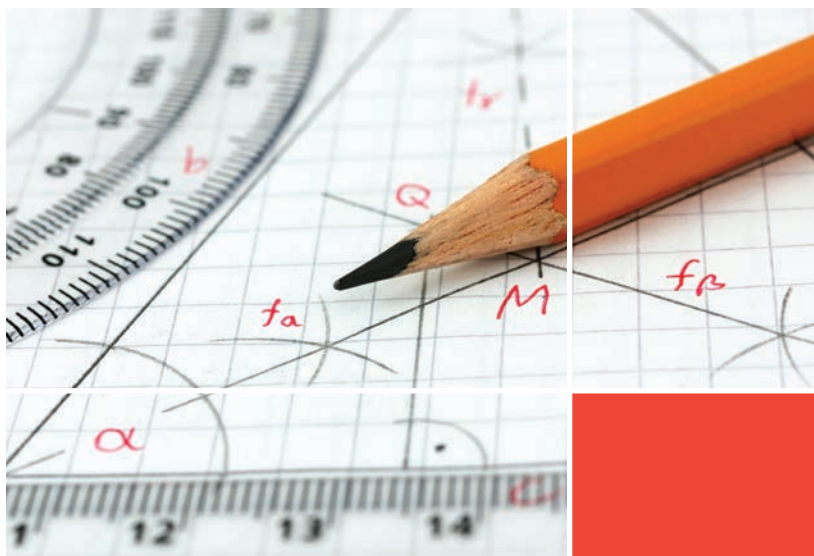


EQAO's Technical Report

for the 2014–2015 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 2014–2015 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 978-1-4606-9826-6, ISSN 1927-7105

© 2017 Queen's Printer for Ontario | Ctrc_report_ne_0317

SPECIAL NOTE FOR INTERPRETING AND USING EQAO RESULTS

In the spring of 2015, due to labour disruptions in English-language public schools by the Elementary Teachers' Federation of Ontario (ETFO) and the Ontario Secondary School Teachers' Federation (OSSTF), the students who participated in the EQAO English-language assessments in 2015 do not reflect the full provincial population for Grades 3, 6 and 9. The students who participated in the primary- and junior-division assessments include those from the English Catholic schools, French Public schools, French Catholic schools, English-language Provincial schools, Private schools, First Nations schools, and international schools. The students who participated in the Grade 9 Mathematics assessment include those from all but three school boards and one school board in which some individual teachers decided not to administer the assessment to their class. These exceptional circumstances caused a drastic difference to the participation in the 2015 Grades 3, 6 and 9 assessments, so analyses of data and reporting of results changed accordingly: The analyses were limited to those students who participated in the assessments, who do not reflect the full provincial population; some of the results in the tables have been blacked out in order to prevent misinterpretation. This special note is to caution readers about the exceptional circumstances in 2015 while they interpret and use the results in this technical report.

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	8
ASSESSMENT ADMINISTRATION	8
The Administration Guides	8
Support for Students with Special Education Needs and English Language Learners: The Guides for Accommodations and Special Provisions	8
EQAO Policies and Procedures	9
QUALITY ASSURANCE	10
ASSESSMENT PARTICIPATION	10
CHAPTER 4: SCORING	12
SCORING IN TRANSITION	12
THE RANGE-FINDING PROCESS	12
Pre-Range Finding	13
Range Finding	13
Overview of the Range-Finding Process	14
PREPARING TRAINING MATERIALS FOR SCORING	14
FIELD-TEST SCORING	14
Training Field-Test Scoring Leaders and Scorers	15
Scoring Open-Response Field-Test Items	15
Developing Additional Scorer-Training Materials Before Scoring Operational Items	15
SCORING OPEN-RESPONSE OPERATIONAL ITEMS	15
Scoring Rooms for Scoring Open-Response Operational Items	16
Training for Scoring Open-Response Operational Items	16
Training of Scoring Leaders and Scoring Supervisors for Scoring Open-Response Operational Items	16
Training of Scorers for Scoring Open-Response Operational Items	17
PROCEDURES AT THE SCORING SITE	18
Students at Risk	18
Inappropriate Content, Cheating and Other Issues	19
Ongoing Daily Training	19
Daily Scoring-Centre Reports for Monitoring the Quality of Open-Response Item Scoring	19
Required Actions: Consequences of the Review and Analysis of Daily Scoring-Centre Data Reports	21
Auditing	22
SCORER VALIDITY AND RELIABILITY	22
Scoring Validity	22
Scorer Reliability (for OSSLT only)	23
CHAPTER 5: EQUATING	24
IRT MODELS	24

EQUATING DESIGN.....	25
CALIBRATION AND EQUATING SAMPLES.....	25
CALIBRATION.....	26
IDENTIFICATION OF ITEMS TO BE EXCLUDED FROM EQUATING.....	26
THE ASSESSMENTS OF READING, WRITING AND MATHEMATICS: PRIMARY AND JUNIOR DIVISIONS	27
Description of the IRT Model	27
Equating Sample: Exclusion Rules	27
Equating Steps.....	28
Eliminating Items and Collapsing of Score Categories.....	28
Equating Results.....	29
THE GRADE 9 ASSESSMENT OF MATHEMATICS.....	31
Description of the IRT Model	31
Equating Sample	31
Equating Steps.....	32
Eliminating Items and the Collapsing of Score Categories.....	32
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.....	35
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.....	58
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT).....	64
THE ONTARIO SECONDARY SCHOOL LITERACY TEST (OSSLT)	66
Classical Test Theory (CTT) Analysis.....	66
Item Response Theory (IRT) Analysis.....	66
Descriptive Item Statistics for Classical Test Theory (CTT) and Item Response Theory (IRT).....	69
DIFFERENTIAL ITEM FUNCTIONING (DIF)	69
The Primary- and Junior-Division Assessments	71
The Grade 9 Mathematics Assessment.....	73
The OSSLT	75
DECISION ACCURACY AND CONSISTENCY	76
Accuracy	77
Consistency	77
Estimation from One Test Form.....	78
The Primary and Junior Assessments.....	79
The Grade 9 Assessment of Mathematics	79
The OSSLT	80
REFERENCES	80
CHAPTER 8: VALIDITY EVIDENCE.....	81
INTRODUCTION	81
The Purposes of EQAO Assessments	81

Conceptual Framework for the Validity Argument.....	81
VALIDITY EVIDENCE BASED ON THE CONTENT OF THE ASSESSMENTS AND THE ASSESSMENT PROCESSES ..	82
Test Specifications for EQAO Assessments	82
Appropriateness of Test Questions.....	82
Quality Assurance in Administration	83
Scoring of Open-Response Items	83
Equating	84
VALIDITY EVIDENCE BASED ON THE TEST CONSTRUCTS AND INTERNAL STRUCTURE.....	84
Test Dimensionality	84
Technical Quality of the Assessments.....	84
VALIDITY EVIDENCE BASED ON EXTERNAL ASSESSMENT DATA	85
Linkages to International Assessment Programs.....	85
VALIDITY EVIDENCE SUPPORTING APPROPRIATE INTERPRETATIONS OF RESULTS.....	86
Setting Standards.....	86
Reporting.....	86
CONCLUSION.....	87
REFERENCES	87
APPENDIX 4.1: SCORING VALIDITY FOR ALL ASSESSMENTS AND INTERRATER	
RELIABILITY FOR OSSLT	89
APPENDIX 7.1: SCORE DISTRIBUTIONS AND ITEM STATISTICS	99

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.

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 at 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 at 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 2014–2015 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; and writing prompts) for a particular assessment are appropriate to the age and grade of the students, the curriculum expectations being measured and the purpose of the assessment. They make suggestions for the inclusion, exclusion or revision of items.

