

EQAO's Technical Report

for the 2013–2014 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 2013–2014 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 978-1-4606-7332-4, ISSN 1927-7105

© 2015 Queen's Printer for Ontario | Ctrc_report_ne_0315

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	4
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	13
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.....	21
SCORER VALIDITY AND RELIABILITY	22
Scoring Validity	22
Scorer Reliability.....	23
CHAPTER 5: EQUATING.....	25
IRT MODELS	25
EQUATING DESIGN.....	26
CALIBRATION AND EQUATING SAMPLES	26

CALIBRATION	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 the 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	38
REPORTING THE RESULTS OF THE GRADE 9 ASSESSMENT OF MATHEMATICS	39
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)	70
The Primary- and Junior-Division Assessments	72
The Grade 9 Mathematics Assessment.....	74
The OSSLT	76
DECISION ACCURACY AND CONSISTENCY.....	77
Accuracy	78
Consistency	78
Estimation from One Test Form.....	79
The Primary and Junior Assessments.....	80
The Grade 9 Assessment of Mathematics	80
The OSSLT	81
REFERENCES	81
CHAPTER 8: VALIDITY EVIDENCE.....	82
INTRODUCTION.....	82
The Purposes of EQAO Assessments.....	82
Conceptual Framework for the Validity Argument.....	82

VALIDITY EVIDENCE BASED ON THE CONTENT OF THE ASSESSMENTS AND THE ASSESSMENT PROCESSES ...	83
Test Specifications for EQAO Assessments	83
Appropriateness of Test Questions.....	83
Quality Assurance in Administration	84
Scoring of Open-Response Items	84
Equating	85
VALIDITY EVIDENCE BASED ON THE TEST CONSTRUCTS AND INTERNAL STRUCTURE.....	85
Test Dimensionality	85
Technical Quality of the Assessments.....	85
VALIDITY EVIDENCE BASED ON EXTERNAL ASSESSMENT DATA	86
Linkages to International Assessment Programs	86
VALIDITY EVIDENCE SUPPORTING APPROPRIATE INTERPRETATIONS OF RESULTS.....	87
Setting Standards.....	87
Reporting.....	87
REFERENCES	88
APPENDIX 4.1: SCORING VALIDITY AND INTERRATER RELIABILITY	89
APPENDIX 7.1: SCORE DISTRIBUTIONS AND ITEM STATISTICS	105

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 tasks 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. Since the tests are not identical, one should avoid making comparisons between the language groups.

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.

EQAO develops separate versions of the Grade 9 Assessment of Mathematics for students in academic and applied courses. 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). 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 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 a five-year cycle, 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 items 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;
- 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 2013–2014 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 try out the items they have developed in their own classes. 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; 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 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 at 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, proper administration procedures);
- 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 only); for the OSSLT, the test can be deferred to a later year for some students. 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. The various administration and accommodation guides may be found on EQAO's Web site, www.eqao.com.

Definition of "Accommodations"

Accommodations are defined in the accommodation 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 School 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 manually into a computer for analysis.

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. 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. 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. 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. (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. Each education officer reads 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 papers and possible papers 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, for which 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.

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 a room are trained to score the items in that room 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 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 using 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 score code (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), which contain responses that are solid examples of student work for a given score 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 to be resolved by an EQAO staff member. The resolution may involve contact with a school to seek clarification.

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 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 validity papers, which were identified during range finding and were scored by an expert. Scorers score up to 10 validity papers a day. Their scores are compared to the scores assigned by the expert. 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 results of the comparisons to determine whether scorers are drifting from the scoring standards (established during scorer training) 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 reading, writing and mathematics items for the primary- and junior-division assessments and mathematics items for the Grade 9 assessment 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 requirements (which are the same as those for scoring validity) 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 aggregated daily and cumulative individual scorer data reports include daily and cumulative validity data, 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. Scoring leaders and supervisors document all retraining as well as decisions about retention or dismissal of a scorer.

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 10 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 2013–2014, the EQAO target of 95% exact-plus-adjacent agreement for validity was met for all items for the primary and junior assessments.

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 2014. 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 but one item on the English-language academic version of the assessment.

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 2013–2014, the EQAO target of 95% exact-plus-adjacent agreement for validity was met for all but one item in English short writing.

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 2013–2014, the EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for all but two reading items, all but two writing prompts and all but one mathematics item. The aggregate reliability estimates ranged from 95.7–99.1%.

The Grade 9 Assessment of Mathematics: Academic and Applied

In 2013–2014, the EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for all but two items, and the aggregate reliability estimates ranged from 97.5–98.7%.

The Ontario Secondary School Literacy Test (OSSLT)

The EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for all but one reading item, for all but three short-writing prompts and for all but four long-writing prompts. The aggregate reliability estimates ranged from 94.2–97.7%.

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. The equating processes conducted by EQAO staff are replicated by an external contractor to ensure accuracy.

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 2012–2013 to 2013–2014, only equating procedures were used in 2013–2014.

The following sections describe the Item Response Theory (IRT) models, equating design, equating samples and calibration procedures used during the 2013–2014 school year for the various EQAO assessments.

IRT Models

Item-response models define the relationship between an unobserved construct or proficiency (θ , or theta) 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 proficiency 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 proficiency 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 proficiency 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 proficiency 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. The following steps are used in equating for the EQAO assessments:

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.

Calibration and Equating Samples

For each assessment, EQAO uses a set of exclusion rules to select calibration and equating samples. The exclusion rules ensure that the samples are 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. The equating and calibration samples are identical for the current assessment year; for the previous year, the calibration sample was reduced further by excluding students who did not answer any of the field-test questions that were brought forward to the operational test for the current year.

Calibration

Calibration is the process of estimating the item parameters that determine the relationship between proficiency 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 for 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 & 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 very low proficiency 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 2012–2013 and 2013–2014:

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 2012–2013 population who responded to both the 2013–2014 operational-test items and the field-test items that had been brought forward to form the 2013–2014 operational tests and who were not excluded by the rules stated above (calibration sample);
2. students from the 2012–2013 population who wrote the operational test and who were not excluded by the rules stated above (equating sample) and
3. students from the 2013–2014 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 2012–2013 and 2013–2014 Primary- and Junior-Division Assessments (English and French)

Assessment	Number of Students		
	2013–2014 Calibration and Equating Sample	2012–2013 Calibration Sample	2012–2013 Equating Sample
Primary Reading (English)	98 905	27 514	100 257
Junior Reading (English)	104 322	29 869	108 373
Primary Reading (French)	6 883	2 274	6 686
Junior Reading (French)	5 925	1 905	5 765
Primary Writing (English)	99 385	45 999	100 610
Junior Writing (English)	104 345	48 420	108 479
Primary Writing (French)	6 919	4 036	6 719
Junior Writing (French)	5 915	3 855	5 778
Primary Mathematics (English)	95 162	70 567	96 375
Junior Mathematics (English)	106 274	105 896	110 425
Primary Mathematics (French)	6 976	6 351	6 789
Junior Mathematics (French)	5 999	5 852	5 870

Equating Steps

In equating the 2012–2013 and 2013–2014 tests, the forward-fixed common-item-parameter non-equivalent group design was implemented as follows:

1. The 2013–2014 operational items were calibrated independently to obtain item parameter estimates and student-proficiency scores for the 2013–2014 calibration and equating sample.

2. The 2012–2013 operational items were calibrated (using the calibration sample) together with the field-test items that were brought forward to the 2013–2014 operational assessments. In this calibration, the item parameter estimates of the field-test items were fixed at the values obtained from the 2013–2014 calibration runs (Step 1).
3. The 2012–2013 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 2012–2013 equating sample from the levels assigned in 2012–2013. 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 2013–2014.
6. The operational item parameter estimates of 2013–2014 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 2013–2014 student theta values, students were assigned to levels, and the percentage of students at each performance level was determined.

Eliminating Items and the Collapsing of Score Categories

For the primary- and junior-division assessments, 8 multiple-choice items and four open-response items across the 12 assessment components were excluded from equating. Seven multiple-choice reading items were excluded because they had been modified between the field- and operational-test administrations. One reading item was dropped due to an error in the response options. One multiple-choice mathematics item was excluded because its format had been changed. Long-writing prompts were not field-tested in the previous year and were excluded from equating. 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 (2013–2014)

Assessment	No. of Items Excluded from Equating*		No. of Items Dropped from the Assessment
	Multiple-Choice	Open-Response	
Primary Reading (English)	3	0	1
Junior Reading (English)	0	0	0
Primary Reading (French)	1	0	0
Junior Reading (French)	3	0	0
Primary Writing (English)	0	1	0
Junior Writing (English)	0	1	0
Primary Writing (French)	0	1	0
Junior Writing (French)	0	1	0
Primary Mathematics (English)	0	0	0
Junior Mathematics (English)	1	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 2012–2013 and 2013–2014 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 0.97 for Levels 3 and 4, -0.67 for Levels 2 and 3, -1.88 for Levels 1 and 2 and -2.98 for “not enough evidence for Level 1” (NE1) and Level 1.

Since the 2012–2013 and 2013–2014 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	
		2013	2014		2013	2014
Number of Students		100 257	98 905		6686	6883
Level 4		14.6%	14.3%		39.1%	42.5%
Level 3	0.97	61.1%	63.2%	0.24	44.4%	42.3%
Level 2	-0.67	20.4%	19.7%	-0.98	15.3%	13.9%
Level 1	-1.88	3.5%	2.4%	-2.25	1.2%	1.2%
NE1	-2.98	0.4%	0.4%	-3.46	0.1%	0.1%
% of Students at or Above the Provincial Standard		75.7%	77.4%		83.5%	84.8%

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

	Junior Reading (English)			Junior Reading (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2013	2014		2013	2014
Number of Students		108 373	104 322		5,765	5,925
Level 4		15.7%	14.6%		32.4%	38.6%
Level 3	0.96	69.1%	72.4%	0.30	61.7%	56.5%
Level 2	-1.04	14.2%	12.0%	-1.61	5.9%	4.9%
Level 1	-2.39	1.0%	0.9%	-3.12	0.1%	0.0%
NE1	-3.57	0.0%	0.0%	-3.95	0.0%	0.0%
% of Students at or Above the Provincial Standard		84.8%	87.0%		94.0%	95.1%

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	
		2013	2014		2013	2014
Number of Students		100 610	99 385		6719	6919
Level 4		7.7%	7.1%		26.5%	26.8%
Level 3	1.38	74.9%	76.3%	0.59	60.1%	57.5%
Level 2	-0.91	16.8%	16.0%	-0.96	12.3%	14.2%
Level 1	-2.39	0.4%	0.5%	-1.96	1.0%	1.2%
NE1	-3.27	0.1%	0.1%	-2.65	0.1%	0.3%
% of Students at or Above the Provincial Standard		82.7%	83.4%		86.6%	84.3%

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	
		2013	2014		2013	2014
Number of Students		108 479	104 345		5778	5915
Level 4		15.4%	14.6%		25.3%	31.1%
Level 3	0.97	69.3%	71.9%	0.44	66.1%	62.6%
Level 2	-1.02	14.8%	13.1%	-1.45	7.7%	5.9%
Level 1	-2.43	0.4%	0.4%	-2.38	0.8%	0.3%
NE1	-3.36	0.1%	0.1%	-3.25	0.1%	0.1%
% of Students at or Above the Provincial Standard		84.7%	86.5%		91.8%	93.8%

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	
		2013	2014		2013	2014
Number of Students		96 375	95 162		6789	6976
Level 4		14.2%	14.5%		26.4%	27.0%
Level 3	1.00	59.8%	59.5%	0.62	55.6%	56.2%
Level 2	-0.59	24.1%	23.9%	-0.94	17.0%	16.2%
Level 1	-2.02	1.8%	1.9%	-2.37	0.9%	0.5%
NE1	-2.77	0.2%	0.2%	-3.20	0.0%	0.1%
% of Students at or Above the Provincial Standard		74.0%	74.0%		82.0%	83.2%

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	
		2013	2014		2013	2014
Number of Students		110 425	106 274		5870	5999
Level 4		15.6%	14.7%		46.3%	49.2%
Level 3	1.01	48.7%	47.2%	0.05	39.5%	38.5%
Level 2	-0.27	28.9%	30.0%	-1.15	13.9%	12.0%
Level 1	-1.37	6.7%	8.0%	-2.52	0.2%	0.2%
NE1	-3.00	0.1%	0.1%	-2.91	0.1%	0.1%
% of Students at or Above the Provincial Standard		64.3%	61.9%		85.8%	87.7%

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 proficiency 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 very low proficiency 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 2012–2013 and 2013–2014 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 2012–2013 and the 2013–2014 assessments.

Table 5.9 Number of Grade 9 Students in the Equating Samples

Version	2012–2013				2013–2014	
	English		French		English	French
	Calibration Sample	Equating Sample	Calibration Sample	Equating Sample	Calibration and Equating Sample	Calibration and Equating Sample
Academic	87 865	94 679	3468	3769	93 431	3947
Applied	28 364	33 301	1152	1190	32 122	1294

Equating Steps

The calibration and equating of the 2012–2013 and 2013–2014 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 2013–2014 winter and spring samples.
2. Calibration and equating were conducted for the 2012–2013 winter and spring samples (including the field-test items that were brought forward to the 2013–2014 operational test). In this calibration, the item parameter estimates of the field-test items that were brought forward to the 2013–2014 operational tests were fixed at the values obtained from the 2013–2014 calibration runs (Step 1). The parameter estimates of the 2012–2013 operational items that were repeated on the 2013–2014 tests were also fixed at the values obtained from the 2013–2014 calibration runs (Step 1). The 2012–2013 equating samples (operational items only) were scored using the scaled 2012–2013 operational-item parameter estimates.
3. The percentage of students in each achievement level was determined for the 2012–2013 equating sample using the levels assigned in 2012–2013. 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 scores for the 2013–2014 assessments.
5. The parameter estimates for the 2013–2014 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 2013–2014 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, one item was excluded during the equating process due to issues of calibrating this item (see Table 5.10).

Table 5.10 Number of Items Excluded from the Equating Process and Dropped: Grade 9 (2013–2014)

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)	0	0	0
Applied, Spring (French)	1	0	1
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 percentage of students in 2012–2013 and 2013–2014 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 1.22 for Levels 3 and 4, 0.02 for Levels 2 and 3, –1.04 for Levels 1 and 2, and –1.77 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	
		2013	2014		2013	2014
Number of Students		33 301	32 122		1190	1294
Level 4		8.8%	9.7%		9.7%	9.3%
Level 3	1.22	39.0%	41.2%	1.26	43.6%	45.0%
Level 2	0.02	36.8%	35.3%	-0.11	37.8%	37.0%
Level 1	-1.04	12.7%	10.9%	-1.30	7.1%	7.4%
Below Level 1	-1.77	2.6%	2.8%	-1.97	1.8%	1.3%
% of Students at or Above the Provincial Standard		47.9%	50.9%		53.3%	54.3%

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	
		2013	2014		2013	2014
Number of Students		94 679	93 431		3769	3947
Level 4		13.2%	12.2%		6.3%	7.3%
Level 3	1.15	72.3%	73.7%	1.43	76.3%	77.0%
Level 2	-1.02	10.3%	10.5%	-0.96	13.2%	11.9%
Level 1	-1.64	4.0%	3.5%	-1.63	4.2%	3.8%
Below Level 1	-2.69	0.1%	0.1%	-2.65	0.1%	0.0%
% of Students at or Above the Provincial Standard		85.5%	85.8%		82.5%	84.3%

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 very low proficiency 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 proficiency parameters for open-response items.

Equating Sample

First-time eligible students from both publicly funded and private schools were selected for the 2012–2013 and 2013–2014 equating samples. Students in schools included in the 2013 equating sample (April administration) were used to create the 2013 and 2014 equating samples.

Therefore, the 2013 and 2014 equating samples were drawn from comparable populations. 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 2012–2013 and 2013–2014 tests.

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

OSSLT	2013		2014	
	English	French	English	French
First-Time Eligible	83 324	4306	81 553	4245

Equating Steps

The following steps were implemented to calibrate and equate the 2013 and 2014 OSSLT:

1. The parameter estimates of the operational items administered in 2014 were calibrated using the 2014 equating sample.
2. The operational items that formed the 2013 test and the field-test items brought forward to the 2014 test were recalibrated using the 2013 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 2013 test, obtained in Step 2, were used to score the 2013 equating sample data.
4. The percentage of successful students was determined for the 2013 equating sample from the student results reported in 2013. The theta-value cut point that replicated this percentage was identified for the distribution of scores in the 2013 equating sample.
5. This theta value was applied to student scores for the 2014 assessment to determine which students would be successful. The results are presented in Table 5.14.

Scale Score

The reporting scale scores for the 2014 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

One multiple-choice English-language reading item was dropped from the assessment because this item has two possible answers. One multiple-choice writing item on the French-language test was excluded from the equating process due to modifications made to the items between the 2013 field-test and the 2014 operational-test administrations.

The score 1.0 in the scoring rubric for long-writing prompts was collapsed with the score 1.5 for topic development and the use of conventions 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 2013 and 2014 are reported for English-language and French-language students. For example, the equated cut score for the English-language test was -0.64 . The percentage of successful students in the equating samples was 82.9% in 2013 and 84.1% in 2014.

Table 5.14 Equating Results for the OSSLT

	English-Language			French-Language		
	Theta Cut Point 2013–2014	Equating Sample 2013	Equating Sample 2014	Theta Cut Point 2013–2014	Equating Sample 2013	Equating Sample 2014
No. of Students		83 324	81 553		4 306	4 245
% Successful	-0.64	82.9%	84.1%	-0.75	90.1%	89.9%
Unsuccessful		17.1%	15.9%		9.9%	10.1%

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 annual provincial reports for education stakeholders and the general public. The reports for the 2013–2014 English-language assessments are available on www.eqao.com:

- *EQAO's Provincial Elementary School Report: Results of the 2013–2014 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, 2013–2014.*

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 10 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. percentages based on 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. percentages based on 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.

In 2012, EQAO introduced EQAO Reporting, an interactive Web-based reporting application that enables school principals to access their school's EQAO data and to link achievement data to contextual and attitudinal data. This application was made available to elementary school principals in 2012 and to secondary school principals in 2013. Since all of the data previously provided in the detailed school and board reports can be generated in EQAO Reporting, EQAO is phasing out the detailed reports. The full set of detailed reports was provided to secondary school principals in 2013 because that was the first year that EQAO Reporting was available to principals.

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) and access to the EQAO Reporting application, which enables them to view the results for all schools in the board and to link achievement data with demographic data. The directors also receive school lists with achievement results over time, for convenient reference.

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 and
- 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.

In EQAO Reporting, principals can generate the following types of data for the province, board and school:

- overall jurisdictional results for each component of an assessment;
- longitudinal data showing jurisdictional results over time;
- overall jurisdictional results for each assessment component by gender and other relevant characteristics (e.g., English language learners, special education needs, French Immersion);
- results for sub-groups of students based on contextual, achievement or attitudinal data;
- areas of strength and areas for improvement with respect to sections of the curriculum;
- data for individual items and collections of items, with a link to the actual items;
- 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, www.eqao.com.

In addition, schools receive the Item Information Report: Student Roster, which provides item results for each student who has completed each assessment and summary item statistics for the school, board and province. The data for individual students are also provided in data files. Results by exceptionality category and for students receiving each type of accommodation are provided for each school board and the province.

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.

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. Mathematics assessment results achieved in Grades 3 and 6, if available, are printed on the Grade 9 ISR.

The provincial, board and school reports provide the following:

- overall jurisdictional results for the academic and applied courses;
- longitudinal data showing jurisdictional results over time;
- overall jurisdictional results by gender and other relevant characteristics (e.g., English language learners, special education needs, semester);
- results by exceptionality category and results for students receiving each type of accommodation (board and province only);
- areas of strength and areas for improvement with respect to the curriculum expectations;
- cohort-tracking results from Grade 6 to Grade 9 (provincial results are provided for tracking students from Grade 3 to Grade 6 to Grade 9);
- contextual data and
- student questionnaire results.

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 for students receiving each type of accommodation (board and province only);
- 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 with respect to the curriculum expectations;
- cohort-tracking results from Grade 3 to 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. The Web-based EQAO Reporting application has a professional-development component built into it that provides directions on how to use the application and guidelines for using data for school improvement planning. EQAO delivers workshops on interpreting EQAO data and on using these data appropriately for school improvement planning.

EQAO also produces the following resource documents:

- *Using Data to Promote Student Success: A Brief Guide to Assist School Administrators in Interpreting Their Data*;
- “Guide to Using the Item Information Report: Student Roster” and
- “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 2013–2014 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 46.3% (for the reading component of the English-language primary-division assessment) to 65.7% (for math component of the French-language primary-division assessment).

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 0.88 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 4.6% to 4.7% of the possible maximum score for reading, 6.7% to 7.5% of the possible maximum score for writing and 6.2% 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. 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.

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)	35	25	10	65	118 392	0	61	30.07	8.63	0.88	2.99
Junior Reading (English)	36	26	10	66	124 355	0	65	39.57	8.99	0.88	3.11
Primary Reading (French)	36	26	10	66	7 723	1	62	37.63	8.92	0.89	2.96
Junior Reading (French)	36	26	10	66	6 746	7	65	41.71	8.80	0.88	3.05
Primary Writing (English)	14	8	6*	29	118 573	0	29	17.75	5.00	0.81	2.18
Junior Writing (English)	14	8	6*	29	124 321	0	29	17.71	4.89	0.81	2.13
Primary Writing (French)	14	8	6*	29	7 685	1	29	17.42	4.73	0.83	1.95
Junior Writing (French)	14	8	6*	29	6 746	1	29	18.95	4.67	0.83	1.93
Primary Mathematics (English)	36	28	8	60	124 027	0	60	39.03	11.26	0.89	3.73
Junior Mathematics (English)	36	28	8	60	124 167	0	60	37.32	12.04	0.90	3.81
Primary Mathematics (French)	36	28	8	60	7 750	3	60	39.39	11.09	0.88	3.84
Junior Mathematics (French)	36	28	8	60	6 742	5	60	39.07	12.10	0.90	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 proficiency 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)	118 392	-3.82	3.38	-0.14	1.01
Junior Reading (English)	124 355	-3.92	3.37	-0.18	1.03
Primary Reading (French)	7 723	-3.82	3.19	-0.10	0.99
Junior Reading (French)	6 746	-3.91	3.11	-0.14	1.01
Primary Writing (English)	118 573	-3.56	2.50	-0.10	0.95
Junior Writing (English)	124 321	-3.71	2.47	-0.18	0.99
Primary Writing (French)	7 685	-3.41	2.72	-0.05	0.95
Junior Writing (French)	6 746	-3.74	2.44	-0.14	0.99
Primary Mathematics (English)	124 027	-3.79	2.46	-0.14	1.01
Junior Mathematics (English)	124 167	-3.72	2.43	-0.17	1.03
Primary Mathematics (French)	7 750	-3.56	2.35	-0.09	0.98
Junior Mathematics (French)	6 743	-3.64	2.23	-0.12	1.00

Note. SD = standard deviation.

The Test Characteristic Curves (TCCs) and the distributions 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 proficiency score in IRT to a student proficiency 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 35%. 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 0.30. 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)

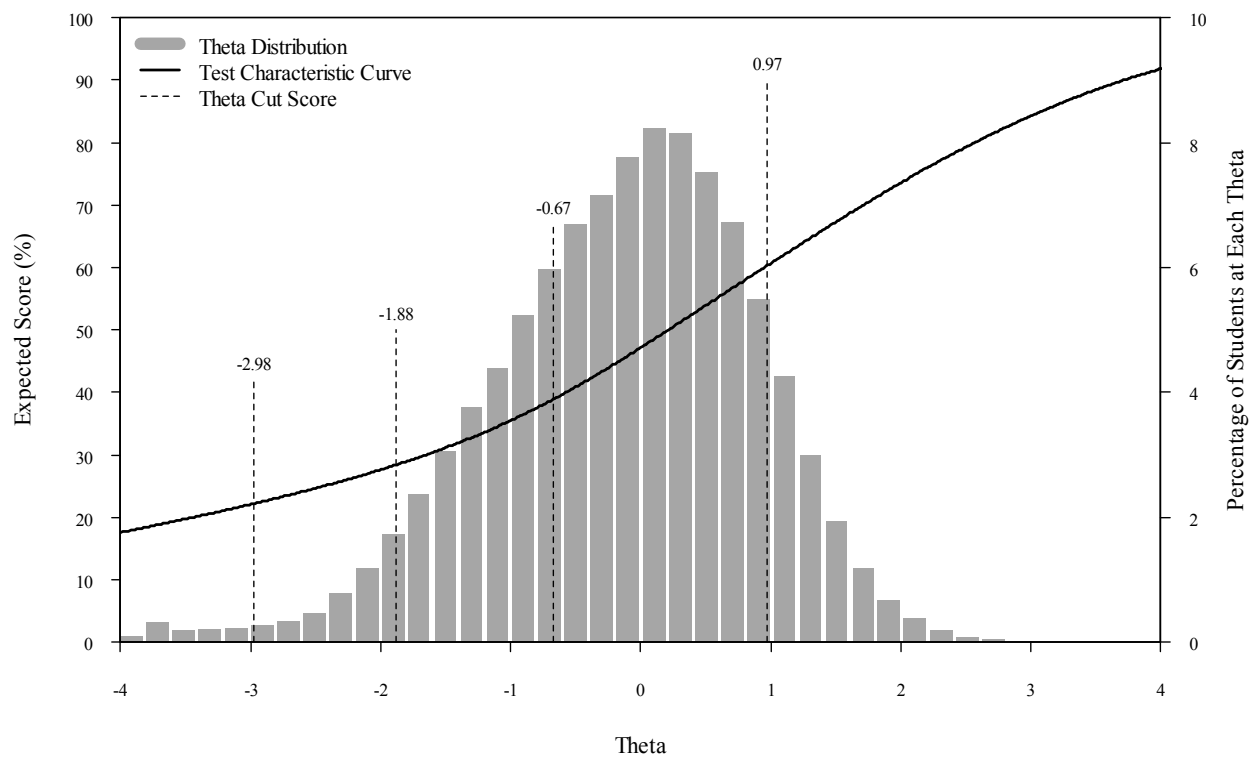


Figure 7.2 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (English)

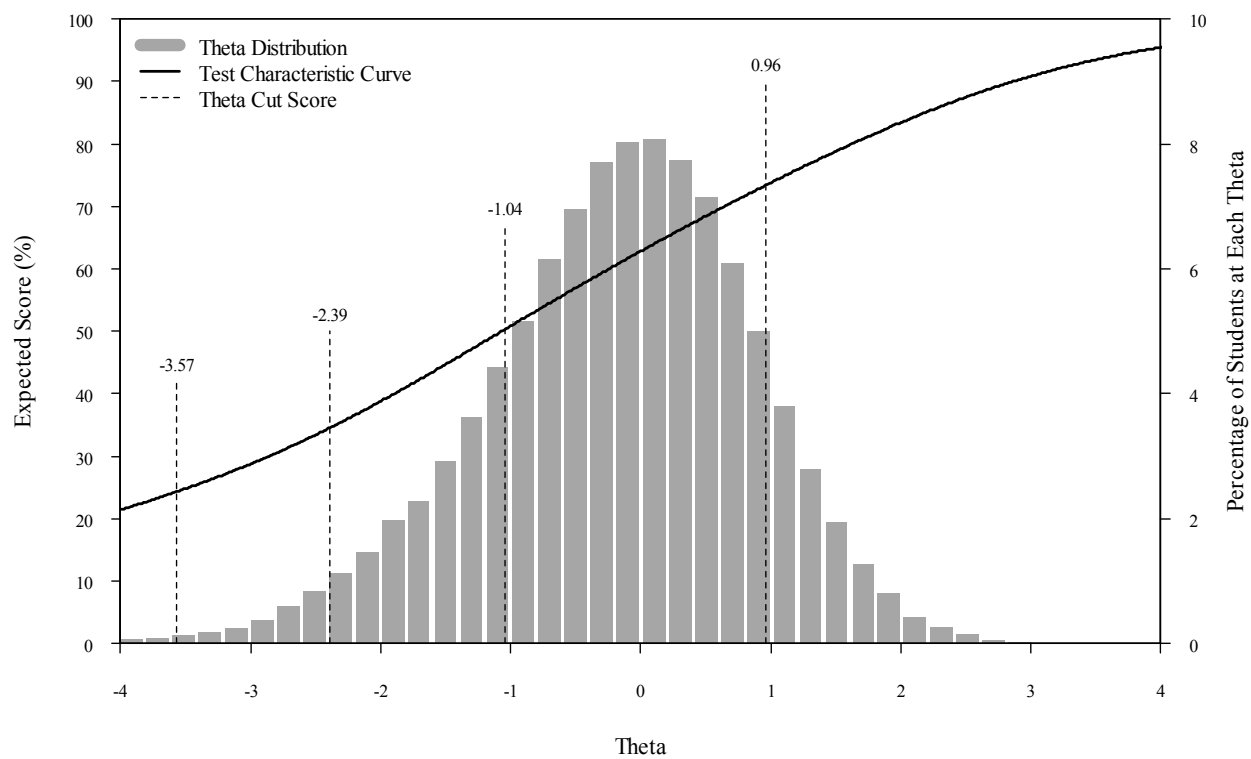


Figure 7.3 Test Characteristic Curve and Distribution of Thetas for Reading: Primary Division (French)

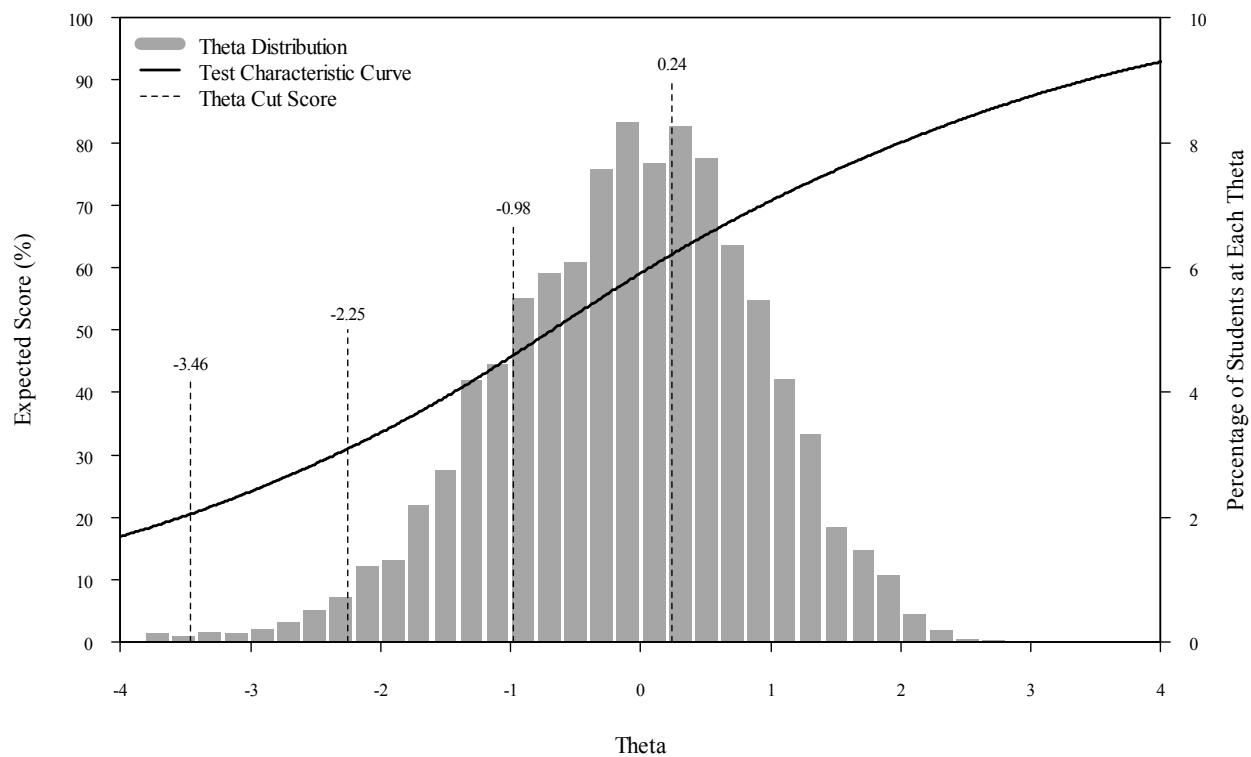


Figure 7.4 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (French)

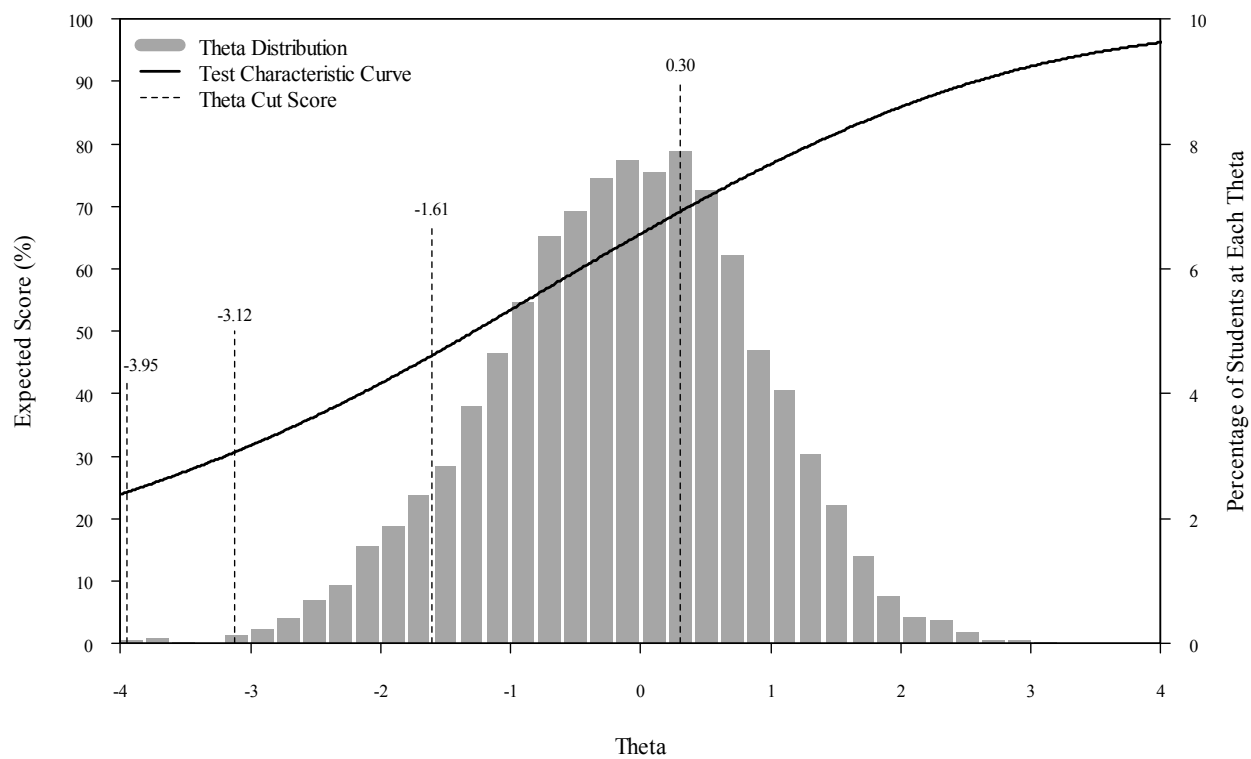


Figure 7.5 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (English)

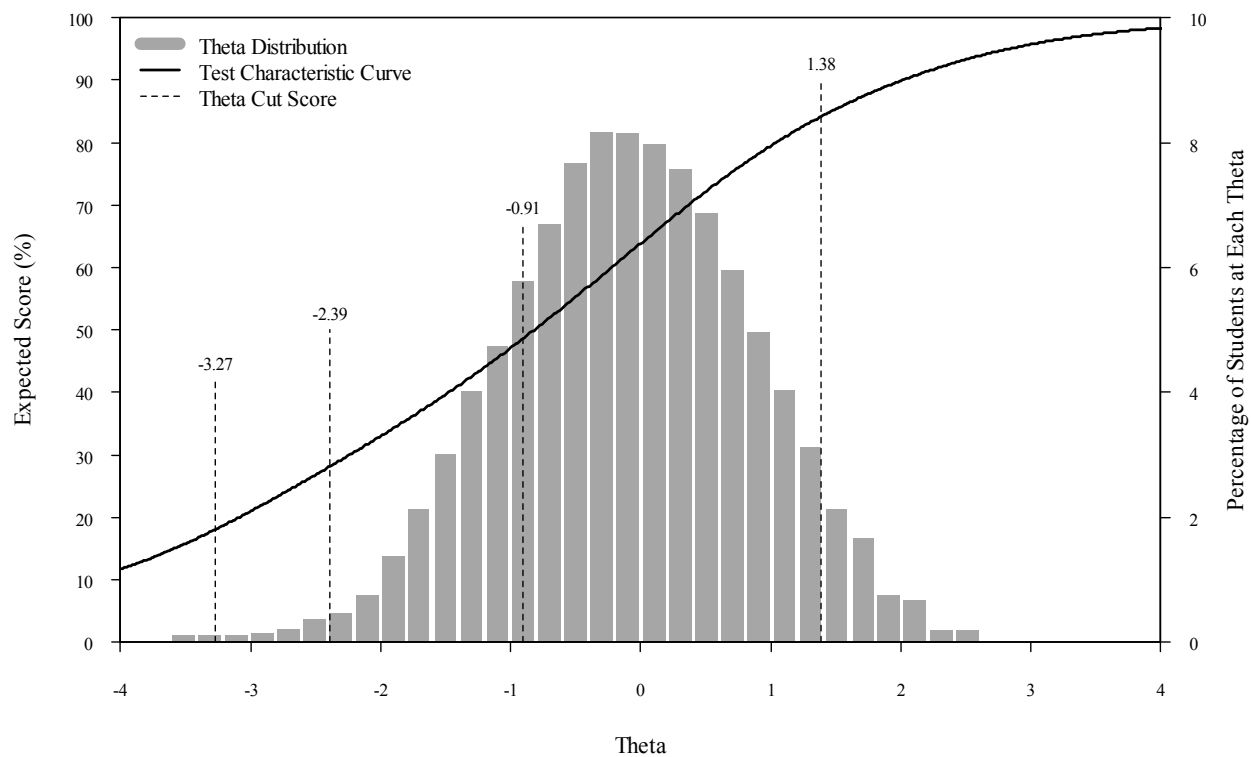


Figure 7.6 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (English)

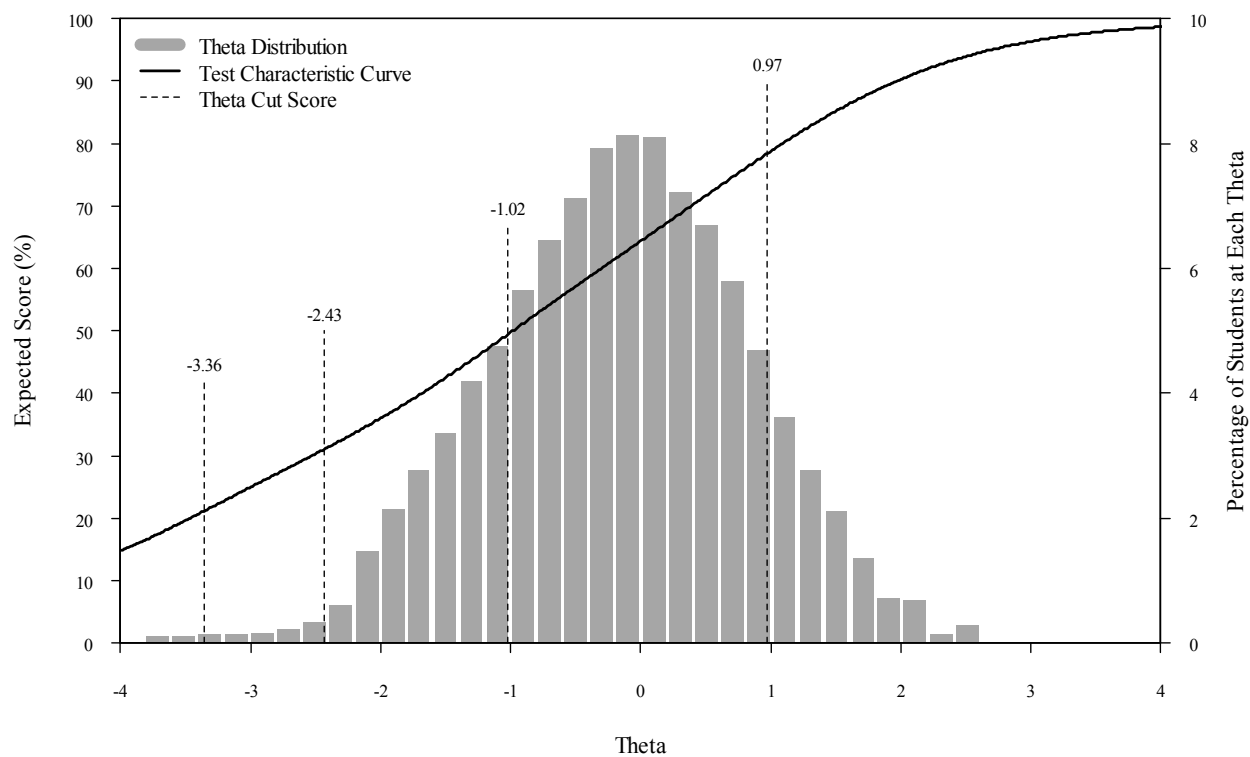


Figure 7.7 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (French)

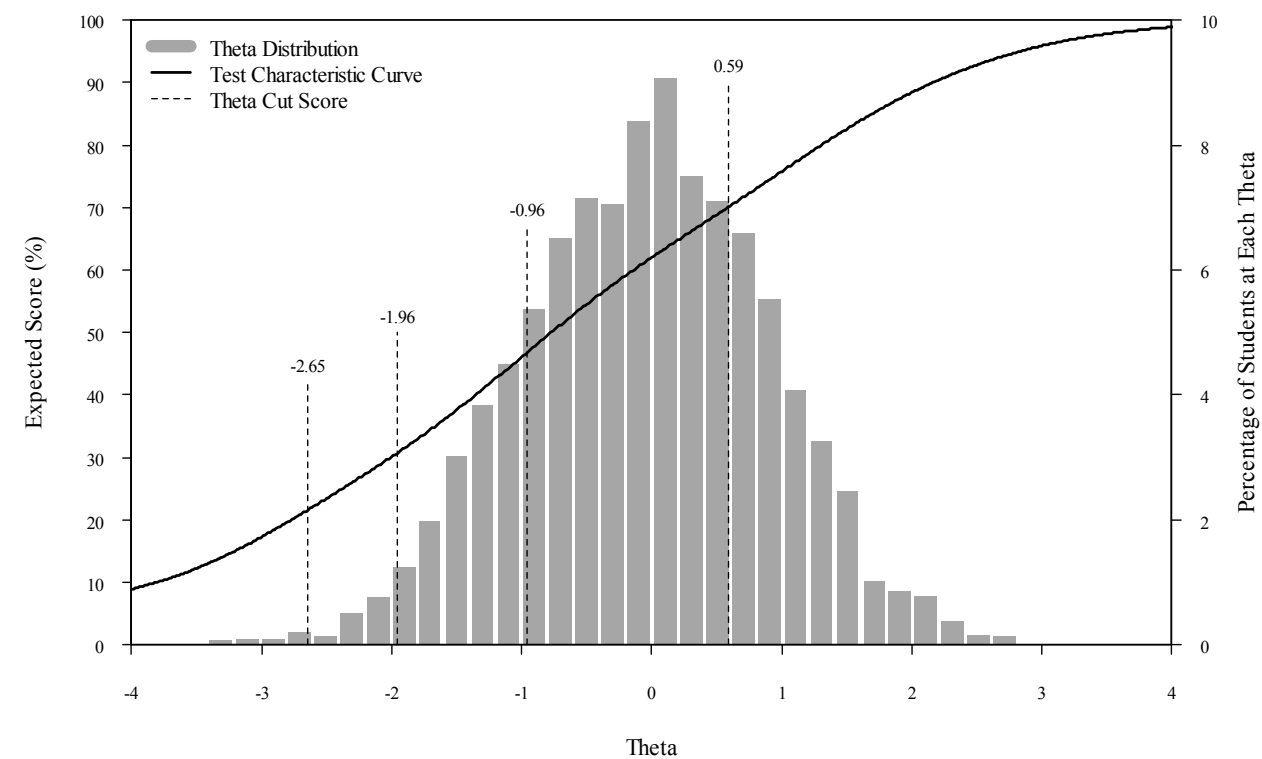


Figure 7.8 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (French)

