0.0000	B-
19840_V	n/a	12	-0.20	0.0000	B-	-0.25	0.0000	B-
18378_T	I	1	-0.27	0.0000	C-	-0.23	0.0000	B-
18378_V	I	1	-0.32	0.0000	C-	-0.29	0.0000	C-
13498_262	IV	6	-0.09	0.0023		-0.05	0.0902	
13499_262	IV	7	-0.07	0.0007		-0.05	0.0402	
18021_T	V	1	-0.10	0.0000		-0.08	0.0173	
18021_V	V	1	-0.02	0.0586		-0.03	0.5678	

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

Table 7.1.84a Gender-Based DIF Statistics for Multiple-Choice Items: TPCL (French)

Item Code	Section	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
13196_232	s. o.	1	-0.01	-0.83	0.82	
13198_232	s. o.	2	2.23	1.77	2.69	C+
13201_232	s. o.	3	0.77	0.37	1.17	
13202_232	s. o.	4	0.09	-0.39	0.58	
13200_232	s. o.	5	1.77	0.88	2.66	B+
13294	s. o.	7	0.13	-0.23	0.49	
13305	s. o.	8	-1.49	-2.08	-0.89	B-
13289	s. o.	9	0.49	0.10	0.88	
13300	s. o.	10	0.37	-0.06	0.79	
13158_227	s. o.	13	0.38	-0.20	0.96	
13167_227	s. o.	14	0.22	-0.10	0.54	
13161_227	s. o.	15	0.59	0.25	0.92	
13157_227	s. o.	16	-0.85	-1.80	0.11	
13165_227	s. o.	17	0.75	0.44	1.07	
13168_227	s. o.	18	0.27	-0.30	0.84	
13293	II	1	0.36	-0.07	0.79	
13287	II	2	0.06	-0.28	0.39	
10387	II	3	-1.36	-2.38	-0.34	B-
13306	II	4	-0.52	-0.86	-0.18	
13223_235	III	1	0.98	0.65	1.31	

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

Table 7.1.84a Gender-Based DIF Statistics for Multiple-Choice Items: TPCL (French) (continued)

Item Code	Section	Sequence	Sample 1			
			Δ	Lower Limit	Upper Limit	DIF Level
13218_235	III	2	1.15	0.35	1.96	B+
13228_235	III	3	0.75	0.37	1.13	
13230_235	III	4	0.11	-0.24	0.46	
13217_235	III	5	0.66	0.30	1.01	
13221_235	III	6	1.29	0.83	1.74	B+
13225_235	III	7	1.38	0.97	1.79	B+
13229_235	III	8	-0.28	-0.65	0.09	
13227_235	III	9	1.03	0.64	1.42	B+
13233_236	IV	1	0.79	0.39	1.19	
13234_236	IV	2	0.58	0.16	0.99	
13240_236	IV	3	-0.69	-1.16	-0.22	
13237_236	IV	4	-0.12	-0.66	0.42	
13238_236	IV	5	1.30	0.97	1.64	B+
13149_225_12	VI	1	-0.31	-0.71	0.10	
13471_225_12	VI	2	1.30	0.92	1.68	B+
13134_225_12	VI	3	1.04	0.66	1.42	B+
13139_225_12	VI	4	-0.09	-0.45	0.26	
13130_225_12	VI	5	0.89	0.55	1.23	
13152_225_12	VI	6	0.67	0.31	1.03	

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

Table 7.1.84b Gender-Based DIF Statistics for Open-Response Items: TPCL (French)

Item Code	Section	Sequence	Sample 1		
			Effect Size	p-Value	DIF Level
18325_232	s. o.	6	-0.03	0.0388	
18181_227	s. o.	19	0.02	0.0006	
13278_T	s. o.	11	-0.08	0.0002	
13278_V	s. o.	11	0.00	0.1015	
19077_T	s. o.	12	-0.13	0.0000	
19077_V	s. o.	12	-0.12	0.0001	
19076_T	I	1	-0.23	0.0000	B-
19076_V	I	1	-0.22	0.0000	B-
13242_236	IV	6	-0.27	0.0000	C-
18328_236	IV	7	-0.23	0.0000	B-
13272_T	V	1	-0.11	0.0000	
13272_V	V	1	-0.04	0.0013	