The EQAO Sensitivity Committee

The Sensitivity Committee, which considers all four EQAO assessments, comprises 8–10 Ontario educators who serve for terms of one to five years. About 4–8 members meet in focused subgroups once a year to make recommendations that will assist EQAO in ensuring the fairness of all field-test and operational items being proposed for its assessments. They provide expert advice from a specialized equity perspective to ensure that assessment materials are fair for a wide range of students. The members of the Sensitivity Committee possess expertise in and current experience with equity issues in education (issues related to the diversity of Ontario students, students with special education needs and English language learners).

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

Field Testing

Field testing of assessment materials ensures that assessment items selected for future operational assessments are psychometrically sound and fair for all students. Field testing also provides data to equate each year's assessment with the previous year's assessment, so 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); 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, and 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 strategies, completion of mathematics homework; home environment (e.g., time spent doing extra-curricular activities, “screen time,” language(s) spoken at home by student); 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., 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 of school.

CHAPTER 3: TEST ADMINISTRATION AND PARTICIPATION

Assessment Administration

To ensure consistent and fair practice across the province in the administration of the assessments, EQAO publishes an administration guide and a guide for accommodations and special provisions 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:

- sign language or oral interpreter
- contracted, uncontracted and Unified English Braille versions 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
- MP3 audio version plus a set of regular-print booklets
- MP3 audio 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 statistical analyses of student response data to identify student response patterns to multiple-choice items that suggest the possibility of collusion between two or more students.
- 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.

Assessment Participation

The year of 2015 was atypical with regard to the administration of EQAO assessments. The administration of the OSSLT proceeded as planned, but in the spring of 2015, labour disruptions by the Elementary Teachers' Federation of Ontario (ETFO) and the Ontario Secondary School Teachers' Federation (OSSTF) caused EQAO's primary and junior division assessments to not be administered in any English-language public elementary schools, representing approximately 68% of the student body in those grades. Elementary school students in the English Catholic, French Public and French Catholic systems participated in the assessments as usual. Students in English-language Provincial schools, Private schools, First Nations schools, as well as international schools, also participated. In addition, the labour action resulted in EQAO's spring

administration of the Grade 9 Assessment of Mathematics to not be administered to students in three English Public School Boards; labour action in another two school boards was such that the Boards gave individuals teachers the responsibility to decide whether or not to administer the assessments. There were major reporting implications as a result of these labour actions.

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.

Scoring in Transition

EQAO is moving toward online scoring for all of its assessments. In 2014–2015, the OSSLT followed the traditional paper-based approach to scoring; the primary, junior and Grade 9 assessments employed online scoring. Online scoring was conducted in a distributed fashion, in that scorers coded student work from home; scoring supervisors, under EQAO direction, oversaw the scoring process from a central location in downtown Toronto. Online scoring incorporated all of the rigorous procedures used in paper-based scoring but included many enhancements.

Following is a description of the traditional paper-based scoring process formerly used for all EQAO assessments (only the OSSLT in 2015). 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. Education officers read through 25 booklets, or more if necessary, to see if there is a range of responses and if the item or prompt worked with students. The pre-range finding process for items or tasks does not proceed unless there is a range of responses.
4. Typically, booklets are sorted into four piles based on the range of responses: approximately 20 low, 20 medium, 20 high and 25 of mixed range. The booklets chosen for the piles represent the full range of student responses, including off-topic, incorrect, typical and unusual responses. The mixed pile is determined after the other three piles.
5. Items and tasks that have been left unanswered (“blanks”) or that are difficult to read due to poor handwriting or light ink are not selected for pre-range finding.
6. A cover sheet for each range, showing item, task and booklet numbers, is printed and labelled “high,” “medium,” “low” or “mixed.”

Range Finding

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

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

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 (for the OSSLT only), 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 2014–2015, the EQAO target of 95% exact-plus-adjacent agreement for validity was met for all but one item in mathematics, and for all items in reading and writing. The aggregate validity estimates for exact-plus-adjacent agreement ranged from 98–100%.

The Grade 9 Assessment of Mathematics: Academic and Applied

The Grade 9 Assessment of Mathematics has separate English-language and French-language versions for students enrolled in academic and applied courses. The assessment is administered in January for students in mathematics courses in the first semester and in June for students in second-semester and full-year courses. The scoring validity estimates for the Grade 9 Assessment of Mathematics are presented in Tables 4.1.13–4.1.16 of Appendix 4.1 for both administrations. The tables present statistics for each open-response item and the aggregate for open-response items for each administration. They also include aggregate statistics across the winter and spring administrations, because both were scored during the same scoring session in July 2015. 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 assessment. The aggregate validity estimates ranged from 98.5–100%.

The Ontario Secondary School Literacy Test (OSSLT)

The scoring validity estimates for the OSSLT are reported in Tables 4.1.17–4.1.20 of Appendix 4.1. For each test, four reading items were scored with three-point rubrics, and two long-writing prompts were scored with a six-point rubric for topic development and a four-point rubric for conventions. Two short-writing prompts were scored with a three-point rubric for topic development and a two-point rubric for conventions, which were combined into a five-point rubric for the purposes of validity statistics. Aggregate statistics are provided separately for reading items, short-writing prompts, long-writing prompts and all writing prompts. In 2014–2015, the EQAO target of 95% exact-plus-adjacent agreement for validity was met for all items. The aggregate validity estimates ranged from 96.5–99.4%.

Scorer Reliability (for OSSLT only)

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. To determine the reliability of open-response scoring for the OSSLT, 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 OSSLT are presented in Tables 4.1.21–4.1.24 of Appendix 4.1. The EQAO target of 95% exact-plus-adjacent agreement for interrater reliability was met for all but one reading item, for none of the short-writing prompts and for half of the long-writing prompts. The aggregate reliability estimates ranged from 90.8–97.7%.

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 2013–2014 to 2014–2015, only equating procedures were used in 2014–2015.

The following sections describe the Item Response Theory (IRT) models, equating design, equating samples and calibration procedures used during the 2014–2015 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.

With the exceptional circumstances in 2015 due to labour disruptions in English-language public schools, data used for equating and other analyses in 2015 do not reflect the provincial population for Grades 3, 6 and 9. To ensure the accuracy of equating results, EQAO's psychometric team conducted sensitivity analyses to determine which students from the 2013–2014 assessments should be used in the equating of 2014–2015 assessments. This study was conducted prior to the 2014–2015 administration of the primary- and junior-division and Grade 9 assessments, using data from the 2012–2013 and 2013–2014 administrations of these assessments. Three methods (see below) were explored using different equating samples, with results from Method 1 as the benchmark against which results from Methods 2 and 3 were compared. In the descriptions below, “full” data refer to the population data that would result if all or almost all students in the province participated in the assessments; “partial” data refer to the data that would result as in the 2015 administration (see the section of “Assessment Participation” in Chapter 3).

Method 1 – Full to full: equating 2014 full population data to 2013 full population data;

Method 2 – Partial to full: equating 2014 partial data to 2013 full population data;

Method 3 – Partial to partial: equating 2014 partial data to 2013 partial data.

The results of this study demonstrated that using different equating samples produced very similar results in terms of the percentage of students classified at each performance level for the 2013–2014 assessments. This indicated that equating results are not sensitive to which students are included in the equating samples and that the current equating methods can work well in dealing with the exceptional circumstances in 2015. Based on the results of this study, the partial-to-full method (that is, Method 2) was chosen for the 2014–2015 equating as it produced results more similar to those in Method 1.