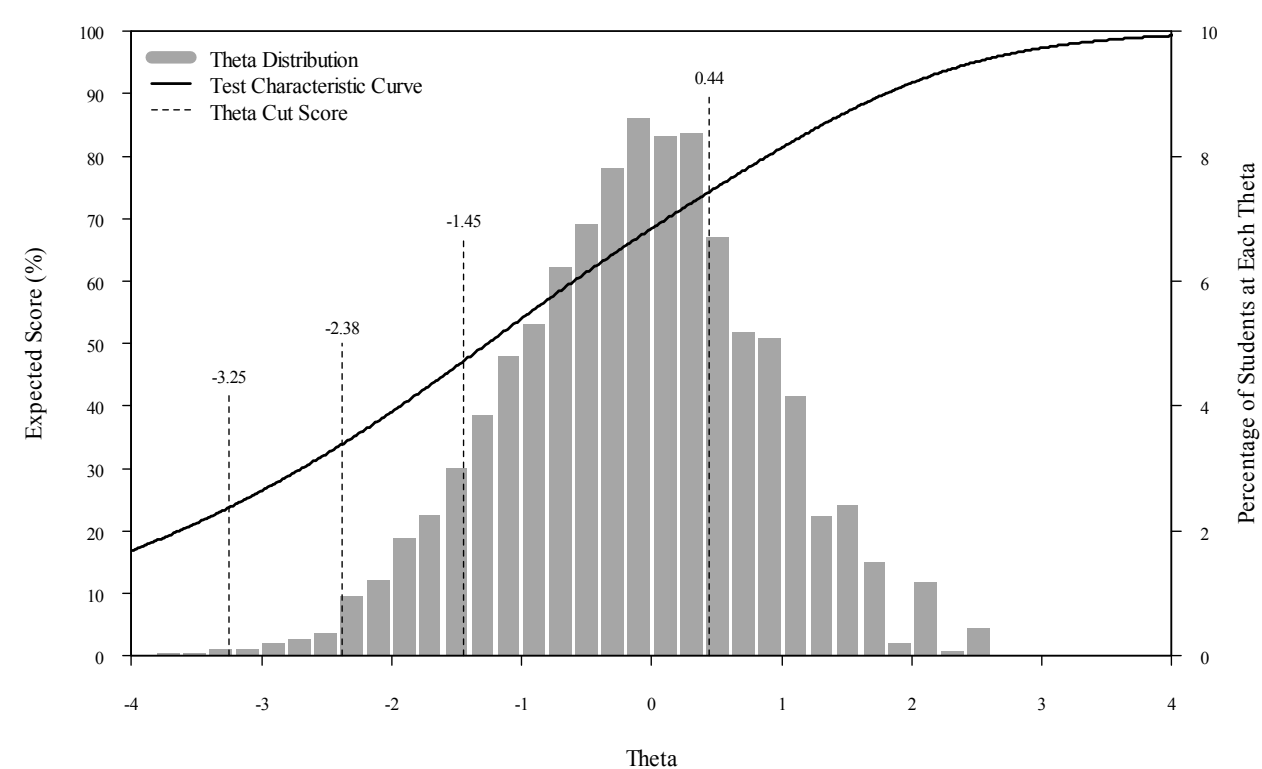


Figure 7.9 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (English)

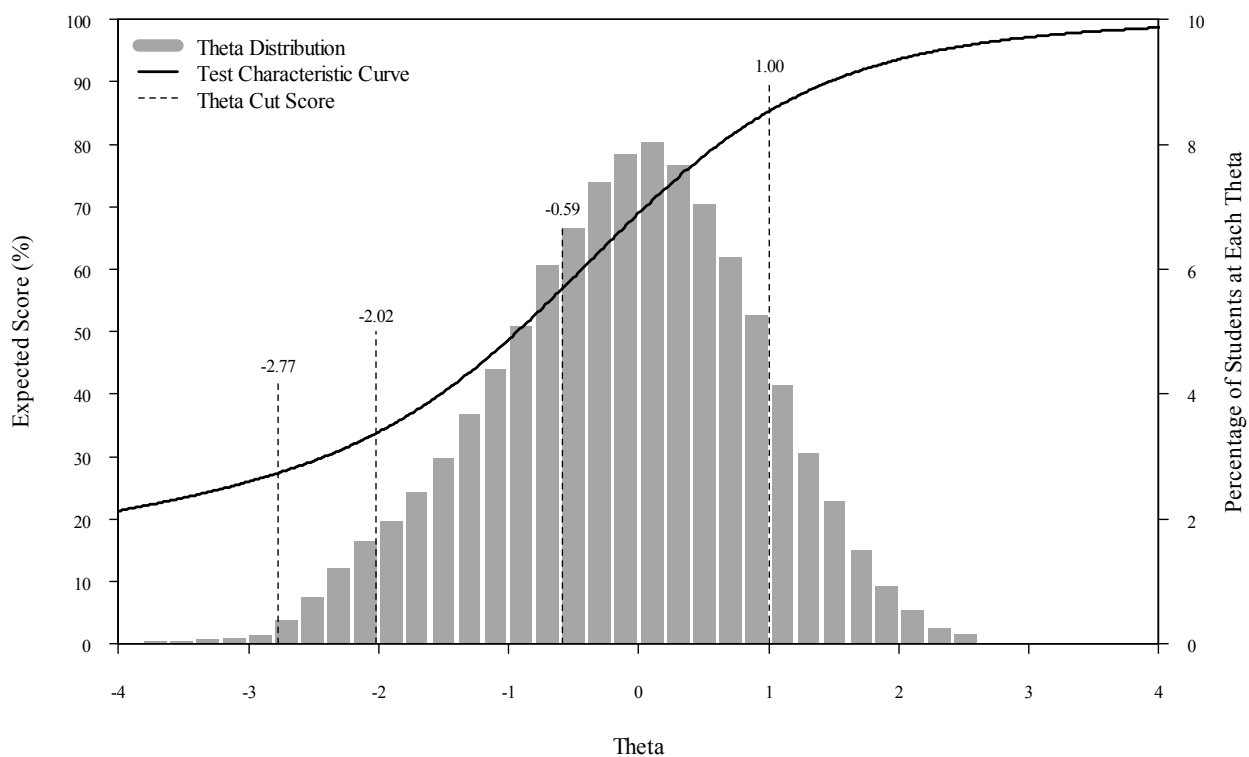


Figure 7.10 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (English)

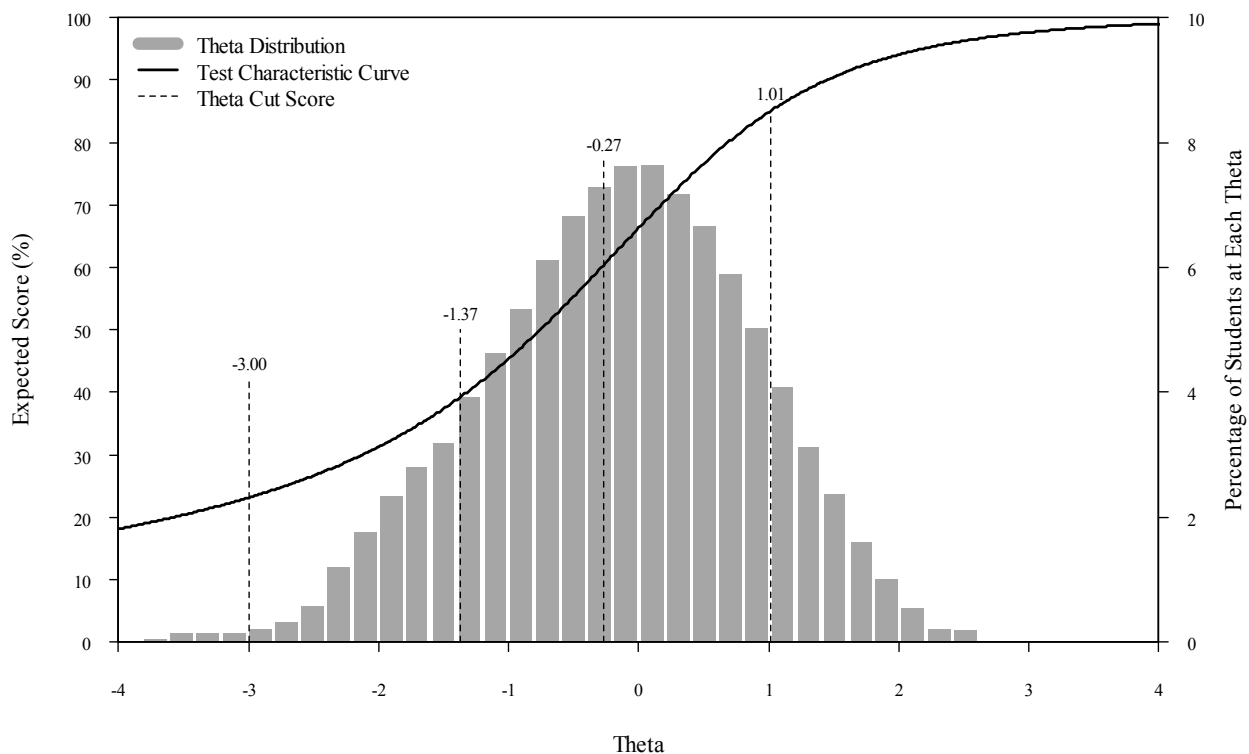


Figure 7.11 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (French)

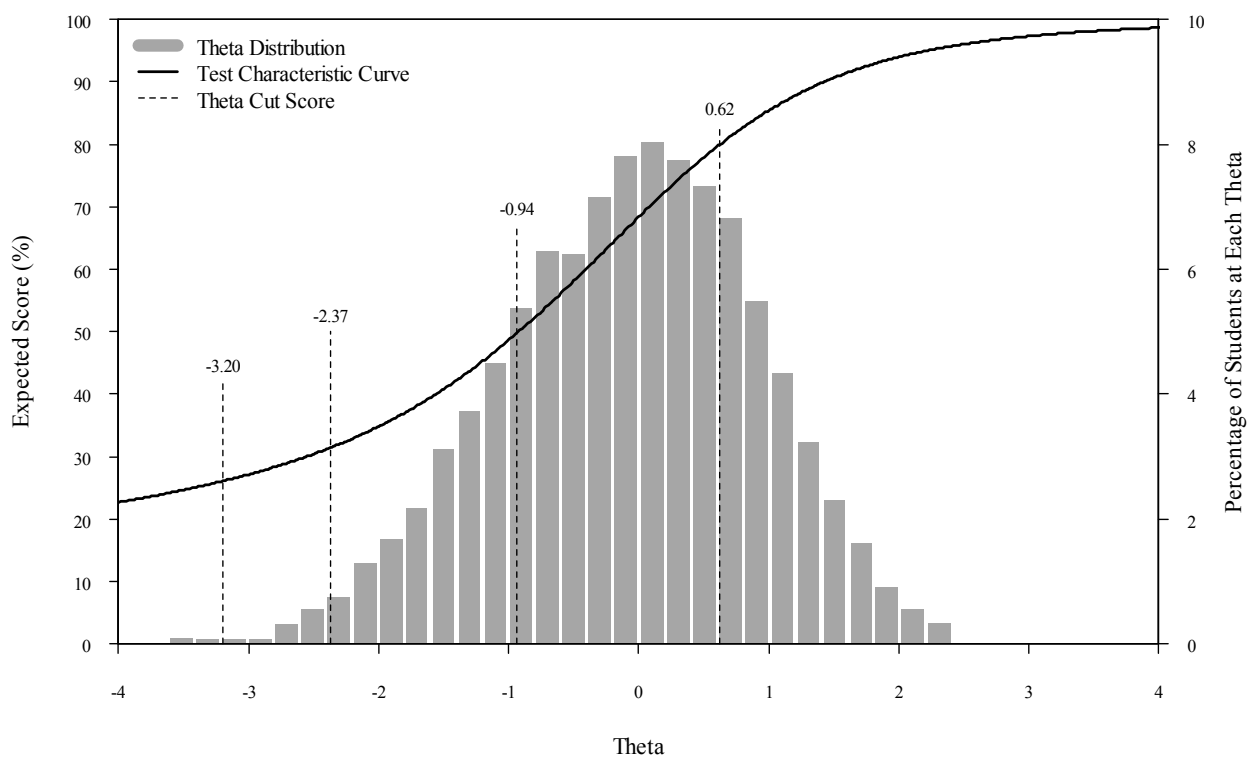


Figure 7.12 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (French)

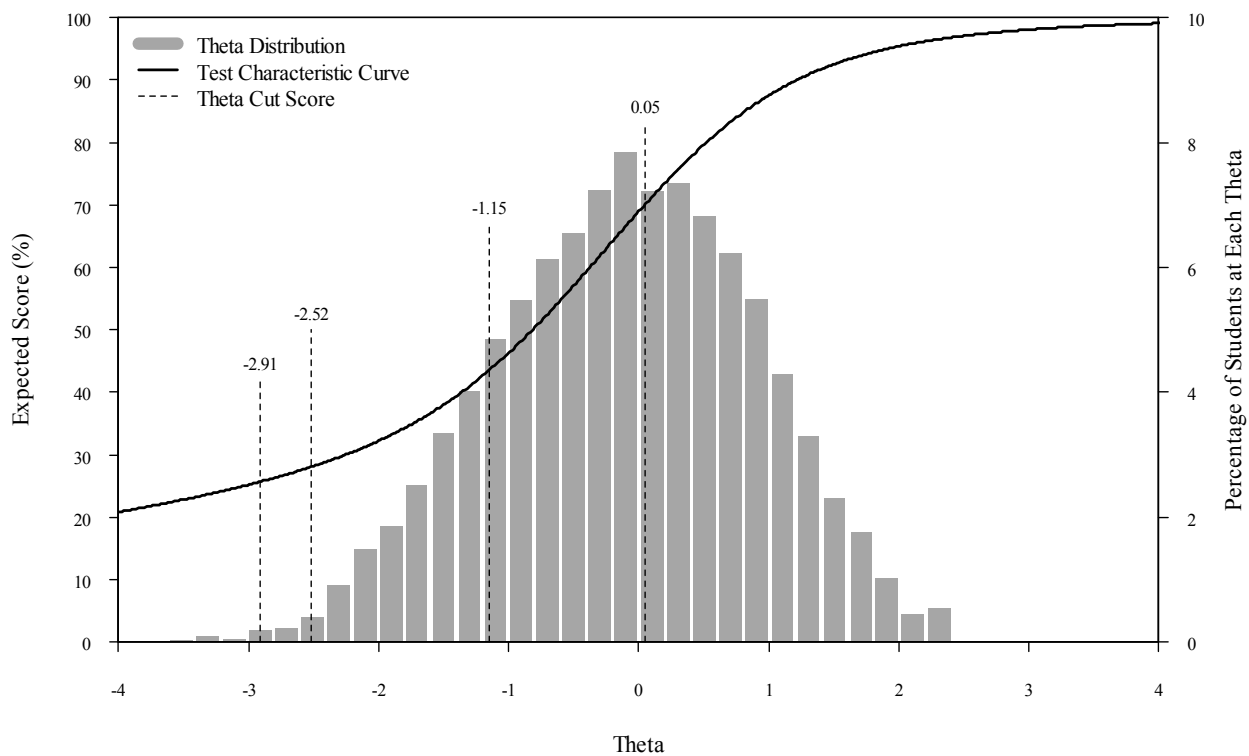


Figure 7.13 Test Information Function for Reading: Primary Division (English)

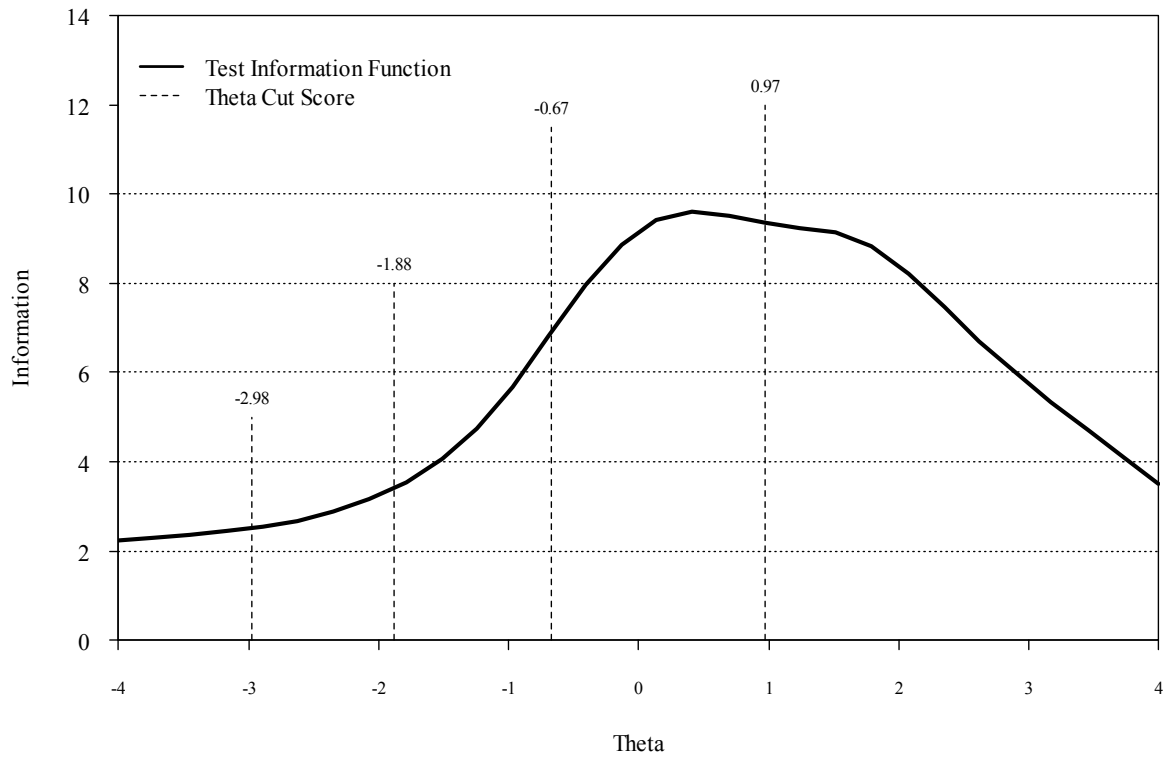


Figure 7.14 Test Information Function for Reading: Junior Division (English)

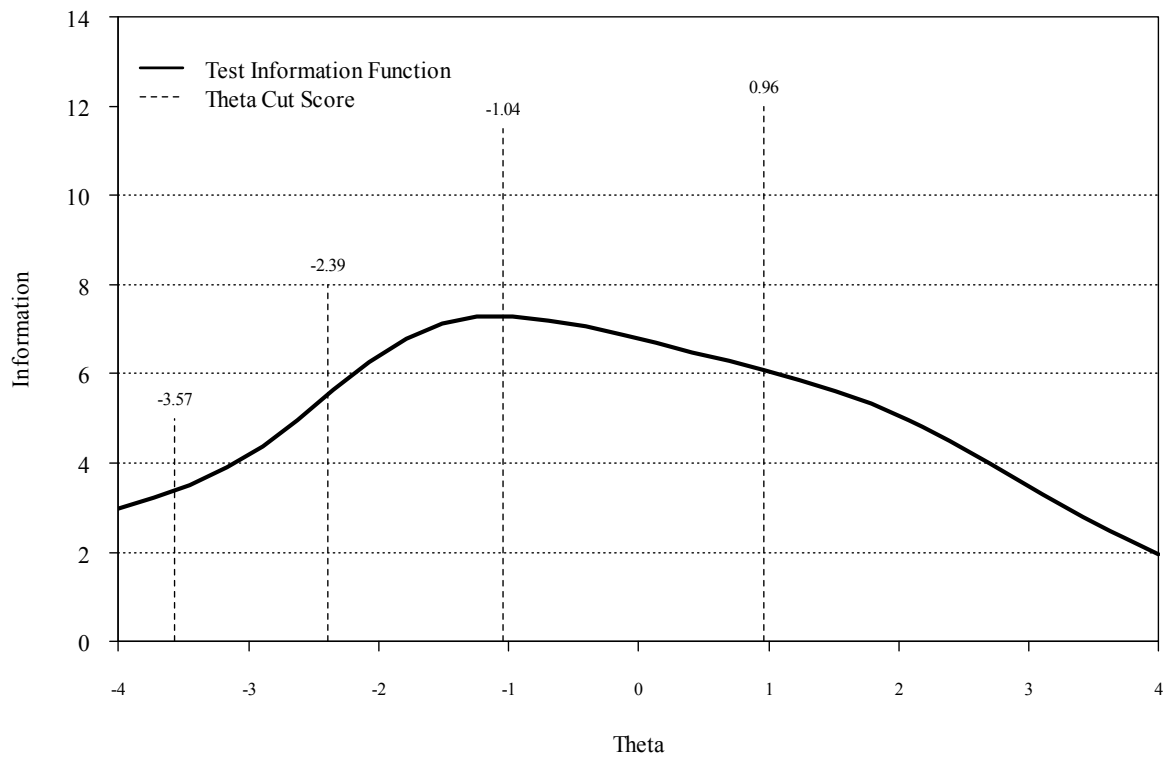


Figure 7.15 Test Information Function for Reading: Primary Division (French)

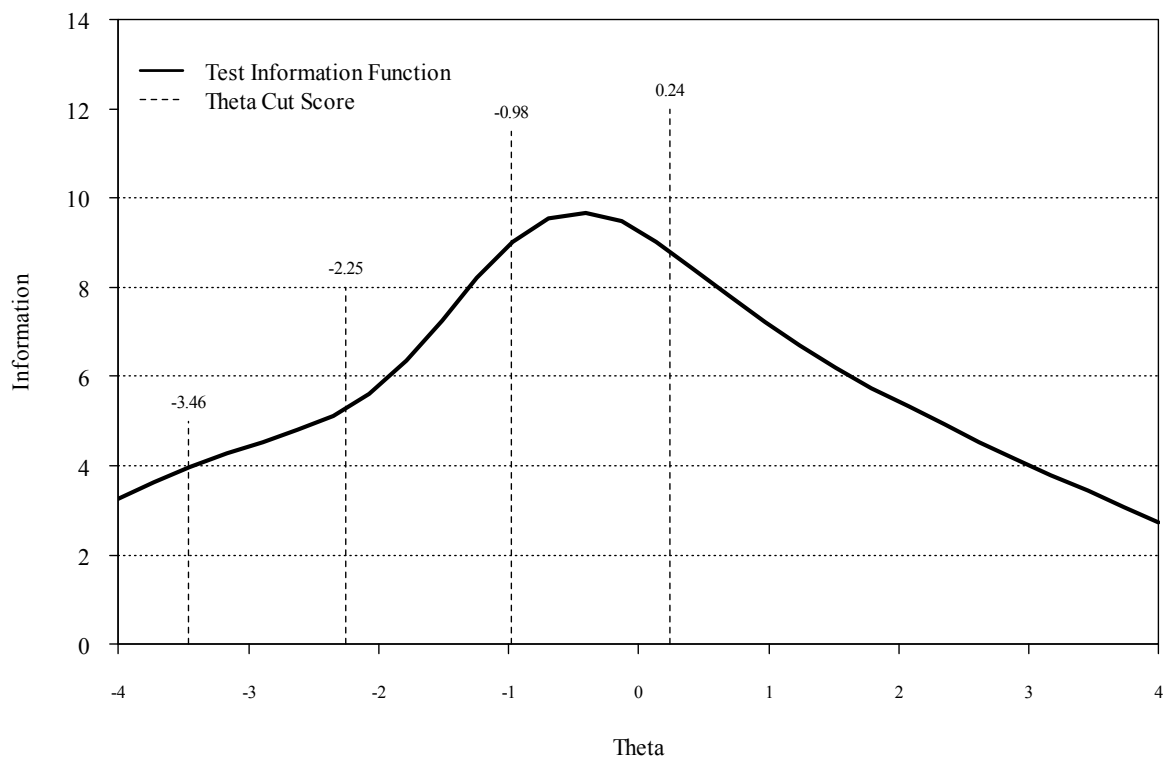


Figure 7.16 Test Information Function for Reading: Junior Division (French)

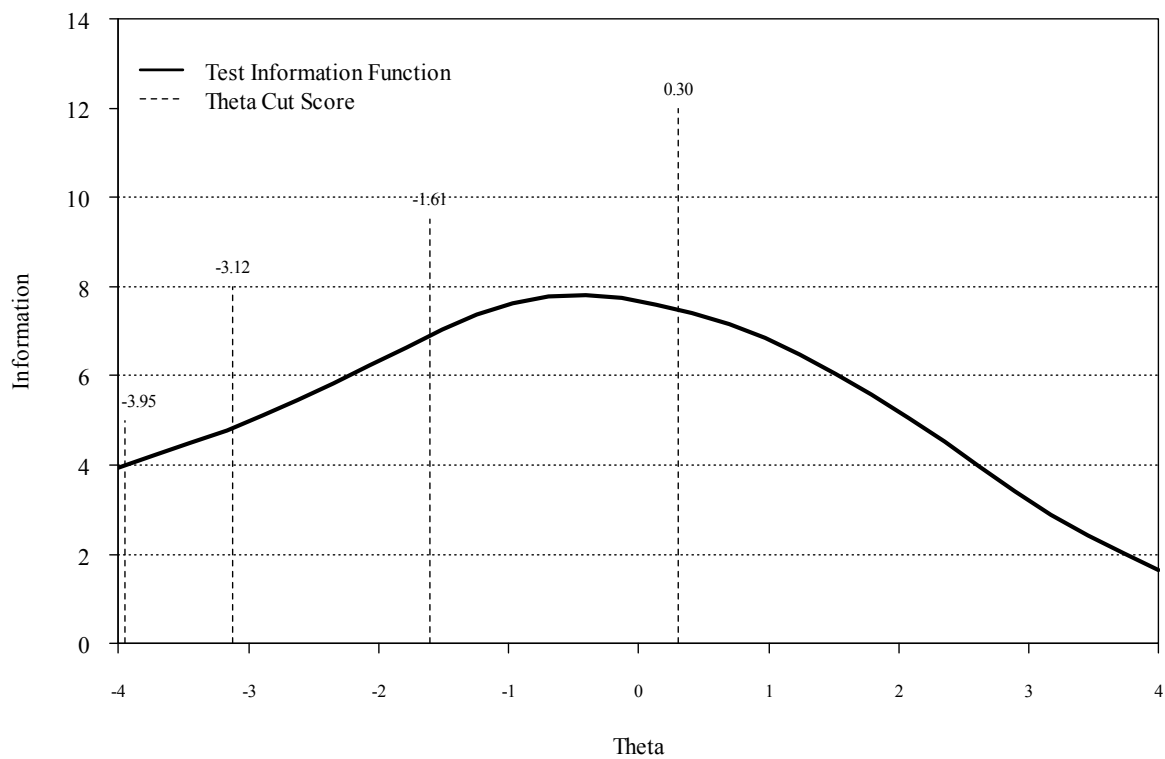


Figure 7.17 Test Information Function for Writing: Primary Division (English)

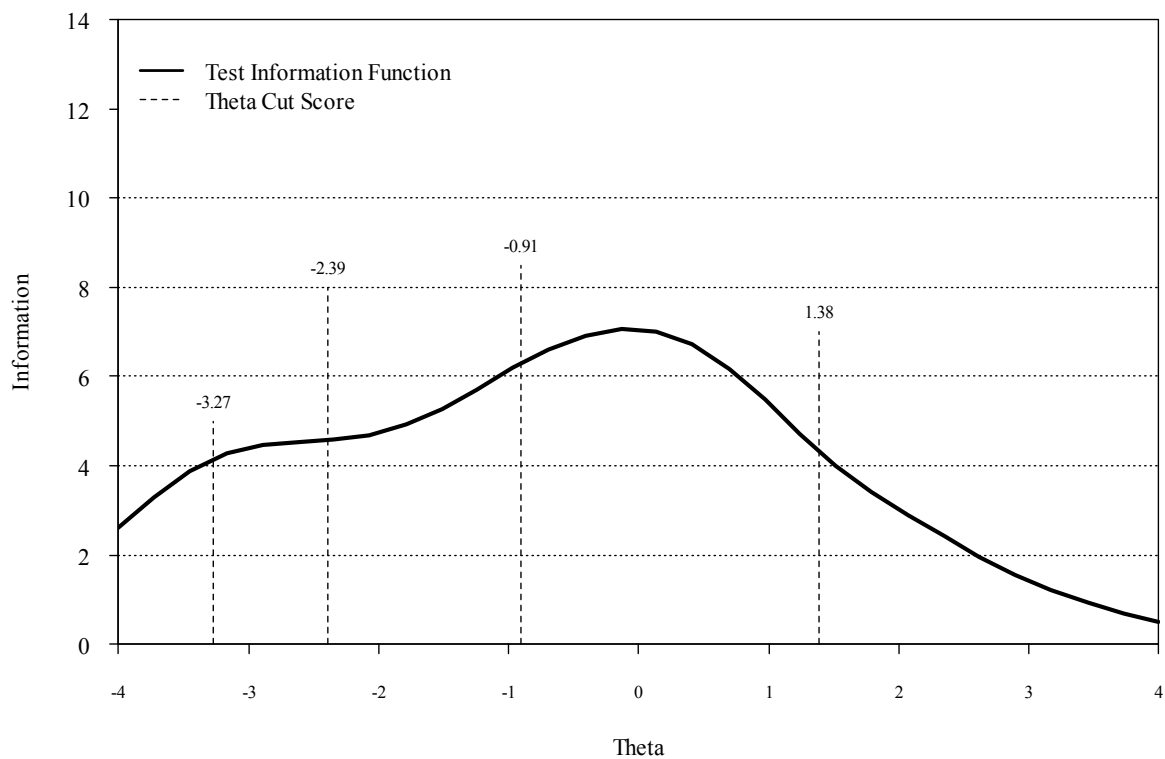


Figure 7.18 Test Information Function for Writing: Junior Division (English)

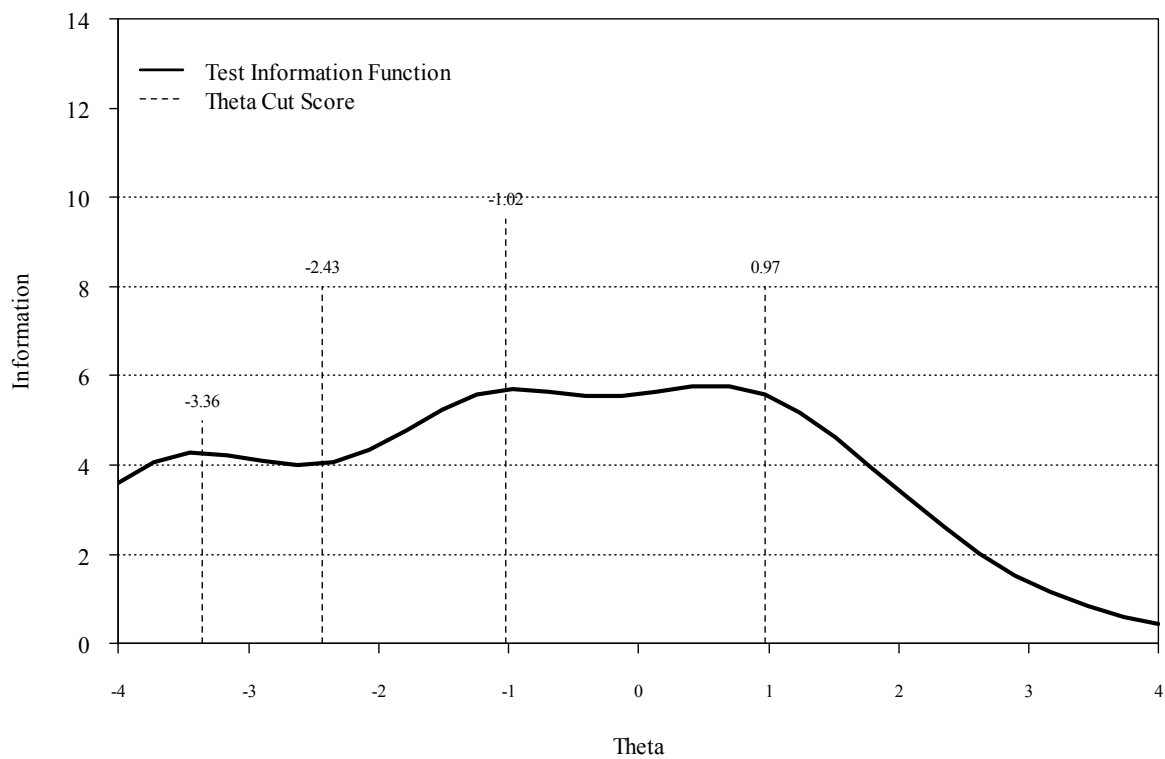


Figure 7.19 Test Information Function for Writing: Primary Division (French)

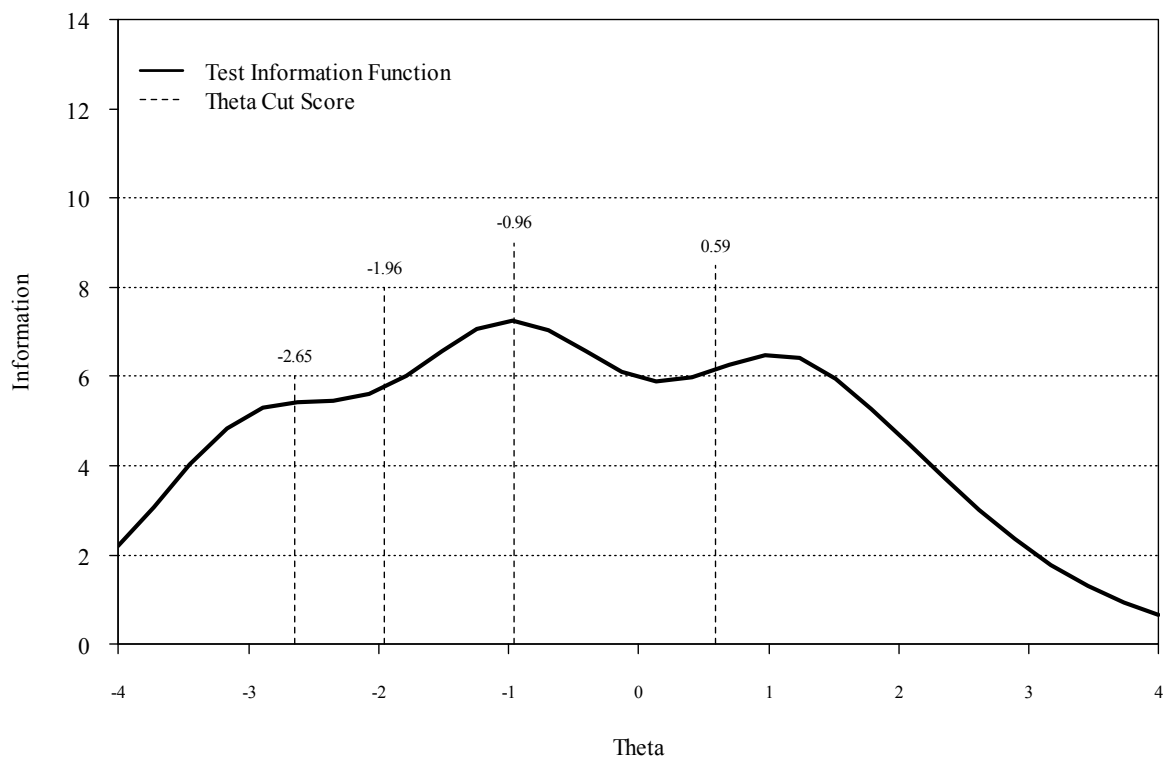


Figure 7.20 Test Information Function for Writing: Junior Division (French)

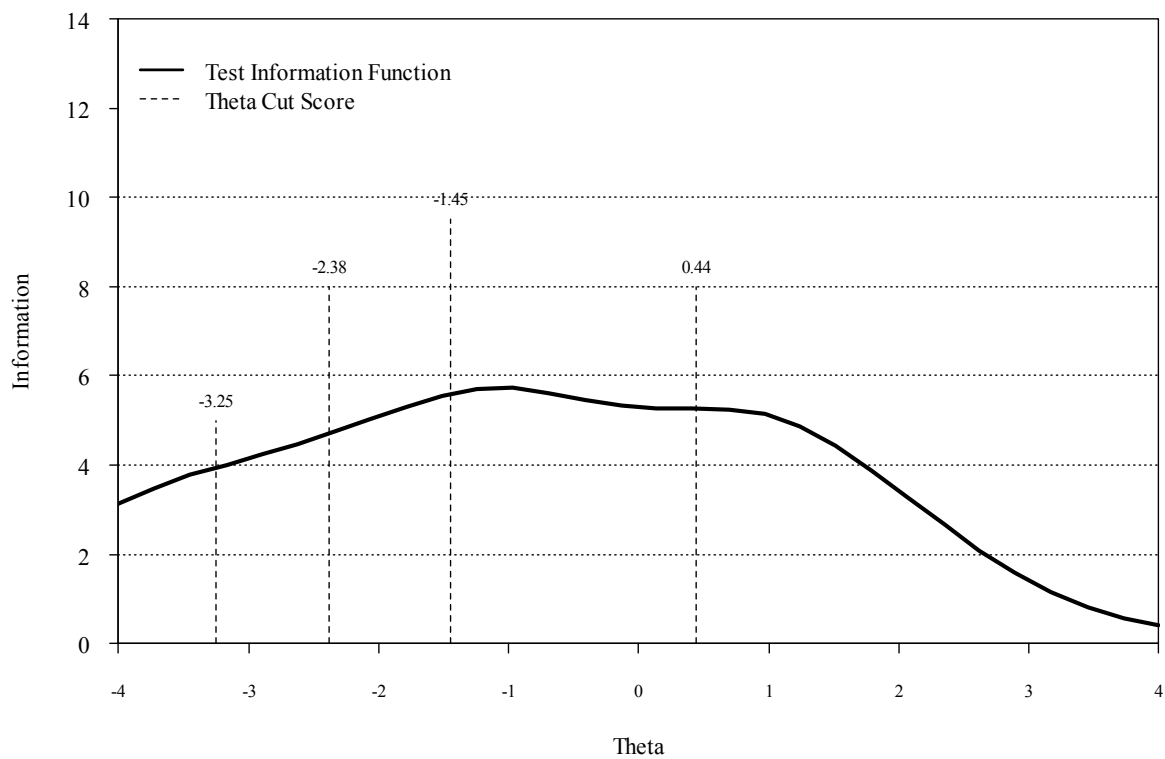


Figure 7.21 Test Information Function for Mathematics: Primary Division (English)

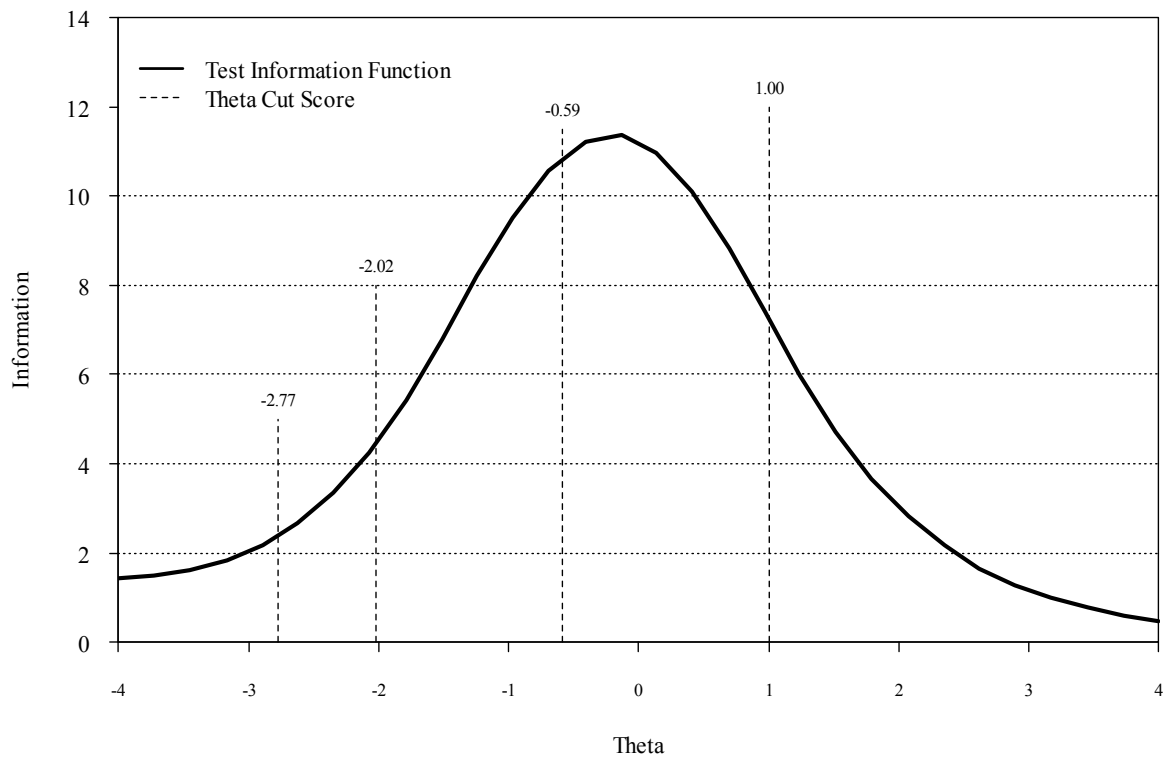


Figure 7.22 Test Information Function for Mathematics: Junior Division (English)

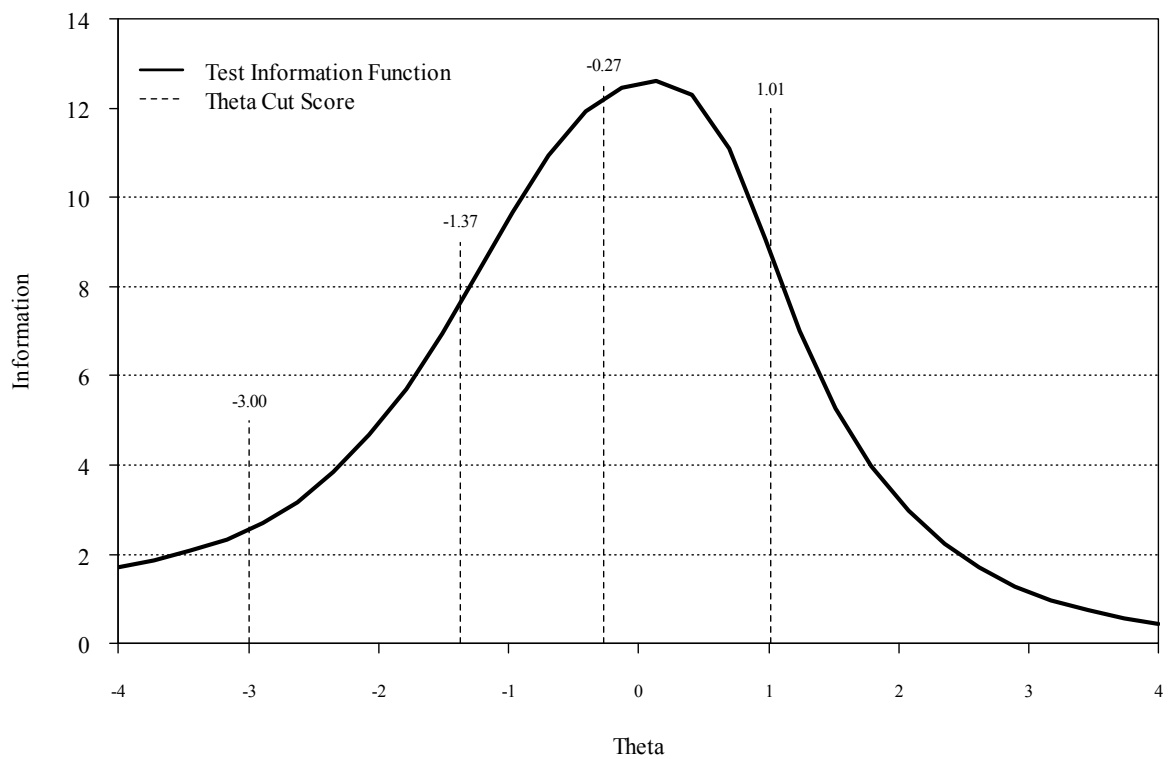


Figure 7.23 Test Information Function for Mathematics: Primary Division (French)

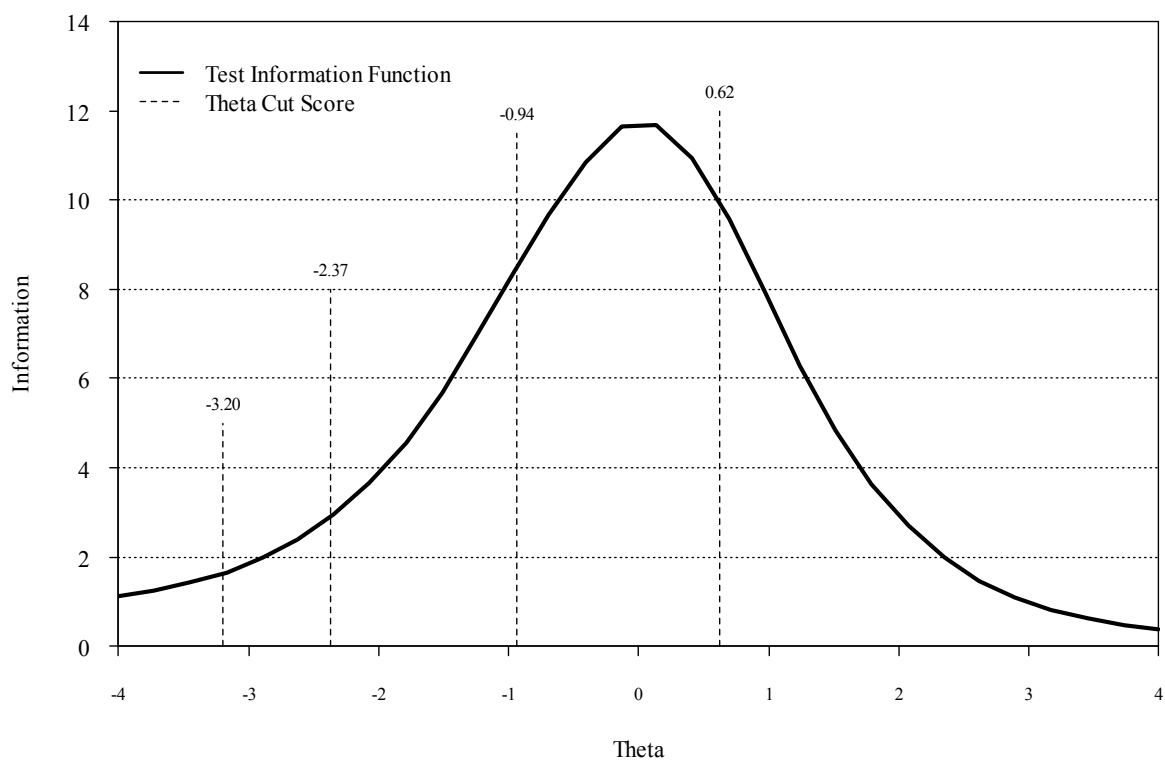
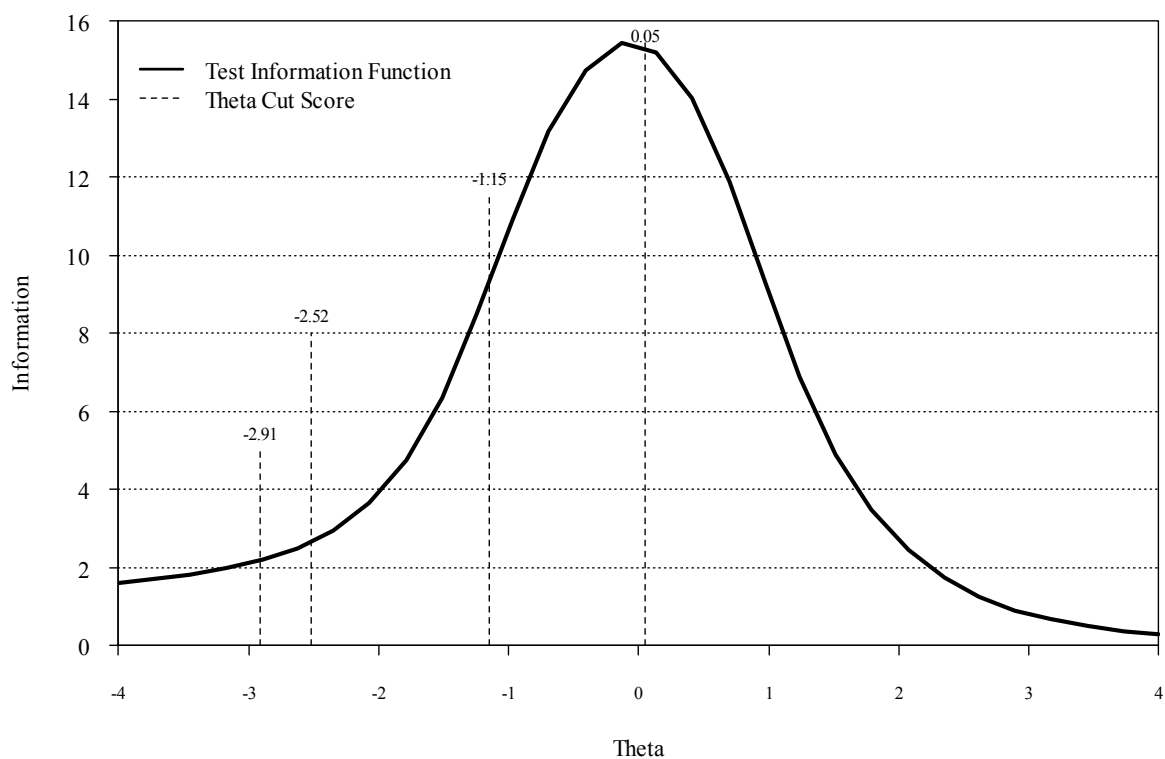


Figure 7.24 Test Information Function for Mathematics: Junior Division (French)



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)	35	Min.	0.20	0.19	0.36	-2.08
		Max.	0.92	0.53	1.49	2.05
		Mean	0.55	0.37 [†]	0.72	0.03
		SD	0.18	0.09	0.26	0.90
Junior Reading (English)	36	Min.	0.37	0.17	0.27	-3.33
		Max.	0.97	0.53	1.00	1.43
		Mean	0.70	0.34	0.57	-1.01
		SD	0.16	0.10	0.16	0.91
Primary Reading (French)	36	Min.	0.39	0.19	0.34	-2.60
		Max.	0.94	0.53	1.21	1.19
		Mean	0.65	0.39	0.71	-0.46
		SD	0.14	0.08	0.21	0.73
Junior Reading (French)	36	Min.	0.39	0.15	0.25	-3.29
		Max.	0.94	0.57	1.03	2.15
		Mean	0.69	0.35	0.57	-0.96
		SD	0.13	0.12	0.18	0.99

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.	0.43	0.22	0.34	-3.07
		Max.	0.90	0.62	1.08	1.38
		Mean	0.64	0.49 [†]	0.70	-0.74
		SD	0.13	0.16	0.25	1.26
Junior Writing (English)	14	Min.	0.55	0.26	0.33	-2.38
		Max.	0.78	0.64	1.13	-0.05
		Mean	0.63	0.49	0.66	-0.91
		SD	0.07	0.16	0.28	0.67
Primary Writing (French)	14	Min.	0.49	0.32	0.54	-1.10
		Max.	0.76	0.61	1.21	-0.03
		Mean	0.65	0.52	0.83	-0.64
		SD	0.09	0.12	0.23	0.39
Junior Writing (French)	14	Min.	0.53	0.31	0.48	-1.92
		Max.	0.83	0.62	1.13	-0.40
		Mean	0.69	0.51	0.74	-1.15
		SD	0.09	0.12	0.19	0.53
Primary Mathematics (English)	36	Min.	0.47	0.25	0.34	-2.47
		Max.	0.88	0.61	1.06	0.53
		Mean	0.66	0.41	0.68	-0.78
		SD	0.10	0.12	0.20	0.77
Junior Mathematics (English)	36	Min.	0.38	0.20	0.37	-1.89
		Max.	0.86	0.68	1.85	1.25
		Mean	0.63	0.41	0.70	-0.57
		SD	0.13	0.17	0.28	0.87
Primary Mathematics (French)	36	Min.	0.40	0.17	0.26	-3.77
		Max.	0.93	0.55	1.40	0.82
		Mean	0.66	0.40	0.73	-0.86
		SD	0.13	0.11	0.31	1.10
Junior Mathematics (French)	36	Min.	0.39	0.14	0.25	-3.13
		Max.	0.92	0.65	1.58	1.83
		Mean	0.64	0.44	0.83	-0.64
		SD	0.13	0.14	0.31	1.09

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 2013–2014 the mean percentages ranged from 58.50% (English, winter) to 59.81% (French, winter) for applied mathematics. The mean percentages ranged from 65.77% (French, spring) to 70.42% (French, winter) for academic mathematics.