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

Table 7.1.85a SLL-Based DIF Statistics for Multiple-Choice 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
18158_269	n/a	1	0.39	-0.07	0.85		0.36	-0.10	0.81	
18159_269	n/a	2	3.20	2.62	3.78	C+	3.98	3.37	4.59	C+
18160_269	n/a	3	1.83	1.47	2.19	C+	2.66	2.30	3.03	C+
18161_269	n/a	4	0.47	-0.11	1.04		1.59	0.97	2.21	B+
18162_269	n/a	5	-0.33	-0.64	-0.01		-0.55	-0.88	-0.22	
17970	n/a	7	0.26	-0.09	0.61		0.07	-0.30	0.43	
17980	n/a	8	0.10	-0.21	0.42		0.25	-0.07	0.57	
17996	n/a	9	-0.13	-0.70	0.43		-0.76	-1.38	-0.13	
17995	n/a	10	0.15	-0.17	0.47		0.02	-0.31	0.35	
17236_265	n/a	13	-0.49	-0.88	-0.10		0.31	-0.10	0.72	
17246_265	n/a	14	-0.24	-0.58	0.11		-0.56	-0.92	-0.20	
17357_265	n/a	15	0.77	0.44	1.11		1.01	0.67	1.34	B+
18065_265	n/a	16	-0.23	-0.56	0.11		-0.43	-0.78	-0.08	
18072_265	n/a	17	-0.62	-0.96	-0.27		-1.14	-1.51	-0.77	B-
18075_265	n/a	18	-0.84	-1.16	-0.52		-1.17	-1.50	-0.84	B-
17968	II	1	-0.11	-0.46	0.24		-0.16	-0.54	0.21	
17977	II	2	1.05	0.65	1.45	B+	1.27	0.87	1.68	B+
17992	II	3	-0.37	-0.69	-0.05		-0.74	-1.08	-0.41	
18002	II	4	1.51	1.09	1.94	C+	1.66	1.21	2.11	C+
18068_255	III	1	0.97	0.62	1.32		0.88	0.52	1.24	
18069_255	III	2	0.75	0.41	1.09		0.26	-0.08	0.60	
18070_255	III	3	-1.05	-1.38	-0.72	B-	-0.76	-1.10	-0.43	
18071_255	III	4	-0.02	-0.36	0.31		0.15	-0.19	0.48	
18073_255	III	5	0.75	0.41	1.08		0.33	-0.01	0.68	
18074_255	III	6	1.60	1.22	1.98	C+	1.40	1.01	1.79	B+
18076_255	III	7	1.62	1.15	2.09	C+	2.11	1.65	2.58	C+
18077_255	III	8	0.32	-0.04	0.68		0.33	-0.03	0.69	
18078_255	III	9	-0.51	-0.87	-0.16		-1.14	-1.53	-0.76	B-
13495_262	IV	1	1.76	1.41	2.11	C+	1.84	1.48	2.19	C+
13494_262	IV	2	1.34	0.99	1.69	B+	1.66	1.31	2.00	C+
13508_262	IV	3	-0.92	-1.40	-0.44		-0.60	-1.12	-0.07	
13502_262	IV	4	1.50	1.12	1.88	C+	0.62	0.23	1.01	
13492_262	IV	5	0.74	0.42	1.06		0.64	0.31	0.97	
18052_254_21	VI	1	1.55	1.17	1.92	C+	1.38	1.00	1.76	B+
18054_254_21	VI	2	-1.13	-1.48	-0.77	B-	-1.02	-1.40	-0.63	B-
18056_254_21	VI	3	0.36	0.01	0.71		0.44	0.08	0.79	
19524_254	VI	4	0.41	-0.01	0.83		1.05	0.60	1.49	B+
18059_254_21	VI	5	0.07	-0.30	0.44		0.53	0.15	0.90	
18061_254_21	VI	6	-0.38	-0.70	-0.07		-1.06	-1.39	-0.72	B-

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

Table 7.1.85b SLL-Based DIF Statistics for Open-Response 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
18163_269	n/a	6	-0.22	0.0000	B-	-0.23	0.0000	B-
17237_265	n/a	19	-0.40	0.0000	C-	-0.46	0.0000	C-
18022_T	n/a	11	0.01	0.8266		0.01	0.9366	
18022_V	n/a	11	0.03	0.0760		0.05	0.0074	
19840_T	n/a	12	-0.12	0.0003		-0.11	0.0006	
19840_V	n/a	12	0.05	0.0002		0.08	0.0000	
18378_T	I	1	-0.24	0.0000	B-	-0.24	0.0000	B-
18378_V	I	1	0.03	0.0004		0.12	0.0000	
13498_262	IV	6	-0.06	0.0000		-0.08	0.0000	
13499_262	IV	7	-0.07	0.0000		-0.06	0.1221	
18021_T	V	1	-0.12	0.0005		-0.15	0.0000	
18021_V	V	1	0.08	0.0000		-0.02	0.0002	

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

**Education Quality and
Accountability Office**



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

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

© 2013 Queen's Printer for Ontario | Ctrc_report_ne_0813