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 2013–2014 and 2014–2015:

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 2013–2014 population who responded to both the 2014–2015 operational-test items and the field-test items that had been brought forward to form the 2014–2015 operational tests and who were not excluded by the rules stated above (calibration sample);

2. students from the 2013–2014 population who wrote the operational test and who were not excluded by the rules stated above (equating sample) and
3. students from the 2014–2015 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 2013–2014 and 2014–2015 Primary- and Junior-Division Assessments (English and French)

Assessment	Number of Students		
	2014–2015 Calibration and Equating Sample	2013–2014 Calibration Sample	2013–2014 Equating Sample
Primary Reading (English)	29 268	26 916	97 175
Junior Reading (English)	30 682	28 972	102 567
Primary Reading (French)	6 219	1 939	5 930
Junior Reading (French)	5 276	1 740	5 154
Primary Writing (English)	29 376	45 726	97 676
Junior Writing (English)	30 707	46 850	102 582
Primary Writing (French)	6 251	3 717	5 963
Junior Writing (French)	5 276	3 016	5 153
Primary Mathematics (English)	28 814	88 397	93 378
Junior Mathematics (English)	31 189	90 237	104 514
Primary Mathematics (French)	6 297	5 992	6 010
Junior Mathematics (French)	5 376	4 558	5 227

Equating Steps

In equating the 2013–2014 and 2014–2015 tests, the forward-fixed common-item-parameter non-equivalent group design was implemented as follows:

1. The 2014–2015 operational items were calibrated independently to obtain item parameter estimates and student-proficiency scores for the 2014–2015 calibration and equating sample.
2. The 2013–2014 operational items were calibrated (using the calibration sample) together with the field-test items that were brought forward to the 2014–2015 operational assessments. In this calibration, the item parameter estimates of the field-test items were fixed at the values obtained from the 2014–2015 calibration runs (Step 1).
3. The 2013–2014 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 2013–2014 equating sample from the levels assigned in 2013–2014. 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 2014–2015.
6. The operational item parameter estimates of 2014–2015 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 2014–2015 student theta values, students were assigned to levels, and the percentage of students at each performance level was determined.

Eliminating Items and Collapsing of Score Categories

For the primary- and junior-division assessments, seven multiple-choice items and four open-response items across the 12 assessment components were excluded from equating. The seven reading items were modified between field- and operational-test administrations. 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 (2014–2015)

Assessment	No. of Items Excluded from Equating*		No. of Items Dropped from the Assessment
	Multiple-Choice	Open-Response	
Primary Reading (English)	0	0	0
Junior Reading (English)	1	0	0
Primary Reading (French)	4	0	0
Junior Reading (French)	2	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)	0	0	0
Primary Mathematics (French)	0	0	0
Junior Mathematics (French)	0	0	0

Note. *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 2013–2014 and 2014–2015 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 0.92 for Levels 3 and 4, -0.71 for Levels 2 and 3, -1.96 for Levels 1 and 2 and -3.10 for “not enough evidence for Level 1” (NE1) and Level 1.

Since the 2013–2014 and 2014–2015 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	
		2014	2015*		2014	2015
Number of Students		97 175			5930	6219
Level 4		14.2%			42.9%	39.2%
Level 3	0.92	63.2%		0.28	41.9%	47.7%
Level 2	-0.71	19.8%		-1.09	13.8%	12.6%
Level 1	-1.96	2.5%		-2.54	1.3%	0.5%
NE1	-3.10	0.4%		-3.62	0.1%	0.0%
% of Students at or Above the Provincial Standard		77.4%			84.8%	86.9%

Note. *Due to exceptional circumstances that caused partial participation in the English-language assessment, the 2015 results were not presented in the table to avoid misinterpretation.

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	
		2014	2015*		2014	2015
Number of Students		102 567			5154	5276
Level 4		14.6%			39.0%	28.8%
Level 3	0.97	72.4%		0.55	56.1%	66.8%
Level 2	-1.15	12.0%		-1.60	4.8%	4.5%
Level 1	-2.57	0.9%		-3.28	0.0%	0.0%
NE1	-3.66	0.0%		-4.21	0.0%	0.0%
% of Students at or Above the Provincial Standard		87.0%			95.1%	95.5%

Note. *Due to exceptional circumstances that caused partial participation in the English-language assessment, the 2015 results were not presented in the table to avoid misinterpretation.

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	
		2014	2015*		2014	2015
Number of Students		97 676			5 963	6 251
Level 4		7.1%			26.6%	19.0%
Level 3	1.43	76.3%		0.79	57.9%	64.1%
Level 2	-0.88	16.0%		-0.94	14.0%	15.6%
Level 1	-2.37	0.5%		-2.04	1.3%	1.1%
NE1	-3.13	0.1%		-2.79	0.3%	0.2%
% of Students at or Above the Provincial Standard		83.4%			84.4%	83.1%

Note. *Due to exceptional circumstances that caused partial participation in the English-language assessment, the 2015 results were not presented in the table to avoid misinterpretation.

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	
		2014	2015*		2014	2015
Number of Students		102 582			5 153	5 276
Level 4		14.5%			31.3%	20.6%
Level 3	0.79	72.0%		0.76	62.3%	69.8%
Level 2	-1.12	13.1%		-1.21	5.9%	9.1%
Level 1	-2.51	0.4%		-2.17	0.4%	0.5%
NE1	-3.42	0.1%		-3.07	0.1%	0.1%
% of Students at or Above the Provincial Standard		86.5%			93.6%	90.4%

Note. *Due to exceptional circumstances that caused partial participation in the English-language assessment, the 2015 results were not presented in the table to avoid misinterpretation.

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	
		2014	2015*		2014	2015
Number of Students		93 378			6010	6297
Level 4		14.4%			26.7%	27.9%
Level 3	1.04	59.5%		0.58	56.5%	58.6%
Level 2	-0.54	24.0%		-1.04	16.2%	13.1%
Level 1	-1.98	1.9%		-2.51	0.5%	0.5%
NE1	-2.75	0.2%		-3.20	0.1%	0.0%
% of Students at or Above the Provincial Standard		73.9%			83.2%	86.4%

Note. *Due to exceptional circumstances that caused partial participation in the English-language assessment, the 2015 results were not presented in the table to avoid misinterpretation.

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	
		2014	2015*		2014	2015
Number of Students		104 461			5227	5376
Level 4		14.7%			46.3%	51.0%
Level 3	0.94	47.1%		0.01	39.5%	39.0%
Level 2	-0.30	30.0%		-1.25	13.9%	9.7%
Level 1	-1.38	8.1%		-2.59	0.2%	0.2%
NE1	-3.01	0.1%		-3.02	0.1%	0.1%
% of Students at or Above the Provincial Standard		61.8%			87.9%	90.0%

Note. *Due to exceptional circumstances that caused partial participation in the English-language assessment, the 2015 results were not presented in the table to avoid misinterpretation.

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 2013–2014 and 2014–2015 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

- a) did not attend a publicly funded school;
- b) were home-schooled;
- c) completed no work on an item (by booklet and item type);
- d) received accommodations, except for English language learners who received the accommodation for setting, and
- e) 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 2013–2014 and the 2014–2015 assessments.

Table 5.9 Number of Grade 9 Students in the Equating Samples

Version	2013–2014				2014–2015	
	English		French		English	French
	Calibration Sample	Equating Sample	Calibration Sample	Equating Sample	Calibration and Equating Sample	Calibration and Equating Sample
Academic	82 505	93 435	3 594	3 948	85 728	3 857
Applied	27 422	32 127	1 297	1 298	28 378	1 084

Equating Steps

The calibration and equating of the 2013–2014 and 2014–2015 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 for the 2014–2015 winter and spring samples.
2. Calibration (concurrent) and equating were conducted for the 2013–2014 winter and spring samples (including the field-test items that were brought forward to the 2014–2015 operational test). In this calibration, the item parameter estimates of the field-test items that were brought forward to the 2014–2015 operational tests were fixed at the values obtained from the 2014–2015 calibration runs (Step 1). The parameter estimates of the 2013–2014 operational items that were repeated on the 2014–2015 tests were also fixed at the values obtained from the 2014–2015 calibration runs (Step 1). The 2013–2014 equating samples (operational items only) were scored using the scaled 2013–2014 operational-item parameter estimates.
3. The percentage of students in each achievement level was determined for the 2013–2014 equating sample using the levels assigned in 2013–2014. 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 2014–2015 assessments.
5. The parameter estimates for the 2014–2015 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 2014–2015 student theta values to classify students to achievement levels, and the percentage of students at each performance level was determined.