Cronbach's alpha estimates ranged from 0.82 to 0.84 for applied mathematics and from 0.86 to 0.87 for academic mathematics. The corresponding SEMs were likewise similar, ranging from 7.18% to 7.36% of the possible maximum score for applied mathematics and from 6.64% to 7.08% 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	16 788	0	52	30.42	9.09	0.84	3.59
Applied, Spring (English)	31	24	7	52	19 970	0	52	30.46	9.11	0.84	3.66
Academic, Winter (English)	31	24	7	52	42 497	3	52	35.15	9.65	0.87	3.54
Academic, Spring (English)	31	24	7	52	52 681	1	52	35.90	9.33	0.86	3.50
Applied, Winter (French)	31	24	7	52	339	9	49	31.10	8.78	0.82	3.68
Applied, Spring (French)	30	23	7	51	1 170	8	50	31.06	8.72	0.84	3.50
Academic, Winter (French)	31	24	7	52	1 152	9	52	36.62	8.83	0.86	3.32
Academic, Spring (French)	31	24	7	52	2 886	6	52	34.20	9.44	0.86	3.54

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 -0.05 to -0.10 for the applied version of the mathematics assessment and from -0.01 to 0.16 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)	16 788	-3.45	2.84	-0.07	0.97
Applied, Spring (English)	19 970	-3.49	2.74	-0.05	0.96
Academic, Winter (English)	42 497	-3.53	2.17	-0.05	0.96
Academic, Spring (English)	52 681	-3.58	2.25	0.01	0.94
Applied, Winter (French)	339	-2.75	2.12	-0.10	0.93
Applied, Spring (French)	1 170	-2.65	2.66	-0.04	0.95
Academic, Winter (French)	1 152	-2.85	2.34	0.16	0.94
Academic, Spring (French)	2 886	-3.39	2.36	-0.08	0.94

Note. SD = standard deviation.

The TCCs and the distribution of student thetas are displayed in Figures 7.25 to 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 applied version provided most of its information around the 2/3 cut points, whereas the academic version provided most of its 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)

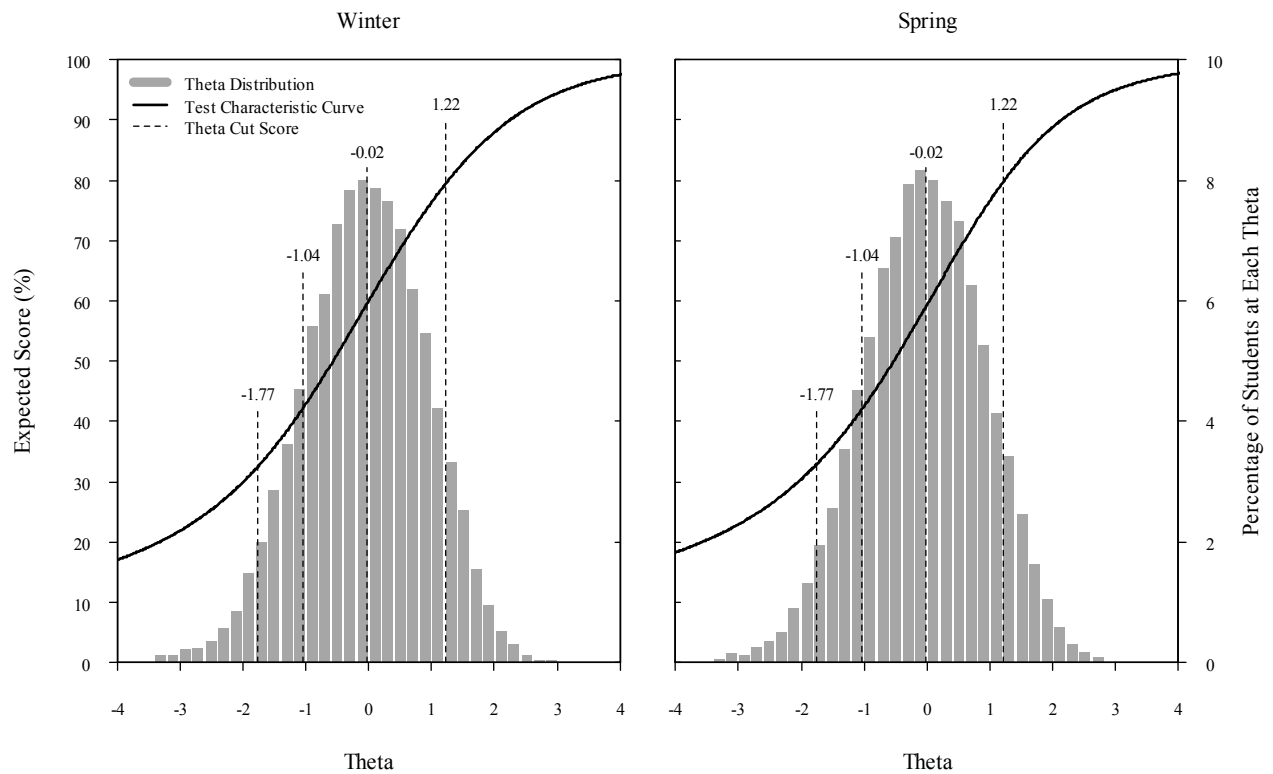


Figure 7.26 Test Characteristic Curves and Distributions of Thetas: Grade 9 Academic Math (English)

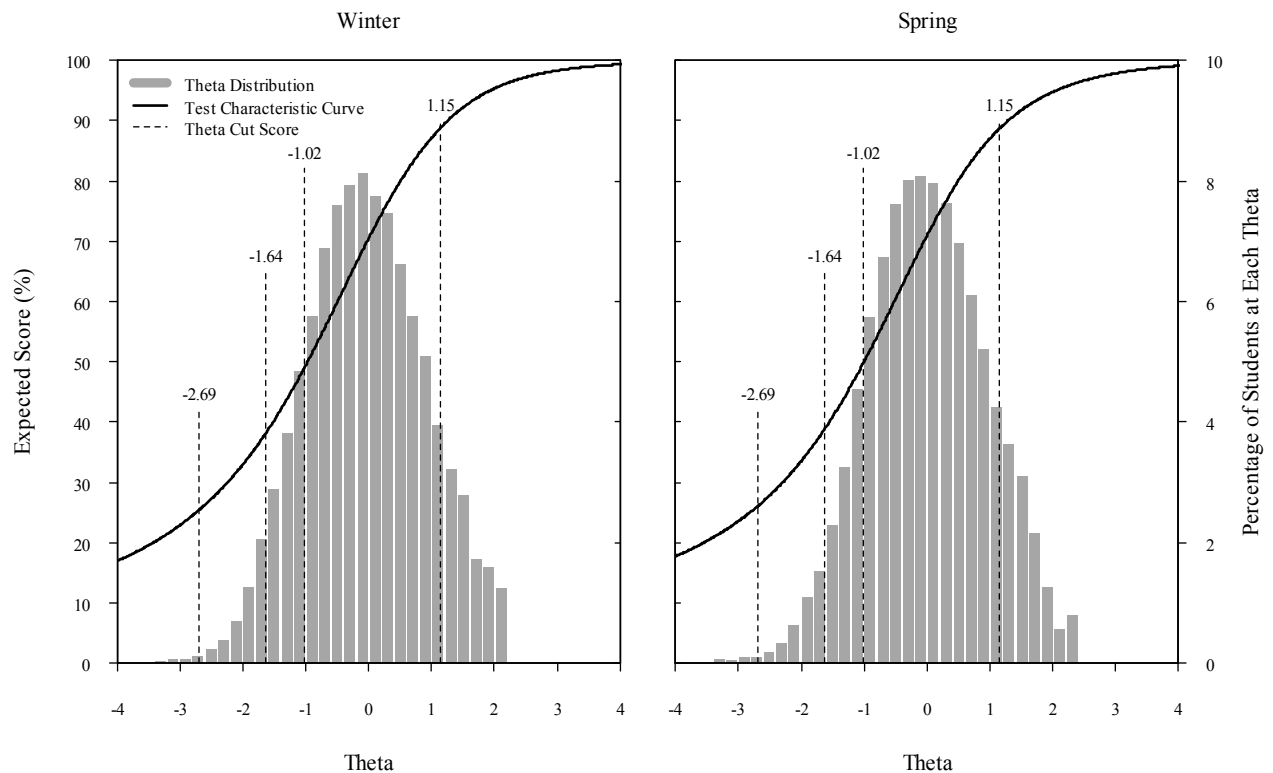


Figure 7.27 Test Characteristic Curves and Distributions of Thetas: Grade 9 Applied Math (French)

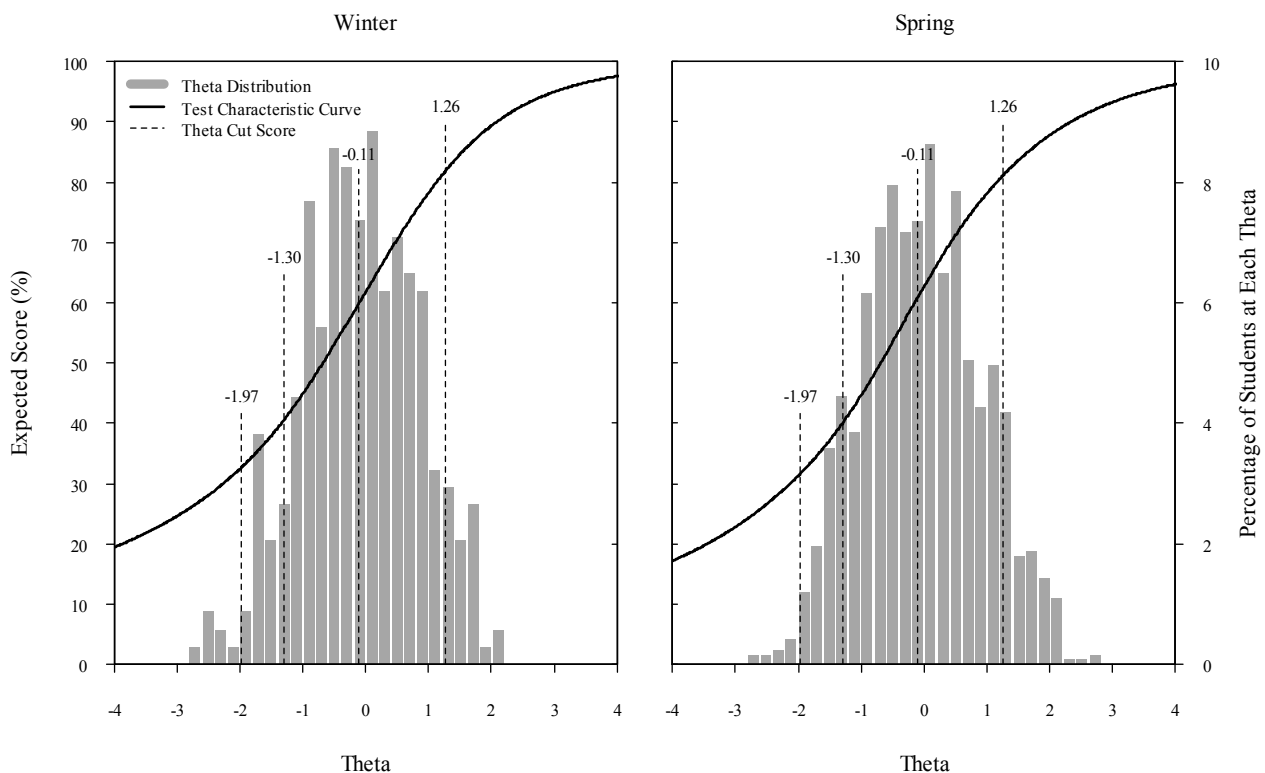


Figure 7.28 Test Characteristic Curves and Distributions of Thetas: Grade 9 Academic Math (French)

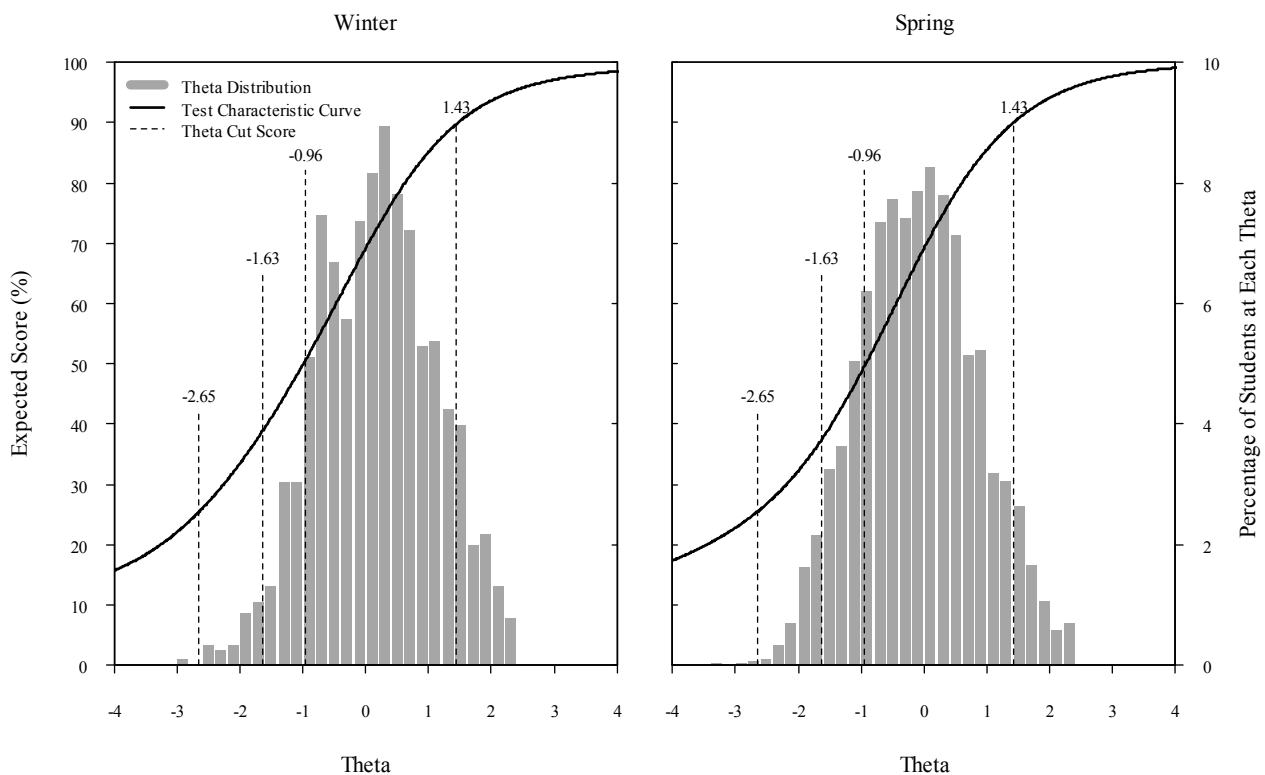


Figure 7.29 Test Information Functions: Grade 9 Applied Math (English)

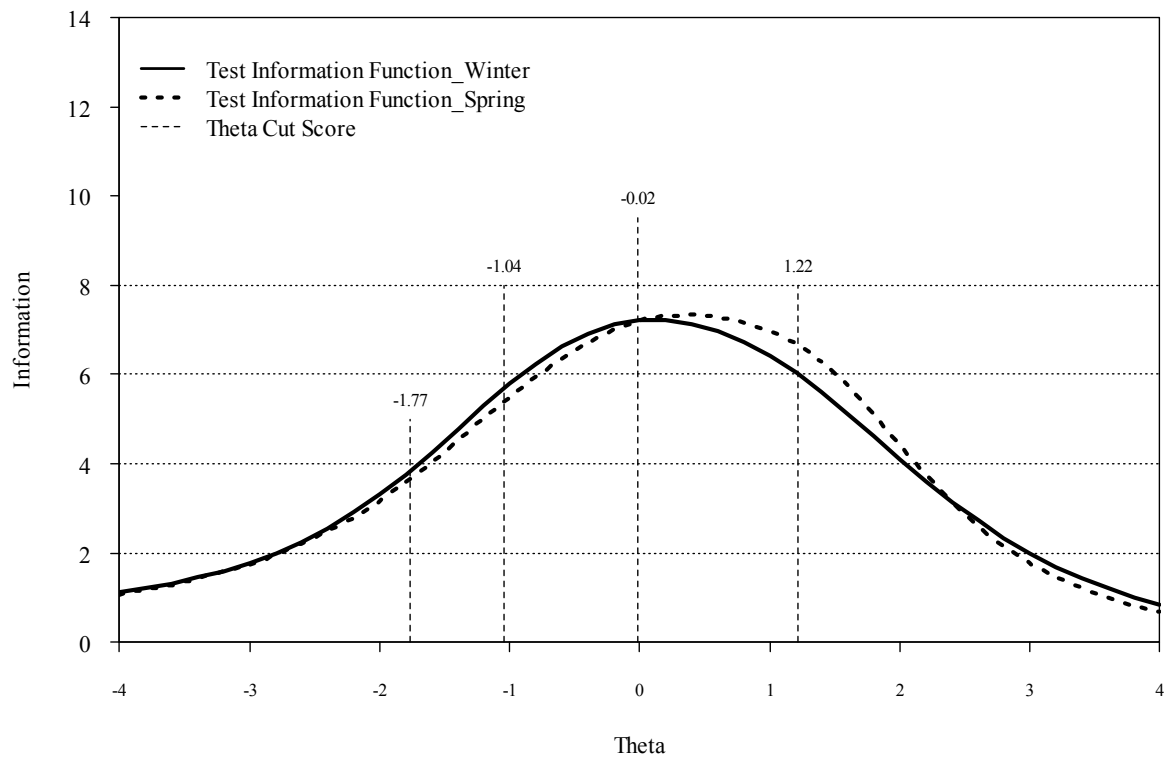


Figure 7.30 Test Information Functions: Grade 9 Academic Math (English)

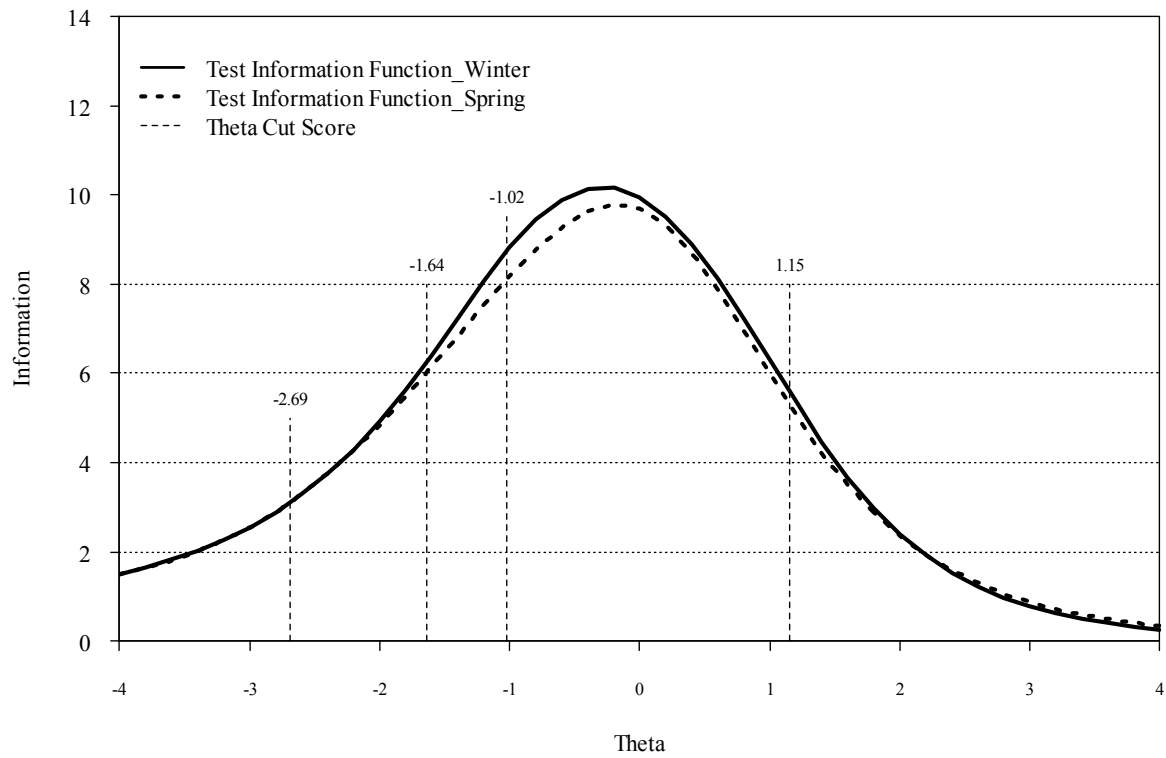


Figure 7.31 Test Information Functions: Grade 9 Applied Math (French)

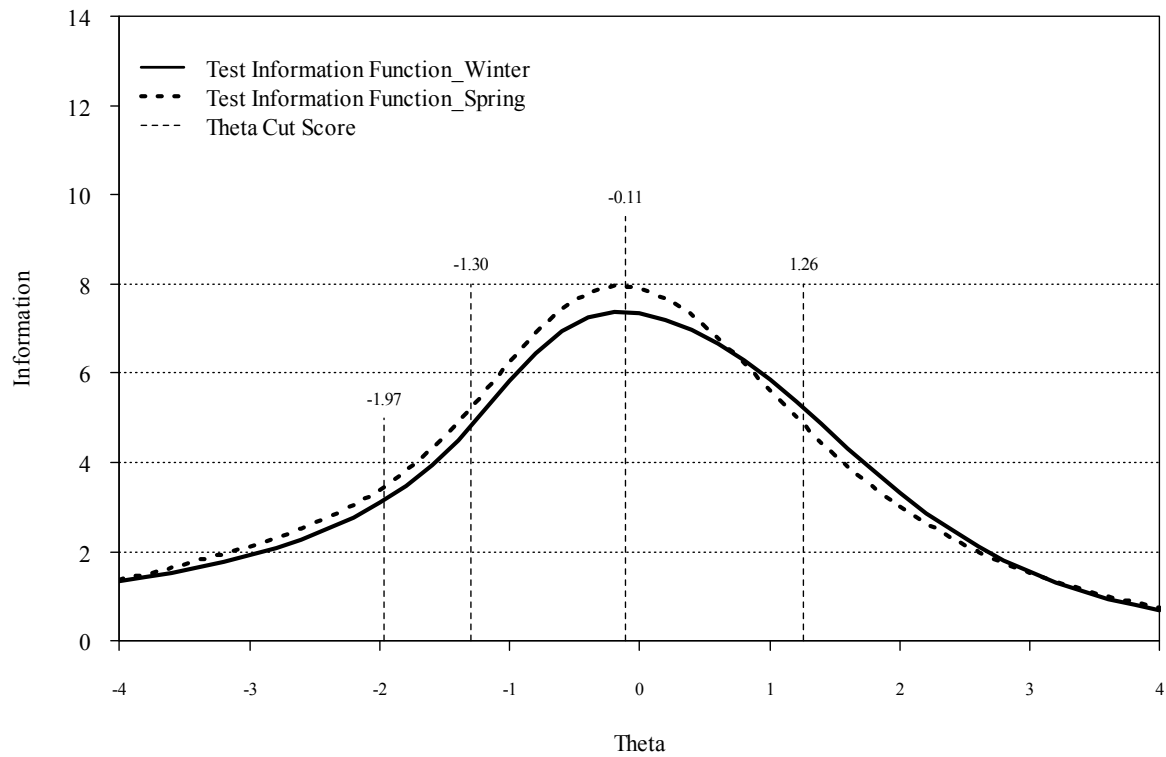
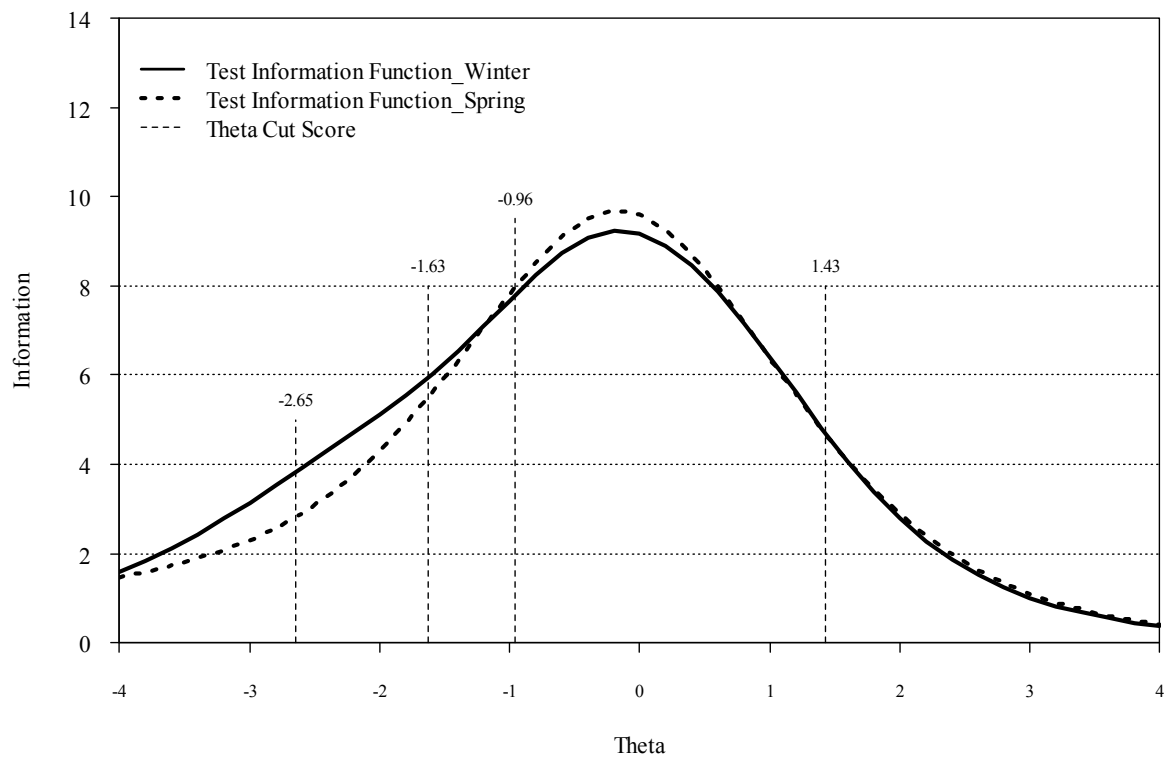


Figure 7.32 Test Information Functions: Grade 9 Academic Math (French)



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)		Min.	30.70	0.16	-2.35	0.29
	31	Max.	86.19	0.62	1.80	1.04
		Mean	59.56	0.37†	-0.19	0.61
		SD	15.20	0.12	1.03	0.19
Applied, Spring (English)		Min.	28.63	0.14	-2.59	0.25
	31	Max.	83.59	0.63	1.54	1.33
		Mean	59.20	0.36	-0.26	0.60
		SD	15.20	0.13	1.09	0.25
Academic, Winter (English)		Min.	45.05	0.24	-2.30	0.38
	31	Max.	90.30	0.71	0.70	1.14
		Mean	69.47	0.42	-0.81	0.68
		SD	12.37	0.13	0.74	0.22
Academic, Spring (English)		Min.	47.33	0.23	-2.30	0.31
	31	Max.	91.94	0.65	0.70	1.05
		Mean	69.54	0.41	-0.77	0.66
		SD	10.87	0.12	0.73	0.21
Applied, Winter (French)		Min.	33.45	0.07	-4.30	0.11
	31	Max.	90.20	0.69	1.64	1.37
		Mean	59.62	0.34	-0.38	0.54
		SD	13.37	0.15	1.25	0.23
Applied, Spring (French)		Min.	37.07	0.12	-2.28	0.19
	30	Max.	89.28	0.68	1.95	1.23
		Mean	61.20	0.37	-0.21	0.59
		SD	13.49	0.14	1.02	0.25
Academic, Winter (French)		Min.	42.22	0.08	-1.89	0.14
	31	Max.	88.06	0.64	2.01	1.08
		Mean	67.70	0.40	-0.36	0.65
		SD	10.71	0.12	0.87	0.21
Academic, Spring (French)		Min.	41.14	0.26	-2.03	0.28
	31	Max.	85.89	0.66	1.08	1.02
		Mean	62.61	0.41	-0.43	0.67
		SD	11.96	0.12	0.88	0.19

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 81.0% for English-language students and 77.4% for French-language students.

Cronbach's alpha estimates are 0.89 and 0.88 for English and French, respectively. The corresponding SEMs are 3.8% 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	46	38	4	2	2	80	131 712	1.0	80.0	64.8	9.07	0.89	3.03
French	47	39	4	2	2	81	5 165	10.0	81.0	62.7	9.16	0.88	3.22

Note. MC = multiple choice; OR = open response (reading); SW = short writing; LW = long writing; SD = standard deviation; R = Cronbach's alpha; 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. 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	131 712	-3.88	2.71	0.22	0.91
French	5 165	-3.77	2.95	0.29	0.89

The TCCs and the distribution of student thetas are displayed in Figures 7.33 and 7.34 for the English-language and French-language students, respectively. The TCCs follow the expected S-shaped distribution. The distribution of student thetas is plotted on the TCC graphs, with the right vertical scale indicating the percentage of students at each theta value. The TIF plots for the English-language and French-language 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)

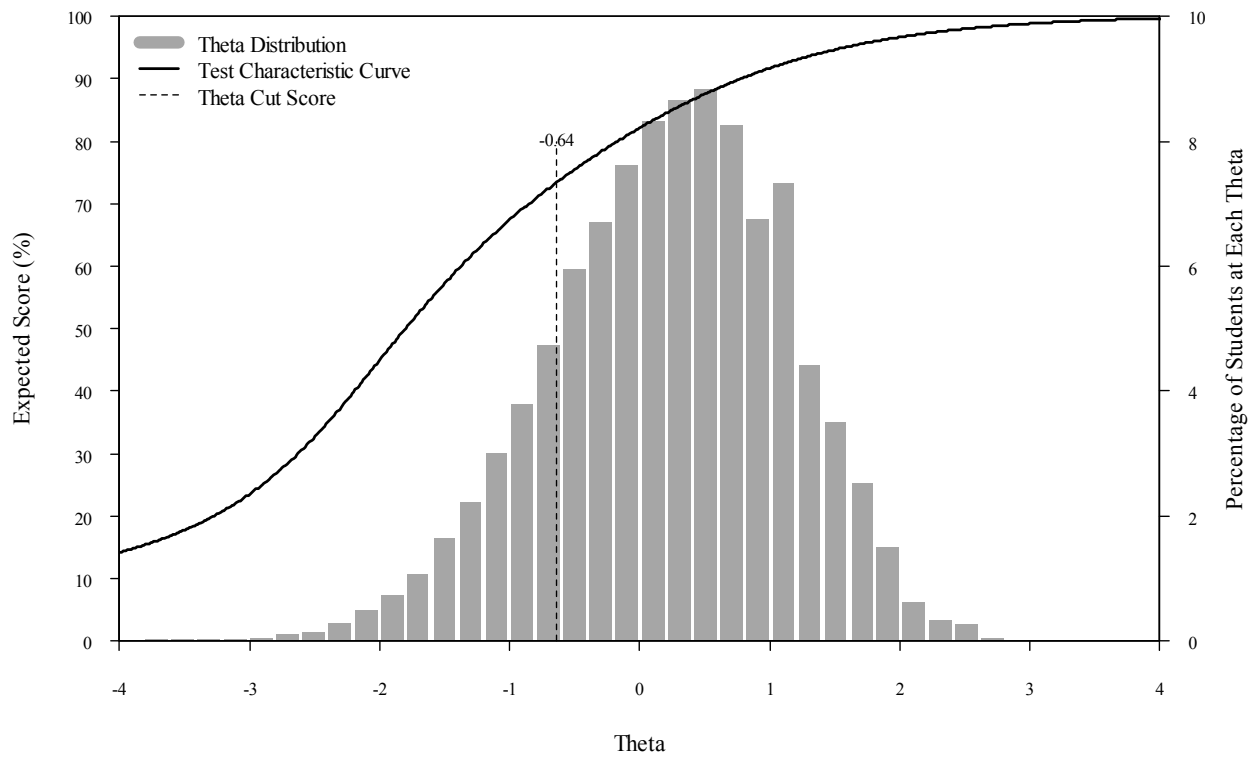


Figure 7.34 Test Characteristic Curve and Distribution of Theta: OSSLT (French)

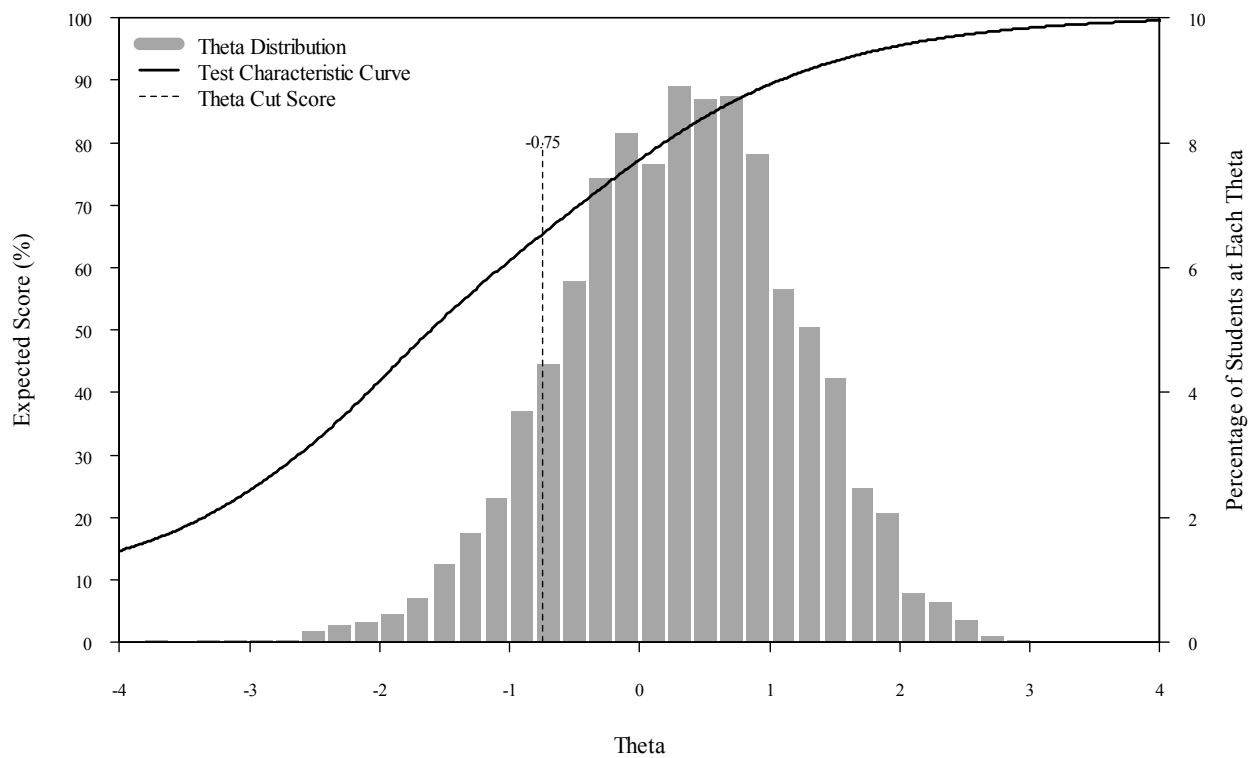


Figure 7.35 Test Information Function: OSSLT (English)

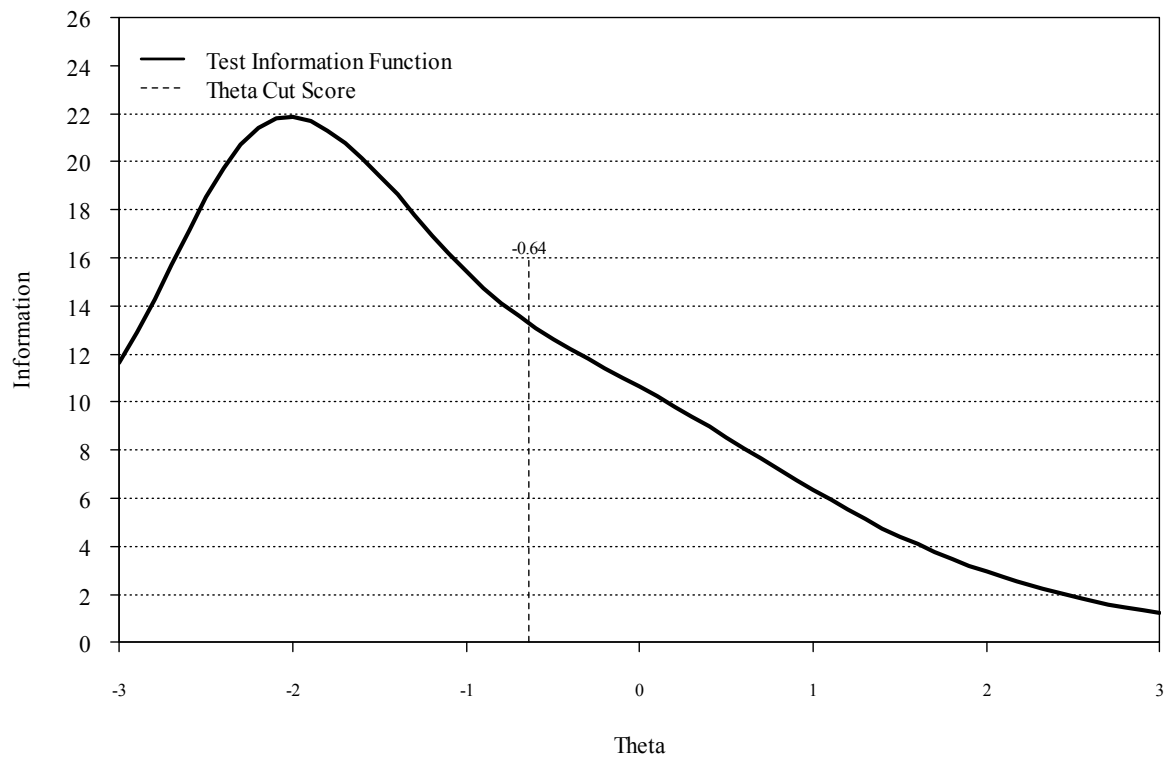
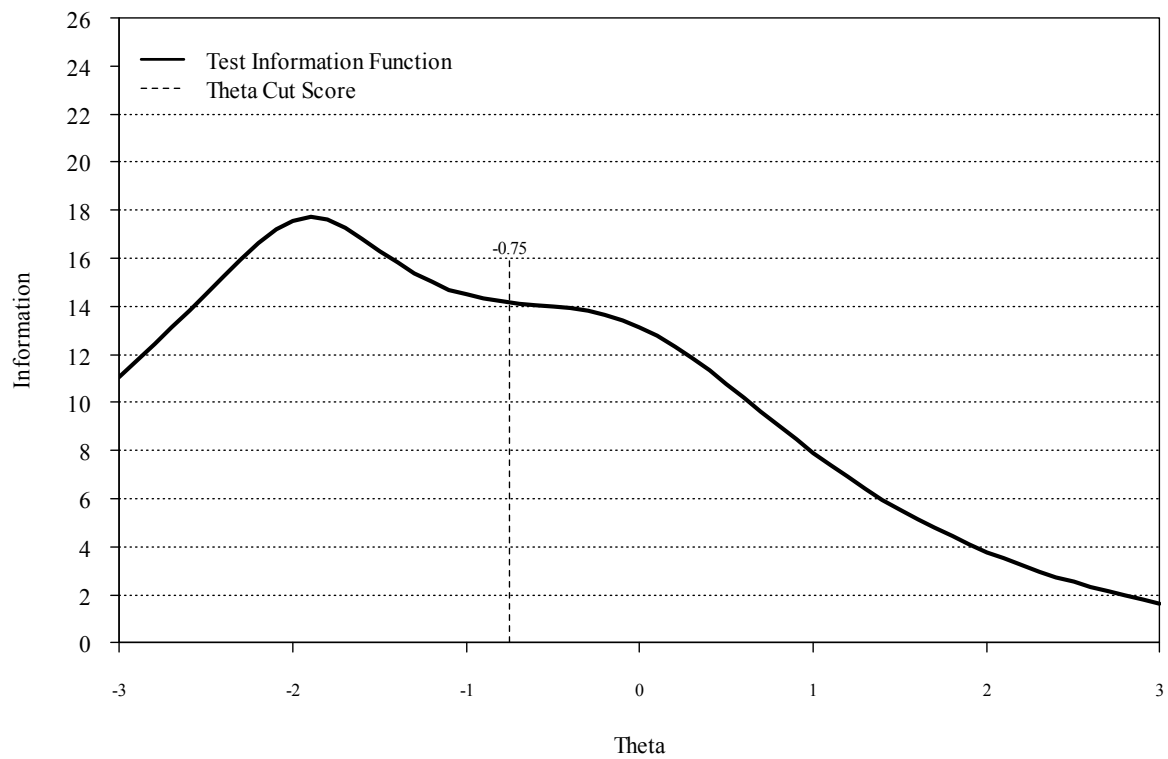


Figure 7.36 Test Information Function: OSSLT (French)



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 value for the minimum point-biserial correlation for the English- and French-language tests suggests 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	46	Min.	0.58	0.15	-2.98
		Max.	0.96	0.59	0.23
		Mean	0.81	0.35 [†]	-1.35
		SD	0.10	0.09	0.80
French	47	Min.	0.48	0.16	-3.05
		Max.	0.96	0.58	1.12
		Mean	0.77	0.33 [†]	-1.03
		SD	0.11	0.10	0.94

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 (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 2013–2014 assessments for gender- and 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 Multiple-Choice 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 multiple-choice 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, Canada, 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 Multiple-Choice 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 open-response items, the SMD between the reference and focal groups was used in conjunction with the MH approach. The SMD compares the means of the reference and focal groups, adjusting for the differences in the distribution of the reference- and focal-group members across the values of the matching variable. The SMD has the following form:

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

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

Table 7.12 DIF Classification Rules for Open-Response Items

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

For each assessment, except for the French-language Grade 9 version, two random samples of 2000 examinees were selected from the provincial student population. 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.