Eliminating Items and the Collapsing of Score Categories

For the Grade 9 mathematics assessments, no 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)	0	0	0
Academic, Winter (French)	0	0	0
Academic, Spring (French)	0	0	0

Equating Results

The equating results for the applied version of the Grade 9 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 2013–2014 and 2014–2015 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 1.16 for Levels 3 and 4, -0.03 for Levels 2 and 3, -1.02 for Levels 1 and 2, and -1.74 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	
		2014	2015*		2014	2015
Number of Students		32 122			1294	1084
Level 4		9.7%			9.3%	8.2%
Level 3	1.16	41.2%		1.33	45.0%	42.9%
Level 2	-0.03	35.3%		-0.02	37.0%	38.8%
Level 1	-1.02	10.9%		-1.25	7.4%	8.7%
Below Level 1	-1.74	2.8%		-1.91	1.3%	1.4%
% of Students at or Above the Provincial Standard		50.9%			54.3%	51.1%

Note. *Due to exceptional circumstances that caused partial participation in the assessment in the English-language assessment, the 2015 results were not presented in the table to avoid misinterpretation.

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	
		2014	2015*		2014	2015
Number of Students		93 431			3947	3857
Level 4		12.2%			7.3%	6.4%
Level 3	1.11	73.7%		1.44	77.0%	76.8%
Level 2	-1.00	10.5%		-0.95	11.9%	12.6%
Level 1	-1.60	3.5%		-1.62	3.8%	4.1%
Below Level 1	-2.65	0.1%		-2.65	0.0%	0.0%
% of Students at or Above the Provincial Standard		85.8%			84.3%	83.2%

Note. *Due to exceptional circumstances that caused partial participation in the assessment in the English-language assessment, the 2015 results were not presented in the table to avoid misinterpretation.

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 2013–2014 and 2014–2015 equating samples. The following categories of students were excluded from the equating samples:

- students with no work or incomplete work on a major section of the test;
- 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;
- previously eligible students;
- students who were exempted, deferred or taking the Ontario Secondary School Literacy Course (OSSLC) and
- students who were home-schooled.

Table 5.13 presents the number of first-time eligible students in the OSSLT equating samples for the 2013–2014 and 2014–2015 tests.

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

OSSLT	2014		2015	
	English	French	English	French
First-Time Eligible	125 584	3 684	123 427	3 778

Equating Steps

The following steps were implemented to calibrate and equate the 2014 and 2015 OSSLT:

1. The parameter estimates of the operational items administered in 2015 were calibrated using the 2015 equating sample.
2. The operational items that formed the 2014 test and the field-test items brought forward to the 2015 test were recalibrated using the 2014 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 2014 test, obtained in Step 2, were used to score the 2014 equating sample data.
4. The percentage of successful students was determined for the 2014 equating sample from the student results reported in 2014. The theta-value cut point that replicated this percentage was identified for the distribution of scores in the 2014 equating sample.
5. This theta value was applied to student scores for the 2015 assessment to determine which students would be successful. The results are presented in Table 5.14.

Scale Score

The reporting scale scores for the 2015 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 and two multiple-choice French-language items were excluded from the equating process due to modifications made to the items between the 2014 field-test and the 2015 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 2014 and 2015 are reported for English-language and French-language students. For example, the equated cut score for the English-language test was -0.71 . The percentage of successful students in the equating samples was 85.6% in 2014 and 84.7% in 2015.

Table 5.14 Equating Results for the OSSLT

	English-Language			French-Language		
	Theta Cut Point 2014–2015	Equating Sample 2014	Equating Sample 2015	Theta Cut Point 2014–2015	Equating Sample 2014	Equating Sample 2015
No. of Students		125 584	123 427		3 684	3 778
% Successful	-0.71	85.6%	84.7%	-1.00	90.2%	90.1%
Unsuccessful		14.4%	15.3%		9.8%	9.9%

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 typically publishes annual provincial reports for education stakeholders and the general public. The report for the 2014–2015 English-language OSSLT assessment is available at www.eqao.com, as well as the corresponding report for the French-language assessments:

- *EQAO's Provincial Report on the Results of the 2014–2015 Ontario Secondary School Literacy Test.*

Provincial results for EQAO's 2014–2015 primary- and junior-division, and Grade 9 mathematics assessments for the English-language school system are not available, because not all schools in that system participated in the assessments due to labour disruptions. However, provincial-level reports for the French-language assessments are available.

EQAO posts school and board results at 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 and boards (where data is available):

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 and board (where data is available).

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 reports can be generated in EQAO Reporting, EQAO has phased out the secure detailed reports.

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.

Where data is available, results for the teacher and principal questionnaires are reported at the board level only for the English-language school system. For the French-language school system, results for the teacher and principal questionnaires are reported at both the provincial and the board levels.

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 and board. 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 and board (where data is available).

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 (the provincial reports are only available for the French-language school system), 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 for the English-language school system, and cohort tracking results from Grade 3 to Grade 6 to Grade 9 for the French-language school system;
- 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 document: “Guide to Using the Item Information Report: Student Roster.”

CHAPTER 7: STATISTICAL AND PSYCHOMETRIC SUMMARIES

A variety of statistical and psychometric analyses were conducted for the 2014–2015 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 53.8% (for the reading component of the French-language primary-division assessment) to 69.8% (for the 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 0.86 to 0.87 for reading, 0.81 to 0.82 for writing and 0.88 to 0.90 for mathematics. The corresponding standard errors of measurement range from 4.4% to 4.8% of the possible maximum score for reading, 6.6% to 7.2% of the possible maximum score for writing, and 5.8% 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)	36	26	10	66	34 971	1	62	36.25	8.93	0.87	3.18
Junior Reading (English)	36	26	10	66	37 020	1	64	39.40	8.45	0.86	3.12
Primary Reading (French)	36	26	10	66	8 107	3	62	35.49	8.45	0.87	3.07
Junior Reading (French)	36	26	10	66	6 903	6	60	39.99	8.16	0.87	2.91
Primary Writing (English)	14	8	6*	29	35 216	0	29	17.75	4.93	0.82	2.09
Junior Writing (English)	14	8	6*	29	37 192	0	29	18.36	4.61	0.81	2.01
Primary Writing (French)	14	8	6*	29	8 122	0	29	16.73	4.50	0.82	1.91
Junior Writing (French)	14	8	6*	29	6 923	1	29	17.00	4.49	0.82	1.90
Primary Mathematics (English)	36	28	8	60	35 689	1	60	40.33	10.76	0.89	3.57
Junior Mathematics (English)	36	28	8	60	36 977	0	60	38.12	11.41	0.90	3.61
Primary Mathematics (French)	36	28	8	60	8 124	3	60	41.88	10.04	0.88	3.48
Junior Mathematics (French)	36	28	8	60	6 912	2	60	38.11	12.11	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)	34 971	-3.87	3.21	-0.12	0.99
Junior Reading (English)	37 020	-3.92	3.32	-0.18	1.02
Primary Reading (French)	8 107	-3.80	3.39	-0.10	0.97
Junior Reading (French)	6 903	-3.94	2.75	-0.15	1.01
Primary Writing (English)	35 216	-3.58	2.50	-0.08	0.94
Junior Writing (English)	37 192	-3.76	2.43	-0.16	0.98
Primary Writing (French)	8 122	-3.50	2.87	-0.06	0.94
Junior Writing (French)	6 923	-3.84	2.80	-0.11	0.97
Primary Mathematics (English)	35 862	-3.83	2.45	-0.13	1.00
Junior Mathematics (English)	37 066	-3.79	2.53	-0.20	1.04
Primary Mathematics (French)	8 127	-3.75	2.39	-0.10	1.00
Junior Mathematics (French)	6 922	-3.64	2.47	-0.30	1.07

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 45%. 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 0.10. The precision of the scores is greatest at this point. The theta cut points used to assign students to achievement levels are also marked to show the amount of information at each cut point.