The Primary- and Junior-Division Assessments

For the reading, writing and mathematics components of the 2013–2014 primary- and junior-division assessments for both languages, the summaries of the number of items that showed statistically significant gender-based DIF with at least a B-level or C-level effect size in both the two samples are reported in Tables 7.13 and 7.14, respectively. The numbers in the “boys” and “girls” columns indicate the number of DIF items favouring boys and girls.

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 k = 36*	1 (MC) 0 (OR)	0 (MC) 0 (OR)	2 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (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)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)
Math k = 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)	2 (MC) 0 (OR)	1 (MC) 0 (OR)

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

* For reading component of the English-language primary-division assessment, $k = 35$, as one item was dropped.

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 k = 36*	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Writing 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)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	2 (MC) 0 (OR)	0 (MC) 0 (OR)

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

* For the reading component of the English-language primary-division assessment, $k = 35$, as one item was dropped.

Of the 343 items comprising the primary- and junior-division assessments (one item was dropped from the reading component of the assessment), 15 items were found to have gender-based DIF. The majority of these items had B-level DIF (11), and only four items had C-level DIF. Of the 15 multiple-choice items showing B- or C-level DIF, 12 favoured the boys, while three favoured the girls. No open-response items were found to show DIF. Overall, more items were found to favour boys than girls. The mathematics component of the French-language junior-division assessment had the largest number of gender-based DIF items (three).

The summaries for SLL-based DIF are reported in Tables 7.15 and 7.16. Six out of the 343 items across all the assessments showed SLL-based DIF, with five having B-level DIF and one having C-level DIF. Five multiple-choice DIF items favoured non-SLL students, and one multiple-choice DIF item favoured SLL students. No open-response item showed SLL-based DIF. The reading component of the English-language junior-division assessment was found to have the largest number of SLL-based DIF items (three).

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)	3 (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)	1 (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)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)

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

* For the reading component of the English-language primary-division assessment, $k = 35$, as one item was dropped.

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

Component	Primary English		Junior English		Primary French		Junior French	
	Non-SLLs	SLLs	Non-SLLs	SLLs	Non-SLLs	SLLs	Non-SLLs	SLLs
Reading k = 36*	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Writing 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.

* For the reading component of the English-language primary-division assessment, $k = 35$, as one item was dropped.

All items identified as having B-level or C-level DIF on the primary- and junior-division assessments were reviewed by the assessment team. Since the reviewers did not identify any apparent bias in the content of these items, they were not removed from the calibration, equating and scoring processes.

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 k = 31	0 (MC)	0 (MC)	4 (MC)	2 (MC)
	0 (OR)	0 (OR)	1 (OR)	0 (OR)
Applied, Spring k = 31	1 (MC)	1 (MC)	2 (MC)	1 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Winter k = 31	2 (MC)	0 (MC)	2 (MC)	1 (MC)
	0 (OR)	0 (OR)	1 (OR)	0 (OR)
Academic, Spring k = 31	0 (MC)	1 (MC)	0 (MC)	0 (MC)
	0 (OR)	1 (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 k = 31	0 (MC)	0 (MC)	0 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Applied, Spring k = 31	0 (MC)	0 (MC)	0 (MC)	0 (MC)
	0 (OR)	0 (OR)	1 (OR)	0 (OR)
Academic, Winter k = 31	0 (MC)	0 (MC)	1 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Spring k = 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 247 items across eight Grade 9 math assessments (one item was dropped from one assessment), 17 multiple-choice items showed B-level gender-based DIF. Of the 17 B-level gender-based multiple-choice DIF items, 11 favoured boys and 6 favoured girls. Three open-response items showed B-level gender-based DIF, with two favouring boys and one favouring girls. Two multiple-choice items showed C-level gender-based DIF: all favoured boys. One open-response item showed C-level gender-based DIF, which favoured boys. Overall, more multiple-choice items were found to favour boys.

Across the Grade 9 math assessments, 8–10 multiple-choice items were used in both versions of the winter and spring administrations. Eight items that were repeated across eight assessments showed gender-based DIF. Of the eight items, three showed DIF in the winter administration, three showed DIF in the spring administration, and two were found having DIF across both administrations.

SLL-based DIF was conducted only for the English-language courses. There are eight multiple-choice B-level SLL-based DIF items. Of these items, four items favoured SLL students, and four favoured non-SLL students. Three open-response items showed B-level SLL-based DIF. One favoured SLL students, and two favoured non-SLL students. Two multiple-choice items showed C-level SLL-based DIF, and all favoured SLL students. No open-response items showed C-level SLL-based DIF.

Two repeated items showed SLL-based DIF; one item showed DIF in the winter administration, and one item showed DIF in the spring administration. Overall, more SLL-based DIF items were found to favour SLL students. More SLL-based DIF items were found for the English-language academic course.

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	1 (MC)	1 (MC)	1 (MC)	2 (MC)
	0 (OR)	0 (OR)	0 (OR)	1 (OR)
Spring k = 31	0 (MC)	1 (MC)	2 (MC)	0 (MC)
	0 (OR)	1 (OR)	1 (OR)	0 (OR)

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

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

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

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

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 calibration, equating and scoring processes.

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 = 50)		French (k = 51)	
Males	Females	Males	Females
8 (MC)	1 (MC) 1 (OR) 1 (SW)	5 (MC) 1 (OR)	1 (OR) 1 (LW)

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

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

English (k = 50)		French (k = 51)	
Males	Females	Males	Females
1 (MC)	0	1 (MC)	0

Note. MC = multiple choice.

There were 11 B-level DIF items on the English-language version of the OSSLT. Eight multiple-choice items favoured the males, and one multiple-choice and one open-response item favoured the females. One short-writing item favoured the females for topic development. One multiple-choice item exhibited C-level DIF.

There were eight B-level DIF items on the French-language version of the OSSLT. Five multiple-choice items favoured the males. One open-response reading item favoured the males, and one open-response reading item favoured the females. One long-writing item favoured the females for use of conventions. One C-level DIF multiple-choice item favoured the males.

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

Three 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 calibration, equating and scoring processes.

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

English (k = 50)			
B-Level DIF		C-Level DIF	
Non-SLLs	SLLs	Non-SLLs	SLLs
6 (MC)	1 (OR) 1 (SW) 1 (LW)	3 (MC)	0

Note. MC = multiple choice; OR = open response; SW = short writing; 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)	0.81	0.74	0.90	0.87
Junior Reading (English)	0.84	0.78	0.92	0.88
Primary Reading (French)	0.82	0.74	0.93	0.89
Junior Reading (French)	0.86	0.80	0.95	0.92
Primary Writing (English)	0.84	0.78	0.89	0.84
Junior Writing (English)	0.81	0.73	0.89	0.84
Primary Writing (French)	0.79	0.70	0.90	0.86
Junior Writing (French)	0.82	0.74	0.92	0.89
Primary Mathematics (English)	0.82	0.75	0.90	0.87
Junior Mathematics (English)	0.79	0.71	0.91	0.87
Primary Mathematics (French)	0.82	0.75	0.91	0.88
Junior Mathematics (French)	0.84	0.77	0.92	0.90

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	0.83	0.83
Applied, Spring (English)	0.70	0.59	0.88	0.83
Academic, Winter (English)	0.83	0.76	0.92	0.89
Academic, Spring (English)	0.84	0.78	0.93	0.90
Applied, Winter (French)	0.73	0.63	0.86	0.81
Applied, Spring (French)	0.75	0.64	0.88	0.83
Academic, Winter (French)	0.87	0.82	0.94	0.92
Academic, Spring (French)	0.85	0.80	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	0.94	0.92

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, & 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, 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 Grade 9 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 or 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 on 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 eight 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 items. The aggregate exact-plus-adjacent validity estimates for the items in each component ranged from 98.6 to 99.7%. For Grade 9 mathematics, the EQAO target of 95% exact-plus-adjacent agreement was met for all but one item in the winter administration of the English-language academic version of the assessment. The aggregate validity estimates ranged from 98.6 to 100%. For the English-language version of the OSSLT, the EQAO target of 95% exact-plus-adjacent agreement was met for all but one short-writing item. The aggregate validity estimates ranged from 97.8 to 99.3% (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 the Grade 9 mathematics assessment, and 0.88 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 0.89 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 all the 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
20203	1(A)	5	1 498	97.0	77.1	19.9	10.2	9.7	3.0
20183	1(A)	6	1 602	97.8	68.9	28.9	15.2	13.7	2.1
20225	1(A)	11	1 518	94.6	81.6	13.0	11.9	1.1	5.5
20229	1(A)	12	1 342	99.9	82.9	17.0	6.7	10.3	0.1
20060	1(B)	5	1 514	99.8	88.3	11.5	3.8	7.7	0.1
20061	1(B)	6	1 626	98.8	73.4	25.4	13.3	12.1	1.2
19924	NR	NR	1 354	100.0	88.6	11.4	8.3	3.1	0.0
19925	NR	NR	1 472	99.6	83.8	15.8	13.5	2.3	0.5
19933	NR	NR	1 443	99.5	89.9	9.6	9.3	0.3	0.5
19935	NR	NR	1 469	99.2	79.6	19.6	8.0	11.6	0.7
Aggregate			14 838	98.6	81.1	17.4	10.1	7.3	1.4

Note. NR = not released.

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
19955	1(A)	5	1 779	98.8	78.9	19.9	15.5	4.4	1.3
19956	1(A)	6	1 731	99.1	69.8	29.3	12.9	16.4	0.9
20005	1(A)	11	1 700	99.0	79.2	19.8	9.3	10.5	1.0
20004	1(A)	12	1 615	99.3	81.0	18.3	10.6	7.7	0.7
20032	1(B)	5	1 670	98.9	80.7	18.2	9.9	8.3	1.0
20031	1(B)	6	1 690	98.9	78.9	20.0	12.0	8.0	1.1
19747	NR	NR	1 799	99.5	79.3	20.2	11.3	8.9	0.4
19750	NR	NR	1 705	98.8	68.9	29.9	18.9	11.0	1.2
19968	NR	NR	1 646	95.6	73.1	22.5	15.2	7.3	4.4
19965	NR	NR	1 803	98.6	79.9	18.7	11.6	7.1	1.4
Aggregate			17 138	98.7	77.0	21.7	12.7	9.0	1.3

Note. NR = not released.

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
19126	1(A)	5	287	100.0	89.9	10.1	9.8	0.3	0.0
19127	1(A)	6	268	99.2	79.1	20.1	17.5	2.6	0.7
19132	1(A)	11	260	100.0	98.8	1.2	1.2	0.0	0.0
19133	1(A)	12	319	99.4	92.2	7.2	5.6	1.6	0.6
11217	1(B)	5	215	100.0	93.0	7.0	4.7	2.3	0.0
11218	1(B)	6	235	100.1	97.9	2.2	1.3	0.9	0.0
17687	NR	NR	142	100.0	93.7	6.3	3.5	2.8	0.0
17688	NR	NR	156	99.3	94.2	5.1	4.5	0.6	0.6
19138	NR	NR	158	99.4	96.8	2.6	1.3	1.3	0.6
19139	NR	NR	143	98.6	84.6	14.0	10.5	3.5	1.4
Aggregate			2183	99.6	91.8	7.8	6.3	1.5	0.4

Note. NR = not released.

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
19377	1(A)	5	118	100.0	79.7	20.3	1.7	18.6	0.0
19378	1(A)	6	172	100.0	69.8	30.2	18.6	11.6	0.0
19404	1(A)	11	170	99.4	81.2	18.2	10.6	7.6	0.6
19403	1(A)	12	144	100.1	90.3	9.8	4.9	4.9	0.0
19207	1(B)	5	194	99.5	73.2	26.3	13.4	12.9	0.5
19209	1(B)	6	223	99.6	72.2	27.4	11.7	15.7	0.4
19409	NR	NR	130	100.0	89.2	10.8	2.3	8.5	0.0
19410	NR	NR	129	99.3	72.9	26.4	17.1	9.3	0.8
19200	NR	NR	65	100.0	75.4	24.6	16.9	7.7	0.0
19199	NR	NR	63	100.1	93.7	6.4	1.6	4.8	0.0
Aggregate			1408	99.7	78.4	21.4	10.5	10.9	0.3

Note. NR = not released.

Table 4.1.5 Validity Estimates for Writing: Primary Division (English)

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
22817 T	2(D)	7	1600	99.7	89.1	10.6	8.7	1.9	0.3
22817 V	2(D)	7	1685	99.8	97.7	2.1	1.1	0.9	0.2
Aggregate Long Writing			3285	99.8	93.4	6.4	4.9	1.4	0.2
19895 T	NR	NR	1176	99.6	84.8	14.8	5.8	9.0	0.4
19895 V	NR	NR	1206	99.8	88.6	11.2	7.5	3.7	0.2
19879 T	2(C)	13	1521	99.9	88.0	11.8	7.8	4.1	0.1
19879 V	2(C)	13	1513	99.0	87.4	11.6	6.0	5.6	1.0
Aggregate Short Writing			5416	99.6	87.2	12.4	6.8	5.6	0.4
Aggregate All Items			8701	99.6	89.3	10.4	6.1	4.2	0.4

Note. NR = not released.

Table 4.1.6 Validity Estimates for Writing: Junior Division (English)

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
22724 T	2(D)	7	2 623	99.2	76.4	22.8	12.3	10.5	0.8
22724 V	2(D)	7	2 437	99.8	80.6	19.2	7.1	12.1	0.2
Aggregate Long Writing			5 060	99.5	78.5	21.0	9.7	11.3	0.5
19756 T	NR	NR	1 617	99.2	78.2	21.0	9.2	11.9	0.8
19756 V	NR	NR	1 691	99.6	83.7	15.9	5.1	10.8	0.4
16534 T	2(C)	13	1 820	99.0	70.4	28.6	13.0	15.6	1.0
16534 V	2(C)	13	1 983	99.9	77.6	22.3	10.2	12.1	0.1
Aggregate Short Writing			7 111	99.4	77.5	22.0	9.4	12.6	0.6
Aggregate All Items			12 171	99.5	77.8	21.6	9.5	12.1	0.5

Note. NR = not released.

Table 4.1.7 Validity Estimates for Writing: Primary Division (French)

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
22852 T	2(D)	7	176	98.9	75.0	23.9	11.9	11.9	1.1
22852 V	2(D)	7	165	100.0	91.5	8.5	4.8	3.6	0.0
Aggregate Long Writing			341	99.4	83.3	16.2	8.4	7.8	0.6
17700 T	NR	NR	125	100.0	75.2	24.8	11.2	13.6	0.0
17700 V	NR	NR	128	99.2	89.8	9.4	2.3	7.0	0.8
17701 T	2(C)	13	173	99.4	81.5	17.9	5.8	12.1	0.6
17701 V	2(C)	13	154	99.4	75.3	24.0	20.8	3.2	0.6
Aggregate Short Writing			580	99.5	80.5	19.0	10.0	9.0	0.5
Aggregate All Items			921	99.5	81.4	18.1	9.5	8.6	0.5

Note. NR = not released.

Table 4.1.8 Validity Estimates for Writing: Junior Division (French)

Item Code	Booklet (Section)	Sequence	No. of Scores	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
26298 T	2(D)	7	218	99.1	72.5	26.6	18.3	8.3	0.9
26298 V	2(D)	7	171	99.4	63.7	35.7	21.6	14.0	0.6
Aggregate Long Writing			389	99.2	68.1	31.1	20.0	11.1	0.8
23464 T	NR	NR	159	100.0	74.2	25.8	15.1	10.7	0.0
23464 V	NR	NR	172	98.8	74.4	24.4	5.2	19.2	1.2
19532 T	2(C)	13	152	100.0	82.9	17.1	3.9	13.2	0.0
19532 V	2(C)	13	149	98.7	80.5	18.1	8.1	10.1	1.3
Aggregate Short Writing			632	99.4	78.0	21.4	8.1	13.3	0.6
Aggregate All Items			1021	99.3	74.7	24.6	12.1	12.6	0.7

Note. NR = not released.

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
16773	3(1)	8	725	100.0	97.5	2.5	0.0	2.5	0.0
15133	3(1)	9	931	100.0	98.7	1.3	0.9	0.4	0.0
19299	NR	NR	654	99.8	97.4	2.4	0.0	2.4	0.2
16875	NR	NR	933	100.0	98.7	1.3	0.5	0.8	0.0
19334	3(2)	10	718	100.0	86.9	13.1	2.2	10.9	0.0
10987	3(2)	11	1052	99.9	91.1	8.8	1.4	7.4	0.1
11980	NR	NR	1093	98.9	95.6	3.3	1.0	2.3	1.1
19664	NR	NR	869	99.0	95.4	3.6	0.2	3.3	1.0
Aggregate			6975	99.7	95.2	4.5	0.8	3.7	0.3

Note. NR = not released.

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
12717	3(1)	8	780	99.6	92.4	7.2	2.8	4.4	0.4
11307	3(1)	9	755	99.7	97.7	2.0	0.5	1.5	0.3
11358	NR	NR	750	100.0	95.1	4.9	2.0	2.9	0.0
15067	NR	NR	857	99.6	90.7	9.0	4.4	4.6	0.4
15071	3(2)	10	995	99.4	92.3	7.1	4.7	2.4	0.6
17150	3(2)	11	910	99.8	94.8	4.9	0.7	4.3	0.2
20495	NR	NR	983	100.0	93.0	7.0	4.8	2.2	0.0
22532	NR	NR	728	99.5	90.4	9.1	3.6	5.5	0.5
Aggregate			6758	99.7	93.3	6.5	3.0	3.4	0.3

Note. NR = not released.

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
16444	3(1)	8	98	99.0	81.6	17.3	13.3	4.1	1.0
19826	3(1)	9	95	100.0	90.5	9.5	2.1	7.4	0.0
19706	NR	NR	80	100.0	87.5	12.5	6.3	6.3	0.0
16516	NR	NR	80	95.0	90.0	5.0	1.3	3.8	5.0
20594	3(2)	10	95	100.0	92.6	7.4	3.2	4.2	0.0
19688	3(2)	11	83	100.0	96.4	3.6	2.4	1.2	0.0
23534	NR	NR	96	100.0	100.0	0.0	0.0	0.0	0.0
14666	NR	NR	84	98.8	95.2	3.6	3.6	0.0	1.2
Aggregate			711	99.2	91.7	7.5	4.1	3.4	0.8

Note. NR = not released.

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
20866	3(1)	8	63	100.0	96.8	3.2	3.2	0.0	0.0
12755	3(1)	9	73	100.0	98.6	1.4	1.4	0.0	0.0
20193	3(1)	10	66	98.5	89.4	9.1	3.0	6.1	1.5
20210	3(1)	11	72	98.6	87.5	11.1	2.8	8.3	1.4
20869	NR	4	100	98.0	80.0	18.0	17.0	1.0	2.0
13356	NR	8	66	98.5	89.4	9.1	4.5	4.5	1.5
16360	NR	12	77	98.7	87.0	11.7	5.2	6.5	1.3
20868	NR	15	89	98.9	84.3	14.6	13.5	1.1	1.1
Aggregate			606	98.8	88.4	10.4	7.1	3.3	1.2

Note. NR = not released.

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	15610	21	190	99.5	82.6	16.8	3.2	13.7	0.5
	19645	31	145	99.3	84.8	14.5	6.9	7.6	0.0
	19433	8	67	100.0	95.5	4.5	3.0	1.5	0.0
	19434	NR	260	99.6	95.0	4.6	4.2	0.4	0.4
	19626	NR	163	98.8	82.2	16.6	10.4	6.1	1.2
	19625	NR	86	100.0	97.7	2.3	2.3	0.0	0.0
	19662	NR	76	100.0	96.1	3.9	3.9	0.0	0.0
Aggregate			987	99.5	89.4	10.1	5.2	5.0	0.4
Spring	19658	9	113	100.0	97.3	2.7	1.8	0.9	0.0
	19661	23	133	99.2	93.2	6.0	3.0	3.0	0.8
	19643	22	253	99.2	91.3	7.9	7.5	0.4	0.8
	15569	30	161	100.0	99.4	0.6	0.0	0.6	0.0
	19640	37	163	99.4	95.7	3.7	3.7	0.0	0.6
	19624	NR	164	100.0	86.6	13.4	6.7	6.7	0.0
	19444	NR	121	99.2	84.3	14.9	9.1	5.8	0.8
Aggregate			1108	99.5	92.5	7.0	4.8	2.3	0.5
Aggregate Across Administrations			2095	99.5	91.0	8.5	5.0	3.5	0.4

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	19486	23	434	97.7	78.6	19.1	9.2	9.9	2.3
	15703	30	769	100.0	97.5	2.5	1.3	1.2	0.0
	19484	14	495	98.2	90.7	7.5	3.6	3.8	1.6
	19504	NR	512	99.6	93.9	5.7	3.9	1.8	0.4
	19587	NR	641	99.7	90.6	9.0	4.7	4.4	0.3
	23351	NR	468	93.8	85.5	8.3	5.8	2.6	1.9
	19572	NR	480	100.0	95.6	4.4	1.5	2.9	0.0
Aggregate			3799	98.6	91.1	7.5	4.0	3.5	0.8
Spring	23754	6	699	97.0	89.3	7.7	7.6	0.1	3.0
	19488	31	373	99.7	94.6	5.1	1.6	3.5	0.3
	15700	13	861	99.5	83.0	16.5	6.4	10.1	0.3
	19569	22	653	97.9	86.7	11.2	5.2	6.0	2.1
	19608	NR	715	99.9	91.7	8.1	2.9	5.2	0.1
	19588	NR	345	99.4	88.1	11.3	1.7	9.6	0.6
	19505	NR	392	98.5	88.8	9.7	6.4	3.3	1.5
Aggregate			4038	98.8	88.3	10.5	5.0	5.5	1.2
Aggregate Across Administrations			7837	98.7	89.7	9.0	4.5	4.6	1.0

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	20986	17	8	100.0	100.0	0.0	0.0	0.0	0.0
	20411	8	9	100.0	88.9	11.1	11.1	0.0	0.0
	20447	26	12	100.0	91.7	8.3	8.3	0.0	0.0
	18496	NR	14	100.0	92.9	7.1	0.0	7.1	0.0
	20369	NR	8	100.0	100.0	0.0	0.0	0.0	0.0
	20449	NR	9	100.0	77.8	22.2	22.2	0.0	0.0
	20391	NR	10	100.0	90.0	10.0	0.0	10.0	0.0
Aggregate			70	100.0	91.4	8.6	5.7	2.9	0.0
Spring	20243	16	30	100.0	100.0	0.0	0.0	0.0	0.0
	15375	9	25	100.0	100.0	0.0	0.0	0.0	0.0
	14458	27	42	100.0	100.0	0.0	0.0	0.0	0.0
	21787	NR	44	100.0	100.0	0.0	0.0	0.0	0.0
	20448	NR	32	100.0	90.6	9.4	3.1	6.3	0.0
	15329	NR	57	100.0	98.2	1.8	0.0	1.8	0.0
	20429	NR	63	100.0	100.0	0.0	0.0	0.0	0.0
Aggregate			293	100.0	98.6	1.4	0.3	1.0	0.8
Aggregate Across Administrations			363	100.0	97.2	2.8	1.4	1.4	0.7

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	18489	30	36	100.0	88.9	11.1	8.3	2.8	0.0
	20347	21	35	100.0	94.3	5.7	2.9	2.9	0.0
	15245	16	42	100.0	92.9	7.1	0.0	7.1	0.0
	15439	8	48	97.9	95.8	2.1	0.0	2.1	2.1
	15395	NR	41	100.0	95.1	4.9	0.0	4.9	0.0
	20307	NR	31	100.0	67.7	32.3	25.8	6.5	0.0
	15441	NR	34	100.0	100.0	0.0	0.0	0.0	0.0
Aggregate			267	99.6	91.4	8.2	4.5	3.7	0.4
Spring	20291	7	58	100.0	87.9	12.1	6.9	5.2	0.0
	20306	22	63	100.0	90.5	9.5	0.0	9.5	0.0
	10064	15	101	100.0	96.0	4.0	1.0	3.0	0.0
	15459	31	95	100.0	98.9	1.1	1.1	0.0	0.0
	18498	NR	83	100.0	94.0	6.0	0.0	6.0	0.0
	18490	NR	73	100.0	94.5	5.5	2.7	2.7	0.0
	20289	NR	88	100.0	94.3	5.7	2.3	3.4	0.0
Aggregate			561	100.0	94.3	5.7	1.8	3.9	0.0
Aggregate Across Administrations			828	99.9	93.4	6.5	2.7	3.9	0.1

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
20824_499	IV	5	9 013	98.6	79.9	18.6	9.6	9.1	1.4
20559_499	IV	6	7 226	99.8	89.6	10.3	4.8	5.5	0.2
18681_482	NR	NR	8 268	99.1	85.5	13.6	6.6	7.0	0.9
18649_495	NR	NR	7 907	99.8	89.1	10.7	5.2	5.4	0.2
Aggregate			32 414	99.3	85.7	13.5	6.7	6.8	0.7

Note. NR = not released.

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
23368 T	I	1	24 365	99.0	77.7	21.3	12.7	8.6	1.0
23368 V	I	1	24 365	98.3	74.3	24.0	12.4	11.6	1.7
23374 T	NR	NR	19 956	98.1	71.5	26.6	13.5	13.2	1.9
23374 V	NR	NR	19 956	99.5	81.7	17.8	10.8	7.0	0.5
Aggregate Long Writing			88 642	98.7	76.3	22.5	12.4	10.1	1.3
19479 T & V	V	1	9 436	98.2	74.7	23.4	11.4	12.1	1.8
19478 T & V	NR	NR	11 556	90.9	59.4	31.5	14.1	17.4	9.1
Aggregate Short Writing			20 992	94.2	66.3	27.9	12.9	15.0	5.8
Aggregate All Items			109 634	97.9	74.4	23.5	12.5	11.0	2.1

Note. NR = not released.

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
18880 444	IV	6	319	100.0	88.1	11.9	7.5	4.4	0.0
18881 444	IV	7	160	99.4	89.4	10.0	10.0	0.0	0.6
20678 442	NR	NR	183	98.4	91.8	6.6	6.6	0.0	1.6
20636 440	NR	NR	259	96.1	90.0	6.2	5.8	0.4	3.9
Aggregate			921	98.5	89.6	8.9	7.3	1.6	1.5

Note. NR = not released.

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
21122 T	I	1	1124	99.2	74.2	25.0	11.8	13.2	0.8
21122 V	I	1	1124	96.6	72.7	23.9	9.9	14.1	3.4
21123 T	NR	NR	779	98.5	82.9	15.5	8.0	7.6	1.5
21123 V	NR	NR	779	97.9	76.5	21.4	11.8	9.6	2.1
Aggregate Long Writing			3806	98.0	76.0	22.0	10.5	11.6	2.0
18959 T & V	V	1	498	96.4	81.1	15.3	3.8	11.4	3.6
18934 T & V	NR	NR	456	96.9	73.7	23.2	16.7	6.6	3.1
Aggregate Short Writing			954	96.6	77.6	19.1	10.0	9.1	3.4
Aggregate All Items			4760	97.8	76.3	21.4	10.4	11.1	2.2

Note. NR = not released.

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
20203	1(A)	5	1 082	96.2	57.7	38.5	3.8
20183	1(A)	6	1 035	94.5	61.4	33.0	5.5
20225	1(A)	11	894	98.0	74.3	23.7	2.0
20229	1(A)	12	857	98.0	78.6	19.4	2.0
20060	1(B)	5	1 214	98.2	81.3	16.9	1.8
20061	1(B)	6	1 155	98.7	73.6	25.1	1.3
19924	NR	NR	1 894	96.8	66.5	30.3	3.2
19925	NR	NR	1 834	96.1	67.0	29.2	3.9
19933	NR	NR	1 010	96.2	78.6	17.6	3.8
19935	NR	NR	991	95.6	73.2	22.4	4.4
Aggregate			11 966	96.8	70.5	26.2	3.2

Note. NR = not released.

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
19955	1(A)	5	1 211	95.4	62.1	33.3	4.6
19956	1(A)	6	1 237	93.9	56.0	37.9	6.1
20005	1(A)	11	863	97.2	62.3	34.9	2.8
20004	1(A)	12	904	97.1	60.8	36.3	2.9
20032	1(B)	5	918	97.2	63.4	33.8	2.8
20031	1(B)	6	909	96.1	67.1	29.0	3.9
19747	NR	NR	2 994	95.5	57.6	37.9	4.5
19750	NR	NR	3 040	94.9	56.5	38.4	5.1
19968	NR	NR	1 071	94.8	63.3	31.5	5.2
19965	NR	NR	992	98.5	65.5	33.0	1.5
Aggregate			14 139	95.7	60.1	35.6	4.3

Note. NR = not released.

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
19126	1(A)	5	120	100.0	86.7	13.3	0.0
19127	1(A)	6	129	99.2	65.1	34.1	0.8
19132	1(A)	11	217	100.0	78.8	21.2	0.0
19133	1(A)	12	189	100.0	78.8	21.2	0.0
11217	1(B)	5	151	97.4	72.2	25.2	2.6
11218	1(B)	6	141	100.0	78.0	22.0	0.0
17687	NR	NR	97	99.0	78.4	20.6	1.0
17688	NR	NR	90	96.7	92.2	4.4	3.3
19138	NR	NR	123	98.4	90.2	8.1	1.6
19139	NR	NR	129	99.2	84.5	14.7	0.8
Aggregate			1386	99.1	79.8	19.3	0.9

Note. NR = not released.

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
19377	1(A)	5	113	100.0	74.3	25.7	0.0
19378	1(A)	6	86	96.5	53.5	43.0	3.5
19404	1(A)	11	90	97.8	61.1	36.7	2.2
19403	1(A)	12	103	95.1	54.4	40.8	4.9
19207	1(B)	5	165	98.2	70.3	27.9	1.8
19209	1(B)	6	150	97.3	54.0	43.3	2.7
19409	NR	NR	85	98.8	74.1	24.7	1.2
19410	NR	NR	85	98.8	52.9	45.9	1.2
19200	NR	NR	31	93.5	61.3	32.3	6.5
19199	NR	NR	32	100.0	71.9	28.1	0.0
Aggregate			940	97.8	62.6	35.2	2.2

Note. NR = not released.

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
22817_T	2(D)	7	1593	97.7	67.5	30.1	2.3
22817_V	2(D)	7	1530	98.4	73.7	24.7	1.6
Aggregate Long Writing			3123	98.1	70.6	27.4	1.9
19895_T	NR	NR	886	96.6	69.6	27.0	3.4
19895_V	NR	NR	937	99.5	82.2	17.3	0.5
19879_T	2(C)	13	1028	97.2	73.2	24.0	2.8
19879_V	2(C)	13	1048	98.1	76.7	21.4	1.9
Aggregate Short Writing			3899	97.8	75.4	22.4	2.2
Aggregate All Items			7022	97.9	73.8	24.1	2.1

Note. NR = not released.

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
22724_T	2(D)	7	1664	97.2	64.2	33.0	2.8
22724_V	2(D)	7	1762	98.9	71.3	27.5	1.1
Aggregate Long Writing			3426	98.0	67.8	30.3	2.0
19756_T	NR	NR	934	96.1	60.8	35.3	3.9
19756_V	NR	NR	902	97.2	64.3	32.9	2.8
16534_T	2(C)	13	1050	98.7	64.8	33.9	1.3
16534_V	2(C)	13	965	98.9	59.1	39.8	1.1
Aggregate Short Writing			3851	97.7	62.2	35.5	2.3
Aggregate All Items			7277	97.8	64.1	33.7	2.2

Note. NR = not released.

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
22852_T	2(D)	7	106	92.5	59.4	33.0	7.5
22852_V	2(D)	7	110	96.4	63.6	32.7	3.6
Aggregate Long Writing			216	94.4	61.5	32.9	5.6
17700_T	NR	NR	103	99.0	78.6	20.4	1.0
17700_V	NR	NR	103	100.0	81.6	18.4	0.0
17701_T	2(C)	13	207	93.7	70.5	23.2	6.3
17701_V	2(C)	13	216	97.2	66.2	31.0	2.8
Aggregate Short Writing			629	97.5	74.2	23.3	2.5
Aggregate All Items			845	96.5	70.0	26.5	3.5

Note. NR = not released.

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

Item Code	Booklet (Section)	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
26298_T	2(D)	7	80	97.5	62.5	35.0	2.5
26298_V	2(D)	7	104	100.0	68.3	31.7	0.0
Aggregate Long Writing			184	98.8	65.4	33.4	1.3
23464_T	NR	NR	94	100.0	56.4	43.6	0.0
23464_V	NR	NR	86	100.0	67.4	32.6	0.0
19532_T	2(C)	13	197	97.0	58.9	38.1	3.0
19532_V	2(C)	13	198	98.5	66.7	31.8	1.5
Aggregate Short Writing			575	98.9	62.3	36.5	1.1
Aggregate All Items			759	98.8	63.4	35.5	1.2

Note. NR = not released.

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
16773	3(1)	8	1 299	99.9	92.7	7.2	0.1
15133	3(1)	9	1 196	100.0	96.0	4.0	0.0
19299	NR	NR	1 334	99.6	85.7	13.9	0.4
16875	NR	NR	1 195	99.2	89.8	9.5	0.8
19334	3(2)	10	1 482	97.1	67.5	29.6	2.9
10987	3(2)	11	1 315	97.2	74.3	22.9	2.8
11980	NR	NR	1 325	98.4	83.6	14.8	1.6
19664	NR	NR	1 436	96.7	83.0	13.7	3.3
Aggregate			10 582	98.5	83.6	14.9	1.5

Note. NR = not released.

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
12717	3(1)	8	1 430	96.2	80.1	16.1	3.8
11307	3(1)	9	1 445	97.9	80.9	17.0	2.1
11358	NR	NR	1 445	97.8	77.2	20.6	2.2
15067	NR	NR	1 395	97.5	79.1	18.4	2.5
15071	3(2)	10	1 567	99.6	93.9	5.7	0.4
17150	3(2)	11	1 611	98.6	81.2	17.4	1.4
22532	NR	NR	1 700	99.4	90.1	9.4	0.6
20495	NR	NR	1 573	99.2	89.1	10.1	0.8
Aggregate			12 166	98.3	84.2	14.1	1.7

Note. NR = not released.

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
16444	3(1)	8	217	97.7	82.0	15.7	2.3
19826	3(1)	9	222	97.7	85.1	12.6	2.3
16516	NR	NR	225	96.9	87.6	9.3	3.1
19706	NR	NR	225	93.8	83.1	10.7	6.2
20594	3(2)	10	252	97.6	75.0	22.6	2.4
19688	3(2)	11	257	99.2	90.3	8.9	0.8
23534	NR	NR	251	100.0	90.0	10.0	0.0
14666	NR	NR	257	98.4	83.3	15.2	1.6
Aggregate			1906	97.7	84.6	13.2	2.3

Note. NR = not released.

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
20866	3(1)	8	193	96.9	87.0	9.8	3.1
12755	3(1)	9	189	96.3	88.9	7.4	3.7
20193	3(1)	10	193	98.4	86.0	12.4	1.6
20210	3(1)	11	189	99.5	81.5	18.0	0.5
20869	NR	4	293	98.0	87.7	10.2	2.0
13356	NR	8	308	97.1	81.8	15.3	2.9
16360	NR	12	303	98.0	76.6	21.5	2.0
20868	NR	15	298	98.0	84.9	13.1	2.0
Aggregate			1966	97.8	83.9	13.8	2.2

Note. NR = not released.

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	15610	21	274	97.1	83.9	13.1	2.9
	19645	31	288	97.2	83.3	13.9	2.8
	19433	8	143	99.3	76.2	23.1	0.7
	19434	NR	247	96.4	77.3	19.0	3.6
	19626	NR	284	97.2	75.7	21.5	2.8
	19625	NR	133	100.0	91.0	9.0	0.0
	19662	NR	140	99.3	82.9	16.4	0.7
Aggregate			1509	97.7	81.0	16.7	2.3
Spring	19658	9	217	97.7	86.2	11.5	2.3
	19661	23	210	99.0	89.0	10.0	1.0
	19643	22	186	98.9	82.3	16.7	1.1
	15569	30	226	100.0	88.9	11.1	0.0
	19640	NR	224	96.4	84.4	12.1	3.6
	19624	NR	200	97.0	63.0	34.0	3.0
	19444	NR	215	95.3	77.7	17.7	4.7
Aggregate			1478	97.8	81.9	15.9	2.2
Aggregate Across Administrations			2987	97.7	81.4	16.3	2.3

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	19486	23	785	95.7	72.4	23.3	4.3
	15703	30	571	98.9	86.2	12.8	1.1
	19484	14	685	96.6	72.4	24.2	3.4
	19504	NR	667	97.5	78.1	19.3	2.5
	19587	NR	699	97.3	85.3	12.0	2.7
	23351	NR	764	98.6	84.0	14.5	1.4
	19572	NR	760	99.1	84.1	15.0	0.9
Aggregate			4 931	97.6	80.2	17.4	2.4
Spring	23754	6	902	96.2	89.7	6.5	3.8
	19488	31	1 043	99.0	88.7	10.4	1.0
	15700	13	676	95.6	73.1	22.5	4.4
	19569	22	749	97.1	69.3	27.8	2.9
	19608	NR	723	99.3	81.1	18.3	0.7
	19588	NR	1 061	98.7	77.9	20.7	1.3
	19505	NR	1 041	96.0	75.3	20.7	4.0
Aggregate			6 195	97.5	79.8	17.7	2.5
Aggregate Across Administrations			11 126	97.5	80.0	17.6	2.5

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	20986	17	20	100.0	80.0	20.0	0.0
	20411	8	19	89.5	84.2	5.3	10.5
	20447	26	14	100.0	85.7	14.3	0.0
	18496	NR	19	100.0	100.0	0.0	0.0
	20369	NR	22	100.0	90.9	9.1	0.0
	20449	NR	13	100.0	84.6	15.4	0.0
	20391	NR	15	100.0	86.7	13.3	0.0
Aggregate			122	98.4	87.7	10.7	1.6
Spring	20243	16	35	94.3	80.0	14.3	5.7
	15375	9	34	97.1	91.2	5.9	2.9
	14458	27	29	100.0	96.6	3.4	0.0
	21787	NR	31	96.8	93.5	3.2	3.2
	20448	NR	32	100.0	93.8	6.3	0.0
	15329	NR	27	100.0	74.1	25.9	0.0
	20429	NR	24	100.0	100.0	0.0	0.0
Aggregate			212	98.1	89.6	8.5	1.9
Aggregate Across Administrations			334	98.2	88.7	9.6	1.8

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	18489	30	46	97.8	93.5	4.3	2.2
	20347	21	47	100.0	80.9	19.1	0.0
	15245	16	34	97.1	85.3	11.8	2.9
	15439	8	32	100.0	93.8	6.3	0.0
	15395	NR	44	100.0	93.2	6.8	0.0
	20307	NR	46	95.7	80.4	15.2	4.3
	15441	NR	36	100.0	100.0	0.0	0.0
Aggregate			285	98.6	89.1	9.5	1.4
Spring	20291	7	78	97.4	93.6	3.8	2.6
	20306	22	74	100.0	91.9	8.1	0.0
	10064	15	66	100.0	89.4	10.6	0.0
	15459	31	63	96.8	92.1	4.8	3.2
	18498	NR	73	100.0	98.6	1.4	0.0
	18490	NR	72	98.6	84.7	13.9	1.4
	20289	NR	62	98.4	87.1	11.3	1.6
Aggregate			488	98.8	91.2	7.6	1.2
Aggregate Across Administrations			773	98.7	90.2	8.5	1.3

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
20824 499	IV	5	174 769	96.9	57.8	39.1	3.1
20559 499	IV	6	174 764	98.6	65.0	33.6	1.4
18681 482	NR	NR	174 772	97.4	57.1	40.3	2.6
18649 495	NR	NR	174 765	97.7	62.7	35.1	2.3
Aggregate			699 070	97.7	60.7	37.0	2.3

Note. NR = not released.

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
23368 T	I	1	174 787	94.1	49.6	44.5	5.9
23368 V	I	1	174 787	97.1	62.0	35.1	2.9
23374 T	NR	NR	174 795	89.8	44.7	45.2	10.2
23374 V	NR	NR	174 795	97.3	62.0	35.3	2.7
Aggregate Long Writing			699 164	94.6	54.6	40.0	5.4
19479 T & V	V	1	174 770	93.2	52.7	40.5	6.8
19478 T & V	NR	NR	174 771	93.6	55.4	38.2	6.4
Aggregate Short Writing			349 541	93.4	54.1	39.3	6.6
Aggregate All Items			1 048 705	94.2	54.4	39.8	5.8

Note. NR = not released.

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
18880 444	IV	6	6 052	98.8	67.0	31.8	1.2
18881 444	IV	7	6 052	99.0	67.7	31.2	1.0
20678 442	NR	NR	6 053	98.5	74.8	23.7	1.5
20636 440	NR	NR	6 053	91.8	69.7	22.1	8.2
Aggregate			24 210	97.0	69.8	27.2	3.0

Note. NR = not released.

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
21122_T	I	1	6 052	94.3	50.8	43.5	5.7
21122_V	I	1	6 052	98.3	59.9	38.4	1.7
21123_T	NR	NR	6 053	90.1	45.7	44.4	9.9
21123_V	NR	NR	6 053	97.9	57.6	40.3	2.1
Aggregate Long Writing			24 210	95.2	53.5	41.6	4.8
18959_T & V	V	1	6 052	95.0	54.7	40.3	5.0
18934_T & V	NR	NR	6 053	93.7	51.0	42.7	6.3
Aggregate Short Writing			12 105	94.3	52.8	41.5	5.7
Aggregate All Items			36 315	94.9	53.3	41.6	5.1

Note. NR = not released.

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 (SLLs). 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-division 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 7.1.25a–7.1.48b for the primary- and junior-division assessments, Tables 7.1.65a–7.1.76b for the Grade 9 Assessment of Mathematics and Tables 7.1.83a–7.1.85b 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
20211	1(A)	1	R1.0	C	3	0.62	0.28	-0.41	0.41
20215	1(A)	2	R3.0	C	2	N/A	N/A	N/A	N/A
20214	1(A)	3	R2.0	I	3	0.42	0.30	0.88	0.76
20213	1(A)	4	R3.0	I	4	0.43	0.42	0.59	0.99
20203	1(A)	5	R1.0	C	4*	0.45 (1.79)	0.45	-0.24	0.43
20183	1(A)	6	R1.0	C	4*	0.42 (1.67)	0.51	0.13	0.51
20230	1(A)	7	R2.0	C	2	0.29	0.21	2.05	0.63
20231	1(A)	8	R3.0	I	4	0.51	0.45	0.26	1.26
20232	1(A)	9	R3.0	E	1	0.58	0.41	-0.05	0.90
20228	1(A)	10	R1.0	I	3	0.69	0.40	-0.67	0.73
20225	1(A)	11	R1.0	C	4*	0.39 (1.56)	0.43	0.33	0.48
20229	1(A)	12	R2.0	I	4*	0.30 (1.19)	0.41	1.23	0.58
20063	1(B)	1	R1.0	I	4	0.86	0.31	-2.08	0.61
20067	1(B)	2	R3.0	C	3	0.74	0.39	-0.93	0.71
20066	1(B)	3	R3.0	E	4	0.73	0.45	-0.79	0.92
20064	1(B)	4	R2.0	C	1	0.47	0.26	0.70	0.61
20060	1(B)	5	R1.0	E	4*	0.40 (1.59)	0.38	0.01	0.41
20061	1(B)	6	R1.0	C	4*	0.41 (1.62)	0.44	0.18	0.48
19954	NR	NR	R1.0	C	2	0.55	0.43	0.03	0.88
19943	NR	NR	R1.0	I	1	0.80	0.40	-1.27	0.79
19934	NR	NR	R1.0	E	1	0.91	0.32	-2.07	0.84
19952	NR	NR	R1.0	I	3	0.56	0.40	0.02	0.76
19960	NR	NR	R2.0	C	3	0.58	0.36	-0.09	0.61
19980	NR	NR	R3.0	C	3	0.55	0.39	0.10	0.75
19967	NR	NR	R3.0	I	4	0.67	0.53	-0.40	1.42
19969	NR	NR	R3.0	I	2	0.51	0.41	0.28	0.89
19973	NR	NR	R2.0	C	3	0.74	0.31	-1.15	0.50
19957	NR	NR	R2.0	I	3	0.61	0.31	-0.22	0.51
19924	NR	NR	R1.0	C	4*	0.45 (1.81)	0.55	-0.15	0.64
19925	NR	NR	R2.0	I	4*	0.40 (1.60)	0.54	0.23	0.66
19932	NR	NR	R3.0	I	2	0.52	0.40	0.26	0.89
19927	NR	NR	R1.0	I	1	0.20	0.19	1.68	1.49
19929	NR	NR	R2.0	I	4	0.61	0.24	-0.21	0.36
19928	NR	NR	R1.0	I	4	0.37	0.22	1.34	0.69
19933	NR	NR	R1.0	C	4*	0.35 (1.38)	0.47	0.63	0.58
19935	NR	NR	R2.0	C	4*	0.35 (1.39)	0.52	0.70	0.65

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. C = connections; E = explicit; I = implicit; R = reading; NR = not released; N/A = not applicable. Item 20215 was removed from the assessment.