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

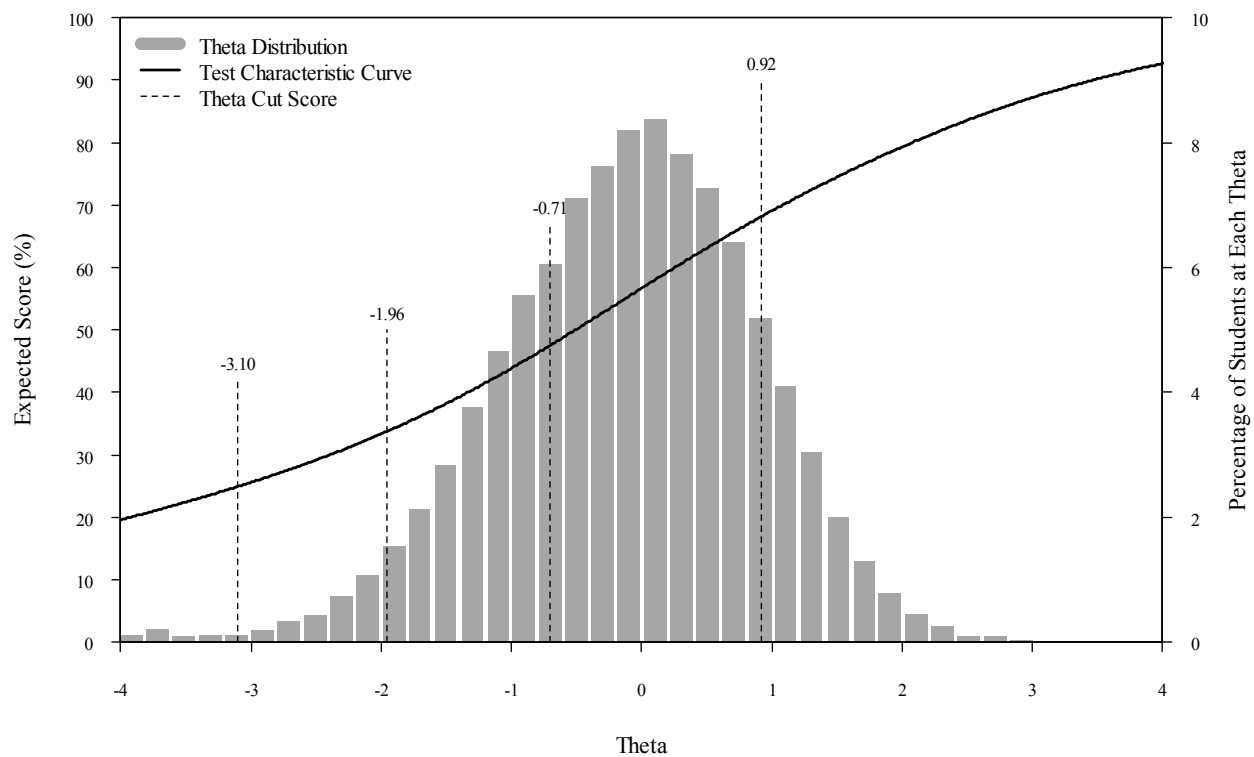


Figure 7.2 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (English)

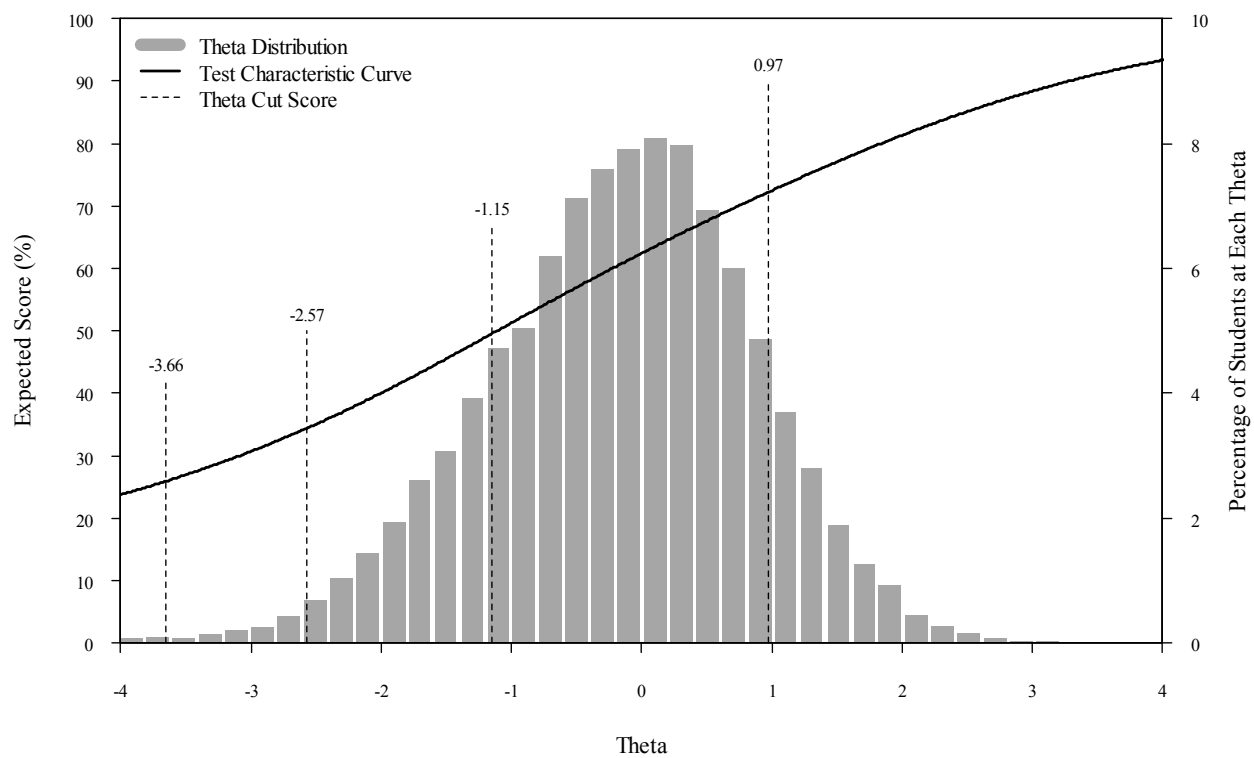


Figure 7.3 Test Characteristic Curve and Distribution of Thetas for Reading: Primary Division (French)

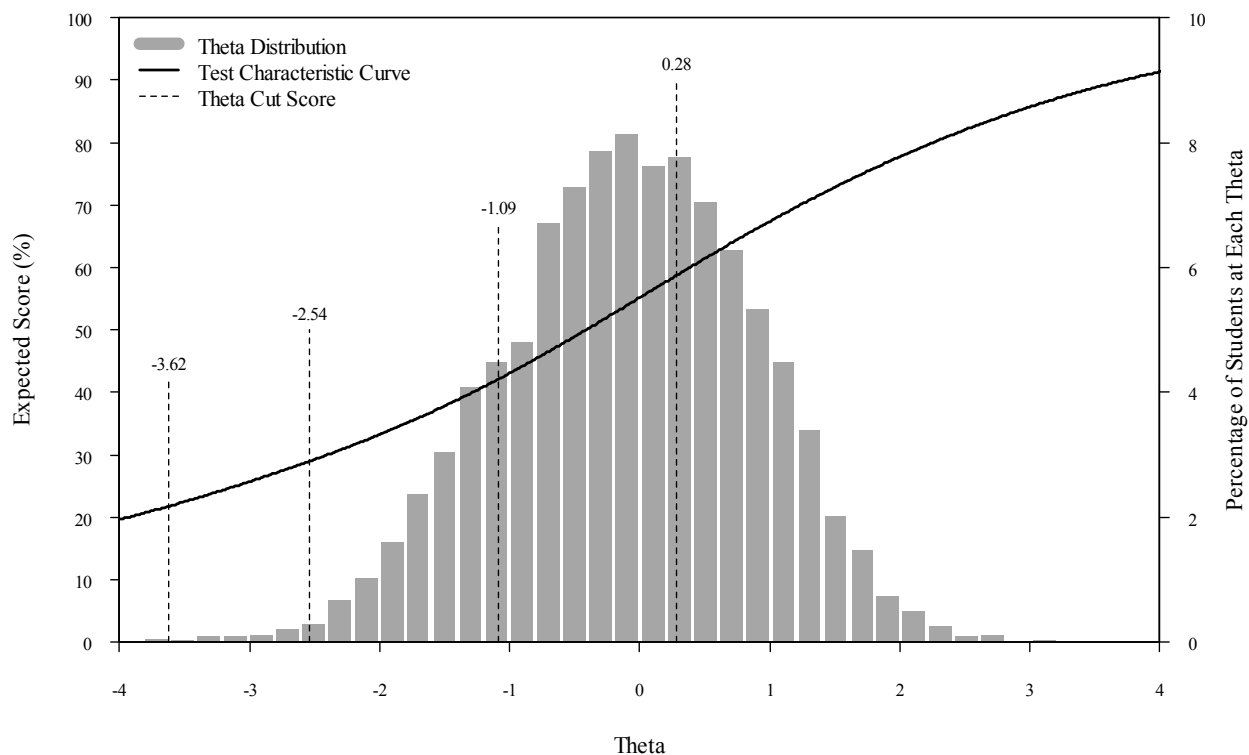


Figure 7.4 Test Characteristic Curve and Distribution of Thetas for Reading: Junior Division (French)

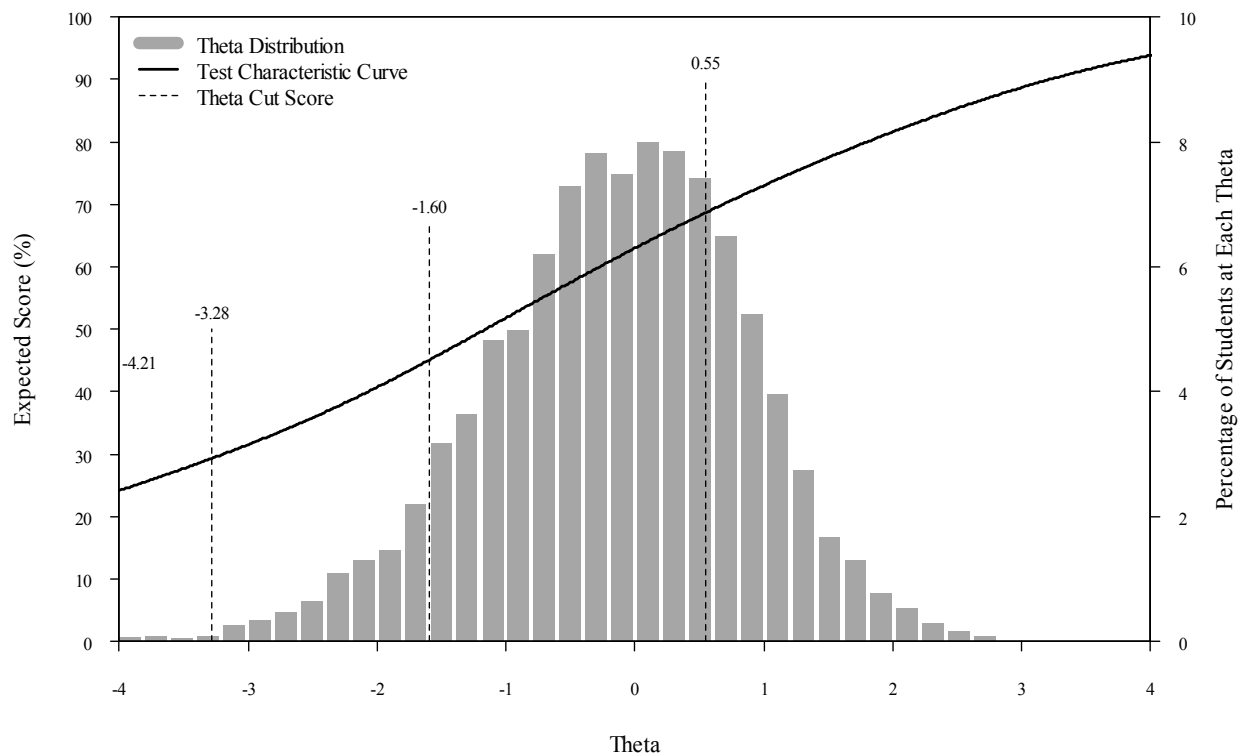


Figure 7.5 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (English)

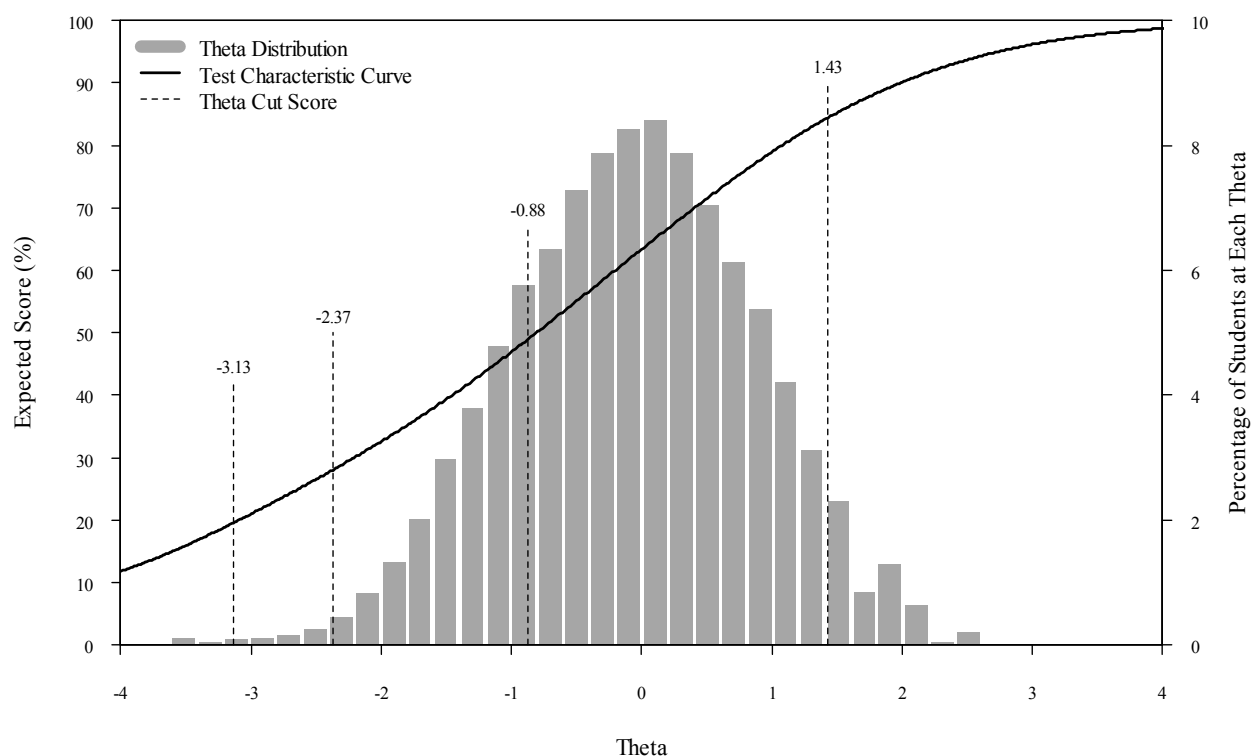


Figure 7.6 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (English)