*Maximum score code for open-response items.

() = mean score for open-response 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
20203	1(A)	5	% of Students	0.70	0.99	36.80	45.04	14.25	2.22
			Parameters			-5.62	-0.73	1.74	3.67
20183	1(A)	6	% of Students	1.46	1.86	44.19	36.32	14.00	2.16
			Parameters			-4.13	-0.21	1.53	3.32
20225	1(A)	11	% of Students	1.07	0.91	49.45	39.65	8.04	0.87
			Parameters			-5.01	0.00	2.32	4.04
20229	1(A)	12	% of Students	2.15	2.93	73.99	17.98	2.76	0.19
			Parameters			-3.63	1.38	2.82	4.37
20060	1(B)	5	% of Students	0.49	1.01	46.67	44.50	6.25	1.08
			Parameters			-6.64	-0.18	2.77	4.13
20061	1(B)	6	% of Students	0.86	0.92	45.73	42.36	9.15	0.98
			Parameters			-5.44	-0.19	2.26	4.08
19924	NR	NR	% of Students	0.75	0.89	36.42	43.07	16.74	2.13
			Parameters			-4.31	-0.59	1.28	3.03
19925	NR	NR	% of Students	1.48	1.41	47.99	36.56	11.10	1.47
			Parameters			-3.77	-0.03	1.62	3.13
19933	NR	NR	% of Students	1.09	3.30	62.88	24.13	7.65	0.95
			Parameters			-3.78	0.73	1.99	3.56
19935	NR	NR	% of Students	1.79	3.90	57.82	29.16	6.60	0.73
			Parameters			-3.24	0.51	2.04	3.47

Note. The total number of students is 118 392. NR = not released.

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
19966	1(A)	1	R2.0	C	2	0.74	0.39	-1.14	0.65
19958	1(A)	2	R1.0	E	2	0.95	0.27	-3.33	0.65
19962	1(A)	3	R3.0	I	1	0.90	0.38	-2.30	0.82
19964	1(A)	4	R1.0	I	4	0.58	0.25	0.05	0.39
19955	1(A)	5	R1.0	C	4*	0.53 (2.11)	0.51	-0.80	0.54
19956	1(A)	6	R2.0	I	4*	0.50 (2.00)	0.55	-0.51	0.56
20009	1(A)	7	R3.0	C	1	0.36	0.23	1.43	0.63
20007	1(A)	8	R1.0	I	3	0.78	0.35	-1.54	0.53
20006	1(A)	9	R1.0	E	4	0.56	0.26	0.17	0.44
20011	1(A)	10	R2.0	C	3	0.56	0.23	0.15	0.35
20005	1(A)	11	R1.0	C	4*	0.55 (2.21)	0.47	-1.06	0.43
20004	1(A)	12	R1.0	C	4*	0.58 (2.30)	0.53	-1.10	0.52
20037	1(B)	1	R3.0	I	3	0.84	0.40	-1.69	0.75
20036	1(B)	2	R2.0	I	1	0.71	0.21	-1.59	0.27
20034	1(B)	3	R1.0	C	3	0.84	0.36	-1.76	0.66
20033	1(B)	4	R1.0	I	2	0.82	0.36	-1.60	0.63
20032	1(B)	5	R1.0	C	4*	0.60 (2.39)	0.52	-1.16	0.52
20031	1(B)	6	R1.0	C	4*	0.50 (1.98)	0.50	-0.67	0.50
19757	NR	NR	R1.0	I	2	0.86	0.34	-2.07	0.61
19780	NR	NR	R1.0	I	4	0.73	0.29	-1.49	0.37
19753	NR	NR	R1.0	E	4	0.76	0.27	-1.53	0.42
19783	NR	NR	R3.0	C	3	0.61	0.42	-0.22	0.80
19751	NR	NR	R1.0	E	1	0.90	0.40	-1.88	1.00
19772	NR	NR	R3.0	E	4	0.68	0.23	-1.09	0.30
19755	NR	NR	R1.0	I	1	0.81	0.36	-1.46	0.67
19763	NR	NR	R2.0	I	2	0.79	0.31	-1.89	0.42
19767	NR	NR	R2.0	I	3	0.80	0.38	-1.51	0.64
19759	NR	NR	R1.0	I	3	0.86	0.44	-1.70	0.94
19747	NR	NR	R1.0	I	4*	0.49 (1.97)	0.50	-0.54	0.51
19750	NR	NR	R2.0	I	4*	0.49 (1.95)	0.54	-0.27	0.56
19974	NR	NR	R2.0	I	4	0.69	0.40	-0.82	0.63
19978	NR	NR	R1.0	I	2	0.51	0.29	0.44	0.57
19971	NR	NR	R1.0	E	1	0.73	0.35	-0.96	0.62
19976	NR	NR	R3.0	C	4	0.70	0.37	-0.80	0.64
19968	NR	NR	R2.0	I	4*	0.47 (1.86)	0.49	-0.06	0.50
19965	NR	NR	R1.0	C	4*	0.46 (1.83)	0.51	-0.01	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; NR = not released.

*Maximum score code for open-response items.

() = mean score for open-response 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
19955	1(A)	5	% of Students	0.43	0.43	17.66	55.66	21.17	4.65
			Parameters			-4.93	-1.96	1.01	2.70
19956	1(A)	6	% of Students	1.1	1.11	30.27	38.58	22.77	6.17
			Parameters			-4.23	-0.93	0.72	2.38
20005	1(A)	11	% of Students	0.5	0.44	19.41	43.63	29.66	6.36
			Parameters			-5.84	-1.88	0.55	2.93
20004	1(A)	12	% of Students	0.61	0.24	17.82	40.19	33.21	7.93
			Parameters			-5.21	-1.79	0.21	2.38
20032	1(B)	5	% of Students	0.51	0.42	12.63	41.67	35.85	8.93
			Parameters			-4.79	-2.25	0.08	2.31
20031	1(B)	6	% of Students	0.82	0.46	29.74	44.63	18.71	5.65
			Parameters			-5.23	-1.07	1.06	2.56
19747	NR	NR	% of Students	0.41	1.08	30.92	41.80	21.17	4.62
			Parameters			-5.01	-0.98	1.02	2.83
19750	NR	NR	% of Students	0.75	2.04	25.81	48.13	19.87	3.40
			Parameters			-3.93	-1.22	1.13	2.95
19968	NR	NR	% of Students	1.65	1.08	33.07	42.02	19.71	2.46
			Parameters			-4.11	-0.81	1.22	3.48
19965	NR	NR	% of Students	1.64	0.77	29.73	51.82	14.19	1.86
			Parameters			-3.88	-1.01	1.56	3.30

Note. The total number of students is 124 355. NR = not released.

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
19129	1(A)	1	B	I	3	0.38	0.33	1.03	0.84
19130	1(A)	2	C	I	2	0.67	0.52	-0.44	1.20
19128	1(A)	3	A	L	4	0.66	0.23	-0.73	0.34
19131	1(A)	4	C	L	1	0.67	0.36	-0.54	0.62
19126	1(A)	5	A	L	4*	0.55 (2.18)	0.48	-0.60	0.58
19127	1(A)	6	A	L	4*	0.53 (2.10)	0.55	-0.41	0.65
19135	1(A)	7	B	L	4	0.84	0.43	-1.35	1.01
19136	1(A)	8	C	E	3	0.46	0.19	1.19	0.37
19137	1(A)	9	C	I	2	0.84	0.45	-1.32	1.02
19134	1(A)	10	A	I	1	0.79	0.34	-1.43	0.59
19132	1(A)	11	A	I	4*	0.52 (2.07)	0.46	-0.46	0.53
19133	1(A)	12	A	L	4*	0.47 (1.87)	0.52	0.14	0.57
11212	1(B)	1	B	L	2	0.71	0.29	-0.96	0.45
11215	1(B)	2	C	L	1	0.57	0.39	0.06	0.76
11213	1(B)	3	C	E	3	0.75	0.31	-1.16	0.51
11211	1(B)	4	A	I	1	0.55	0.38	0.13	0.74
11217	1(B)	5	A	I	4*	0.52 (2.07)	0.49	-0.29	0.49
11218	1(B)	6	A	L	4*	0.54 (2.14)	0.39	-0.61	0.48
17680	NR	NR	B	I	1	0.70	0.37	-0.72	0.63
17679	NR	NR	B	L	4	0.77	0.42	-1.02	0.83
17682	NR	NR	C	I	1	0.82	0.48	-1.10	1.21
17686	NR	NR	C	L	2	0.72	0.38	-0.73	0.73
17683	NR	NR	C	I	1	0.93	0.28	-2.60	0.67
17684	NR	NR	C	L	1	0.60	0.39	-0.11	0.75
17685	NR	NR	C	L	3	0.51	0.36	0.38	0.67
17678	NR	NR	A	I	4	0.66	0.41	-0.42	0.80
17681	NR	NR	C	I	3	0.54	0.37	0.23	0.94
17677	NR	NR	A	E	2	0.73	0.45	-0.75	0.91
17687	NR	NR	A	L	4*	0.53 (2.12)	0.43	-0.42	0.47
17688	NR	NR	A	I	4*	0.41 (1.62)	0.53	0.54	0.61
19140	NR	NR	A	I	3	0.72	0.37	-0.76	0.69
19141	NR	NR	A	I	1	0.73	0.43	-0.79	0.84
19143	NR	NR	C	I	4	0.68	0.44	-0.46	0.96
19142	NR	NR	B	I	4	0.70	0.43	-0.61	0.89
19138	NR	NR	A	I	4*	0.51 (2.05)	0.44	0.01	0.58
19139	NR	NR	B	L	4*	0.45 (1.80)	0.48	0.68	0.54

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; NR = not released.

*Maximum score code for open-response items.

() = mean score for open-response 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
19126	1(A)	5	% of Students	0.51	0.47	9.39	63.34	23.35	2.95
			Parameters			-4.12	-2.46	1.06	3.11
19127	1(A)	6	% of Students	1.16	0.67	16.45	56.63	20.46	4.64
			Parameters			-3.52	-1.61	1.01	2.47
19132	1(A)	11	% of Students	0.47	0.30	17.08	59.22	20.64	2.28
			Parameters			-4.83	-1.87	1.31	3.54
19133	1(A)	12	% of Students	2.02	1.10	31.69	42.26	21.19	1.74
			Parameters			-3.59	-0.69	1.18	3.66
11217	1(B)	5	% of Students	1.28	1.04	23.75	42.77	27.33	3.83
			Parameters			-4.05	-1.23	0.83	3.28
11218	1(B)	6	% of Students	0.57	0.30	7.23	71.43	17.84	2.62
			Parameters			-4.54	-2.99	1.60	3.47
17687	NR	NR	% of Students	0.48	1.12	11.42	63.57	20.22	3.17
			Parameters			-4.15	-2.35	1.39	3.41
17688	NR	NR	% of Students	1.31	2.87	45.89	34.38	14.74	0.81
			Parameters			-3.31	-0.03	1.51	4.01
19138	NR	NR	% of Students	0.80	1.15	9.20	71.50	16.39	0.97
			Parameters			-3.40	-2.25	1.72	3.97
19139	NR	NR	% of Students	1.36	2.95	23.59	59.99	11.71	0.39
			Parameters			-3.16	-1.18	2.19	4.85

Note. The total number of students is 7723. NR = not released.

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
19380	1(A)	1	B	L	4	0.80	0.30	-1.64	0.51
19382	1(A)	2	C	I	2	0.74	0.38	-1.03	0.64
19379	1(A)	3	A	I	3	0.83	0.32	-2.05	0.50
19381	1(A)	4	C	I	4	0.70	0.27	-1.01	0.40
19377	1(A)	5	A	I	4*	0.66 (2.62)	0.52	-1.67	0.64
19378	1(A)	6	A	L	4*	0.62 (2.46)	0.51	-1.70	0.54
19405	1(A)	7	A	E	3	0.93	0.23	-2.95	0.57
19408	1(A)	8	C	L	1	0.39	0.17	2.15	0.32
19407	1(A)	9	B	L	3	0.75	0.31	-1.37	0.45
19406	1(A)	10	A	I	2	0.67	0.23	-0.92	0.32
19404	1(A)	11	A	L	4*	0.58 (2.32)	0.46	-1.46	0.49
19403	1(A)	12	A	L	4*	0.61 (2.44)	0.49	-1.46	0.48
19210	1(B)	1	A	E	1	0.87	0.19	-3.29	0.34
19214	1(B)	2	B	I	2	0.58	0.33	-0.09	0.48
19216	1(B)	3	C	I	3	0.62	0.40	-0.26	0.77
19212	1(B)	4	A	I	4	0.69	0.28	-0.95	0.40
19207	1(B)	5	A	I	4*	0.58 (2.30)	0.55	-1.13	0.65
19209	1(B)	6	A	L	4*	0.54 (2.14)	0.50	-0.92	0.48
19417	NR	NR	B	I	2	0.57	0.37	0.01	0.62
19419	NR	NR	C	I	4	0.80	0.29	-1.97	0.41
19411	NR	NR	A	E	1	0.88	0.33	-1.89	0.74
19416	NR	NR	A	I	2	0.62	0.32	-0.38	0.45
19418	NR	NR	C	I	2	0.52	0.42	0.23	0.80
19420	NR	NR	C	L	3	0.79	0.40	-1.17	0.82
19412	NR	NR	A	I	1	0.76	0.48	-0.88	1.03
19414	NR	NR	A	I	1	0.63	0.19	-0.51	0.28
19413	NR	NR	A	E	2	0.83	0.40	-1.32	0.89
19415	NR	NR	A	I	1	0.65	0.20	-0.81	0.25
19409	NR	NR	A	I	4*	0.58 (2.32)	0.58	-0.98	0.79
19410	NR	NR	A	L	4*	0.59 (2.35)	0.60	-0.84	0.78
19202	NR	NR	A	L	3	0.62	0.34	-0.31	0.54
19204	NR	NR	C	L	1	0.65	0.39	-0.52	0.63
19201	NR	NR	A	I	3	0.38	0.24	1.27	0.66
19203	NR	NR	B	I	2	0.72	0.33	-0.91	0.58
19199	NR	NR	A	I	4*	0.62 (2.46)	0.57	-1.04	0.71
19200	NR	NR	B	L	4*	0.60 (2.39)	0.49	-0.96	0.51

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; NR = not released.

*Maximum score code for open-response items.

() = mean score for open-response 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
19377	1(A)	5	% of Students	0.06	0.06	4.23	37.29	50.58	7.78
			Parameters			-5.31	-3.21	-0.51	2.36
19378	1(A)	6	% of Students	0.13	0.22	8.23	44.73	38.71	7.97
			Parameters			-6.39	-2.85	0.01	2.44
19404	1(A)	11	% of Students	0.19	0.04	11.66	48.96	34.17	4.97
			Parameters			-6.81	-2.57	0.41	3.13
19403	1(A)	12	% of Students	0.28	0.16	9.80	43.01	39.11	7.63
			Parameters			-5.86	-2.68	0.00	2.69
19207	1(B)	5	% of Students	0.18	0.24	13.65	47.79	32.12	6.03
			Parameters			-5.29	-2.00	0.37	2.40
19209	1(B)	6	% of Students	0.33	0.49	23.30	43.01	27.29	5.58
			Parameters			-5.70	-1.49	0.67	2.84
19409	NR	NR	% of Students	0.19	0.13	11.03	52.67	28.27	7.69
			Parameters			-4.26	-2.03	0.43	1.93
19410	NR	NR	% of Students	0.74	0.25	11.48	47.88	30.55	9.10
			Parameters			-3.55	-1.91	0.29	1.81
19199	NR	NR	% of Students	0.48	0.19	9.18	40.17	43.44	6.55
			Parameters			-4.18	-2.16	-0.13	2.33
19200	NR	NR	% of Students	0.91	0.21	12.86	39.23	39.99	6.80
			Parameters			-4.54	-2.10	0.02	2.80

Note. The total number of students is 6746. NR = not released.

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
19895_T	NR	NR	W2.0	T	4*	0.54 (2.17)	0.61	-0.35	0.84
19895_V	NR	NR	W3.0	V	3*	0.66 (1.99)	0.62	-1.07	1.06
19879_T	2(C)	13	W2.0	T	4*	0.56 (2.22)	0.60	-0.42	0.89
19879_V	2(C)	13	W3.0	V	3*	0.64 (1.93)	0.62	-1.09	1.08
19889	2(C)	14	W2.0	T	4	0.76	0.32	-1.38	0.52
19885	2(C)	15	W1.0	T	2	0.81	0.31	-1.74	0.52
19890	2(C)	16	W1.0	T	3	0.73	0.40	-1.03	0.66
19878	2(C)	17	W3.0	V	3	0.58	0.39	-0.16	0.72
19870	2(D)	11	W2.0	T	2	0.90	0.26	-3.07	0.47
19876	2(D)	10	W1.0	T	2	0.62	0.33	-0.35	0.51
19884	2(D)	8	W1.0	T	3	0.44	0.22	1.38	0.34
19898	2(D)	9	W3.0	V	4	0.43	0.28	0.90	0.56
22817_T	2(D)	7	W2.0	T	4*	0.54 (2.14)	0.56	-0.41	0.68
22817_V	2(D)	7	W3.0	V	3*	0.68 (2.05)	0.62	-1.32	1.00

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 items.

() = mean score for open-response items. NR = not released.

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
19895_T	NR	NR	% of Students	1.08	2.65	16.57	45.01	27.90	6.79
			Parameters			-2.43	-1.36	0.47	1.92
19895_V	NR	NR	% of Students	1.08	0.44	27.62	41.19	29.66	
			Parameters			-3.05	-0.73	0.57	
19879_T	2(C)	13	% of Students	0.72	1.44	15.16	46.11	31.37	5.20
			Parameters			-2.84	-1.43	0.40	2.18
19879_V	2(C)	13	% of Students	0.72	0.31	33.36	36.88	28.73	
			Parameters			-3.35	-0.50	0.58	
22817_T	2(D)	7	% of Students	0.71	1.84	20.02	43.05	29.16	5.22
			Parameters			-3.26	-1.26	0.50	2.38
22817_V	2(D)	7	% of Students	0.71	0.25	26.14	39.52	33.37	
			Parameters			-3.54	-0.84	0.42	

Note. The total number of students is 118 573. NR = not released.

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
19756_T	NR	NR	W2.0	T	4*	0.55 (2.18)	0.60	-0.54	0.72
19756_V	NR	NR	W3.0	V	3*	0.66 (1.97)	0.60	-1.18	0.98
16534_T	2(C)	13	W2.0	T	4*	0.55 (2.18)	0.59	-0.64	0.79
16534_V	2(C)	13	W3.0	V	3*	0.63 (1.89)	0.61	-1.17	1.00
19739	2(C)	14	W2.0	T	3	0.57	0.36	-0.06	0.59
19754	2(C)	17	W1.0	T	2	0.65	0.32	-0.62	0.48
19781	2(C)	15	W1.0	T	3	0.63	0.26	-0.59	0.34
19806	2(C)	16	W3.0	V	2	0.70	0.38	-0.93	0.62
17724	2(D)	10	W3.0	V	3	0.78	0.27	-2.38	0.33
19737	2(D)	8	W1.0	T	3	0.65	0.28	-0.73	0.37
19745	2(D)	11	W1.0	T	1	0.69	0.30	-1.02	0.42
20996	2(D)	9	W3.0	V	2	0.57	0.31	-0.05	0.49
22724_T	2(D)	7	W2.0	T	4*	0.57 (2.26)	0.63	-0.80	0.94
22724_V	2(D)	7	W3.0	V	3*	0.67 (2.02)	0.64	-1.40	1.13

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 items.

() = mean score for open-response items. NR = not released.

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
19756_T	NR	NR	% of Students	0.73	1.95	19.19	43.31	27.24	7.57
			Parameters			-3.07	-1.42	0.46	1.87
19756_V	NR	NR	% of Students	0.73	0.63	19.86	59.58	19.20	
			Parameters			-3.26	-1.36	1.08	
16534_T	2(C)	13	% of Students	0.68	0.40	19.55	45.47	28.41	5.49
			Parameters			-3.77	-1.40	0.47	2.14
16534_V	2(C)	13	% of Students	0.68	0.26	27.19	53.62	18.26	
			Parameters			-3.60	-1.00	1.08	
22724_T	2(D)	7	% of Students	0.55	0.34	17.67	43.52	30.37	7.54
			Parameters			-3.78	-1.45	0.27	1.76
22724_V	2(D)	7	% of Students	0.55	0.19	23.62	48.26	27.38	
			Parameters			-3.71	-1.11	0.61	

Note. The total number of students is 124 321. NR = not released.

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
17700_T	NR	NR	13	T	4*	0.50 (2.01)	0.61	-0.12	0.88
17700_V	NR	NR	13	V	3*	0.68 (2.03)	0.61	-1.10	1.21
17701_T	2(C)	13	13	T	4*	0.50 (1.98)	0.60	-0.03	0.95
17701_V	2(C)	13	13	V	3*	0.63 (1.89)	0.60	-0.86	1.16
15487	2(C)	17	17	V	2	0.70	0.33	-0.77	0.55
17713	2(C)	15	15	V	4	0.65	0.38	-0.43	0.63
19017	2(C)	14	14	T	3	0.76	0.37	-1.03	0.80
20849	2(C)	16	16	T	2	0.70	0.33	-0.64	0.75
18486	2(D)	10	10	T	4	0.72	0.58	-0.79	0.67
19041	2(D)	11	11	V	1	0.67	0.57	-0.44	0.81
20846	2(D)	9	9	V	1	0.74	0.28	-1.07	0.54
20850	2(D)	8	8	T	3	0.72	0.32	-0.77	0.72
22852_T	2(D)	7	7	T	4*	0.49 (1.96)	0.29	-0.07	0.79
22852_V	2(D)	7	7	V	3*	0.64 (1.92)	0.28	-0.95	1.09

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 items.

() = mean score for open-response items. NR = not released.

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
17700_T	NR	NR	% of Students	0.46	2.07	19.90	54.18	20.38	3.01
			Parameters			-2.76	-1.17	1.02	2.44
17700_V	NR	NR	% of Students	0.46	0.69	17.42	58.89	22.55	0.00
			Parameters			-3.08	-1.20	0.98	0.00
17701_T	2(C)	13	% of Students	0.25	2.56	22.43	52.01	19.90	2.86
			Parameters			-2.58	-0.99	0.99	2.46
17701_V	2(C)	13	% of Students	0.25	1.33	25.62	55.44	17.36	0.00
			Parameters			-2.99	-0.82	1.23	0.00
22852_T	2(D)	7	% of Students	0.50	2.53	23.47	51.57	18.50	3.43
			Parameters			-2.78	-1.07	1.15	2.42
22852_V	2(D)	7	% of Students	0.50	0.86	21.50	61.05	16.09	0.00
			Parameters			-3.16	-1.05	1.36	0.00

Note. The total number of students is 7685. NR = not released.

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
23464_T	NR	NR	A	T	4*	0.59 (2.34)	0.62	-0.65	0.85
23464_V	NR	NR	C	V	3*	0.72 (2.17)	0.58	-1.80	0.81
12540	2(C)	17	A	T	1	0.65	0.31	-0.51	0.48
13022	2(C)	16	B	T	3	0.72	0.40	-0.87	0.73
19536	2(C)	14	B	T	1	0.76	0.33	-1.26	0.54
21002	2(C)	15	C	V	2	0.80	0.34	-1.57	0.58
19532_T	2(C)	13	A	T	4*	0.53 (2.11)	0.59	-0.43	0.93
19532_V	2(C)	13	C	V	3*	0.65 (1.94)	0.60	-1.18	1.13
19533	2(D)	9	A	T	4	0.75	0.40	-1.12	0.64
19549	2(D)	11	B	T	4	0.79	0.40	-1.35	0.69
20702	2(D)	10	A	T	1	0.64	0.38	-0.40	0.65
20870	2(D)	8	C	V	1	0.83	0.32	-1.92	0.56
26298_T	2(D)	7	A	T	4*	0.63 (2.52)	0.61	-1.33	0.83
26298_V	2(D)	7	C	V	3*	0.65 (1.96)	0.56	-1.72	0.83

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 items.

() = mean score for open-response items. NR = not released.

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
23464_T	NR	NR	% of Students	0.31	1.72	10.56	45.66	34.92	6.83
			Parameters			-2.81	-2.00	0.20	2.01
23464_V	NR	NR	% of Students	0.31	0.31	18.20	44.71	36.47	0.00
			Parameters			-4.30	-1.41	0.31	0.00
19532_T	2(C)	13	% of Students	0.27	1.32	14.91	58.08	21.31	4.11
			Parameters			-3.19	-1.60	0.89	2.18
19532_V	2(C)	13	% of Students	0.27	0.46	23.76	56.50	19.02	0.00
			Parameters			-3.58	-1.03	1.07	0.00
26298_T	2(D)	7	% of Students	0.25	0.15	8.06	41.63	39.18	10.72
			Parameters			-4.53	-2.34	-0.10	1.65
26298_V	2(D)	7	% of Students	0.25	0.12	24.31	53.90	21.43	0.00
			Parameters			-5.06	-1.16	1.06	0.00

Note. The total number of students is 6746. NR = not released.

Table 7.1.17 Item Statistics: Primary Mathematics (English)

Item Code	Booklet (Section)	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
							Difficulty	Item-Total Correlation	Location	Slope
16830	3(1)	1	3	KU	N	2	75.15	0.42	-0.95	0.89
19253	3(1)	2	1	AP	P	1	67.94	0.41	-0.58	0.85
11963	3(1)	3	3	AP	N	4	87.76	0.37	-1.96	0.80
19263	3(1)	4	3	AP	N	1	54.65	0.41	0.11	1.01
19340	3(1)	5	1	KU	M	2	69.44	0.36	-0.69	0.72
16785	3(1)	6	3	AP	G	4	58.68	0.39	0.01	0.95
19290	3(1)	7	2	AP	D	3	61.44	0.31	-0.30	0.53
16773	3(1)	8	1	AP	P	4 ⁺	72.50(2.90)	0.61	-1.87	0.48
15133	3(1)	9	1	TH	G	4 ⁺	55.50(2.22)	0.49	-1.18	0.40
23488	3(1)	12	1	KU	G	4	56.48	0.31	0.15	0.80
11991	3(1)	13	3	TH	N	1	49.76	0.32	0.38	0.83
19337	NR	NR	2	AP	N	2	68.75	0.37	-0.56	0.77
19293	NR	NR	2	TH	N	4	64.85	0.39	-0.42	0.77
16823	NR	NR	1	AP	M	3	67.51	0.43	-0.52	0.82
19299	NR	NR	1	TH	M	4 ⁺	56.00(2.24)	0.60	-0.99	0.53
19267	NR	NR	2	KU	P	2	69.77	0.33	-0.75	0.59
19333	NR	NR	3	AP	D	1	77.64	0.28	-1.42	0.51
16875	NR	NR	2	TH	D	4 ⁺	66.25(2.65)	0.61	-1.66	0.45
19334	3(2)	10	3	AP	D	4 ⁺	62.25(2.49)	0.48	-1.59	0.36
10987	3(2)	11	1	AP	N	4 ⁺	58.75(2.35)	0.57	-1.27	0.43
10950	3(2)	14	1	KU	P	1	48.76	0.41	0.34	1.06
19325	3(2)	15	1	KU	P	2	72.54	0.37	-0.93	0.63
16715	3(2)	16	2	TH	M	2	47.62	0.35	0.46	0.94
18361	3(2)	17	1	AP	M	3	47.08	0.32	0.53	0.67
19213	3(2)	18	2	KU	M	1	62.80	0.36	-0.33	0.66
15074	NR	NR	1	KU	N	2	59.26	0.39	-0.17	0.79
11980	NR	NR	3	TH	N	4 ⁺	73.75(2.95)	0.60	-1.77	0.46
19342	NR	NR	2	AP	M	2	81.95	0.25	-2.47	0.40
16628	NR	NR	2	KU	M	4	74.99	0.31	-1.34	0.50
10775	NR	NR	1	KU	M	2	78.54	0.43	-1.18	0.88
16845	NR	NR	1	KU	G	3	70.52	0.36	-0.78	0.65
19250	NR	NR	2	AP	G	4	76.60	0.39	-1.06	0.77
19664	NR	NR	3	AP	G	4 ⁺	67.50(2.70)	0.52	-1.76	0.34
10781	NR	NR	1	AP	P	2	71.07	0.33	-0.80	0.61
15145	NR	NR	1	TH	P	3	52.10	0.34	0.30	0.73
19332	NR	NR	2	KU	D	2	80.30	0.46	-1.11	1.03

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

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 1–8: Mathematics* (revised 2005).

⁺Maximum score code for open-response items.

() = mean score for open-response items. NR = not released.

**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
16773	3(1)	8	% of Students	0.68	0.12	18.14	13.70	24.90	42.46
			Parameters			-5.27	-0.71	-1.06	-0.45
15133	3(1)	9	% of Students	0.93	0.15	27.68	33.99	22.86	14.38
			Parameters			-5.91	-0.76	0.67	1.25
19299	NR	NR	% of Students	1.03	0.27	31.94	24.88	24.83	17.05
			Parameters			-4.85	-0.31	0.11	1.07
16875	NR	NR	% of Students	0.91	0.16	29.34	14.15	14.60	40.84
			Parameters			-5.67	0.25	-0.20	-1.03
19334	3(2)	10	% of Students	0.79	0.10	17.59	28.38	38.03	15.11
			Parameters			-6.13	-1.54	-0.61	1.91
10987	3(2)	11	% of Students	1.33	0.28	31.35	21.11	22.52	23.41
			Parameters			-5.39	0.00	-0.15	0.48
11980	NR	NR	% of Students	1.24	0.24	15.40	18.21	16.75	48.16
			Parameters			-4.50	-1.16	-0.27	-1.13
19664	NR	NR	% of Students	1.40	0.47	22.09	17.91	20.96	37.18
			Parameters			-5.44	-0.42	-0.44	-0.76

Note. The total number of students is 124 007. NR = not released.

Table 7.1.19 Item Statistics: Junior Mathematics (English)

Item Code	Booklet (Section)	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
							Difficulty	Item-Total Correlation	Location	Slope
11301	3(1)	1	1	KU	P	4	86.10	0.39	-1.74	0.90
23480	3(1)	2	2	AP	M	1	58.76	0.26	0.01	0.68
11314	3(1)	3	3	AP	N	4	81.62	0.29	-1.89	0.50
12711	3(1)	4	2	AP	D	2	55.24	0.44	0.01	1.09
23506	3(1)	5	2	AP	P	1	80.81	0.40	-1.37	0.82
11320	3(1)	6	3	TH	G	4	42.58	0.30	0.80	0.74
15001	3(1)	7	2	TH	M	1	52.47	0.36	0.23	1.21
12717	3(1)	8	1	TH	P	4 ⁺	67.75(2.71)	0.58	-1.75	0.39
11307	3(1)	9	1	AP	N	4 ⁺	61.25(2.45)	0.55	-1.22	0.40
17180	3(1)	12	1	KU	P	3	74.54	0.40	-0.96	0.80
20471	NR	NR	1	KU	N	4	74.74	0.35	-1.04	0.69
22493	NR	NR	2	AP	N	3	55.59	0.34	0.07	0.71
23481	NR	NR	1	AP	M	2	58.95	0.24	-0.12	0.41
17105	NR	NR	3	AP	G	4	68.10	0.39	-0.63	0.76
11358	NR	NR	1	AP	G	4 ⁺	63.75(2.55)	0.65	-1.15	0.53
15066	NR	NR	3	TH	D	2	44.84	0.31	0.62	0.89
23503	NR	NR	1	KU	D	3	48.12	0.31	0.48	0.64
15067	NR	NR	1	TH	D	4 ⁺	66.25(2.65)	0.59	-1.76	0.39
15071	3(2)	10	3	AP	D	4 ⁺	47.00(1.88)	0.53	-0.90	0.39
17150	3(2)	11	3	TH	G	4 ⁺	58.50(2.34)	0.66	-0.84	0.65
14980	3(2)	13	2	KU	N	3	85.99	0.33	-1.85	0.70
15013	3(2)	14	1	TH	G	2	54.52	0.30	0.16	0.70
20467	3(2)	15	3	AP	D	1	77.54	0.44	-1.09	0.97
20521	3(2)	16	2	KU	M	3	48.73	0.25	0.64	0.46
11295	3(2)	17	2	AP	M	4	55.09	0.41	0.03	0.87
22461	3(2)	18	1	TH	N	3	67.72	0.30	-0.68	0.55
11294	NR	NR	3	TH	N	3	57.74	0.20	0.06	0.37
22532	NR	NR	3	TH	N	4 ⁺	65.50(2.62)	0.68	-1.09	0.53
20458	NR	NR	2	KU	M	3	74.69	0.38	-0.96	0.78
17137	NR	NR	2	TH	M	3	41.96	0.40	0.52	1.85
20495	NR	NR	2	AP	M	4 ⁺	62.75(2.51)	0.67	-1.16	0.62
20512	NR	NR	1	KU	G	1	86.37	0.33	-1.87	0.73
17143	NR	NR	1	TH	P	3	51.36	0.29	0.35	0.62
12678	NR	NR	2	AP	P	2	69.81	0.34	-0.78	0.62
11342	NR	NR	1	AP	P	4	38.21	0.27	1.25	0.56
15059	NR	NR	2	AP	D	3	70.36	0.40	-0.78	0.76

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

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 1–8: Mathematics* (revised 2005).

⁺Maximum score code for open-response items.

() = mean score for open-response items. NR = not released.

**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
12717	3(1)	8	% of Students	1.23	0.54	22.47	20.87	12.89	41.99
			Parameters			-5.16	-0.74	0.48	-1.57
11307	3(1)	9	% of Students	2.01	0.53	20.14	29.16	25.64	22.52
			Parameters			-4.27	-1.34	0.08	0.65
11358	NR	NR	% of Students	1.85	0.80	25.51	14.65	28.67	28.52
			Parameters			-3.87	-0.20	-0.91	0.39
15067	NR	NR	% of Students	1.14	0.42	25.86	17.69	15.48	39.42
			Parameters			-5.64	-0.25	-0.02	-1.15
15071	3(2)	10	% of Students	2.18	0.66	57.55	12.43	3.31	23.86
			Parameters			-5.47	2.03	2.18	-2.35
17150	3(2)	11	% of Students	2.40	0.70	22.04	34.76	18.43	21.68
			Parameters			-3.20	-1.15	0.55	0.43
22532	NR	NR	% of Students	3.84	1.22	28.03	11.40	10.80	44.69
			Parameters			-3.12	0.16	-0.10	-1.27
20495	NR	NR	% of Students	1.43	0.36	30.01	15.62	20.88	31.69
			Parameters			-4.07	-0.18	-0.41	0.01

Note. The total number of students in writing the Mathematics component of the English-language junior-division assessment is 124 073. NR = not released.

Table 7.1.21 Item Statistics: Primary Mathematics (French)

Item Code	Booklet (Section)	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
							Difficulty	Item-Total Correlation	Location	Slope
19815	3(1)	1	2	CC	A	3	90.59	0.33	-2.10	0.85
23536	3(1)	2	4	CC	N	2	72.34	0.38	-0.73	0.83
14570	3(1)	3	1	CC	N	4	80.88	0.39	-1.12	1.00
19776	3(1)	4	3	MA	A	2	49.74	0.40	0.34	1.14
19669	3(1)	5	2	CC	M	4	56.62	0.47	0.01	1.20
11237	3(1)	6	1	MA	G	4	64.12	0.37	-0.29	0.83
16335	3(1)	7	1	HP	T	3	66.92	0.36	-0.46	0.69
16444	3(1)	8	2	MA	N	4 ⁺	64.50(2.58)	0.54	-2.07	0.37
19826	3(1)	9	2	HP	T	4 ⁺	51.25(2.05)	0.52	-0.94	0.38
14585	NR	NR	3	HP	N	1	40.00	0.35	0.82	1.01
19711	NR	NR	4	MA	N	3	82.10	0.36	-1.39	0.76
11230	NR	NR	2	MA	M	1	67.38	0.36	-0.48	0.77
16516	NR	NR	4	HP	M	4 ⁺	61.00(2.44)	0.53	-1.75	0.36
16424	NR	NR	2	MA	G	3	68.74	0.37	-0.56	0.71
19706	NR	NR	2	HP	G	4 ⁺	71.00(2.84)	0.45	-2.15	0.26
19799	NR	NR	1	CC	A	2	86.76	0.18	-3.77	0.28
20694	NR	NR	2	HP	A	1	73.34	0.39	-0.76	0.84
19800	NR	NR	2	MA	T	4	59.34	0.47	-0.07	1.23
20594	3(2)	10	2	MA	G	4 ⁺	68.25(2.73)	0.46	-2.11	0.29
19688	3(2)	11	1	HP	A	4 ⁺	66.25(2.65)	0.49	-1.71	0.33
19829	3(2)	12	3	CC	N	2	67.54	0.36	-0.51	0.68
19808	3(2)	13	3	MA	N	3	71.26	0.40	-0.63	0.89
20740	3(2)	14	4	CC	M	1	82.25	0.17	-2.79	0.29
20873	3(2)	15	1	HP	G	2	44.36	0.31	0.69	0.91
20713	3(2)	16	1	MA	M	1	62.81	0.30	-0.23	0.59
19676	3(2)	17	2	HP	N	1	47.06	0.30	0.64	0.76
16457	3(2)	18	1	MA	T	3	74.67	0.47	-0.74	1.14
16296	NR	NR	1	MA	N	1	56.01	0.49	0.04	1.40
20709	NR	NR	2	MA	N	2	65.60	0.29	-0.44	0.53
23534	NR	NR	4	HP	N	4 ⁺	78.25(3.13)	0.53	-2.11	0.43
20579	NR	NR	1	HP	M	3	50.44	0.33	0.44	0.75
16373	NR	NR	3	CC	M	2	93.16	0.26	-2.62	0.69
16329	NR	NR	1	CC	G	4	62.74	0.40	-0.25	0.88
20593	NR	NR	3	CC	A	4	43.19	0.35	0.67	1.05
14655	NR	NR	1	CC	T	3	69.92	0.40	-0.57	0.83
14666	NR	NR	1	MA	T	4 ⁺	62.25(2.49)	0.55	-1.41	0.38

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.

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 1–8: Mathematics* (revised 2005).

⁺Maximum score code for open-response items.

() = mean score for open-response items. NR = not released.

**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
16444	3(1)	8	% of Students	0.37	0.13	31.45	13.24	19.17	35.64
			Parameters			-7.72	0.78	-0.72	-0.62
19826	3(1)	9	% of Students	1.38	1.41	45.56	17.23	12.48	21.94
			Parameters			-5.32	1.23	0.65	-0.33
16516	NR	NR	% of Students	0.70	0.09	39.18	5.15	25.10	29.78
			Parameters			-7.33	2.88	-2.69	0.14
19706	NR	NR	% of Students	0.93	1.07	21.80	8.31	26.47	41.42
			Parameters			-6.47	1.46	-2.78	-0.82
20594	3(2)	10	% of Students	1.28	0.25	22.53	11.98	28.95	35.02
			Parameters			-6.93	0.65	-1.98	-0.19
19688	3(2)	11	% of Students	1.50	0.30	17.10	27.13	22.59	31.39
			Parameters			-5.29	-1.50	0.22	-0.27
23534	NR	NR	% of Students	0.90	0.13	8.39	18.46	20.91	51.21
			Parameters			-4.58	-2.16	-0.61	-1.11
14666	NR	NR	% of Students	1.68	0.36	31.39	12.15	24.68	29.75
			Parameters			-5.52	1.04	-1.25	0.09

Note. The total number of students is 7748. NR = not released.

Table 7.1.23 Item Statistics: Junior Mathematics (French)

Item Code	Booklet (Section)	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
							Difficulty	Item-Total Correlation	Location	Slope
11538	3(1)	2	1	MA	N	4	46.67	0.34	0.56	0.84
16314	3(1)	3	2	CC	G	1	71.60	0.43	-0.64	1.00
14725	3(1)	4	1	MA	M	4	49.35	0.31	0.47	0.81
20106	3(1)	5	1	HP	T	4	39.24	0.14	1.83	0.39
14694	3(1)	6	1	CC	A	3	64.23	0.38	-0.34	0.80
20866	3(1)	8	3	MA	N	4 ⁺	75.00(3.00)	0.59	-1.92	0.44
12755	3(1)	9	1	HP	A	4 ⁺	74.75(2.99)	0.42	-3.13	0.25
20193	3(1)	10	1	MA	G	4 ⁺	49.50(1.98)	0.48	-1.12	0.33
20210	3(1)	11	1	HP	T	4 ⁺	74.75(2.99)	0.49	-3.05	0.29
20862	3(1)	12	1	HP	G	3	70.81	0.41	-0.60	0.97
14765	3(1)	13	2	MA	T	4	41.49	0.47	0.55	1.58
14739	3(1)	16	3	CC	M	1	58.82	0.48	-0.08	1.19
18011	3(1)	17	1	MA	G	3	70.98	0.45	-0.60	1.10
20115	NR	NR	1	CC	N	1	92.14	0.23	-2.86	0.57
20184	NR	NR	2	MA	N	2	72.06	0.37	-0.73	0.81
14718	NR	NR	2	HP	G	2	44.62	0.35	0.64	0.84
20147	NR	NR	1	MA	A	3	73.94	0.46	-0.75	1.10
20864	NR	NR	2	CC	T	2	84.27	0.35	-1.69	0.73
11643	3(2)	1	2	CC	N	4	72.48	0.49	-0.68	1.21
12775	3(2)	7	1	HP	M	2	51.91	0.37	0.28	1.00
15910	3(2)	14	2	CC	M	3	58.67	0.39	-0.04	0.88
20157	3(2)	15	2	HP	A	4	57.78	0.36	0.00	0.68
11539	3(2)	18	3	MA	N	4	75.80	0.40	-0.93	0.86
11646	NR	NR	3	HP	N	4	52.85	0.49	0.13	1.23
20869	NR	NR	1	HP	N	4 ⁺	70.50(2.82)	0.65	-1.33	0.53
20171	NR	NR	1	CC	M	2	72.62	0.45	-0.69	1.03
20102	NR	NR	2	HP	M	3	50.78	0.34	0.38	0.81
16338	NR	NR	3	MA	M	3	69.42	0.47	-0.52	1.13
13356	NR	NR	1	MA	M	4 ⁺	61.00(2.44)	0.65	-1.36	0.54
11481	NR	NR	1	CC	G	1	75.81	0.43	-0.85	1.04
20146	NR	NR	2	CC	G	4	83.78	0.38	-1.59	0.76
16360	NR	NR	2	HP	G	4 ⁺	55.75(2.23)	0.61	-1.08	0.48
20205	NR	NR	2	MA	A	1	63.03	0.47	-0.27	1.06
20868	NR	NR	2	MA	A	4 ⁺	74.25(2.97)	0.63	-1.80	0.60
20148	NR	NR	1	MA	T	1	48.76	0.35	0.49	0.73
20206	NR	NR	2	HP	T	2	49.74	0.46	0.28	1.40

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

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 1–8: Mathematics* (revised 2005).

⁺Maximum score code for open-response items.

() = mean score for open-response items. NR = not released.

**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
20866	3(1)	8	% of Students	0.86	0.24	15.62	18.42	11.69	53.17
			Parameters			-4.85	-1.19	0.21	-1.85
12755	3(1)	9	% of Students	0.49	0.19	14.21	18.44	18.72	47.95
			Parameters			-8.52	-1.54	-0.29	-2.16
20193	3(1)	10	% of Students	1.36	0.47	51.86	10.25	18.65	17.40
			Parameters			-6.80	2.68	-1.00	0.64
20210	3(1)	11	% of Students	0.36	0.16	21.79	10.56	12.42	54.71
			Parameters			-9.29	0.72	-0.69	-2.93
20869	NR	NR	% of Students	2.46	0.55	22.93	12.89	11.50	49.67
			Parameters			-3.61	-0.18	-0.16	-1.39
13356	NR	NR	% of Students	0.83	0.24	32.56	23.41	7.36	35.60
			Parameters			-5.14	-0.29	1.28	-1.30
16360	NR	NR	% of Students	1.65	0.46	39.15	19.36	11.94	27.44
			Parameters			-4.84	0.33	0.70	-0.49
20868	NR	NR	% of Students	0.76	0.07	6.69	31.60	15.95	44.93
			Parameters			-4.06	-2.60	0.22	-0.76

Note. The total number of students is 6741. NR = not released.

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.

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
20211	1(A)	1	-0.24	-0.56	0.09		-0.20	-0.52	0.12	
20215	1(A)	2	NA	NA	NA		NA	NA	NA	
20214	1(A)	3	0.47	0.15	0.80		0.70	0.38	1.01	
20213	1(A)	4	0.52	0.18	0.86		0.55	0.21	0.89	
20230	1(A)	7	0.37	0.03	0.71		0.35	0.01	0.69	
20231	1(A)	8	-0.22	-0.56	0.12		-0.31	-0.65	0.04	
20232	1(A)	9	1.01	0.67	1.36	B+	1.05	0.71	1.39	B+
20228	1(A)	10	0.53	0.18	0.88		0.72	0.35	1.08	
20063	1(B)	1	-0.23	-0.67	0.22		-0.57	-1.03	-0.11	
20067	1(B)	2	0.17	-0.21	0.54		0.03	-0.33	0.40	
20066	1(B)	3	-0.05	-0.43	0.33		0.08	-0.30	0.46	
20064	1(B)	4	-0.06	-0.37	0.26		-0.03	-0.34	0.28	
19954	NR	NR	-0.20	-0.53	0.14		-0.24	-0.58	0.10	
19943	NR	NR	-0.26	-0.68	0.15		0.17	-0.25	0.58	
19934	NR	NR	-0.27	-0.82	0.29		-0.46	-1.02	0.09	
19952	NR	NR	-0.02	-0.35	0.32		0.03	-0.31	0.36	
19960	NR	NR	-0.33	-0.66	-0.00		-0.56	-0.89	-0.24	
19980	NR	NR	0.03	-0.31	0.36		-0.09	-0.42	0.24	
19967	NR	NR	-0.25	-0.63	0.13		-0.13	-0.51	0.25	
19969	NR	NR	-0.29	-0.63	0.04		-0.50	-0.83	-0.17	
19973	NR	NR	-0.01	-0.38	0.35		0.26	-0.10	0.63	
19957	NR	NR	-0.24	-0.56	0.08		-0.23	-0.55	0.09	
19932	NR	NR	0.80	0.47	1.13		0.73	0.40	1.05	
19927	NR	NR	0.73	0.34	1.11		0.40	0.01	0.78	
19929	NR	NR	-0.21	-0.53	0.10		-0.20	-0.52	0.11	
19928	NR	NR	0.65	0.33	0.97		0.55	0.23	0.86	

Note. B = moderate DIF; C = large DIF; - = favouring female students; + = favouring male students; NR = not released. Item 20215 was removed from the assessment.