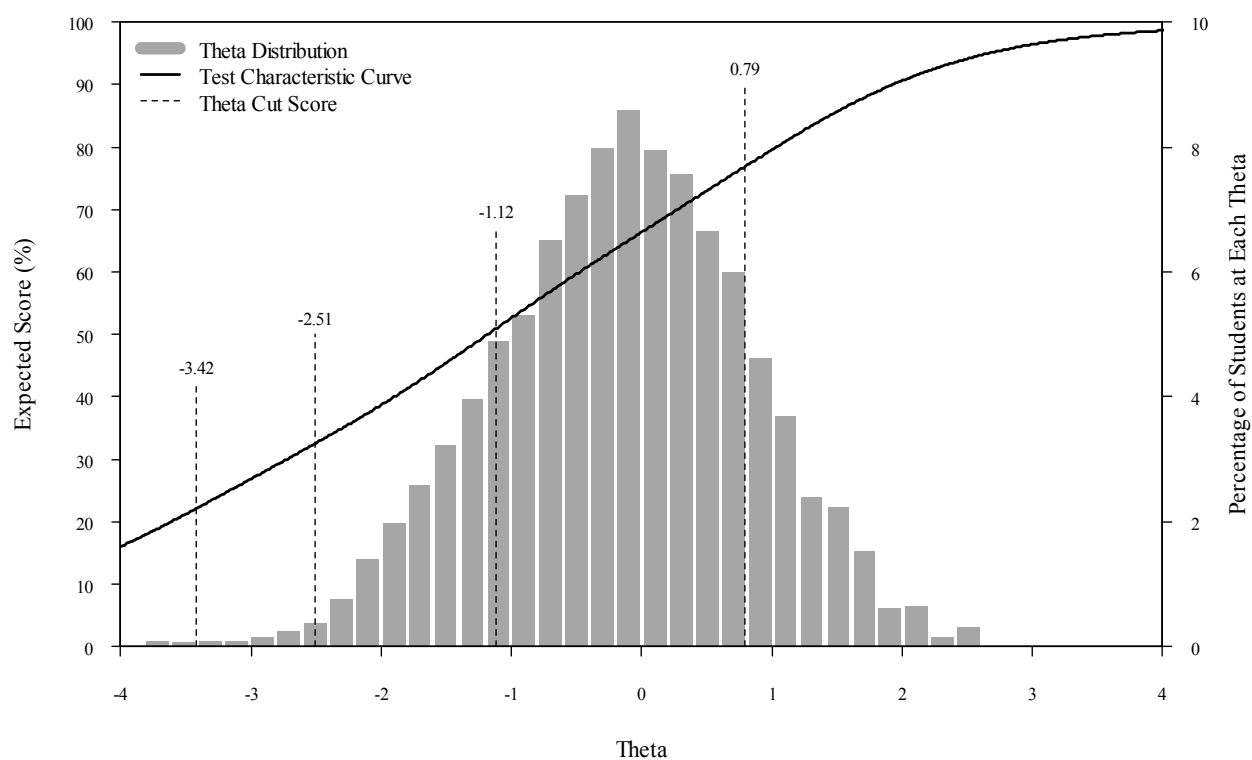


Figure 7.7 Test Characteristic Curve and Distribution of Thetas for Writing: Primary Division (French)

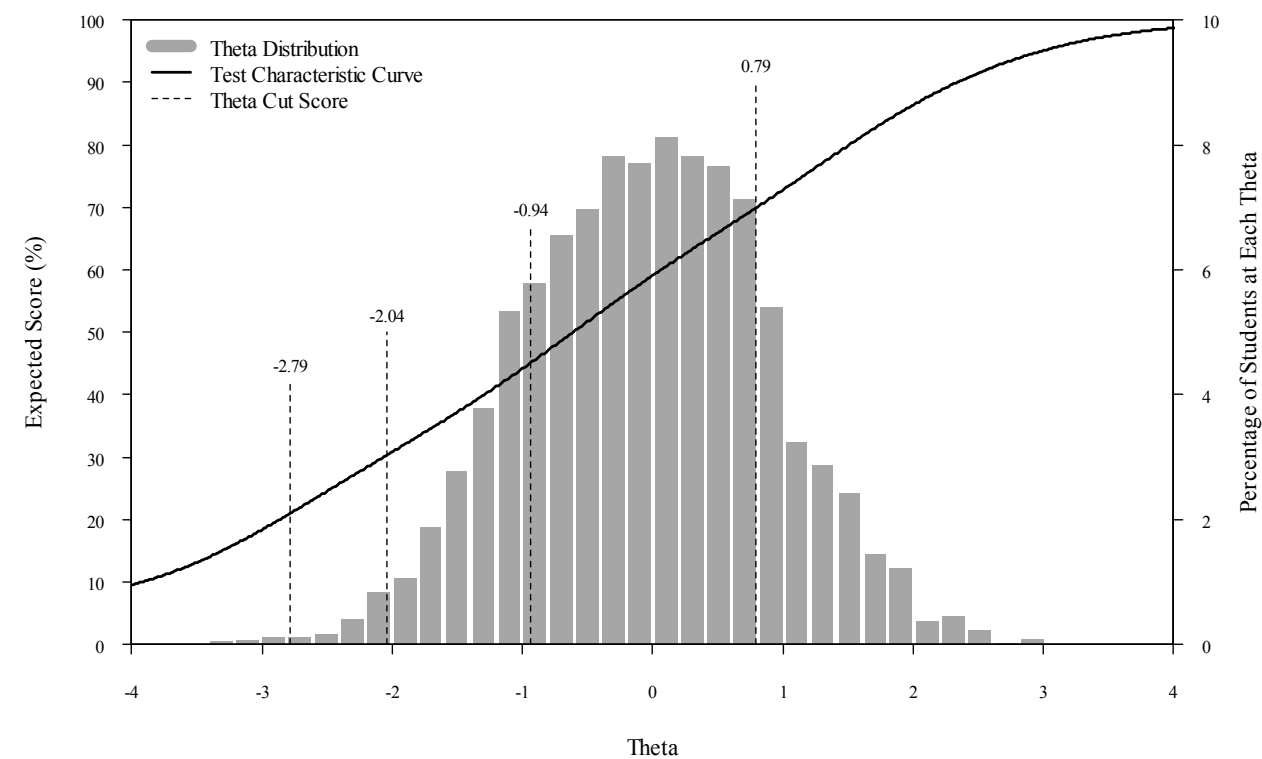


Figure 7.8 Test Characteristic Curve and Distribution of Thetas for Writing: Junior Division (French)