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
20203	1(A)	5	0.07	0.13		0.08	0.06	
20183	1(A)	6	0.09	0.03		0.12	0.00	
20225	1(A)	11	0.00	0.05		0.03	0.34	
20229	1(A)	12	-0.03	0.10		-0.05	0.25	
20060	1(B)	5	0.10	0.00		0.12	0.00	
20061	1(B)	6	0.14	0.00		0.10	0.02	
19924	NR	NR	-0.09	0.05		-0.11	0.00	
19925	NR	NR	-0.03	0.20		0.03	0.64	
19933	NR	NR	0.10	0.00		0.20	0.00	B-
19935	NR	NR	0.07	0.07		0.11	0.00	

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

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
19966	1(A)	1	-0.52	-0.90	-0.15		-0.54	-0.92	-0.16	
19958	1(A)	2	0.24	-0.47	0.95		0.37	-0.38	1.12	
19962	1(A)	3	0.51	-0.06	1.09		0.42	-0.13	0.96	
19964	1(A)	4	0.48	0.17	0.79		0.32	0.00	0.63	
20009	1(A)	7	1.12	0.79	1.45	B+	1.20	0.87	1.53	B+
20007	1(A)	8	0.71	0.31	1.10		0.41	0.03	0.80	
20006	1(A)	9	-0.24	-0.55	0.07		0.00	-0.31	0.32	
20011	1(A)	10	-0.50	-0.81	-0.19		-0.53	-0.83	-0.22	
20037	1(B)	1	0.72	0.27	1.16		0.41	-0.04	0.86	
20036	1(B)	2	-0.41	-0.75	-0.07		-0.22	-0.55	0.11	
20034	1(B)	3	1.36	0.90	1.81	B+	1.56	1.09	2.02	C+
20033	1(B)	4	0.21	-0.21	0.62		0.30	-0.11	0.72	
19757	NR	NR	-0.15	-0.62	0.32		-0.02	-0.48	0.43	
19780	NR	NR	-0.29	-0.64	0.07		-0.23	-0.58	0.12	
19753	NR	NR	-0.14	-0.50	0.22		0.13	-0.24	0.50	
19783	NR	NR	-0.51	-0.84	-0.17		-0.30	-0.64	0.04	
19751	NR	NR	0.07	-0.48	0.61		-0.29	-0.84	0.26	
19772	NR	NR	-0.57	-0.90	-0.23		-0.53	-0.86	-0.20	
19755	NR	NR	0.33	-0.08	0.74		0.53	0.12	0.94	
19763	NR	NR	-0.02	-0.40	0.36		-0.07	-0.45	0.32	
19767	NR	NR	-0.03	-0.43	0.38		0.09	-0.31	0.50	
19759	NR	NR	-0.47	-0.96	0.02		-0.15	-0.65	0.35	
19974	NR	NR	0.79	0.42	1.15		0.45	0.10	0.81	
19978	NR	NR	1.40	1.08	1.72	B+	0.94	0.62	1.26	
19971	NR	NR	0.83	0.47	1.19		1.02	0.65	1.38	B+
19976	NR	NR	1.76	1.39	2.12	C+	1.79	1.42	2.16	C+

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

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
19955	1(A)	5	0.05	0.30		0.01	0.22	
19956	1(A)	6	0.07	0.07		0.07	0.07	
20005	1(A)	11	0.04	0.40		0.11	0.00	
20004	1(A)	12	0.13	0.00		0.14	0.00	
20032	1(B)	5	0.08	0.00		0.05	0.31	
20031	1(B)	6	0.07	0.01		0.04	0.11	
19747	NR	NR	0.00	0.52		0.03	0.39	
19750	NR	NR	-0.04	0.31		-0.02	0.30	
19968	NR	NR	0.03	0.06		0.02	0.05	
19965	NR	NR	0.16	0.00		0.08	0.02	

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

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
19129	1(A)	1	-0.34	-0.67	-0.01		-0.09	-0.42	0.25	
19130	1(A)	2	-0.43	-0.81	-0.05		-0.77	-1.15	-0.38	
19128	1(A)	3	-0.44	-0.76	-0.11		-0.25	-0.57	0.08	
19131	1(A)	4	0.02	-0.33	0.36		0.01	-0.33	0.35	
19135	1(A)	7	0.01	-0.45	0.47		-0.10	-0.56	0.37	
19136	1(A)	8	0.63	0.32	0.94		0.65	0.34	0.96	
19137	1(A)	9	0.41	-0.05	0.87		0.29	-0.17	0.75	
19134	1(A)	10	0.18	-0.22	0.57		0.47	0.08	0.86	
11212	1(B)	1	0.12	-0.22	0.46		-0.07	-0.41	0.28	
11215	1(B)	2	0.29	-0.04	0.63		0.15	-0.19	0.48	
11213	1(B)	3	0.06	-0.30	0.42		0.30	-0.06	0.67	
11211	1(B)	4	0.84	0.51	1.17		0.78	0.45	1.11	
17680	NR	NR	-0.26	-0.61	0.09		-0.35	-0.70	0.00	
17679	NR	NR	-0.10	-0.50	0.30		0.03	-0.37	0.43	
17682	NR	NR	-0.39	-0.84	0.07		-0.92	-1.38	-0.46	
17686	NR	NR	0.06	-0.31	0.42		0.06	-0.31	0.43	
17683	NR	NR	-0.72	-1.33	-0.11		-0.65	-1.27	-0.03	
17684	NR	NR	0.01	-0.32	0.35		0.22	-0.12	0.55	
17685	NR	NR	0.09	-0.23	0.42		0.23	-0.09	0.55	
17678	NR	NR	-0.05	-0.40	0.30		-0.07	-0.42	0.28	
17681	NR	NR	0.25	-0.08	0.58		0.22	-0.11	0.55	
17677	NR	NR	-0.28	-0.66	0.09		-0.23	-0.60	0.15	
19140	NR	NR	0.62	0.27	0.98		0.64	0.28	1.01	
19141	NR	NR	-0.02	-0.40	0.36		0.05	-0.32	0.43	
19143	NR	NR	-0.30	-0.66	0.06		-0.12	-0.48	0.24	
19142	NR	NR	0.12	-0.25	0.48		0.01	-0.35	0.38	

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

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
19126	1(A)	5	0.14	0.00		0.12	0.00	
19127	1(A)	6	0.14	0.00		0.12	0.00	
19132	1(A)	11	0.10	0.00		0.08	0.01	
19133	1(A)	12	0.03	0.00		0.01	0.07	
11217	1(B)	5	0.02	0.66		-0.02	0.93	
11218	1(B)	6	0.05	0.50		0.02	0.84	
17687	NR	NR	-0.08	0.02		-0.11	0.00	
17688	NR	NR	-0.02	0.22		0.02	0.79	
19138	NR	NR	0.06	0.16		0.01	0.93	
19139	NR	NR	0.06	0.27		0.04	0.49	

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

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
19380	1(A)	1	-0.01	-0.41	0.38		-0.01	-0.40	0.39	
19382	1(A)	2	0.24	-0.13	0.61		0.31	-0.06	0.69	
19379	1(A)	3	0.11	-0.30	0.53		0.05	-0.37	0.47	
19381	1(A)	4	0.89	0.54	1.23		0.83	0.48	1.18	
19405	1(A)	7	0.48	-0.11	1.07		0.35	-0.25	0.96	
19408	1(A)	8	0.58	0.27	0.90		0.74	0.43	1.06	
19407	1(A)	9	0.08	-0.28	0.44		0.12	-0.24	0.49	
19406	1(A)	10	0.44	0.12	0.77		0.29	-0.04	0.62	
19210	1(B)	1	-0.57	-1.03	-0.11		-0.50	-0.96	-0.04	
19214	1(B)	2	0.03	-0.29	0.35		-0.12	-0.45	0.20	
19216	1(B)	3	0.33	-0.01	0.68		0.63	0.29	0.98	
19212	1(B)	4	0.25	-0.09	0.58		0.05	-0.29	0.39	
19417	NR	NR	0.38	0.05	0.71		0.55	0.22	0.88	
19419	NR	NR	-0.31	-0.70	0.09		-0.16	-0.56	0.24	
19411	NR	NR	0.35	-0.15	0.84		0.68	0.19	1.16	
19416	NR	NR	-0.04	-0.36	0.29		0.10	-0.23	0.43	
19418	NR	NR	1.50	1.15	1.84	B+	1.40	1.06	1.75	B+
19420	NR	NR	-0.43	-0.84	-0.02		-0.55	-0.96	-0.14	
19412	NR	NR	0.90	0.48	1.32		0.76	0.34	1.17	
19414	NR	NR	0.44	0.12	0.75		0.37	0.06	0.69	
19413	NR	NR	0.27	-0.17	0.71		0.16	-0.28	0.60	
19415	NR	NR	0.19	-0.13	0.51		0.32	-0.01	0.64	
19202	NR	NR	0.97	0.63	1.30		0.88	0.55	1.22	
19204	NR	NR	0.19	-0.16	0.53		0.15	-0.19	0.50	
19201	NR	NR	0.72	0.40	1.05		0.94	0.62	1.27	
19203	NR	NR	-0.53	-0.88	-0.18		-0.51	-0.86	-0.15	

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

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
19377	1(A)	5	0.11	0.00		0.12	0.00	
19378	1(A)	6	0.16	0.00		0.18	0.00	B-
19404	1(A)	11	0.15	0.00		0.13	0.00	
19403	1(A)	12	0.08	0.02		0.01	0.56	
19207	1(B)	5	0.12	0.00		0.17	0.00	
19209	1(B)	6	0.11	0.00		0.12	0.00	
19409	NR	NR	0.05	0.10		0.05	0.11	
19410	NR	NR	0.01	0.04		0.01	0.05	
19199	NR	NR	0.09	0.01		0.09	0.01	
19200	NR	NR	0.05	0.03		0.05	0.10	

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

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
20211	1(A)	1	0.36	0.04	0.67		0.34	0.03	0.66	
20215	1(A)	2								
20214	1(A)	3	0.61	0.29	0.93		0.82	0.50	1.14	
20213	1(A)	4	1.01	0.67	1.36	B+	0.76	0.42	1.09	
20230	1(A)	7	-0.16	-0.49	0.17		-0.04	-0.37	0.30	
20231	1(A)	8	0.35	0.02	0.69		0.42	0.08	0.77	
20232	1(A)	9	0.73	0.40	1.07		0.79	0.46	1.12	
20228	1(A)	10	0.27	-0.08	0.62		0.48	0.13	0.82	
20063	1(B)	1	-0.96	-1.43	-0.49		-0.74	-1.20	-0.29	
20067	1(B)	2	0.10	-0.27	0.46		0.08	-0.28	0.44	
20066	1(B)	3	0.46	0.09	0.83		0.74	0.37	1.11	
20064	1(B)	4	0.06	-0.25	0.37		0.15	-0.16	0.46	
19954	NR	NR	0.50	0.16	0.84		0.28	-0.05	0.62	
19943	NR	NR	0.57	0.17	0.98		0.16	-0.25	0.56	
19934	NR	NR	-0.43	-0.96	0.11		-0.02	-0.56	0.51	
19952	NR	NR	-0.52	-0.85	-0.19		0.14	-0.19	0.47	
19960	NR	NR	-0.25	-0.57	0.07		-0.24	-0.56	0.09	
19980	NR	NR	0.58	0.26	0.91		0.23	-0.10	0.55	
19967	NR	NR	-0.43	-0.82	-0.04		-0.30	-0.69	0.09	
19969	NR	NR	0.01	-0.31	0.34		0.30	-0.03	0.63	
19973	NR	NR	0.07	-0.29	0.43		-0.15	-0.50	0.20	
19957	NR	NR	0.61	0.29	0.93		0.62	0.30	0.94	
19932	NR	NR	1.03	0.70	1.36	B+	0.73	0.40	1.05	
19927	NR	NR	0.15	-0.24	0.53		-0.12	-0.50	0.27	
19929	NR	NR	-0.11	-0.42	0.20		0.19	-0.12	0.50	
19928	NR	NR	0.94	0.61	1.27		0.59	0.27	0.91	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
20203	1(A)	5	0.14	0.00		0.09	0.00	
20183	1(A)	6	0.02	0.15		0.03	0.50	
20225	1(A)	11	0.11	0.00		0.05	0.27	
20229	1(A)	12	0.04	0.24		-0.01	0.84	
20060	1(B)	5	0.02	0.20		0.03	0.72	
20061	1(B)	6	0.11	0.00		0.11	0.00	
19924	NR	NR	0.01	0.72		-0.02	0.41	
19925	NR	NR	0.08	0.01		0.05	0.32	
19933	NR	NR	0.17	0.00	B-	0.10	0.00	
19935	NR	NR	0.13	0.00		0.09	0.00	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19966	1(A)	1	0.92	0.56	1.28		0.76	0.40	1.12	
19958	1(A)	2	1.06	0.39	1.73	B+	0.88	0.19	1.57	
19962	1(A)	3	1.01	0.49	1.52	B+	0.85	0.34	1.37	
19964	1(A)	4	0.85	0.54	1.16		0.50	0.20	0.81	
20009	1(A)	7	0.47	0.15	0.79		0.37	0.05	0.68	
20007	1(A)	8	0.21	-0.17	0.58		0.16	-0.22	0.54	
20006	1(A)	9	0.04	-0.27	0.35		0.20	-0.11	0.50	
20011	1(A)	10	-0.33	-0.64	-0.03		-0.14	-0.45	0.17	
20037	1(B)	1	1.63	1.21	2.05	C+	2.06	1.64	2.49	C+
20036	1(B)	2	0.35	0.02	0.68		0.21	-0.13	0.54	
20034	1(B)	3	1.45	1.04	1.86	B+	1.29	0.87	1.71	B+
20033	1(B)	4	0.41	0.02	0.79		0.84	0.45	1.24	
19757	NR	NR	0.30	-0.13	0.73		0.50	0.06	0.93	
19780	NR	NR	1.30	0.97	1.64	B+	1.25	0.91	1.58	B+
19753	NR	NR	0.13	-0.23	0.49		0.05	-0.31	0.40	
19783	NR	NR	0.43	0.09	0.76		-0.01	-0.35	0.33	
19751	NR	NR	0.42	-0.09	0.94		0.29	-0.24	0.82	
19772	NR	NR	1.43	1.11	1.74	B+	1.68	1.36	1.99	C+
19755	NR	NR	0.51	0.12	0.90		0.83	0.45	1.22	
19763	NR	NR	-0.01	-0.38	0.36		-0.23	-0.60	0.15	
19767	NR	NR	-0.15	-0.55	0.25		-0.21	-0.60	0.18	
19759	NR	NR	-0.06	-0.51	0.40		0.27	-0.22	0.76	
19974	NR	NR	-0.16	-0.52	0.19		0.02	-0.34	0.38	
19978	NR	NR	0.31	0.00	0.63		0.45	0.13	0.77	
19971	NR	NR	0.13	-0.22	0.48		0.21	-0.15	0.56	
19976	NR	NR	0.11	-0.24	0.45		0.00	-0.35	0.35	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19955	1(A)	5	-0.02	0.46		-0.04	0.26	
19956	1(A)	6	0.02	0.10		0.06	0.10	
20005	1(A)	11	0.06	0.00		0.14	0.00	
20004	1(A)	12	0.07	0.00		0.05	0.10	
20032	1(B)	5	0.11	0.00		0.03	0.17	
20031	1(B)	6	0.07	0.03		0.05	0.13	
19747	NR	NR	0.11	0.00		0.11	0.00	
19750	NR	NR	0.11	0.00		0.08	0.02	
19968	NR	NR	0.10	0.00		0.10	0.00	
19965	NR	NR	0.13	0.00		0.14	0.00	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19129	1(A)	1	0.20	-0.14	0.53		0.05	-0.29	0.39	
19130	1(A)	2	0.30	-0.08	0.68		0.34	-0.04	0.72	
19128	1(A)	3	0.23	-0.10	0.55		-0.01	-0.34	0.31	
19131	1(A)	4	-0.28	-0.62	0.05		-0.16	-0.51	0.18	
19135	1(A)	7	0.13	-0.32	0.59		-0.16	-0.62	0.29	
19136	1(A)	8	-0.28	-0.58	0.03		-0.08	-0.39	0.23	
19137	1(A)	9	0.86	0.41	1.30		0.52	0.08	0.97	
19134	1(A)	10	0.26	-0.13	0.65		0.44	0.04	0.83	
11212	1(B)	1	0.22	-0.12	0.56		0.06	-0.28	0.41	
11215	1(B)	2	-0.02	-0.35	0.31		-0.07	-0.41	0.26	
11213	1(B)	3	0.50	0.14	0.86		0.66	0.30	1.02	
11211	1(B)	4	-0.03	-0.35	0.30		0.06	-0.27	0.39	
17680	NR	NR	-0.02	-0.37	0.34		-0.05	-0.41	0.31	
17679	NR	NR	-0.19	-0.60	0.21		0.10	-0.31	0.50	
17682	NR	NR	0.65	0.21	1.09		0.54	0.09	0.99	
17686	NR	NR	0.11	-0.24	0.47		-0.10	-0.46	0.26	
17683	NR	NR	-0.38	-0.97	0.22		-0.39	-1.00	0.21	
17684	NR	NR	-0.21	-0.55	0.13		-0.17	-0.51	0.17	
17685	NR	NR	-0.19	-0.51	0.14		0.12	-0.21	0.45	
17678	NR	NR	-0.15	-0.50	0.20		-0.28	-0.63	0.07	
17681	NR	NR	0.67	0.34	1.00		0.52	0.19	0.85	
17677	NR	NR	0.00	-0.38	0.38		0.41	0.02	0.79	
19140	NR	NR	-0.01	-0.37	0.35		0.08	-0.29	0.44	
19141	NR	NR	-0.17	-0.55	0.21		-0.31	-0.70	0.07	
19143	NR	NR	0.30	-0.05	0.66		-0.05	-0.41	0.31	
19142	NR	NR	0.22	-0.14	0.59		0.33	-0.04	0.70	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19126	1(A)	5	-0.04	0.55		-0.07	0.21	
19127	1(A)	6	-0.06	0.03		-0.03	0.34	
19132	1(A)	11	0.03	0.01		0.01	0.00	
19133	1(A)	12	-0.02	0.66		-0.04	0.08	
11217	1(B)	5	0.01	0.33		0.00	0.69	
11218	1(B)	6	0.00	0.91		-0.01	0.45	
17687	NR	NR	0.02	0.29		0.01	0.35	
17688	NR	NR	0.05	0.06		0.05	0.21	
19138	NR	NR	-0.01	0.59		0.00	0.59	
19139	NR	NR	0.00	0.59		-0.04	0.20	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19380	1(A)	1	0.47	-0.05	1.00		0.93	0.38	1.48	
19382	1(A)	2	0.46	-0.04	0.97		0.21	-0.30	0.72	
19379	1(A)	3	0.34	-0.22	0.91		0.65	0.07	1.23	
19381	1(A)	4	0.19	-0.27	0.65		0.26	-0.20	0.73	
19405	1(A)	7	0.36	-0.48	1.20		0.26	-0.58	1.10	
19408	1(A)	8	0.50	0.07	0.94		0.45	0.01	0.89	
19407	1(A)	9	-0.22	-0.72	0.27		0.36	-0.14	0.87	
19406	1(A)	10	-0.06	-0.51	0.39		0.11	-0.34	0.56	
19210	1(B)	1	0.18	-0.43	0.80		0.40	-0.23	1.03	
19214	1(B)	2	0.33	-0.13	0.78		-0.08	-0.53	0.37	
19216	1(B)	3	-0.19	-0.67	0.29		-0.32	-0.80	0.16	
19212	1(B)	4	-0.14	-0.61	0.32		-0.05	-0.52	0.43	
19417	NR	NR	0.00	-0.46	0.45		0.08	-0.38	0.54	
19419	NR	NR	0.22	-0.32	0.76		0.47	-0.06	1.01	
19411	NR	NR	-0.62	-1.27	0.03		-0.25	-0.92	0.43	
19416	NR	NR	0.53	0.08	0.98		0.44	-0.02	0.90	
19418	NR	NR	0.38	-0.09	0.85		0.47	-0.01	0.95	
19420	NR	NR	-0.37	-0.92	0.19		-0.04	-0.60	0.53	
19412	NR	NR	-0.11	-0.65	0.44		0.41	-0.14	0.97	
19414	NR	NR	1.24	0.80	1.68	B+	0.98	0.54	1.41	
19413	NR	NR	-0.04	-0.64	0.56		-0.41	-1.01	0.19	
19415	NR	NR	0.15	-0.30	0.60		-0.17	-0.62	0.27	
19202	NR	NR	0.06	-0.40	0.52		0.08	-0.38	0.55	
19204	NR	NR	0.07	-0.41	0.55		0.10	-0.38	0.59	
19201	NR	NR	-0.09	-0.54	0.36		-0.21	-0.67	0.24	
19203	NR	NR	-0.21	-0.71	0.29		-0.49	-0.99	0.01	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19377	1(A)	5	0.06	0.10		0.05	0.31	
19378	1(A)	6	0.06	0.56		0.01	0.96	
19404	1(A)	11	0.04	0.15		0.04	0.86	
19403	1(A)	12	0.02	0.83		0.01	0.34	
19207	1(B)	5	0.05	0.50		0.01	0.20	
19209	1(B)	6	0.05	0.14		0.04	0.60	
19409	NR	NR	0.09	0.07		0.12	0.00	
19410	NR	NR	0.06	0.77		-0.02	0.17	
19199	NR	NR	0.01	0.68		0.03	0.87	
19200	NR	NR	0.00	0.73		0.00	0.93	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19878	2(C)	17	0.15	0.49	-0.19		-0.09	0.25	-0.43	
19885	2(C)	15	-0.33	0.08	-0.73		-0.27	0.14	-0.69	
19889	2(C)	14	-0.17	0.20	-0.54		0.38	0.77	-0.00	
19890	2(C)	16	-0.84	-0.46	-1.22		-0.56	-0.18	-0.94	
19870	2(D)	11	0.28	0.81	-0.25		-0.49	0.04	-1.01	
19876	2(D)	10	-0.30	0.04	-0.63		-0.27	0.07	-0.60	
19884	2(D)	8	-0.04	0.27	-0.36		-0.49	-0.18	-0.81	
19898	2(D)	9	0.36	0.69	0.04		0.17	0.49	-0.16	

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
19895 T	NR	NR	-0.05	0.01		-0.03	0.54	
19895 V	NR	NR	-0.03	0.21		-0.03	0.08	
19879 T	2(C)	13	0.06	0.00		0.05	0.02	
19879 V	2(C)	13	0.02	0.70		-0.01	0.18	
22817 T	2(D)	7	-0.03	0.07		0.00	0.97	
22817 V	2(D)	7	-0.06	0.01		-0.10	0.00	

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

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
19739	2(C)	14	-1.33	-0.98	-1.69	B-	-1.17	-0.82	-1.52	B-
19754	2(C)	17	-0.80	-0.45	-1.15		-1.14	-0.79	-1.49	B-
19781	2(C)	15	-0.08	0.26	-0.41		-0.18	0.16	-0.52	
19806	2(C)	16	-0.26	0.11	-0.63		-0.57	-0.20	-0.95	
17724	2(D)	10	-0.47	-0.06	-0.87		-0.03	0.37	-0.42	
19737	2(D)	8	-0.45	-0.11	-0.80		-0.39	-0.05	-0.73	
19745	2(D)	11	-0.72	-0.35	-1.08		-0.36	0.00	-0.71	
20996	2(D)	9	0.22	0.55	-0.11		-0.24	0.09	-0.58	

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
19756 T	NR	NR	-0.10	0.00		-0.09	0.00	
19756 V	NR	NR	-0.05	0.03		-0.05	0.02	
16534 T	2(C)	13	-0.03	0.59		-0.07	0.00	
16534 V	2(C)	13	-0.04	0.00		-0.07	0.00	
22724 T	2(D)	7	-0.07	0.03		-0.03	0.26	
22724 V	2(D)	7	-0.06	0.01		-0.07	0.00	

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

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
15487	2(C)	17	-0.87	-0.35	-1.39		-0.34	0.16	-0.83	
17713	2(C)	15	0.38	0.88	-0.13		-0.44	0.04	-0.91	
19017	2(C)	14	-0.42	0.17	-1.01		-0.59	-0.04	-1.14	
20849	2(C)	16	0.21	0.76	-0.34		-0.35	0.16	-0.86	
18486	2(D)	10	-0.37	0.17	-0.91		-0.02	0.49	-0.53	
19041	2(D)	11	-0.46	0.09	-1.00		0.00	0.50	-0.50	
20846	2(D)	9	-0.16	0.38	-0.70		-0.26	0.25	-0.76	
20850	2(D)	8	0.16	0.71	-0.38		0.05	0.57	-0.47	

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
17700_T	NR	NR	-0.02	0.46		0.00	0.49	
17700_V	NR	NR	-0.06	0.05		0.00	0.07	
17701_T	2(C)	13	-0.06	0.20		-0.02	0.33	
17701_V	2(C)	13	0.04	0.04		-0.02	0.84	
22852_T	2(D)	7	-0.03	0.44		-0.03	0.03	
22852_V	2(D)	7	-0.02	0.07		-0.03	0.00	

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

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
19536	2(C)	14	-0.03	0.52	-0.58		0.64	1.18	0.10	
21002	2(C)	15	-0.23	0.37	-0.84		-0.26	0.32	-0.84	
13022	2(C)	16	-0.70	-0.14	-1.26		-0.99	-0.44	-1.55	
12540	2(C)	17	-0.17	0.32	-0.66		-0.21	0.27	-0.69	
20870	2(D)	8	-0.48	0.17	-1.14		0.06	0.70	-0.57	
19533	2(D)	9	0.45	1.03	-0.13		-0.03	0.53	-0.58	
20702	2(D)	10	-1.02	-0.50	-1.55	B-	-1.20	-0.69	-1.71	B-
19549	2(D)	11	-0.82	-0.20	-1.45		-0.90	-0.29	-1.51	

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
23464_T	NR	NR	-0.04	0.07		-0.02	0.28	
23464_V	NR	NR	-0.07	0.02		-0.07	0.07	
19532_T	2(C)	13	-0.02	0.65		-0.04	0.00	
19532_V	2(C)	13	-0.01	0.05		-0.07	0.11	
26298_T	2(D)	7	-0.08	0.14		-0.03	0.43	
26298_V	2(D)	7	-0.03	0.20		-0.05	0.05	

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

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
19878	2(C)	17	0.29	0.63	-0.05		0.19	0.54	-0.15	
19885	2(C)	15	0.52	0.92	0.13		0.39	0.79	-0.02	
19889	2(C)	14	-0.07	0.30	-0.44		0.21	0.59	-0.16	
19890	2(C)	16	-0.08	0.30	-0.45		-0.49	-0.11	-0.87	
19870	2(D)	11	0.86	1.37	0.34		-0.07	0.43	-0.57	
19876	2(D)	10	-0.10	0.23	-0.43		0.15	0.49	-0.18	
19884	2(D)	8	0.27	0.58	-0.05		0.41	0.72	0.09	
19898	2(D)	9	-0.02	0.30	-0.34		0.05	0.36	-0.27	

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
19895_T	NR	NR	0.08	0.01		0.07	0.00	
19895_V	NR	NR	-0.08	0.01		-0.05	0.09	
19879_T	2(C)	13	0.01	0.57		0.02	0.47	
19879_V	2(C)	13	-0.06	0.03		-0.04	0.20	
22817_T	2(D)	7	-0.01	0.55		-0.03	0.73	
22817_V	2(D)	7	-0.08	0.00		-0.08	0.00	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19739	2(C)	14	-0.19	0.14	-0.52		-0.31	0.03	-0.64	
19754	2(C)	17	0.21	0.54	-0.12		0.78	1.12	0.45	
19781	2(C)	15	0.46	0.79	0.14		-0.03	0.30	-0.35	
19806	2(C)	16	-0.02	0.34	-0.37		-0.12	0.24	-0.47	
17724	2(D)	10	1.41	1.79	1.04	B+	1.26	1.63	0.89	B+
19737	2(D)	8	-0.07	0.26	-0.39		0.13	0.46	-0.20	
19745	2(D)	11	-0.29	0.05	-0.63		0.23	0.58	-0.12	
20996	2(D)	9	-0.01	0.31	-0.33		0.28	0.60	-0.04	

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
19756_T	NR	NR	-0.04	0.06		-0.02	0.34	
19756_V	NR	NR	-0.03	0.49		-0.04	0.35	
16534_T	2(C)	13	-0.03	0.74		-0.04	0.13	
16534_V	2(C)	13	-0.03	0.55		-0.01	0.68	
22724_T	2(D)	7	0.01	0.81		-0.02	0.51	
22724_V	2(D)	7	0.01	0.58		-0.03	0.27	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
15487	2(C)	17	-0.03	0.41	-0.48		0.40	0.83	-0.03	
17713	2(C)	15	0.15	0.58	-0.28		-0.42	-0.01	-0.82	
19017	2(C)	14	-0.01	0.49	-0.51		0.45	0.92	-0.02	
20849	2(C)	16	0.04	0.50	-0.42		-0.04	0.39	-0.47	
18486	2(D)	10	-0.06	0.40	-0.51		0.19	0.62	-0.25	
19041	2(D)	11	0.06	0.52	-0.39		-0.33	0.09	-0.76	
20846	2(D)	9	0.40	0.87	-0.07		-0.12	0.32	-0.55	
20850	2(D)	8	-0.14	0.32	-0.60		0.28	0.73	-0.16	

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
17700 T	NR	NR	0.00	0.16		-0.02	0.25	
17700 V	NR	NR	0.00	0.88		0.01	0.86	
17701 T	2(C)	13	0.01	0.24		0.02	0.05	
17701 V	2(C)	13	0.01	0.88		0.03	0.20	
22852 T	2(D)	7	-0.02	0.40		-0.04	0.19	
22852 V	2(D)	7	-0.01	0.73		-0.02	0.33	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
12540	2(C)	17	0.19	0.67	-0.29		0.17	0.64	-0.31	
13022	2(C)	16	0.19	0.72	-0.35		0.00	0.52	-0.52	
19536	2(C)	14	0.36	0.88	-0.17		0.20	0.71	-0.31	
21002	2(C)	15	-0.10	0.47	-0.67		-0.24	0.31	-0.79	
19533	2(D)	9	0.28	0.83	-0.27		0.44	0.97	-0.09	
19549	2(D)	11	0.01	0.59	-0.56		0.35	0.92	-0.22	
20702	2(D)	10	0.52	1.01	0.03		0.17	0.66	-0.32	
20870	2(D)	8	0.30	0.92	-0.33		0.09	0.69	-0.52	

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
23464 T	NR	NR	0.03	0.14		0.03	0.84	
23464 V	NR	NR	0.02	0.06		0.02	0.56	
19532 T	2(C)	13	-0.08	0.07		0.00	0.70	
19532 V	2(C)	13	0.01	0.16		-0.05	0.21	
26298 T	2(D)	7	-0.09	0.00		-0.06	0.28	
26298 V	2(D)	7	-0.01	0.34		-0.04	0.43	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
16830	3(1)	1	-0.07	-0.46	0.32		0.01	-0.38	0.40	
19253	3(1)	2	-0.17	-0.52	0.19		-0.03	-0.39	0.33	
11963	3(1)	3	-0.11	-0.61	0.38		-0.06	-0.55	0.44	
19263	3(1)	4	-0.22	-0.55	0.11		-0.51	-0.85	-0.17	
19340	3(1)	5	0.32	-0.03	0.67		0.65	0.30	1.00	
16785	3(1)	6	0.09	-0.25	0.42		0.27	-0.06	0.60	
19290	3(1)	7	-0.35	-0.67	-0.03		-0.07	-0.39	0.26	
23488	3(1)	12	-0.34	-0.67	-0.02		0.32	0.00	0.63	
11991	3(1)	13	0.27	-0.05	0.59		0.23	-0.09	0.55	
19337	NR	NR	0.32	-0.03	0.67		0.60	0.25	0.96	
19293	NR	NR	-0.19	-0.53	0.16		0.19	-0.16	0.53	
16823	NR	NR	-0.08	-0.44	0.28		-0.28	-0.63	0.08	
19267	NR	NR	-0.81	-1.16	-0.47		-0.83	-1.17	-0.49	
19333	NR	NR	-0.20	-0.57	0.17		-0.31	-0.68	0.06	
10950	3(2)	14	0.69	0.35	1.02		0.73	0.39	1.08	
19325	3(2)	15	-0.57	-0.93	-0.21		-0.58	-0.94	-0.23	
16715	3(2)	16	0.93	0.61	1.26		1.02	0.69	1.35	B+
18361	3(2)	17	0.14	-0.18	0.45		0.06	-0.25	0.38	
19213	3(2)	18	0.75	0.42	1.08		0.87	0.54	1.20	
15074	NR	NR	1.48	1.14	1.82	B+	1.56	1.22	1.90	C+
19342	NR	NR	-0.69	-1.09	-0.29		-0.50	-0.89	-0.11	
16628	NR	NR	0.10	-0.26	0.45		0.12	-0.25	0.48	
10775	NR	NR	0.74	0.32	1.15		0.91	0.50	1.32	
16845	NR	NR	-0.46	-0.81	-0.11		-0.64	-0.99	-0.29	
19250	NR	NR	-0.36	-0.75	0.03		-0.28	-0.65	0.09	
10781	NR	NR	0.46	0.11	0.80		0.03	-0.32	0.38	
15145	NR	NR	0.34	0.02	0.66		0.23	-0.08	0.55	
19332	NR	NR	-0.18	-0.61	0.26		-0.57	-1.00	-0.14	

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

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
16773	3(1)	8	-0.05	0.01		-0.05	0.00	
15133	3(1)	9	0.04	0.66		0.06	0.03	
19299	NR	NR	0.02	0.32		0.01	0.92	
16875	NR	NR	0.12	0.00		0.12	0.00	
19334	3(2)	10	0.08	0.02		0.15	0.00	
10987	3(2)	11	0.01	0.79		0.00	0.10	
11980	NR	NR	-0.07	0.01		-0.03	0.14	
19664	NR	NR	0.03	0.03		-0.01	0.00	

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

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	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11301	3(1)	1	0.63	0.14	1.12		0.91	0.42	1.40	
23480	3(1)	2	-0.32	-0.64	0.00		-0.09	-0.41	0.23	
11314	3(1)	3	0.45	0.04	0.87		0.37	-0.04	0.77	
12711	3(1)	4	-0.42	-0.77	-0.08		-0.33	-0.68	0.02	
23506	3(1)	5	-0.49	-0.91	-0.06		0.26	-0.16	0.69	
11320	3(1)	6	0.33	0.00	0.65		0.55	0.23	0.87	
15001	3(1)	7	-0.02	-0.35	0.32		0.37	0.04	0.70	
17180	3(1)	12	-0.31	-0.70	0.07		-0.22	-0.61	0.17	
20471	NR	NR	0.54	0.16	0.91		0.24	-0.14	0.62	
22493	NR	NR	-0.03	-0.36	0.30		-0.43	-0.75	-0.10	
23481	NR	NR	0.21	-0.10	0.53		0.47	0.15	0.78	
17105	NR	NR	-0.26	-0.61	0.09		-0.40	-0.75	-0.05	
15066	NR	NR	0.09	-0.23	0.41		0.25	-0.08	0.57	
23503	NR	NR	-0.08	-0.40	0.23		-0.37	-0.69	-0.06	
14980	3(2)	13	-0.11	-0.56	0.34		-0.41	-0.86	0.05	
15013	3(2)	14	-0.07	-0.39	0.25		-0.06	-0.38	0.26	
20467	3(2)	15	0.18	-0.23	0.59		-0.05	-0.46	0.37	
20521	3(2)	16	0.30	-0.02	0.61		0.69	0.39	1.00	
11295	3(2)	17	1.15	0.80	1.49	B+	1.05	0.71	1.39	B+
22461	3(2)	18	0.52	0.19	0.86		0.28	-0.05	0.62	
11294	NR	NR	0.29	-0.01	0.60		0.29	-0.01	0.60	
20458	NR	NR	-0.08	-0.45	0.30		0.14	-0.24	0.52	
17137	NR	NR	0.54	0.19	0.89		0.49	0.13	0.85	
20512	NR	NR	-0.07	-0.54	0.40		-0.16	-0.62	0.31	
17143	NR	NR	0.76	0.44	1.08		0.80	0.48	1.11	
12678	NR	NR	0.22	-0.13	0.58		0.10	-0.25	0.46	
11342	NR	NR	0.46	0.13	0.79		0.33	0.01	0.65	
15059	NR	NR	-0.11	-0.48	0.26		-0.65	-1.02	-0.28	

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

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
12717	3(1)	8	0.08	0.00		0.11	0.00	
11307	3(1)	9	-0.03	0.41		-0.04	0.02	
11358	NR	NR	0.11	0.00		0.04	0.06	
15067	NR	NR	0.11	0.00		0.12	0.00	
15071	3(2)	10	-0.14	0.00		-0.08	0.00	
17150	3(2)	11	0.00	0.53		0.02	0.07	
22532	NR	NR	-0.01	0.06		-0.05	0.00	
20495	NR	NR	0.14	0.00		0.14	0.00	

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

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	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
19815	3(1)	1	0.37	-0.17	0.92		0.50	-0.05	1.05	
23536	3(1)	2	0.23	-0.14	0.60		0.12	-0.24	0.49	
14570	3(1)	3	1.81	1.37	2.24	C+	1.59	1.15	2.03	C+
19776	3(1)	4	0.19	-0.14	0.53		0.30	-0.03	0.64	
19669	3(1)	5	0.60	0.24	0.95		0.49	0.13	0.84	
11237	3(1)	6	0.07	-0.27	0.41		0.06	-0.28	0.40	
16335	3(1)	7	0.17	-0.17	0.52		0.13	-0.21	0.47	
14585	NR	NR	0.31	-0.02	0.65		0.12	-0.21	0.46	
19711	NR	NR	0.24	-0.19	0.67		0.02	-0.40	0.44	
11230	NR	NR	0.72	0.37	1.06		0.69	0.34	1.04	
16424	NR	NR	-0.31	-0.66	0.04		-0.59	-0.94	-0.24	
19799	NR	NR	-1.15	-1.59	-0.70	B-	-0.90	-1.36	-0.44	
20694	NR	NR	-0.26	-0.63	0.11		-0.41	-0.78	-0.04	
19800	NR	NR	0.01	-0.34	0.37		-0.09	-0.45	0.27	
19829	3(2)	12	-0.25	-0.60	0.09		-0.13	-0.47	0.21	
19808	3(2)	13	0.76	0.39	1.12		0.60	0.23	0.97	
20740	3(2)	14	-0.38	-0.78	0.01		-0.42	-0.81	-0.03	
20873	3(2)	15	-0.19	-0.52	0.13		-0.29	-0.61	0.03	
20713	3(2)	16	-0.36	-0.68	-0.03		-0.42	-0.74	-0.09	
19676	3(2)	17	0.10	-0.22	0.42		0.07	-0.25	0.39	
16457	3(2)	18	-0.15	-0.56	0.25		-0.05	-0.45	0.35	
16296	NR	NR	0.74	0.38	1.10		0.42	0.06	0.78	
20709	NR	NR	-0.19	-0.51	0.14		-0.02	-0.35	0.31	
20579	NR	NR	0.43	0.11	0.76		0.03	-0.29	0.35	
16373	NR	NR	-0.56	-1.16	0.04		-0.88	-1.50	-0.26	
16329	NR	NR	-0.74	-1.08	-0.39		-0.55	-0.90	-0.21	
20593	NR	NR	0.23	-0.10	0.56		0.43	0.11	0.76	
14655	NR	NR	-0.32	-0.67	0.04		-0.24	-0.60	0.11	

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

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
16444	3(1)	8	-0.08	0.00		-0.07	0.00	
19826	3(1)	9	0.04	0.00		0.05	0.00	
16516	NR	NR	-0.07	0.00		-0.07	0.00	
19706	NR	NR	0.01	0.00		0.01	0.28	
20594	3(2)	10	0.02	0.00		0.02	0.12	
19688	3(2)	11	0.09	0.01		0.06	0.05	
23534	NR	NR	-0.02	0.13		-0.03	0.29	
14666	NR	NR	0.11	0.00		0.11	0.00	

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

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	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11538	3(1)	2	0.28	-0.05	0.60		0.42	0.09	0.74	
16314	3(1)	3	0.18	-0.19	0.56		0.26	-0.11	0.63	
14725	3(1)	4	0.14	-0.18	0.45		0.07	-0.25	0.39	
20106	3(1)	5	0.38	0.07	0.68		0.17	-0.13	0.47	
14694	3(1)	6	0.17	-0.17	0.51		0.04	-0.30	0.38	
20862	3(1)	12	-0.33	-0.70	0.04		-0.21	-0.58	0.15	
14765	3(1)	13	-0.07	-0.44	0.29		-0.11	-0.47	0.25	
14739	3(1)	16	1.85	1.47	2.22	C+	1.92	1.55	2.29	C+
18011	3(1)	17	-0.08	-0.46	0.29		-0.08	-0.45	0.29	
20115	NR	NR	0.00	-0.57	0.56		0.40	-0.16	0.96	
20184	NR	NR	-0.11	-0.48	0.25		-0.32	-0.68	0.04	
14718	NR	NR	-0.42	-0.74	-0.09		-0.31	-0.64	0.01	
20147	NR	NR	0.21	-0.18	0.60		0.18	-0.21	0.57	
20864	NR	NR	0.66	0.21	1.11		0.62	0.17	1.06	
11643	3(2)	1	1.50	1.09	1.91	C+	1.61	1.20	2.03	C+
12775	3(2)	7	0.63	0.30	0.97		0.59	0.26	0.92	
15910	3(2)	14	-0.32	-0.66	0.02		-0.46	-0.79	-0.13	
20157	3(2)	15	-0.38	-0.70	-0.05		-0.36	-0.69	-0.04	
11539	3(2)	18	-0.01	-0.40	0.37		-0.06	-0.44	0.33	
11646	NR	NR	-1.13	-1.49	-0.77	B-	-1.03	-1.38	-0.67	B-
20171	NR	NR	0.02	-0.36	0.40		-0.15	-0.53	0.23	
20102	NR	NR	-0.15	-0.48	0.17		0.06	-0.26	0.38	
16338	NR	NR	1.10	0.72	1.48	B+	1.35	0.96	1.73	B+
11481	NR	NR	0.28	-0.12	0.68		0.12	-0.27	0.52	
20146	NR	NR	-0.59	-1.03	-0.15		-0.64	-1.08	-0.21	
20205	NR	NR	-0.32	-0.67	0.04		-0.34	-0.70	0.02	
20148	NR	NR	-0.26	-0.58	0.06		-0.29	-0.60	0.03	
20206	NR	NR	1.46	1.10	1.81	B+	1.47	1.11	1.83	B+

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

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
20866	3(1)	8	-0.01	0.01		0.00	0.00	
12755	3(1)	9	0.08	0.00		0.08	0.00	
20193	3(1)	10	0.03	0.10		0.04	0.27	
20210	3(1)	11	0.15	0.00		0.15	0.00	
20869	NR	NR	-0.07	0.00		-0.08	0.00	
13356	NR	NR	0.07	0.00		0.08	0.00	
16360	NR	NR	0.03	0.04		0.03	0.05	
20868	NR	NR	0.02	0.01		0.02	0.07	

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

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	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
16830	3(1)	1	-1.11	-1.52	-0.70	B-	-1.68	-2.10	-1.26	C-
19253	3(1)	2	-0.64	-1.00	-0.28		-0.75	-1.12	-0.39	
11963	3(1)	3	0.46	-0.03	0.96		0.88	0.38	1.38	
19263	3(1)	4	-0.95	-1.29	-0.61		-0.75	-1.09	-0.41	
19340	3(1)	5	-0.09	-0.44	0.25		-0.24	-0.59	0.11	
16785	3(1)	6	0.21	-0.13	0.55		0.06	-0.28	0.39	
19290	3(1)	7	-0.18	-0.51	0.14		-0.22	-0.55	0.10	
23488	3(1)	12	0.18	-0.14	0.50		0.35	0.03	0.67	
11991	3(1)	13	0.50	0.18	0.83		0.36	0.03	0.68	
19337	NR	NR	-0.07	-0.42	0.27		-0.11	-0.46	0.24	
19293	NR	NR	-0.54	-0.88	-0.19		-0.20	-0.54	0.15	
16823	NR	NR	0.03	-0.33	0.39		-0.15	-0.51	0.21	
19267	NR	NR	0.26	-0.08	0.60		0.09	-0.26	0.43	
19333	NR	NR	0.39	0.03	0.74		0.84	0.47	1.20	
10950	3(2)	14	-0.03	-0.36	0.31		0.08	-0.26	0.42	
19325	3(2)	15	0.20	-0.16	0.56		-0.30	-0.66	0.07	
16715	3(2)	16	0.19	-0.14	0.51		0.17	-0.16	0.49	
18361	3(2)	17	0.67	0.35	0.99		0.98	0.66	1.30	
19213	3(2)	18	0.35	0.03	0.68		0.44	0.11	0.77	
15074	NR	NR	0.06	-0.28	0.39		-0.31	-0.64	0.02	
19342	NR	NR	-0.07	-0.47	0.32		0.23	-0.16	0.62	
16628	NR	NR	0.42	0.06	0.78		-0.03	-0.40	0.34	
10775	NR	NR	-0.92	-1.34	-0.50		-0.97	-1.38	-0.55	
16845	NR	NR	0.06	-0.29	0.41		0.09	-0.26	0.44	
19250	NR	NR	0.50	0.12	0.87		0.63	0.25	1.01	
10781	NR	NR	-0.44	-0.80	-0.09		-0.46	-0.81	-0.11	
15145	NR	NR	-0.31	-0.63	0.01		-0.10	-0.42	0.23	
19332	NR	NR	0.24	-0.18	0.66		0.44	0.00	0.88	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
16773	3(1)	8	-0.04	0.02		-0.06	0.00	
15133	3(1)	9	0.08	0.02		0.04	0.14	
19299	NR	NR	-0.02	0.16		-0.03	0.16	
16875	NR	NR	-0.06	0.00		-0.07	0.00	
19334	3(2)	10	-0.03	0.64		0.01	0.91	
10987	3(2)	11	0.05	0.36		0.07	0.06	
11980	NR	NR	-0.04	0.07		-0.05	0.10	
19664	NR	NR	0.01	0.93		0.03	0.76	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11301	3(1)	1	-0.35	-0.83	0.14		-0.12	-0.62	0.38	
23480	3(1)	2	-0.17	-0.48	0.15		0.31	-0.01	0.63	
11314	3(1)	3	-0.02	-0.43	0.39		-0.39	-0.81	0.02	
12711	3(1)	4	-0.71	-1.06	-0.35		-0.64	-0.99	-0.29	
23506	3(1)	5	-0.08	-0.50	0.34		0.21	-0.21	0.63	
11320	3(1)	6	-0.21	-0.54	0.11		-0.17	-0.49	0.15	
15001	3(1)	7	0.10	-0.23	0.43		0.18	-0.15	0.51	
17180	3(1)	12	0.07	-0.31	0.44		0.25	-0.13	0.64	
20471	NR	NR	-0.06	-0.42	0.30		-0.07	-0.44	0.31	
22493	NR	NR	-0.01	-0.34	0.32		0.12	-0.21	0.44	
23481	NR	NR	0.50	0.18	0.81		0.20	-0.11	0.52	
17105	NR	NR	-0.58	-0.93	-0.23		-0.26	-0.61	0.10	
15066	NR	NR	-0.39	-0.72	-0.06		0.10	-0.22	0.42	
23503	NR	NR	-0.23	-0.55	0.09		-0.20	-0.51	0.11	
14980	3(2)	13	-0.51	-0.96	-0.05		-0.75	-1.22	-0.29	
15013	3(2)	14	0.27	-0.05	0.58		0.34	0.02	0.65	
20467	3(2)	15	1.03	0.63	1.44	B+	0.40	0.00	0.81	
20521	3(2)	16	0.34	0.03	0.64		0.10	-0.20	0.41	
11295	3(2)	17	0.39	0.05	0.73		0.35	0.01	0.69	
22461	3(2)	18	-0.05	-0.38	0.29		0.07	-0.27	0.40	
11294	NR	NR	0.29	-0.01	0.60		0.11	-0.19	0.42	
20458	NR	NR	-0.14	-0.52	0.23		0.08	-0.30	0.45	
17137	NR	NR	0.23	-0.13	0.60		0.26	-0.10	0.62	
20512	NR	NR	-0.26	-0.70	0.18		-0.06	-0.51	0.40	
17143	NR	NR	0.28	-0.03	0.60		0.51	0.19	0.82	
12678	NR	NR	0.13	-0.21	0.48		0.30	-0.05	0.65	
11342	NR	NR	0.31	-0.01	0.63		-0.01	-0.33	0.30	
15059	NR	NR	0.16	-0.19	0.52		0.02	-0.34	0.38	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
12717	3(1)	8	0.00	0.24		-0.01	0.46	
11307	3(1)	9	0.05	0.37		0.03	0.76	
11358	NR	NR	0.06	0.09		0.02	0.43	
15067	NR	NR	0.02	0.28		0.01	0.48	
15071	3(2)	10	-0.04	0.17		-0.01	0.57	
17150	3(2)	11	0.03	0.32		-0.03	0.59	
22532	NR	NR	-0.04	0.03		-0.03	0.49	
20495	NR	NR	-0.04	0.21		-0.05	0.01	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
19815	3(1)	1	-0.38	-0.92	0.16		-0.30	-0.84	0.23	
23536	3(1)	2	-0.34	-0.70	0.02		-0.35	-0.72	0.01	
14570	3(1)	3	-0.06	-0.48	0.36		0.13	-0.30	0.55	
19776	3(1)	4	-0.47	-0.81	-0.12		-0.38	-0.73	-0.04	
19669	3(1)	5	0.20	-0.15	0.55		0.28	-0.07	0.63	
11237	3(1)	6	0.02	-0.32	0.36		0.15	-0.19	0.49	
16335	3(1)	7	-0.24	-0.58	0.10		-0.01	-0.35	0.33	
14585	NR	NR	0.19	-0.15	0.52		-0.06	-0.40	0.28	
19711	NR	NR	0.47	0.06	0.89		0.04	-0.37	0.45	
11230	NR	NR	0.17	-0.17	0.52		-0.03	-0.38	0.31	
16424	NR	NR	0.16	-0.18	0.51		0.22	-0.13	0.56	
19799	NR	NR	0.28	-0.16	0.72		0.51	0.06	0.97	
20694	NR	NR	-0.12	-0.49	0.24		0.13	-0.24	0.50	
19800	NR	NR	-0.14	-0.50	0.22		-0.16	-0.52	0.19	
19829	3(2)	12	-0.24	-0.58	0.10		0.01	-0.34	0.35	
19808	3(2)	13	-0.06	-0.43	0.30		-0.20	-0.57	0.16	
20740	3(2)	14	0.23	-0.17	0.62		-0.01	-0.39	0.38	
20873	3(2)	15	0.05	-0.27	0.38		0.33	0.01	0.66	
20713	3(2)	16	-0.11	-0.44	0.21		0.29	-0.04	0.62	
19676	3(2)	17	-0.31	-0.63	0.01		-0.09	-0.41	0.22	
16457	3(2)	18	0.09	-0.30	0.49		0.00	-0.39	0.40	
16296	NR	NR	-0.11	-0.47	0.25		-0.25	-0.61	0.11	
20709	NR	NR	0.02	-0.30	0.35		0.03	-0.29	0.36	
20579	NR	NR	0.50	0.18	0.82		0.20	-0.12	0.52	
16373	NR	NR	0.91	0.30	1.53		0.74	0.14	1.34	
16329	NR	NR	-0.13	-0.47	0.21		0.03	-0.31	0.37	
20593	NR	NR	-0.30	-0.64	0.03		-0.27	-0.60	0.07	
14655	NR	NR	0.29	-0.06	0.65		0.37	0.02	0.73	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
16444	3(1)	8	-0.03	0.65		-0.04	0.06	
19826	3(1)	9	0.01	0.71		-0.01	0.92	
16516	NR	NR	-0.04	0.14		-0.04	0.12	
19706	NR	NR	0.06	0.08		0.05	0.20	
20594	3(2)	10	0.06	0.21		0.05	0.24	
19688	3(2)	11	0.03	0.11		0.01	0.04	
23534	NR	NR	-0.02	0.89		0.00	0.97	
14666	NR	NR	0.01	0.98		0.01	0.95	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
11538	3(1)	2	-0.21	-0.65	0.24		-0.17	-0.61	0.28	
16314	3(1)	3	0.13	-0.39	0.64		-0.01	-0.51	0.50	
14725	3(1)	4	-0.29	-0.74	0.15		-0.32	-0.76	0.12	
20106	3(1)	5	-0.37	-0.80	0.06		-0.54	-0.98	-0.11	
14694	3(1)	6	0.07	-0.40	0.54		-0.28	-0.74	0.19	
20862	3(1)	12	0.77	0.27	1.27		0.60	0.09	1.10	
14765	3(1)	13	-0.63	-1.13	-0.13		-0.59	-1.08	-0.10	
14739	3(1)	16	-0.49	-0.99	0.02		-0.51	-1.01	-0.02	
18011	3(1)	17	0.19	-0.33	0.70		0.26	-0.26	0.78	
20115	NR	NR	-0.03	-0.78	0.71		0.14	-0.61	0.89	
20184	NR	NR	0.39	-0.11	0.89		0.44	-0.06	0.94	
14718	NR	NR	0.25	-0.20	0.69		-0.30	-0.75	0.16	
20147	NR	NR	0.24	-0.30	0.78		-0.02	-0.55	0.51	
20864	NR	NR	0.23	-0.38	0.83		0.16	-0.44	0.76	
11643	3(2)	1	-0.06	-0.62	0.51		0.08	-0.48	0.63	
12775	3(2)	7	-0.45	-0.91	0.01		-0.38	-0.85	0.08	
15910	3(2)	14	0.08	-0.38	0.54		-0.17	-0.63	0.30	
20157	3(2)	15	0.06	-0.39	0.51		-0.20	-0.65	0.25	
11539	3(2)	18	0.27	-0.25	0.79		0.23	-0.29	0.74	
11646	NR	NR	-0.63	-1.13	-0.14		-0.20	-0.69	0.30	
20171	NR	NR	-0.68	-1.21	-0.15		-0.62	-1.15	-0.09	
20102	NR	NR	0.17	-0.28	0.62		0.22	-0.23	0.66	
16338	NR	NR	0.16	-0.35	0.68		-0.27	-0.78	0.25	
11481	NR	NR	0.03	-0.52	0.57		-0.02	-0.56	0.52	
20146	NR	NR	-0.28	-0.87	0.31		-0.09	-0.69	0.51	
20205	NR	NR	-0.31	-0.82	0.19		-0.18	-0.68	0.31	
20148	NR	NR	-0.38	-0.83	0.07		-0.73	-1.18	-0.29	
20206	NR	NR	-0.51	-1.00	-0.02		-0.65	-1.14	-0.16	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
20866	3(1)	8	0.01	0.46		0.01	0.14	
12755	3(1)	9	0.01	0.78		0.01	0.39	
20193	3(1)	10	0.06	0.03		0.05	0.27	
20210	3(1)	11	0.09	0.13		0.07	0.43	
20869	NR	NR	-0.05	0.27		-0.03	0.51	
13356	NR	NR	0.00	0.82		0.00	0.38	
16360	NR	NR	0.09	0.02		0.10	0.00	
20868	NR	NR	-0.01	0.64		0.02	0.79	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
15516	1	1.01	KU	N	4	72.08	0.41	-0.66	0.83
10239	3	1.05	TH	N	2	62.17	0.27	-0.16	0.48
15540	4	1.06	TH	N	3	58.08	0.35	0.04	0.72
19446	5	2.02	AP	N	4	50.52	0.34	0.51	0.65
14778	6	2.04	KU	N	2	78.41	0.33	-1.13	0.64
19433	8	1.04	TH	N	4 [†]	72.23 (2.89)	0.53	-1.27	0.40
15544	10	1.01	AP	R	2	71.23	0.24	-0.99	0.41
19450	13	2.03	AP	R	1	45.97	0.22	1.11	0.42
19634	14	3.01	KU	R	3	44.45	0.33	0.79	0.80
15558	15	3.04	TH	R	4	69.47	0.36	-0.52	0.68
15521	16	3.05	AP	R	2	75.69	0.32	-0.95	0.64
10133	18	4.03	AP	R	3	50.64	0.35	0.53	0.79
19454	20	4.05	TH	R	2	79.23	0.32	-1.19	0.65
15610	21	2.01	AP	R	4 [†]	77.31 (3.09)	0.44	-2.35	0.37
23641	24	1.03	KU	M	3	77.19	0.26	-1.32	0.47
15526	25	2.02	TH	M	2	44.56	0.19	1.23	0.43
15563	28	3.01	KU	M	3	77.21	0.39	-0.94	0.80
19456	29	3.02	AP	M	4	53.17	0.36	0.34	0.85
19645	31	3.02	AP	M	4 [†]	40.01 (1.60)	0.48	0.51	0.46
15517	NR	1.03	AP	N	1	48.83	0.30	0.63	0.65
19422	NR	2.07	KU	N	2	59.43	0.31	0.03	0.59
19434	NR	2.08	AP	N	4 [†]	51.52 (2.06)	0.49	-0.17	0.33
14799	NR	2.01	KU	R	3	86.19	0.31	-1.66	0.70
23365	NR	2.02	AP	R	3	56.56	0.33	0.18	0.63
14818	NR	4.01	KU	R	1	71.18	0.41	-0.55	0.88
19619	NR	4.06	TH	R	4	34.21	0.16	1.80	0.62
19625	NR	3.04	AP	R	4 [†]	52.99 (2.12)	0.55	-0.73	0.51
19626	NR	4.01	TH	R	4 [†]	63.12 (2.52)	0.45	-1.16	0.29
17183	NR	2.03	TH	M	1	34.17	0.32	1.22	1.04
15561	NR	2.05	AP	M	3	30.70	0.25	1.71	0.79
19662	NR	2.05	TH	M	4 [†]	57.77 (2.31)	0.51	-0.71	0.39

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

See overall expectations for the associated strand in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005).

[†]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
19433	8	% of Students	3.65	0.86	13.87	16.88	17.66	47.08
		Parameters			-2.73	-0.93	-0.27	-1.17
15610	21	% of Students	0.33	0.15	6.34	19.92	30.01	43.26
		Parameters			-5.42	-2.62	-0.96	-0.40
19645	31	% of Students	8.85	1.20	36.92	39.81	9.38	3.84
		Parameters			-2.30	-0.15	2.34	2.17
19434	NR	% of Students	11.34	2.81	24.18	22.16	20.42	19.08
		Parameters			-1.60	-0.08	0.31	0.69
19625	NR	% of Students	1.31	0.30	36.39	25.27	21.90	14.83
		Parameters			-4.55	0.08	0.38	1.18
19626	NR	% of Students	4.69	0.94	19.35	28.12	10.67	36.22
		Parameters			-3.30	-1.16	1.95	-2.13
19662	NR	% of Students	4.11	0.95	26.65	22.45	23.83	22.01
		Parameters			-3.37	-0.12	-0.01	0.65

Note. The total number of students is 14 617. NR = not released.

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

Item Code	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
15516	1	1.01	KU	N	4	74.17	0.38	-0.66	0.83
15553	2	1.03	AP	N	1	40.15	0.32	1.05	0.81
15540	4	1.06	TH	N	3	60.41	0.35	0.04	0.72
23638	7	2.08	KU	N	2	57.18	0.35	0.17	0.74
19658	9	2.04	AP	N	4 [†]	68.53 (2.74)	0.52	-1.09	0.42
15544	10	1.01	AP	R	2	72.56	0.25	-0.99	0.41
19631	11	2.01	KU	R	4	77.72	0.34	-1.00	0.70
19425	12	2.02	AP	R	3	69.83	0.21	-0.90	0.35
15558	15	3.04	TH	R	4	69.38	0.34	-0.52	0.68
15521	16	3.05	AP	R	2	76.24	0.32	-0.95	0.64
19428	17	4.01	KU	R	3	83.59	0.22	-2.08	0.42
10133	18	4.03	AP	R	3	49.11	0.34	0.53	0.79
19429	19	4.04	TH	R	1	56.54	0.23	0.30	0.39
19643	22	3.01	AP	R	4 [†]	45.25 (1.81)	0.40	-0.35	0.34
19661	23	4.06	TH	R	4 [†]	60.62 (2.42)	0.53	-1.00	0.35
15526	25	2.02	TH	M	2	44.97	0.23	1.23	0.43
10247	26	2.04	AP	M	4	28.63	0.36	1.36	1.33
15588	27	2.05	TH	M	1	36.90	0.27	1.24	0.90
15563	28	3.01	KU	M	3	78.18	0.37	-0.94	0.80
15569	30	2.03	TH	M	4 [†]	72.29 (2.89)	0.56	-1.12	0.53
15545	NR	1.04	TH	N	2	70.17	0.31	-0.62	0.58
18952	NR	2.01	KU	N	2	66.25	0.18	-0.64	0.30
19447	NR	2.02	AP	N	3	42.70	0.18	1.42	0.44
19640	NR	1.05	TH	N	4 [†]	60.03 (2.40)	0.54	-0.79	0.41
19614	NR	2.02	AP	R	1	42.43	0.30	1.03	0.67
19652	NR	3.03	KU	R	2	49.37	0.34	0.56	0.84
14830	NR	4.05	TH	R	1	30.35	0.29	1.54	0.96
19624	NR	2.03	AP	R	4 [†]	63.54 (2.54)	0.44	-2.06	0.37
15560	NR	1.02	KU	M	4	79.75	0.14	-2.59	0.25
19432	NR	3.01	AP	M	4	56.33	0.37	0.22	0.84
19444	NR	3.02	AP	M	4 [†]	51.98 (2.08)	0.54	-0.49	0.40

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

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005).

[†]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
19658	9	% of Students	3.76	0.58	13.84	21.61	23.78	36.44
		Parameters			-2.66	-1.20	-0.24	-0.24
19643	22	% of Students	2.37	0.98	38.15	41.94	7.30	9.27
		Parameters			-4.85	-0.29	3.38	0.35
19661	23	% of Students	4.24	0.68	33.01	14.29	10.23	37.54
		Parameters			-3.93	1.03	0.60	-1.72
15569	30	% of Students	3.26	0.46	10.07	17.33	31.07	37.81
		Parameters			-2.35	-1.35	-0.87	0.11
19640	NR	% of Students	4.15	1.13	25.19	22.74	17.69	29.10
		Parameters			-3.12	-0.24	0.42	-0.22
19624	NR	% of Students	0.18	0.03	13.09	38.94	27.82	19.94
		Parameters			-7.56	-2.23	0.53	1.01
19444	NR	% of Students	7.02	1.00	39.57	17.62	6.06	28.74
		Parameters			-3.02	0.97	1.77	-1.67

Note. The total number of students is 17 505. NR = not released.

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

Item Code	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
15686	1	1.01	KU	N	3	90.30	0.25	-2.30	0.64
19578	2	1.03	AP	N	1	68.88	0.24	-0.74	0.41
15689	3	2.02	AP	N	3	50.17	0.29	0.70	0.60
23363	5	2.07	KU	N	1	73.10	0.30	-0.88	0.57
15635	7	1.01	TH	R	1	77.28	0.35	-1.13	0.68
14890	8	2.01	KU	R	2	80.02	0.37	-1.11	0.80
15619	10	2.03	AP	R	4	82.64	0.29	-1.45	0.64
19559	12	3.03	AP	R	4	57.30	0.34	0.11	0.68
19484	14	3.04	TH	R	4 [†]	63.26 (2.53)	0.58	-1.05	0.46
15357	15	1.03	AP	G	1	58.90	0.48	-0.02	1.05
19560	17	2.02	AP	G	1	59.88	0.38	-0.04	0.77
19581	19	3.02	KU	G	4	77.69	0.36	-1.02	0.73
15658	20	3.03	TH	G	3	73.24	0.37	-0.78	0.71
19486	23	3.05	TH	G	4 [†]	66.64 (2.67)	0.54	-1.41	0.38
15696	24	2.01	KU	M	1	57.24	0.43	0.02	1.01
19583	26	2.03	AP	M	2	63.91	0.26	-0.35	0.42
15697	27	2.06	TH	M	2	77.11	0.37	-1.02	0.74
19565	28	3.01	KU	M	2	89.13	0.30	-1.88	0.75
15703	30	2.02	TH	M	4 [†]	78.68 (3.15)	0.53	-1.80	0.46
19593	NR	2.09	TH	N	3	45.05	0.38	0.65	0.96
23351	NR	2.06	AP	N	4 [†]	45.36 (1.81)	0.51	-0.29	0.44
10263	NR	2.05	TH	R	1	77.96	0.32	-1.15	0.63
19459	NR	3.01	KU	R	3	75.13	0.41	-0.78	0.89
19587	NR	2.04	AP	R	4 [†]	72.86 (2.91)	0.56	-1.58	0.43
23637	NR	1.01	KU	G	3	78.37	0.32	-1.14	0.65
19598	NR	2.04	TH	G	4	45.61	0.42	0.56	1.14
19563	NR	3.04	AP	G	3	76.51	0.46	-0.78	1.10
19504	NR	2.01	AP	G	4 [†]	70.33 (2.81)	0.53	-1.18	0.42
15623	NR	1.02	TH	M	2	80.44	0.28	-1.49	0.53
19500	NR	3.01	AP	M	1	78.19	0.40	-0.93	0.90
19572	NR	3.01	AP	M	4 [†]	62.46 (2.50)	0.61	-0.94	0.53

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

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005).

[†]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
19484	14	% of Students	2.03	1.17	21.90	28.10	12.27	34.53
		Parameters			-3.44	-0.85	0.98	-0.88
19486	23	% of Students	2.44	0.18	25.01	18.11	11.71	42.55
		Parameters			-4.46	-0.04	0.55	-1.70
15703	30	% of Students	0.89	0.40	8.83	19.41	14.84	55.64
		Parameters			-3.74	-1.86	-0.06	-1.55
23351	NR	% of Students	4.36	1.11	46.37	24.96	7.67	15.53
		Parameters			-3.56	0.64	1.89	-0.14
19587	NR	% of Students	1.37	0.58	16.56	20.83	9.40	51.25
		Parameters			-4.05	-1.01	0.83	-2.11
19504	NR	% of Students	2.97	1.47	10.13	24.42	21.64	39.36
		Parameters			-2.23	-1.89	-0.04	-0.55
19572	NR	% of Students	2.60	0.43	24.24	23.52	18.25	30.95
		Parameters			-3.35	-0.52	0.22	-0.10

Note. The total number of students is 41 649. NR = not released.

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

Item Code	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
15686	1	1.01	KU	N	3	91.94	0.25	-2.30	0.64
15689	3	2.02	AP	N	3	47.33	0.28	0.70	0.60
19556	4	2.03	TH	N	2	61.49	0.43	-0.04	1.03
23754	6	2.02	AP	N	4 [†]	66.37 (2.65)	0.51	-1.16	0.31
15635	7	1.01	TH	R	1	80.40	0.33	-1.13	0.68
20978	9	2.01	TH	R	2	51.47	0.44	0.37	1.04
15619	10	2.03	AP	R	4	83.06	0.32	-1.45	0.64
19492	11	3.01	KU	R	4	72.86	0.25	-0.98	0.44
15700	13	2.05	AP	R	4 [†]	66.41 (2.66)	0.56	-1.35	0.49
15357	15	1.03	AP	G	1	60.85	0.45	-0.02	1.05
15638	16	2.01	KU	G	2	84.70	0.28	-1.66	0.60
20892	18	2.02	TH	G	4	68.32	0.31	-0.51	0.56
15658	20	3.03	TH	G	3	74.64	0.35	-0.78	0.71
19001	21	3.04	AP	G	3	85.96	0.32	-1.55	0.74
19569	22	1.01	AP	G	4 [†]	69.14 (2.77)	0.53	-1.26	0.45
15696	24	2.01	KU	M	1	61.38	0.43	0.02	1.01
15624	25	2.02	TH	M	4	52.90	0.41	0.33	0.91
15697	27	2.06	TH	M	2	79.01	0.36	-1.02	0.74
14960	29	3.01	AP	M	2	56.53	0.43	0.16	0.98
19488	31	3.01	AP	M	4 [†]	79.53 (3.18)	0.53	-1.46	0.47
19592	NR	1.04	AP	N	4	65.16	0.31	-0.30	0.57
15688	NR	2.07	KU	N	4	71.67	0.34	-0.67	0.63
19576	NR	2.04	KU	R	1	79.52	0.26	-1.44	0.49
23636	NR	3.04	AP	R	4	71.24	0.37	-0.59	0.72
19588	NR	3.02	TH	R	4 [†]	70.60 (2.82)	0.57	-1.66	0.49
19461	NR	2.03	AP	G	1	56.31	0.33	0.23	0.65
19463	NR	3.01	KU	G	4	77.00	0.37	-0.89	0.77
19505	NR	3.05	TH	G	4 [†]	58.04 (2.32)	0.53	-0.72	0.42
19498	NR	2.06	AP	M	3	72.77	0.22	-1.08	0.39
14954	NR	3.01	KU	M	4	66.43	0.35	-0.34	0.67
19608	NR	2.03	TH	M	4 [†]	72.74 (2.91)	0.51	-1.45	0.51

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

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005).

[†]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
23754	6	% of Students	5.87	1.30	28.72	7.65	4.39	52.08
		Parameters			-3.38	2.13	1.01	-4.41
15700	13	% of Students	0.81	0.14	21.21	19.05	28.87	29.93
		Parameters			-4.81	-0.44	-0.58	0.41
19569	22	% of Students	1.80	0.19	17.68	14.84	32.80	32.70
		Parameters			-3.89	-0.38	-1.16	0.39
19488	31	% of Students	2.88	0.29	8.34	12.36	19.46	56.67
		Parameters			-2.43	-1.30	-0.93	-1.19
19588	NR	% of Students	0.52	0.08	17.21	21.59	20.38	40.22
		Parameters			-5.15	-0.92	-0.10	-0.45
19505	NR	% of Students	3.57	0.43	26.05	25.49	22.68	21.77
		Parameters			-3.45	-0.34	0.26	0.64
19608	NR	% of Students	0.90	0.11	6.14	22.34	41.91	28.61
		Parameters			-3.39	-2.27	-0.96	0.81

Note. The total number of students is 51 782. NR = not released.

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

Item Code	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
15267	1	01	CC	N	4	54.05	0.29	0.31	0.51
14367	2	02	MA	N	2	52.36	0.22	0.39	0.51
15269	3	06	CC	N	3	90.20	0.20	-2.28	0.47
20359	6	10	MA	N	3	56.76	0.26	0.17	0.50
14384	7	12	HP	N	2	57.43	0.29	0.05	0.65
20411	8	02	MA	N	4 [†]	56.76 (2.27)	0.44	-1.18	0.25
9707	10	03	CC	R	1	57.43	0.30	0.11	0.65
20981	11	01	MA	R	2	73.65	0.33	-1.01	0.64
14399	12	07	CC	R	4	39.53	0.22	1.44	0.49
20891	13	09	MA	R	4	51.35	0.43	0.16	0.86
10032	14	13	CC	R	3	50.34	0.25	0.17	0.55
20986	17	13	HP	R	4 [†]	60.73 (2.43)	0.51	-1.11	0.41
15299	18	05	CC	M	3	72.97	0.44	-0.61	0.97
14428	19	05	MA	M	3	50.34	0.22	0.78	0.36
15304	21	14	MA	M	2	66.89	0.25	-0.83	0.45
20416	22	15	HP	M	4	36.49	0.23	1.64	0.57
15348	23	17b	CC	M	1	72.30	0.23	-1.12	0.39
14447	25	17a	HP	M	3	50.68	0.27	0.75	0.52
20447	NR	02	MA	M	4 [†]	77.03 (3.08)	0.47	-1.88	0.38
20373	NR	05	CC	N	1	72.97	0.19	-1.35	0.33
14374	NR	08	MA	N	1	57.43	0.27	0.11	0.50
20391	NR	10	HP	N	4 [†]	53.63 (2.15)	0.53	-0.63	0.50
14394	NR	04	HP	R	4	60.14	0.26	-0.08	0.50
20442	NR	14	MA	R	2	74.66	0.53	-0.67	1.37
15339	NR	11	HP	R	3	75.00	0.07	-4.30	0.11
20369	NR	01	MA	R	4 [†]	73.14 (2.93)	0.48	-1.65	0.37
20449	NR	09	MA	R	4 [†]	46.20 (1.85)	0.50	-0.69	0.46
20395	NR	03	HP	M	1	68.92	0.33	-0.62	0.59
15320	NR	05	HP	M	2	41.55	0.16	1.64	0.38
14443	NR	17b	MA	M	2	33.45	0.23	1.48	0.78
18496	NR	15	HP	M	4 [†]	63.94 (2.56)	0.60	-1.00	0.56

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.

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005).

[†]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
20411	8	% of Students	3.04	3.72	37.50	16.22	1.01	38.51
		Parameters			-4.70	1.68	6.57	-8.27
20986	17	% of Students	1.69	0.34	23.99	23.99	29.05	20.95
		Parameters			-4.56	-0.50	-0.30	0.91
20447	26	% of Students	1.01	0.68	13.51	12.16	20.27	52.36
		Parameters			-4.45	-0.62	-1.12	-1.34
20391	NR	% of Students	1.69	1.01	26.01	38.85	18.92	13.51
		Parameters			-3.68	-0.88	0.98	1.06
20369	NR	% of Students	1.01	1.35	12.16	20.61	20.27	44.59
		Parameters			-3.77	-1.54	-0.23	-1.06
20449	NR	% of Students	2.03		49.66	27.70	2.70	17.91
		Parameters			-4.86	0.53	3.24	-1.67
18496	NR	% of Students	2.36	0.68	19.93	29.39	13.51	34.12
		Parameters			-3.16	-1.01	0.72	-0.55

Note. The total number of students is 296. NR = not released.

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

Item Code	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
14367	2	02	MA	N	2	54.81	0.29	0.39	0.51
15269	3	06	CC	N	3	85.77	0.23	-2.28	0.47
9869	4	06	CC	N	4	41.08	0.35	0.98	0.80
21714	5	07	MA	N	2	46.39	0.13	1.95	0.22
14384	7	12	HP	N	2	59.52	0.35	0.05	0.65
15375	9	12	HP	N	4 [†]	60.70 (2.43)	0.40	-0.91	0.31
20981	11	01	MA	R	2	77.76	0.32	-1.01	0.64
20891	13	09	MA	R	4	57.62	0.41	0.16	0.86
10032	14	13	CC	R	3	59.62	0.30	0.17	0.55
15363	15	14	MA	R	4	54.21	0.39	0.28	0.83
20243	16	03	MA	R	4 [†]	71.07 (2.84)	0.59	-1.30	0.54
15299	18	05	CC	M	3	73.25	0.42	-0.61	0.97
15344	20	09	HP	M	2	53.01	0.26	0.53	0.45
15304	21	14	MA	M	2	71.84	0.27	-0.83	0.45
20416	22	15	HP	M	4	37.07	0.18	1.64	0.57
23625	24	17a	MA	M	2	64.93	0.27	-0.33	0.47
14447	25	17a	HP	M	3	48.10	0.28	0.75	0.52
14458	27	15	HP	M	4 [†]	72.49 (2.90)	0.54	-1.30	0.45
20372	NR	01	CC	N	3	89.28	0.25	-2.18	0.59
14381	NR	12	MA	N	1	43.49	0.36	0.87	0.73
15329	NR	02	MA	N	4 [†]	48.05 (1.92)	0.45	-0.11	0.32
20360	NR	03	CC	R	4	50.20	0.45	0.40	1.23
21748	NR	06	CC	R	1	68.24	0.42	-0.38	0.90
23620	NR	12	HP	R	1	61.02	0.12	-0.14	0.19
20429	NR	07	MA	R	4 [†]	64.95 (2.60)	0.42	-1.04	0.40
20448	NR	13	HP	R	4 [†]	50.48 (2.02)	0.50	-0.30	0.45
20435	NR	04	HP	M	3	69.44	0.48	-0.40	1.11
14425	NR	05	MA	M	3	47.60	0.20	1.10	0.38
15303	NR	17b	CC	M	1	80.66	0.33	-1.18	0.71
21787	NR	01	MA	M	4 [†]	73.27 (2.93)	0.53	-1.24	0.42

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.

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005).

[†]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
15375	9	% of Students	2.00	1.00	19.14	24.75	38.28	14.83
		Parameters			-4.28	-0.86	-0.78	2.27
20243	16	% of Students	1.70	0.20	20.54	16.63	13.23	47.70
		Parameters			-3.74	-0.43	0.06	-1.08
14458	27	% of Students	2.61	0.50	12.83	20.54	18.04	45.49
		Parameters			-2.96	-1.28	-0.03	-0.91
15329	NR	% of Students	6.61	3.91	29.86	31.46	13.23	14.93
		Parameters			-2.50	-0.25	1.87	0.44
20429	NR	% of Students	1.20	0.70	7.21	34.97	40.98	14.93
		Parameters			-3.01	-2.88	-0.27	1.99
20448	NR	% of Students	2.71	0.30	33.17	30.56	25.45	7.82
		Parameters			-3.94	-0.15	0.52	2.37
21787	NR	% of Students	3.61	1.10	15.13	14.73	13.23	52.20
		Parameters			-2.66	-0.60	-0.06	-1.65

Note. The total number of students is 998. NR = not released.

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

Item Code	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
12823	1	01	MA	N	3	59.10	0.32	0.48	0.63
21690	2	07	CC	N	4	58.40	0.27	0.29	0.51
14468	3	08	CC	N	1	73.70	0.29	-0.77	0.52
14477	6	15	HP	N	1	60.51	0.33	0.35	0.74
15439	8	15	MA	N	4 [†]	78.72 (3.15)	0.48	-1.39	0.39
15383	10	02	HP	R	4	65.44	0.40	-0.18	0.79
15428	11	06	CC	R	3	55.58	0.32	0.32	0.67
15234	13	12	MA	R	1	71.77	0.32	-0.55	0.59
9683	14	14	HP	R	4	42.22	0.39	0.94	1.08
15245	16	13	HP	R	4 [†]	68.14 (2.73)	0.48	-1.67	0.38
9950	18	08	MA	G	2	62.01	0.38	0.18	0.73
20339	19	01	HP	G	1	69.92	0.44	-0.27	1.01
18466	20	11	CC	G	2	58.66	0.34	0.22	0.66
20347	21	01	MA	G	4 [†]	67.28 (2.69)	0.57	-0.94	0.58
14526	23	04	HP	M	3	76.78	0.47	-0.36	1.00
20342	25	06	MA	M	2	70.80	0.28	-0.62	0.47
23758	27	18	HP	M	2	46.17	0.20	1.22	0.47
15437	28	20b	CC	M	3	74.23	0.41	-0.41	0.87
14556	29	20a	MA	M	3	79.68	0.37	-0.78	0.81
18489	30	01	HP	M	4 [†]	74.27 (2.97)	0.57	-1.31	0.51
15424	NR	10	MA	N	1	75.90	0.38	-0.68	0.77
14486	NR	18	MA	N	4	69.83	0.27	-0.57	0.45
15395	NR	03	MA	N	4 [†]	88.06 (3.52)	0.46	-1.89	0.47
21728	NR	02	MA	R	1	73.44	0.37	-0.55	0.74
15429	NR	08	MA	R	2	50.66	0.40	0.59	0.95
20307	NR	04	MA	R	4 [†]	76.34 (3.05)	0.52	-1.41	0.43
23611	NR	03	CC	G	4	76.08	0.36	-0.70	0.75
20282	NR	12	HP	G	2	51.72	0.08	2.01	0.14
14527	NR	05	MA	M	4	76.34	0.37	-0.74	0.73
14544	NR	17	CC	M	3	73.88	0.37	-0.55	0.77
15441	NR	20a	HP	M	4 [†]	73.02 (2.92)	0.54	-1.27	0.48

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

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005).

[†]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
15439	8	% of Students	2.90	0.53	4.66	17.77	21.90	52.24
		Parameters			-1.47	-2.62	-0.47	-0.99
15245	16	% of Students	0.44	0.26	14.95	29.46	20.84	34.04
		Parameters			-5.59	-1.45	0.60	-0.23
20347	21	% of Students	1.32	0.35	7.56	37.03	27.44	26.30
		Parameters			-2.66	-2.15	0.35	0.71
18489	30	% of Students	0.97	0.79	8.80	31.57	6.33	51.54
		Parameters			-2.95	-2.07	1.78	-2.01
15395	NR	% of Students	1.14	0.18	2.81	9.41	15.22	71.24
		Parameters			-2.32	-2.42	-1.03	-1.79
20307	NR	% of Students	1.93	0.26	6.68	25.95	13.90	51.28
		Parameters			-2.55	-2.43	0.73	-1.40
15441	NR	% of Students	1.23	0.53	6.07	31.57	19.53	41.07
		Parameters			-2.59	-2.58	0.52	-0.43

Note. The total number of students is 1137. NR = not released.

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

Item Code	Sequence	Overall Curriculum Expectation*	Cognitive Skill	Strand	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
12823	1	01	MA	N	3	48.93	0.28	0.48	0.63
21718	4	11	MA	N	1	75.59	0.36	-0.93	0.73
14475	5	15	MA	N	1	83.24	0.25	-1.81	0.51
14477	6	15	HP	N	1	50.43	0.33	0.35	0.74
20291	7	03	MA	N	4 [†]	85.89 (3.44)	0.47	-2.02	0.40
23613	9	05	MA	R	3	41.14	0.25	1.08	0.65
15383	10	02	HP	R	4	62.88	0.39	-0.18	0.79
15428	11	06	CC	R	3	53.91	0.31	0.32	0.67
21761	12	08	MA	R	2	45.27	0.36	0.66	0.80
10064	15	04	MA	R	4 [†]	72.01 (2.88)	0.57	-1.51	0.45
15388	17	04	CC	G	4	66.51	0.41	-0.39	0.89
9950	18	08	MA	G	2	54.31	0.33	0.18	0.73
18466	20	11	CC	G	2	54.80	0.32	0.22	0.66
20306	22	04	MA	G	4 [†]	73.47 (2.94)	0.51	-1.53	0.47
14526	23	04	HP	M	3	64.23	0.44	-0.36	1.00
15416	24	09	CC	M	2	73.99	0.28	-1.07	0.50
20303	26	15	CC	M	4	72.74	0.30	-0.88	0.57
15437	28	20b	CC	M	3	65.98	0.41	-0.41	0.87
14556	29	20a	MA	M	3	72.85	0.40	-0.78	0.81
15459	31	20b	HP	M	4 [†]	65.37 (2.61)	0.58	-1.11	0.47
20253	NR	07	CC	N	1	44.09	0.32	0.78	0.75
14472	NR	09	CC	N	4	52.28	0.43	0.24	1.00
20289	NR	13	MA	N	4 [†]	70.47 (2.82)	0.52	-1.47	0.48
15449	NR	11	MA	R	1	76.16	0.33	-1.04	0.64
20362	NR	11	HP	R	3	59.43	0.30	-0.03	0.59
18490	NR	15	HP	R	4 [†]	59.15 (2.37)	0.57	-0.88	0.46
20260	NR	05	HP	G	2	65.09	0.45	-0.31	1.02
20301	NR	09	HP	G	4	61.81	0.42	-0.15	0.97
22603	NR	05	MA	M	4	41.57	0.28	1.00	0.68
15423	NR	17	HP	M	2	56.41	0.30	0.14	0.59
18498	NR	18	HP	M	4 [†]	70.77 (2.83)	0.42	-2.03	0.28

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

*See overall expectations for the associated strand in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005).

[†]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
20291	7	% of Students	1.42	0.82	10.85	5.41	4.09	77.40
		Parameters			-3.63	0.09	-0.13	-4.43
10064	15	% of Students	1.25	0.82	17.54	18.75	13.56	48.08
		Parameters			-3.93	-0.81	0.14	-1.44
20306	22	% of Students	1.35	0.14	9.04	17.94	37.15	34.38
		Parameters			-3.52	-1.69	-1.24	0.32
15459	31	% of Students	2.21	1.03	22.60	18.68	20.39	35.09
		Parameters			-3.48	-0.38	-0.26	-0.33
20289	NR	% of Students	0.78	0.43	10.07	22.92	37.22	28.58
		Parameters			-3.84	-1.79	-0.86	0.62
18490	NR	% of Students	2.88	0.93	26.16	27.47	14.77	27.79
		Parameters			-3.40	-0.56	0.77	-0.33
18498	NR	% of Students	1.71	0.28	18.29	18.40	17.30	44.02
		Parameters			-5.63	-0.57	-0.07	-1.83

Note. The total number of students is 2810. 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 7.1.65a–7.1.76b. The DIF results for the applied and academic versions of the English-language assessment are based on two random samples of 2000 examinees. For the French-language assessment, gender-based DIF analysis was conducted for one sample of students. SLL-based DIF analysis was not conducted because no decent sample could be collected. 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 is 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
15516	11101	1	0.60	0.08	1.12		0.79	0.26	1.32	
10239	11101	3	1.07	0.62	1.52	B+	0.92	0.47	1.37	
15540	11101	4	0.95	0.49	1.40		1.08	0.62	1.54	B+
19446	11101	5	-0.32	-0.78	0.14		-0.45	-0.90	0.00	
14778	11101	6	-0.64	-1.18	-0.10		-0.96	-1.51	-0.41	
15544	11101	10	-1.43	-1.91	-0.95	B-	-0.72	-1.20	-0.24	
19450	11101	13	-0.45	-0.88	-0.01		-0.81	-1.25	-0.38	
19634	11101	14	0.34	-0.12	0.79		0.09	-0.36	0.55	
15558	11101	15	0.57	0.07	1.07		0.25	-0.25	0.75	
15521	11101	16	-0.20	-0.73	0.32		-0.04	-0.56	0.48	
10133	11101	18	0.12	-0.34	0.57		-0.18	-0.63	0.27	
19454	11101	20	0.52	-0.02	1.06		-0.29	-0.83	0.25	
23641	11101	24	-0.58	-1.09	-0.06		-0.44	-0.94	0.07	
15526	11101	25	0.76	0.33	1.19		0.98	0.55	1.41	
15563	11101	28	0.54	0.01	1.08		0.34	-0.20	0.88	
19456	11101	29	0.22	-0.25	0.68		-0.07	-0.53	0.39	
15517	11101	NR	0.40	-0.04	0.85		0.33	-0.11	0.77	
19422	11101	NR	-0.68	-1.14	-0.23		-0.87	-1.33	-0.42	
14799	11101	NR	0.75	0.12	1.38		0.80	0.14	1.46	
23365	11101	NR	0.53	0.08	0.98		0.61	0.16	1.05	
14818	11101	NR	0.50	-0.01	1.01		0.22	-0.30	0.73	
19619	11101	NR	0.22	-0.24	0.67		0.57	0.12	1.02	
17183	11101	NR	-0.61	-1.09	-0.14		-0.83	-1.29	-0.36	
15561	11101	NR	-0.36	-0.83	0.10		-0.61	-1.08	-0.14	

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
19433	11101	8	0.03	0.30		0.12	0.00	
15610	11101	21	-0.07	0.07		-0.06	0.15	
19645	11101	31	-0.02	0.70		-0.04	0.74	
19434	11101	NR	0.11	0.01		0.14	0.00	
19625	11101	NR	-0.10	0.00		-0.16	0.00	
19626	11101	NR	-0.14	0.00		-0.08	0.00	
19662	11101	NR	0.09	0.01		0.06	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
15516	11201	1	0.95	0.43	1.47		0.69	0.17	1.21	
15553	11201	2	-0.20	-0.65	0.25		-0.22	-0.68	0.25	
15540	11201	4	1.34	0.88	1.80	B+	1.22	0.76	1.68	B+
23638	11201	7	-0.11	-0.57	0.34		-0.47	-0.92	-0.02	
15544	11201	10	-1.25	-1.74	-0.77	B-	-1.34	-1.83	-0.85	B-
19631	11201	11	0.64	0.11	1.17		0.86	0.31	1.41	
19425	11201	12	0.65	0.19	1.11		-0.14	-0.61	0.33	
15558	11201	15	0.43	-0.06	0.92		-0.17	-0.65	0.32	
15521	11201	16	0.20	-0.31	0.72		0.02	-0.49	0.53	
19428	11201	17	0.51	-0.06	1.09		0.07	-0.48	0.63	
10133	11201	18	-0.01	-0.46	0.44		-0.24	-0.69	0.21	
19429	11201	19	0.32	-0.11	0.76		0.03	-0.40	0.46	
15526	11201	25	0.88	0.44	1.31		1.04	0.60	1.48	B+
10247	11201	26	0.91	0.39	1.43		0.48	-0.03	0.99	
15588	11201	27	-0.26	-0.71	0.20		0.44	-0.02	0.91	
15563	11201	28	0.56	0.00	1.12		0.53	-0.01	1.07	
15545	11201	NR	0.27	-0.22	0.75		0.31	-0.17	0.80	
18952	11201	NR	-0.85	-1.29	-0.40		-0.41	-0.86	0.04	
19447	11201	NR	0.10	-0.33	0.54		-0.09	-0.53	0.34	
19614	11201	NR	-0.12	-0.57	0.33		-0.41	-0.86	0.05	
19652	11201	NR	-0.24	-0.69	0.22		-0.09	-0.54	0.36	
14830	11201	NR	-0.22	-0.71	0.27		-0.06	-0.54	0.42	
15560	11201	NR	-0.13	-0.64	0.38		-0.53	-1.05	0.00	
19432	11201	NR	0.35	-0.10	0.81		0.48	0.02	0.94	

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
19658	11201	9	0.16	0.00		0.16	0.00	
19643	11201	22	0.04	0.72		0.07	0.01	
19661	11201	23	-0.05	0.33		-0.01	0.77	
15569	11201	30	0.12	0.00		0.14	0.00	
19640	11201	NR	-0.03	0.33		-0.10	0.07	
19624	11201	NR	0.08	0.00		0.09	0.02	
19444	11201	NR	-0.11	0.02		-0.03	0.05	

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
15686	12101	1	-0.44	-1.16	0.28		0.00	-0.72	0.72	
19578	12101	2	0.08	-0.38	0.54		-0.73	-1.19	-0.26	
15689	12101	3	0.08	-0.36	0.52		0.40	-0.04	0.85	
23363	12101	5	-0.04	-0.54	0.46		-0.33	-0.81	0.15	
15635	12101	7	1.08	0.54	1.62	B+	1.24	0.71	1.78	B+
14890	12101	8	0.68	0.11	1.25		0.08	-0.50	0.67	
15619	12101	10	0.46	-0.11	1.03		0.43	-0.14	1.00	
19559	12101	12	0.08	-0.38	0.53		0.48	0.04	0.93	
15357	12101	15	-0.48	-0.98	0.03		-0.89	-1.38	-0.41	
19560	12101	17	0.22	-0.25	0.69		0.09	-0.37	0.56	
19581	12101	19	-0.48	-1.02	0.07		-0.80	-1.35	-0.26	
15658	12101	20	1.26	0.75	1.77	B+	1.29	0.77	1.81	B+
15696	12101	24	0.84	0.37	1.32		0.49	0.03	0.96	
19583	12101	26	-0.03	-0.47	0.42		-0.14	-0.59	0.31	
15697	12101	27	-1.04	-1.59	-0.49	B-	0.10	-0.44	0.63	
19565	12101	28	0.15	-0.53	0.84		0.14	-0.57	0.85	
19593	12101	NR	0.87	0.41	1.34		0.54	0.08	0.99	
10263	12101	NR	0.36	-0.17	0.90		0.33	-0.19	0.85	
19459	12101	NR	-0.51	-1.05	0.02		-0.76	-1.30	-0.22	
23637	12101	NR	0.26	-0.26	0.78		-0.48	-1.00	0.04	
19598	12101	NR	0.18	-0.30	0.65		-0.63	-1.10	-0.16	
19563	12101	NR	0.08	-0.48	0.65		0.06	-0.50	0.62	
15623	12101	NR	0.08	-0.47	0.62		0.78	0.24	1.33	
19500	12101	NR	-0.05	-0.60	0.50		-0.52	-1.08	0.04	

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
19484	12101	14	0.03	0.35		-0.02	0.41	
19486	12101	23	-0.04	0.05		-0.07	0.21	
15703	12101	30	0.07	0.00		0.12	0.00	
23351	12101	NR	0.14	0.00		0.13	0.00	
19587	12101	NR	-0.07	0.08		-0.02	0.75	
19504	12101	NR	-0.08	0.00		-0.11	0.00	
19572	12101	NR	0.00	0.92		-0.03	0.46	

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
15686	12201	1	-0.91	-1.69	-0.13		-0.73	-1.52	0.06	
15689	12201	3	-0.18	-0.62	0.26		0.27	-0.17	0.71	
19556	12201	4	0.34	-0.14	0.83		0.90	0.42	1.39	
15635	12201	7	1.44	0.87	2.01	B+	0.91	0.33	1.48	
20978	12201	9	-0.08	-0.56	0.40		0.16	-0.31	0.62	
15619	12201	10	0.46	-0.12	1.04		0.48	-0.12	1.08	
19492	12201	11	-0.32	-0.80	0.16		-0.54	-1.03	-0.06	
15357	12201	15	-0.67	-1.15	-0.18		-1.28	-1.78	-0.77	B-
15638	12201	16	-1.09	-1.69	-0.49	B-	-1.30	-1.92	-0.68	B-
20892	12201	18	0.18	-0.29	0.65		0.14	-0.34	0.61	
15658	12201	20	0.67	0.16	1.17		0.96	0.44	1.48	
19001	12201	21	-0.43	-1.07	0.21		0.43	-0.20	1.06	
15696	12201	24	0.17	-0.31	0.64		0.44	-0.05	0.92	
15624	12201	25	0.41	-0.06	0.88		0.59	0.13	1.06	
15697	12201	27	0.05	-0.49	0.59		-0.21	-0.75	0.33	
14960	12201	29	0.06	-0.42	0.54		-0.32	-0.79	0.15	
19592	12201	NR	-0.37	-0.82	0.09		-0.45	-0.92	0.01	
15688	12201	NR	-0.02	-0.51	0.47		-0.33	-0.82	0.16	
19576	12201	NR	-0.19	-0.71	0.34		-0.40	-0.94	0.15	
23636	12201	NR	0.79	0.30	1.29		1.09	0.58	1.60	B+
19461	12201	NR	0.31	-0.14	0.76		0.73	0.27	1.19	
19463	12201	NR	0.20	-0.35	0.74		-0.09	-0.65	0.46	
19498	12201	NR	0.16	-0.31	0.63		0.89	0.41	1.38	
14954	12201	NR	0.53	0.05	1.01		0.32	-0.15	0.80	

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
23754	12201	6	-0.03	0.41		-0.05	0.52	
15700	12201	13	0.19	0.00	B-	0.14	0.00	
19569	12201	22	0.21	0.00	B-	0.27	0.00	C-
19488	12201	31	0.08	0.04		0.06	0.00	
19588	12201	NR	-0.11	0.01		-0.16	0.00	
19505	12201	NR	-0.12	0.00		-0.12	0.00	
19608	12201	NR	0.11	0.00		0.09	0.01	

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
15267	21101	1	0.24	-0.91	1.39	
14367	21101	2	-0.05	-1.19	1.09	
15269	21101	3	1.53	-0.37	3.44	B+
20359	21101	6	-0.03	-1.19	1.14	
14384	21101	7	-0.02	-1.19	1.15	
9707	21101	10	-0.16	-1.36	1.03	
20981	21101	11	-0.99	-2.28	0.30	
14399	21101	12	-1.23	-2.40	-0.06	B-
20891	21101	13	0.72	-0.52	1.95	
10032	21101	14	0.85	-0.30	2.00	
15299	21101	18	0.92	-0.49	2.32	
14428	21101	19	-0.85	-1.99	0.29	
15304	21101	21	-0.29	-1.49	0.92	
20416	21101	22	0.11	-1.07	1.29	
15348	21101	23	-0.83	-2.10	0.43	
14447	21101	25	1.23	0.07	2.38	B+
20373	21101	29	-0.35	-1.61	0.90	
14374	21101	NR	1.31	0.14	2.48	B+
14394	21101	NR	0.98	-0.21	2.16	
20442	21101	NR	-1.24	-2.84	0.36	B-
15339	21101	NR	0.31	-0.94	1.56	
20395	21101	NR	-0.97	-2.27	0.32	
15320	21101	NR	1.01	-0.14	2.15	B+
14443	21101	NR	0.29	-0.89	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
20411	21101	8	0.13	0.17	
20986	21101	17	-0.22	0.02	B+
20447	21101	26	0.08	0.35	
20391	21101	NR	0.17	0.20	
20369	21101	NR	-0.09	0.25	
20449	21101	NR	-0.01	0.92	
18496	21101	NR	0.03	0.35	

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
14367	21201	2	0.31	-0.32	0.93	
15269	21201	3	-0.36	-1.23	0.52	
9869	21201	4	0.17	-0.49	0.82	
21714	21201	5	-0.20	-0.80	0.39	
14384	21201	7	-0.33	-0.98	0.32	
20981	21201	11	-0.42	-1.18	0.33	
20891	21201	13	-0.64	-1.31	0.04	
10032	21201	14	1.09	0.45	1.74	B+
15363	21201	15	0.55	-0.11	1.21	
15299	21201	18	0.29	-0.46	1.04	
15344	21201	20	-0.12	-0.74	0.50	
15304	21201	21	0.02	-0.68	0.72	
20416	21201	22	0.26	-0.37	0.88	
23625	21201	24	0.46	-0.19	1.10	
14447	21201	25	1.01	0.38	1.63	B+
20372	21201	NR	-1.08	-2.08	-0.08	B-
14381	21201	NR	0.20	-0.45	0.85	
20360	21201	NR	-0.39	-1.07	0.30	
21748	21201	NR	0.26	-0.45	0.97	
23620	21201	NR	-0.07	-0.69	0.54	
20435	21201	NR	0.71	-0.03	1.45	
14425	21201	NR	-0.53	-1.14	0.08	
15303	21201	NR	0.89	0.08	1.70	

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
15375	21201	9	-0.10	0.23	
20243	21201	16	0.06	0.58	
14458	21201	27	0.10	0.14	
15329	21201	NR	0.03	0.18	
20429	21201	NR	0.12	0.14	
20448	21201	NR	-0.32	0.00	C+
21787	21201	NR	0.13	0.06	

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
12823	22101	1	0.40	-0.21	1.00	
21690	22101	2	-0.73	-1.33	-0.13	
14468	22101	3	-1.34	-2.02	-0.66	B-
14477	22101	6	-0.18	-0.80	0.43	
15383	22101	10	1.75	1.08	2.41	C+
15428	22101	11	-0.05	-0.65	0.56	
15234	22101	13	-0.10	-0.76	0.56	
9683	22101	14	-0.03	-0.66	0.60	
9950	22101	18	-0.76	-1.39	-0.13	
20339	22101	19	-0.31	-0.99	0.38	
18466	22101	20	-0.59	-1.20	0.02	
14526	22101	23	0.76	-0.01	1.53	
20342	22101	25	0.69	0.04	1.35	
23758	22101	27	-0.45	-1.03	0.13	
15437	22101	28	-0.12	-0.83	0.60	
14556	22101	29	-0.06	-0.82	0.69	
15424	22101	NR	-0.53	-1.25	0.19	
14486	22101	NR	0.27	-0.37	0.91	
21728	22101	NR	0.03	-0.66	0.73	
15429	22101	NR	1.00	0.38	1.62	B+
23611	22101	NR	-0.84	-1.55	-0.12	
20282	22101	NR	-0.10	-0.66	0.46	
14527	22101	NR	-0.06	-0.78	0.66	
14544	22101	NR	1.10	0.39	1.81	B+

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
15439	22101	8	0.13	0.02	
15245	22101	16	-0.19	0.00	B+
20347	22101	21	0.07	0.29	
18489	22101	30	0.13	0.00	
15395	22101	NR	-0.06	0.06	
20307	22101	NR	-0.04	0.00	
15441	22101	NR	0.00	0.40	

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
12823	22201	1	-0.01	-0.38	0.36	
21718	22201	4	-0.62	-1.06	-0.17	
14475	22201	5	0.03	-0.46	0.51	
14477	22201	6	-0.26	-0.64	0.12	
23613	22201	9	-0.01	-0.39	0.36	
15383	22201	10	1.72	1.31	2.13	C+
15428	22201	11	0.54	0.17	0.92	
21761	22201	12	0.00	-0.38	0.38	
15388	22201	17	-0.62	-1.04	-0.21	
9950	22201	18	-0.35	-0.73	0.03	
18466	22201	20	-0.55	-0.93	-0.17	
14526	22201	23	0.07	-0.35	0.48	
15416	22201	24	0.47	0.05	0.89	
20303	22201	26	-0.64	-1.06	-0.22	
15437	22201	28	0.76	0.34	1.17	
14556	22201	29	0.02	-0.42	0.46	
20253	22201	NR	-0.06	-0.44	0.32	
14472	22201	NR	-0.15	-0.55	0.24	
15449	22201	NR	0.83	0.38	1.27	
20362	22201	NR	0.79	0.41	1.17	
20260	22201	NR	-0.09	-0.51	0.33	
20301	22201	NR	-0.24	-0.65	0.16	
22603	22201	NR	-0.24	-0.61	0.14	
15423	22201	NR	0.29	-0.09	0.66	

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
20291	22201	7	0.01	0.57	
10064	22201	15	0.16	0.00	
20306	22201	22	0.07	0.16	
15459	22201	31	0.01	0.12	
20289	22201	NR	-0.13	0.00	
18490	22201	NR	-0.12	0.00	
18498	22201	NR	-0.04	0.01	

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
15516	11101	1	-0.46	-0.96	0.05		-0.28	-0.79	0.23	
10239	11101	3	0.44	-0.01	0.89		0.39	-0.06	0.84	
15540	11101	4	0.49	0.03	0.95		0.50	0.04	0.97	
19446	11101	5	-0.41	-0.87	0.05		-0.64	-1.10	-0.18	
14778	11101	6	-0.97	-1.53	-0.42		-1.12	-1.68	-0.55	B-
15544	11101	10	-0.13	-0.62	0.35		-0.04	-0.52	0.45	
19450	11101	13	-0.44	-0.88	0.00		-0.21	-0.65	0.23	
19634	11101	14	-0.10	-0.55	0.36		-0.57	-1.04	-0.11	
15558	11101	15	-0.09	-0.58	0.40		0.07	-0.41	0.56	
15521	11101	16	0.45	-0.05	0.95		0.28	-0.22	0.78	
10133	11101	18	-0.13	-0.58	0.33		-0.11	-0.57	0.35	
19454	11101	20	0.86	0.34	1.37		0.67	0.15	1.19	
23641	11101	24	0.18	-0.32	0.67		0.86	0.35	1.38	
15526	11101	25	0.36	-0.08	0.80		-0.15	-0.60	0.29	
15563	11101	28	-0.32	-0.85	0.21		0.03	-0.50	0.57	
19456	11101	29	0.01	-0.45	0.46		-0.14	-0.60	0.32	
15517	11101	NR	-1.04	-1.49	-0.59	B-	-1.34	-1.79	-0.88	B-
19422	11101	NR	-1.93	-2.41	-1.44	C-	-1.66	-2.14	-1.17	C-
14799	11101	NR	1.83	1.22	2.43	C+	1.46	0.86	2.05	B+
23365	11101	NR	0.97	0.52	1.42		0.94	0.49	1.39	
14818	11101	NR	0.02	-0.49	0.53		0.12	-0.39	0.64	
19619	11101	NR	0.04	-0.42	0.49		0.29	-0.16	0.74	
17183	11101	NR	-0.68	-1.16	-0.21		-0.64	-1.12	-0.17	
15561	11101	NR	-0.91	-1.38	-0.43		-0.78	-1.25	-0.31	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19433	11101	8	-0.07	0.09		-0.09	0.00	
15610	11101	21	-0.13	0.00		-0.12	0.03	
19645	11101	31	0.05	0.59		0.03	0.23	
19434	11101	NR	0.14	0.00		0.11	0.00	
19625	11101	NR	-0.03	0.33		-0.07	0.01	
19626	11101	NR	-0.16	0.00		-0.13	0.00	
19662	11101	NR	0.02	0.59		0.02	0.39	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
15516	11201	1	-0.49	-1.00	0.03		-0.13	-0.64	0.39	
15553	11201	2	-1.06	-1.52	-0.60	B-	-1.13	-1.59	-0.66	B-
15540	11201	4	0.09	-0.38	0.56		0.47	0.00	0.93	
23638	11201	7	0.11	-0.35	0.57		0.00	-0.46	0.46	
15544	11201	10	0.37	-0.11	0.85		0.56	0.08	1.04	
19631	11201	11	0.52	-0.01	1.04		0.93	0.40	1.45	
19425	11201	12	0.41	-0.06	0.88		0.21	-0.25	0.67	
15558	11201	15	0.31	-0.18	0.79		0.58	0.10	1.05	
15521	11201	16	-0.42	-0.93	0.10		0.12	-0.39	0.63	
19428	11201	17	-0.14	-0.71	0.43		0.06	-0.51	0.62	
10133	11201	18	-0.26	-0.72	0.20		-0.10	-0.56	0.36	
19429	11201	19	0.15	-0.29	0.59		-0.45	-0.89	-0.01	
15526	11201	25	0.53	0.09	0.97		0.59	0.15	1.04	
10247	11201	26	-0.78	-1.29	-0.26		-1.13	-1.66	-0.60	B-
15588	11201	27	0.24	-0.23	0.72		0.06	-0.41	0.54	
15563	11201	28	-0.21	-0.76	0.34		-0.11	-0.65	0.43	
15545	11201	NR	0.56	0.08	1.04		0.17	-0.30	0.65	
18952	11201	NR	-0.94	-1.42	-0.46		-1.00	-1.47	-0.53	B-
19447	11201	NR	0.58	0.13	1.02		0.70	0.25	1.14	
19614	11201	NR	0.16	-0.29	0.62		0.21	-0.25	0.67	
19652	11201	NR	-0.28	-0.74	0.18		-0.23	-0.68	0.23	
14830	11201	NR	0.51	0.01	1.02		0.30	-0.20	0.81	
15560	11201	NR	-0.04	-0.56	0.47		-0.32	-0.82	0.19	
19432	11201	NR	0.04	-0.42	0.50		-0.13	-0.59	0.34	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
14788	11201	9	0.18	0.00	B-	0.18	0.00	B-
15591	11201	22	-0.01	0.01		0.00	0.03	
15531	11201	23	-0.08	0.04		-0.11	0.00	
15533	11201	30	-0.14	0.00		-0.13	0.00	
15590	11201	NR	-0.02	0.17		-0.06	0.03	
15567	11201	NR	-0.08	0.11		-0.11	0.05	
15569	11201	NR	0.05	0.59		0.07	0.14	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
15686	12101	1	-0.69	-1.47	0.10		-1.09	-1.84	-0.33	B-
19578	12101	2	-1.13	-1.62	-0.64	B-	-1.13	-1.62	-0.65	B-
15689	12101	3	0.27	-0.18	0.72		0.15	-0.29	0.60	
23363	12101	5	-0.73	-1.26	-0.20		-1.07	-1.59	-0.54	B-
15635	12101	7	1.25	0.73	1.78	B+	1.74	1.21	2.27	C+
14890	12101	8	0.08	-0.49	0.65		0.35	-0.22	0.91	
15619	12101	10	-0.02	-0.60	0.56		0.42	-0.16	1.01	
19559	12101	12	-0.46	-0.92	0.00		-0.47	-0.93	0.00	
15357	12101	15	-1.33	-1.85	-0.82	B-	-1.32	-1.83	-0.81	B-
19560	12101	17	0.25	-0.22	0.72		0.48	0.02	0.94	
19581	12101	19	-1.07	-1.62	-0.51	B-	-0.83	-1.38	-0.27	
15658	12101	20	-0.02	-0.55	0.50		0.16	-0.37	0.69	
15696	12101	24	-0.40	-0.88	0.09		-0.21	-0.69	0.27	
19583	12101	26	0.00	-0.46	0.45		0.06	-0.40	0.51	
15697	12101	27	0.00	-0.54	0.53		0.17	-0.38	0.72	
19565	12101	28	0.15	-0.54	0.84		0.77	0.06	1.48	
19593	12101	NR	0.37	-0.10	0.84		0.37	-0.10	0.84	
10263	12101	NR	-0.10	-0.64	0.45		-0.12	-0.67	0.43	
19459	12101	NR	0.27	-0.27	0.81		0.11	-0.43	0.65	
23637	12101	NR	-0.09	-0.63	0.45		0.14	-0.39	0.67	
19598	12101	NR	-0.67	-1.15	-0.19		-0.59	-1.07	-0.12	
19563	12101	NR	0.36	-0.20	0.92		0.46	-0.11	1.02	
15623	12101	NR	0.25	-0.30	0.81		0.41	-0.14	0.96	
19500	12101	NR	0.09	-0.47	0.65		-0.02	-0.58	0.53	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
19484	12101	14	-0.08	0.07		-0.09	0.02	
19486	12101	23	-0.07	0.16		-0.11	0.00	
15703	12101	30	0.01	0.00		0.02	0.00	
23351	12101	NR	0.24	0.00	B-	0.25	0.00	C-
19587	12101	NR	-0.06	0.11		-0.01	0.18	
19504	12101	NR	-0.13	0.00		-0.10	0.01	
19572	12101	NR	0.02	0.31		0.01	0.06	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
15686	12201	1	-0.54	-1.33	0.24		-0.39	-1.17	0.39	
15689	12201	3	-0.37	-0.82	0.07		-0.24	-0.68	0.20	
19556	12201	4	-1.13	-1.61	-0.64	B-	-0.86	-1.34	-0.37	
15635	12201	7	0.75	0.21	1.30		0.85	0.31	1.38	
20978	12201	9	0.74	0.26	1.22		0.54	0.06	1.01	
15619	12201	10	0.21	-0.37	0.79		-0.01	-0.58	0.56	
19492	12201	11	-0.58	-1.07	-0.09		-0.04	-0.54	0.46	
15357	12201	15	-1.63	-2.15	-1.11	C-	-1.70	-2.23	-1.18	C-
15638	12201	16	-0.72	-1.35	-0.08		-0.71	-1.34	-0.09	
20892	12201	18	0.61	0.16	1.07		0.43	-0.03	0.89	
15658	12201	20	0.25	-0.27	0.76		0.49	-0.03	1.00	
19001	12201	21	1.17	0.55	1.79	B+	1.22	0.60	1.83	B+
15696	12201	24	-0.59	-1.07	-0.10		-0.92	-1.41	-0.43	
15624	12201	25	0.24	-0.23	0.70		0.04	-0.43	0.50	
15697	12201	27	-0.24	-0.80	0.31		-0.05	-0.60	0.50	
14960	12201	29	-1.18	-1.67	-0.69	B-	-0.92	-1.41	-0.44	
19592	12201	NR	-0.88	-1.37	-0.39		-0.80	-1.28	-0.32	
15688	12201	NR	-1.33	-1.84	-0.81	B-	-0.75	-1.27	-0.24	
19576	12201	NR	0.85	0.34	1.35		1.31	0.79	1.83	B+
23636	12201	NR	1.16	0.68	1.64	B+	1.16	0.68	1.64	B+
19461	12201	NR	-0.08	-0.54	0.37		0.00	-0.45	0.46	
19463	12201	NR	-0.45	-0.99	0.09		-0.08	-0.61	0.45	
19498	12201	NR	0.77	0.30	1.25		0.62	0.15	1.08	
14954	12201	NR	-0.44	-0.93	0.05		-0.56	-1.04	-0.09	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring Non-SLLs; NR = not released.

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
14884	12201	6	0.14	0.00		0.13	0.00	
20976	12201	13	-0.02	0.15		0.03	0.84	
15630	12201	22	0.04	0.06		0.04	0.16	
15631	12201	31	-0.02	0.34		0.02	0.24	
15700	12201	NR	-0.04	0.02		-0.10	0.03	
15665	12201	NR	-0.22	0.00	B+	-0.20	0.00	B+
15649	12201	NR	-0.16	0.00		-0.12	0.00	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring Non-SLLs; NR = not released.

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
23368 T	I	1	W4	6*	0.79 (4.71)	0.59	-1.36
23368 V	I	1	W3	4*	0.90 (3.61)	0.53	-2.24
19361	II	1	W1	2	0.83	0.36	-1.39
19399	II	2	W3	3	0.91	0.38	-2.25
19360	II	3	W2	1	0.66	0.29	-0.12
19393	II	4	W3	2	0.89	0.38	-2.04
21815 502	III	1	R1	1	0.85	0.31	-1.49
18703 502	III	2	R1	4	0.74	0.33	-0.66
18713 502	III	3	R2	1	0.68	0.25	-0.27
18714 502	III	4	R3	1	0.75	0.34	-0.76
18711 502	III	5	R2	1	0.77	0.15	-0.74
21823 502	III	6	R2	1	0.92	0.31	-2.24
18706 502	III	7	R2	4	0.86	0.41	-1.59
18708 502	III	8	R3	4	0.96	0.28	-2.98
18710 502	III	9	R2	2	0.79	0.32	-0.95
20925 499	IV	1	R3	1	0.84	0.31	-1.35
20818 499	IV	2	R2	3	0.81	0.28	-1.14
20822 499	IV	3	R2	2	0.93	0.33	-2.54
18795 499	IV	4	R2	1	0.90	0.25	-1.95
20824 499	IV	5	R2	3*	0.83 (2.49)	0.50	-1.79
20559 499	IV	6	R3	3*	0.78 (2.34)	0.45	-1.45
19479 T	V	1	W4	3*	0.78 (2.34)	0.41	-1.65
19479 V	V	1	W3	2*	0.95 (1.90)	0.37	-2.52
21187 485	VI	1	R1	3	0.82	0.26	-0.99
20815 485	VI	2	R3	4	0.84	0.39	-1.44
20646 485	VI	3	R2	4	0.70	0.36	-0.47
18774 485	VI	4	R2	2	0.91	0.30	-2.19
20657 485	VI	5	R2	2	0.58	0.47	0.17
20656 485	VI	6	R2	1	0.74	0.40	-0.69
18680 482	NR	NR	R2	1	0.84	0.29	-1.42
20775 482	NR	NR	R2	3	0.93	0.29	-2.49
20777 482	NR	NR	R1	3	0.89	0.32	-1.86

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.

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

NR = not released; R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections; W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development.

*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
18674 482	NR	NR	R1	2	0.93	0.31	-2.44
20778 482	NR	NR	R3	2	0.63	0.39	-0.02
18681 482	NR	NR	R3	3*	0.74 (2.22)	0.38	-1.58
19357	NR	NR	W1	1	0.80	0.29	-1.12
19369	NR	NR	W2	3	0.87	0.31	-1.75
19400	NR	NR	W3	4	0.67	0.39	-0.34
19387	NR	NR	W3	3	0.77	0.38	-0.94
19478 T	NR	NR	W4	3*	0.86 (2.57)	0.30	-1.70
19478 V	NR	NR	W3	2*	0.95 (1.91)	0.31	-2.36
23374 T	NR	NR	W4	6*	0.73 (4.37)	0.56	-0.92
23374 V	NR	NR	W3	4*	0.90 (3.61)	0.53	-2.07
18643 495	NR	NR	R2	4	0.60	0.27	0.23
20770 495	NR	NR	R2	2	0.67	0.26	-0.08
18642 495	NR	NR	R3	3	0.79	0.28	-0.92
20769 495	NR	NR	R1	2	0.73	0.30	-0.57
18644 495	NR	NR	R1	3	0.93	0.28	-2.42
23168 495	NR	NR	R2	1	0.79	0.31	-0.91
18649 495	NR	NR	R2	3*	0.75 (2.25)	0.37	-0.95

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.

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

NR = not released; R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections; W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development.

*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		Insufficient	Inadequate	Off Topic	Missing	Illegible	Score 1.0	Score 2.0	Score 3.0
20824_499	IV	5	% of Students	N/A	N/A	0.40	0.57	0.01	7.00	34.16	57.86
			Parameters						-3.20	-2.43	-1.28
20559_499	IV	6	% of Students	N/A	N/A	0.65	0.83	0.01	6.17	49.20	43.14
			Parameters						-2.88	-2.16	-1.30
19479_T	V	1	% of Students	N/A	N/A	0.13	0.75	0.02	11.28	40.55	47.28
			Parameters						-3.26	-2.57	-1.08
19479_V	V	1	% of Students	0.49	0.17	0.13	0.75	0.02	6.73	91.71	
			Parameters						-3.30	-1.62	
18681_482	NR	NR	% of Students	N/A	N/A	0.30	0.20	0.01	10.53	55.25	33.71
			Parameters						-3.37	-2.76	-1.29
19478_T	NR	NR	% of Students	N/A	N/A	0.58	1.24	0.03	3.73	30.00	64.41
			Parameters						-2.90	-2.03	-1.25
19478_V	NR	NR	% of Students	0.21	0.09	0.58	1.24	0.03	5.07	92.78	
			Parameters						-3.03	-1.13	
18649_495	NR	NR	% of Students	N/A	N/A	0.75	3.97	0.03	8.81	42.99	43.45
			Parameters						-2.23	-1.41	-0.57

Note. The total number of students is 131 660; NR = not released; N/A = not applicable.

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
23368_T	I	1	% of Students	0.02	0.33	0.02	0.08	0.08	0.52	0.85
			Parameters				-2.93*		-2.50	-2.30
23368_V	I	1	% of Students	0.00	0.33	0.02	0.21	0.11	1.32	3.90
			Parameters				-3.48*		-2.88	-2.64
23374_T	NR	NR	% of Students	0.01	0.50	0.02	0.39	0.35	0.90	1.85
			Parameters				-2.77*		-2.35	-1.92
23374_V	NR	NR	% of Students	0.00	0.50	0.02	0.36	0.15	1.07	3.66
			Parameters				-3.29*		-2.67	-2.38
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
23368_T	I	1	% of Students	2.68	5.81	14.64	22.17	26.27	18.90	7.64
			Parameters	-2.12	-1.97	-1.75	-1.49	-1.05	-0.43	0.77
23368_V	I	1	% of Students	12.59	30.79	50.72				
			Parameters	-2.31	-1.92	-1.06				
23374_T	NR	NR	% of Students	4.33	10.64	22.40	25.60	17.19	11.59	4.23
			Parameters	-1.70	-1.49	-1.26	-0.93	-0.42	0.28	1.32
23374_V	NR	NR	% of Students	12.65	31.04	50.56				
			Parameters	-2.11	-1.77	-0.96				

Note. The total number of students is 131 660.

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

NR = not released.

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
21122_T	I	1	W4	6*	0.77 (4.62)	0.58	-1.24
21122_V	I	1	W3	4*	0.84 (3.35)	0.57	-2.07
18858	II	1	W1	4	0.76	0.21	-0.62
18984	II	2	W3	1	0.88	0.23	-1.69
18855	II	3	W2	4	0.96	0.18	-3.05
18885	II	4	W3	4	0.90	0.33	-1.96
18938_447	III	1	R1	4	0.94	0.27	-2.53
18946_447	III	2	R2	4	0.69	0.34	-0.25
18948_447	III	3	R2	1	0.82	0.39	-1.20
18941_447	III	4	R2	1	0.63	0.22	0.21
18945_447	III	5	R3	4	0.88	0.31	-1.68
18939_447	III	6	R1	2	0.76	0.34	-0.69
18951_447	III	7	R3	3	0.48	0.19	1.12
18949_447	III	8	R2	4	0.71	0.16	-0.33
18943_447	III	9	R2	1	0.87	0.43	-1.67
20640_444	IV	1	R2	4	0.78	0.25	-0.78
20647_444	IV	2	R2	3	0.89	0.28	-1.90
20645_444	IV	3	R2	1	0.69	0.32	-0.21
20643_444	IV	4	R3	3	0.87	0.32	-1.66
18878_444	IV	5	R2	2	0.62	0.35	0.16
18880_444	IV	6	R2	3*	0.77 (2.31)	0.38	-1.59
18881_444	IV	7	R3	3*	0.79 (2.37)	0.40	-1.57
18959_T	V	1	W4	3*	0.87 (2.62)	0.38	-1.98
18959_V	V	1	W3	2*	0.85 (1.70)	0.44	-2.26
20608_449_12	VI	1	R2	3	0.89	0.24	-1.74
18825_449_12	VI	2	R1	2	0.67	0.33	-0.04
18831_449_12	VI	3	R2	2	0.84	0.17	-1.20
18833_449_12	VI	4	R2	2	0.67	0.31	-0.08
18832_449_12	VI	5	R3	4	0.83	0.24	-1.13
18827_449_12	VI	6	R2	4	0.83	0.27	-1.24
20679_442	NR	NR	R1	1	0.64	0.34	0.07
18916_442	NR	NR	R2	4	0.51	0.38	0.80
18915_442	NR	NR	R1	2	0.94	0.24	-2.59
18919_442	NR	NR	R3	2	0.91	0.27	-2.11
18921_442	NR	NR	R2	2	0.72	0.32	-0.43
20678_442	NR	NR	R3	3*	0.72 (2.15)	0.36	-1.11

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.

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

NR = not released; R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections; W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development.

*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
18854	NR	NR	W1	1	0.80	0.30	-0.97
18886	NR	NR	W3	3	0.83	0.27	-1.22
18861	NR	NR	W2	2	0.74	0.31	-0.54
18888	NR	NR	W3	1	0.52	0.16	0.93
18934_T	NR	NR	W4	3*	0.82 (2.47)	0.47	-1.66
18934_V	NR	NR	W3	2*	0.87 (1.74)	0.40	-2.18
21123_T	NR	NR	W4	6*	0.72 (4.31)	0.57	-1.11
21123_V	NR	NR	W3	4*	0.80 (3.20)	0.58	-1.86
18929_440	NR	NR	R2	1	0.81	0.37	-1.11
18931_440	NR	NR	R2	1	0.82	0.27	-1.14
18927_440	NR	NR	R3	3	0.77	0.35	-0.72
18925_440	NR	NR	R1	4	0.77	0.44	-0.77
18928_440	NR	NR	R1	2	0.60	0.29	0.28
18930_440	NR	NR	R2	4	0.66	0.24	0.01
20636_440	NR	NR	R2	3*	0.67 (2.01)	0.29	-0.48

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.

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

NR = not released; R1 = understanding explicitly; R2 = understanding implicitly; R3 = making connections; W1 = developing main idea; W2 = organizing information; W3 = using conventions; W4 = topic development.

*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		Insufficient	Inadequate	Off Topic	Missing	Illegible	Score 1.0	Score 2.0	Score 3.0
18880_444	IV	6	% of Students	N/A	N/A	0.41	0.29	0.00	10.21	45.99	43.10
			Parameters						-3.27	-2.62	-1.14
18881_444	IV	7	% of Students	N/A	N/A	0.33	0.48	0.02	3.68	53.27	42.21
			Parameters						-3.02	-2.33	-1.60
18959_T	V	1	% of Students	N/A	N/A	0.12	0.52	0.00	0.41	35.34	63.62
			Parameters						-3.10	-2.29	-2.03
18959_V	V	1	% of Students	0.48	0.14	0.12	0.52	0.00	27.14	71.60	
			Parameters						-3.52	-2.42	
20678_442	NR	NR	% of Students	N/A	N/A	0.76	0.58	0.02	9.01	63.35	26.29
			Parameters						-2.76	-2.12	-1.09
18934_T	NR	NR	% of Students	N/A	N/A	0.27	0.54	0.00	2.52	45.56	51.10
			Parameters						-2.97	-2.22	-1.64
18934_V	NR	NR	% of Students	0.46	0.08	0.27	0.54	0.00	23.36	75.28	
			Parameters						-3.34	-2.15	
20636_440	NR	NR	% of Students	N/A	N/A	2.81	4.82	0.17	29.02	17.30	45.87
			Parameters						-1.79	-0.88	1.05

Note. The total number of students is 5162. NR = not released.

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
21122_T	I	1	% of Students	0.12	0.06	0.02	0.14	0.06	0.45	0.85
			Parameters				-2.88*		-2.44	-2.18
21122_V	I	1	% of Students	0.00	0.06	0.02	0.23	0.25	2.58	9.14
			Parameters				-3.58*		-3.06	-2.76
21123_T	NR	NR	% of Students	0.14	0.00	0.37	0.12	1.07	1.82	0.14
			Parameters				-3.16*		-2.77	-2.30
21123_V	NR	NR	% of Students	0.00	0.14	0.00	0.23	0.48	5.13	14.26
			Parameters				-3.54*		-3.06	-2.73
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
21122_T	I	1	% of Students	2.94	6.04	23.23	22.96	17.73	17.13	8.29
			Parameters	-1.98	-1.84	-1.62	-1.36	-0.78	-0.12	0.83
21122_V	I	1	% of Students	29.19	29.74	28.79				
			Parameters	-2.36	-1.79	-0.34				
21123_T	NR	NR	% of Students	6.59	13.17	25.75	22.61	11.84	11.35	5.19
			Parameters	-1.98	-1.72	-1.35	-0.93	-0.34	0.32	1.15
21123_V	NR	NR	% of Students	32.49	27.02	20.24				
			Parameters	-2.19	-1.43	0.02				

Note. The total number of students is 5162.

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

NR = not released.

Differential Item Functioning (DIF) Analysis Results

The gender- and SLL-based DIF results for the OSSLT are provided in Tables 7.1.83a–7.1.85b. The results for the English-language test are from two random samples of 2000 students for gender and 1500 students for SLL-based. 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
19361	II	1	0.31	-0.10	0.73		0.57	0.12	1.02	
19399	II	2	0.14	-0.43	0.71		0.39	-0.19	0.97	
19360	II	3	-0.03	-0.37	0.30		0.13	-0.20	0.46	
19393	II	4	-0.12	-0.65	0.40		-0.57	-1.10	-0.03	
21815_502	III	1	1.66	1.22	2.09	C+	0.77	0.32	1.22	
18703_502	III	2	0.90	0.54	1.26		0.72	0.35	1.10	
18713_502	III	3	0.31	-0.02	0.64		0.24	-0.10	0.57	
18714_502	III	4	-0.51	-0.87	-0.14		-0.32	-0.70	0.05	
18711_502	III	5	0.46	0.10	0.82		0.38	0.03	0.73	
21823_502	III	6	0.23	-0.33	0.79		-0.16	-0.77	0.45	
18706_502	III	7	0.12	-0.34	0.58		0.28	-0.21	0.76	
18708_502	III	8	1.92	1.16	2.67	C+	1.72	0.96	2.48	B+
18710_502	III	9	1.15	0.76	1.55	B+	1.09	0.69	1.48	B+
20925_499	IV	1	-0.09	-0.50	0.33		0.29	-0.15	0.72	
20818_499	IV	2	-0.36	-0.75	0.03		-0.41	-0.81	-0.01	
20822_499	IV	3	-2.18	-2.90	-1.47	C-	-1.48	-2.14	-0.82	B-
18795_499	IV	4	-0.12	-0.65	0.41		-0.28	-0.79	0.23	
21187_485	VI	1	1.71	1.29	2.13	C+	1.90	1.50	2.30	C+
20815_485	VI	2	1.18	0.75	1.62	B+	1.11	0.65	1.57	B+
20646_485	VI	3	0.33	-0.02	0.68		0.30	-0.06	0.66	
18774_485	VI	4	0.69	0.15	1.23		1.09	0.55	1.63	B+
20657_485	VI	5	0.70	0.34	1.06		0.85	0.48	1.22	
20656_485	VI	6	0.63	0.26	1.00		0.49	0.11	0.87	
18680_482	NR	NR	0.07	-0.35	0.49		0.75	0.32	1.18	
20775_482	NR	NR	1.20	0.62	1.79	B+	0.98	0.37	1.60	
20777_482	NR	NR	1.49	1.01	1.97	B+	0.47	-0.05	1.00	
18674_482	NR	NR	1.04	0.43	1.64	B+	1.70	1.08	2.32	C+
20778_482	NR	NR	0.80	0.46	1.14		0.91	0.56	1.26	
19357	NR	NR	0.57	0.18	0.96		0.67	0.27	1.06	
19369	NR	NR	0.00	-0.45	0.44		-0.16	-0.64	0.31	
19400	NR	NR	0.13	-0.21	0.48		0.52	0.16	0.88	
19387	NR	NR	-0.23	-0.62	0.15		-0.14	-0.54	0.25	
18643_495	NR	NR	1.21	0.89	1.53	B+	1.15	0.82	1.48	B+
20770_495	NR	NR	-0.06	-0.39	0.27		-0.49	-0.83	-0.15	
18642_495	NR	NR	1.49	1.10	1.89	B+	1.37	0.98	1.77	B+
20769_495	NR	NR	1.37	1.01	1.73	B+	1.13	0.76	1.50	B+
18644_495	NR	NR	0.46	-0.13	1.04		0.31	-0.28	0.90	
23168_495	NR	NR	1.42	1.03	1.80	B+	1.87	1.45	2.28	C+

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

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
23368_T	I	1	-0.15	0.00		-0.17	0.00	B-
23368_V	I	1	-0.19	0.00	B-	-0.16	0.00	
20824_499	IV	5	-0.10	0.00		-0.11	0.00	
20559_499	IV	6	-0.21	0.00	B-	-0.20	0.00	B-
19479_T	V	1	-0.09	0.01		-0.08	0.02	
19479_V	V	1	-0.13	0.00		-0.07	0.00	
18681_482	NR	NR	0.05	0.47		-0.03	0.63	
19478_T	NR	NR	-0.20	0.00	B-	-0.20	0.00	B-
19478_V	NR	NR	-0.12	0.00		-0.13	0.00	
23374_T	NR	NR	-0.18	0.00	B-	-0.15	0.00	
23374_V	NR	NR	-0.20	0.00	B-	-0.13	0.00	
18649_495	NR	NR	-0.14	0.00		-0.10	0.00	

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

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
18858	II	1	0.62	0.25	0.98	
18984	II	2	0.69	0.21	1.17	
18855	II	3	0.45	-0.35	1.25	
18885	II	4	-0.90	-1.44	-0.36	
18938_447	III	1	1.45	0.80	2.10	B+
18946_447	III	2	0.80	0.45	1.16	
18948_447	III	3	1.19	0.76	1.63	B+
18941_447	III	4	0.26	-0.06	0.58	
18945_447	III	5	0.41	-0.08	0.91	
18939_447	III	6	0.18	-0.20	0.55	
18951_447	III	7	0.00	-0.32	0.31	
18949_447	III	8	-0.86	-1.20	-0.52	
18943_447	III	9	0.61	0.11	1.12	
20640_444	IV	1	-0.34	-0.71	0.04	
20647_444	IV	2	-0.22	-0.73	0.29	
20645_444	IV	3	0.48	0.13	0.83	
20643_444	IV	4	1.17	0.68	1.65	B+
18878_444	IV	5	0.86	0.52	1.20	
20608_449_12	VI	1	1.47	0.97	1.97	B+
18825_449_12	VI	2	1.80	1.44	2.15	C+

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

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
18831_449_12	VI	3	0.27	-0.15	0.68	
18833_449_12	VI	4	0.93	0.59	1.27	
18832_449_12	VI	5	0.26	-0.15	0.68	
18827_449_12	VI	6	0.28	-0.13	0.70	
20679_442	NR	NR	0.86	0.52	1.20	
18916_442	NR	NR	0.95	0.61	1.28	
18915_442	NR	NR	0.26	-0.39	0.91	
18919_442	NR	NR	0.29	-0.27	0.85	
18921_442	NR	NR	0.23	-0.13	0.59	
18854	NR	NR	0.83	0.43	1.23	
18886	NR	NR	-0.93	-1.35	-0.50	
18861	NR	NR	-0.66	-1.02	-0.29	
18888	NR	NR	0.44	0.13	0.75	
18929_440	NR	NR	0.05	-0.37	0.46	
18931_440	NR	NR	0.12	-0.29	0.53	
18927_440	NR	NR	0.90	0.51	1.29	
18925_440	NR	NR	1.01	0.60	1.42	B+
18928_440	NR	NR	0.42	0.10	0.75	
18930_440	NR	NR	-0.08	-0.41	0.25	

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

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
21122_T	I	1	-0.14	0.00	
21122_V	I	1	-0.23	0.00	B-
18880_444	IV	6	-0.17	0.00	
18881_444	IV	7	-0.18	0.00	B-
18959_T	V	1	-0.13	0.00	
18959_V	V	1	-0.12	0.00	
20678_442	NR	NR	0.21	0.00	B+
18934_T	NR	NR	-0.16	0.00	
18934_V	NR	NR	-0.08	0.00	
21123_T	NR	NR	-0.11	0.00	
21123_V	NR	NR	-0.14	0.00	
20636_440	NR	NR	-0.16	0.00	

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

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
19361	II	1	0.66	0.21	1.10		0.36	-0.13	0.86	
19399	II	2	1.08	0.51	1.65	B+	1.36	0.80	1.93	B+
19360	II	3	0.21	-0.17	0.59		0.11	-0.28	0.50	
19393	II	4	0.39	-0.18	0.95		0.58	0.01	1.15	
21815_502	III	1	-0.82	-1.31	-0.33		-0.80	-1.31	-0.29	
18703_502	III	2	0.71	0.31	1.11		0.96	0.55	1.37	
18713_502	III	3	-0.38	-0.77	0.01		-0.95	-1.36	-0.55	
18714_502	III	4	-0.37	-0.79	0.04		-1.31	-1.77	-0.86	B-
18711_502	III	5	1.17	0.77	1.57	B+	1.48	1.08	1.88	B+
21823_502	III	6	0.95	0.34	1.57		1.19	0.58	1.81	B+
18706_502	III	7	1.79	1.28	2.30	C+	1.97	1.46	2.47	C+
18708_502	III	8	1.82	1.05	2.59	C+	1.72	1.00	2.44	B+
18710_502	III	9	1.29	0.86	1.71	B+	1.50	1.08	1.93	C+
20925_499	IV	1	0.29	-0.18	0.75		1.03	0.55	1.51	B+
20818_499	IV	2	0.63	0.19	1.08		0.60	0.16	1.04	
20822_499	IV	3	1.25	0.60	1.90	B+	1.48	0.81	2.14	B+
18795_499	IV	4	1.68	1.14	2.22	C+	1.64	1.08	2.19	C+
21187_485	VI	1	2.13	1.70	2.56	C+	2.75	2.34	3.17	C+
20815_485	VI	2	0.44	-0.05	0.93		-0.28	-0.80	0.23	
20646_485	VI	3	-0.11	-0.51	0.29		-0.35	-0.76	0.06	
18774_485	VI	4	-0.45	-1.05	0.15		-0.92	-1.56	-0.27	
20657_485	VI	5	-0.19	-0.60	0.22		-0.53	-0.97	-0.10	
20656_485	VI	6	-0.15	-0.56	0.27		0.17	-0.26	0.59	
18680_482	NR	NR	0.08	-0.38	0.53		-0.19	-0.67	0.29	
20775_482	NR	NR	-0.27	-0.92	0.39		-0.74	-1.43	-0.04	
20777_482	NR	NR	0.32	-0.19	0.84		0.91	0.36	1.46	
18674_482	NR	NR	1.15	0.51	1.78	B+	0.92	0.27	1.58	
20778_482	NR	NR	0.39	0.01	0.77		0.76	0.37	1.15	
19357	NR	NR	0.07	-0.37	0.51		-0.34	-0.79	0.12	
19369	NR	NR	-0.09	-0.60	0.43		0.24	-0.28	0.76	
19400	NR	NR	-0.32	-0.71	0.07		-0.83	-1.24	-0.42	
19387	NR	NR	0.27	-0.16	0.69		-0.45	-0.89	0.00	
18643_495	NR	NR	0.18	-0.19	0.55		-0.38	-0.75	0.00	
20770_495	NR	NR	0.31	-0.06	0.68		0.21	-0.16	0.59	
18642_495	NR	NR	1.40	0.97	1.83	B+	1.04	0.62	1.47	B+
20769_495	NR	NR	-0.06	-0.47	0.35		0.35	-0.06	0.76	
18644_495	NR	NR	0.12	-0.55	0.80		0.17	-0.51	0.85	
23168_495	NR	NR	0.06	-0.38	0.49		-0.35	-0.81	0.11	

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

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
23368_T	I	1	-0.24	0.00	B-	-0.27	0.00	C-
23368_V	I	1	0.08	0.02		0.19	0.00	B+
20824_499	IV	5	-0.01	0.48		-0.11	0.01	
20559_499	IV	6	-0.11	0.00		-0.15	0.00	
19479_T	V	1	-0.21	0.00	B-	-0.20	0.00	B-
19479_V	V	1	0.07	0.10		0.13	0.00	
18681_482	NR	NR	-0.13	0.00		-0.09	0.01	
19478_T	NR	NR	-0.05	0.14		-0.20	0.00	B-
19478_V	NR	NR	0.02	0.03		0.05	0.01	
23374_T	NR	NR	-0.13	0.00		-0.06	0.13	
23374_V	NR	NR	0.07	0.03		0.14	0.00	
18649_495	NR	NR	-0.20	0.00	B-	-0.31	0.00	C-

Note. B = moderate DIF; C = large DIF; - = favouring SLLs; + = favouring non-SLLs; NR = not released.