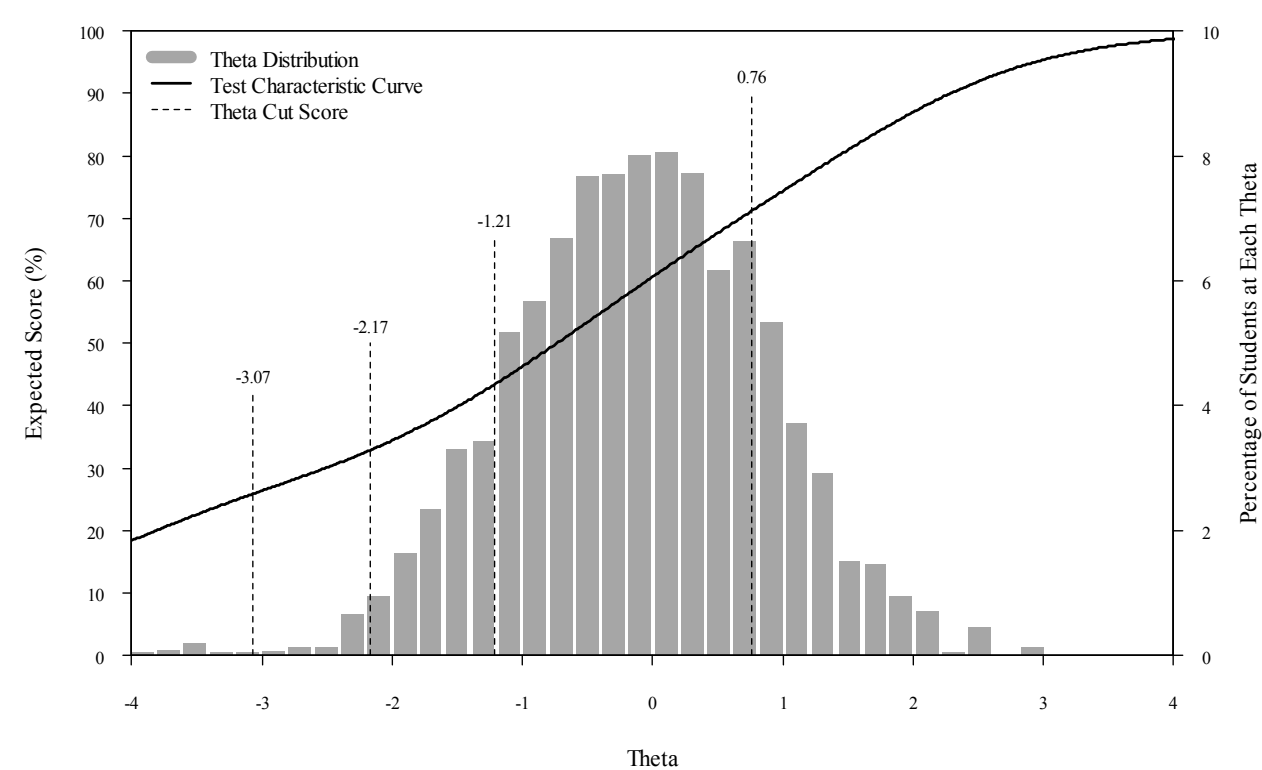


Figure 7.9 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (English)

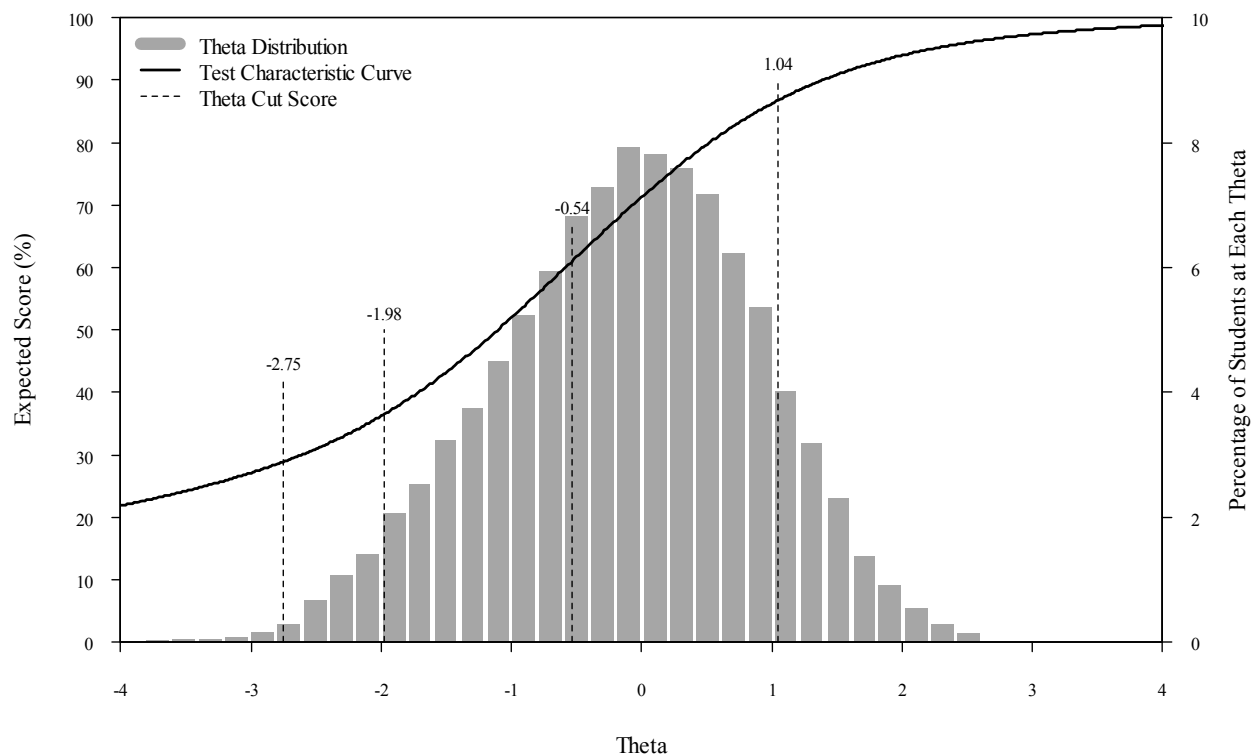


Figure 7.10 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (English)

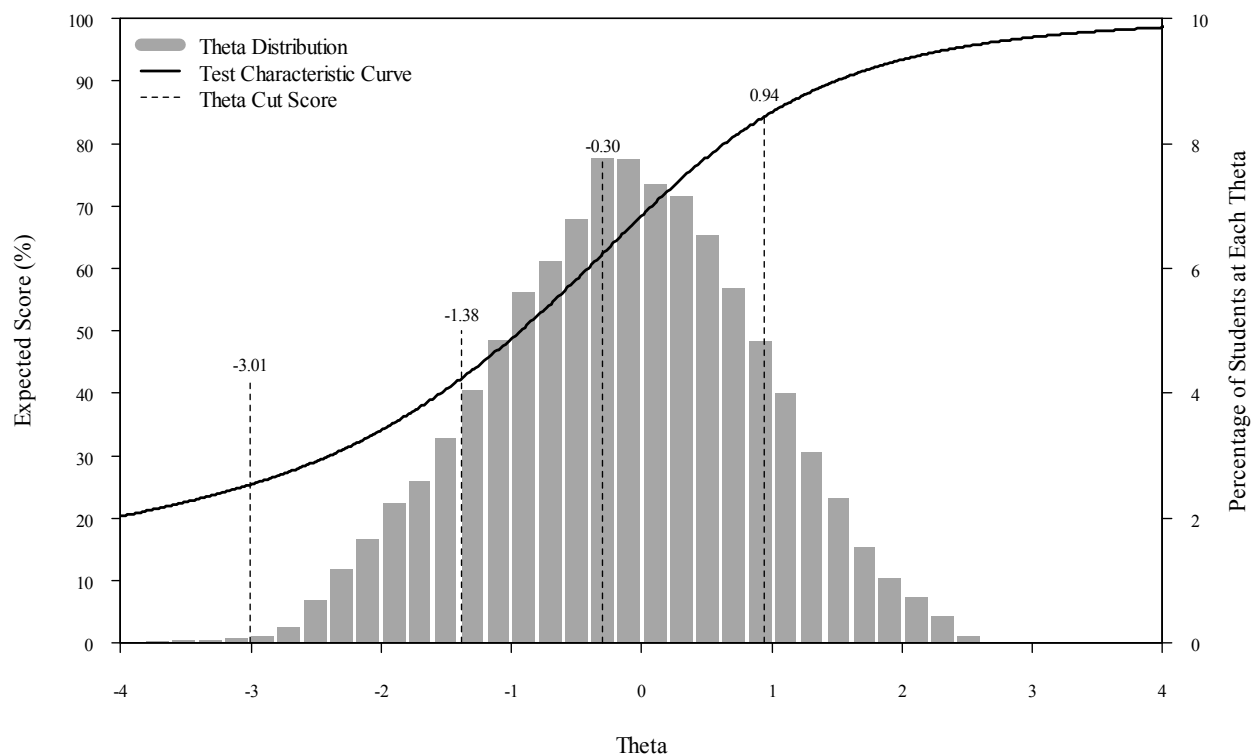


Figure 7.11 Test Characteristic Curve and Distribution of Thetas for Mathematics: Primary Division (French)

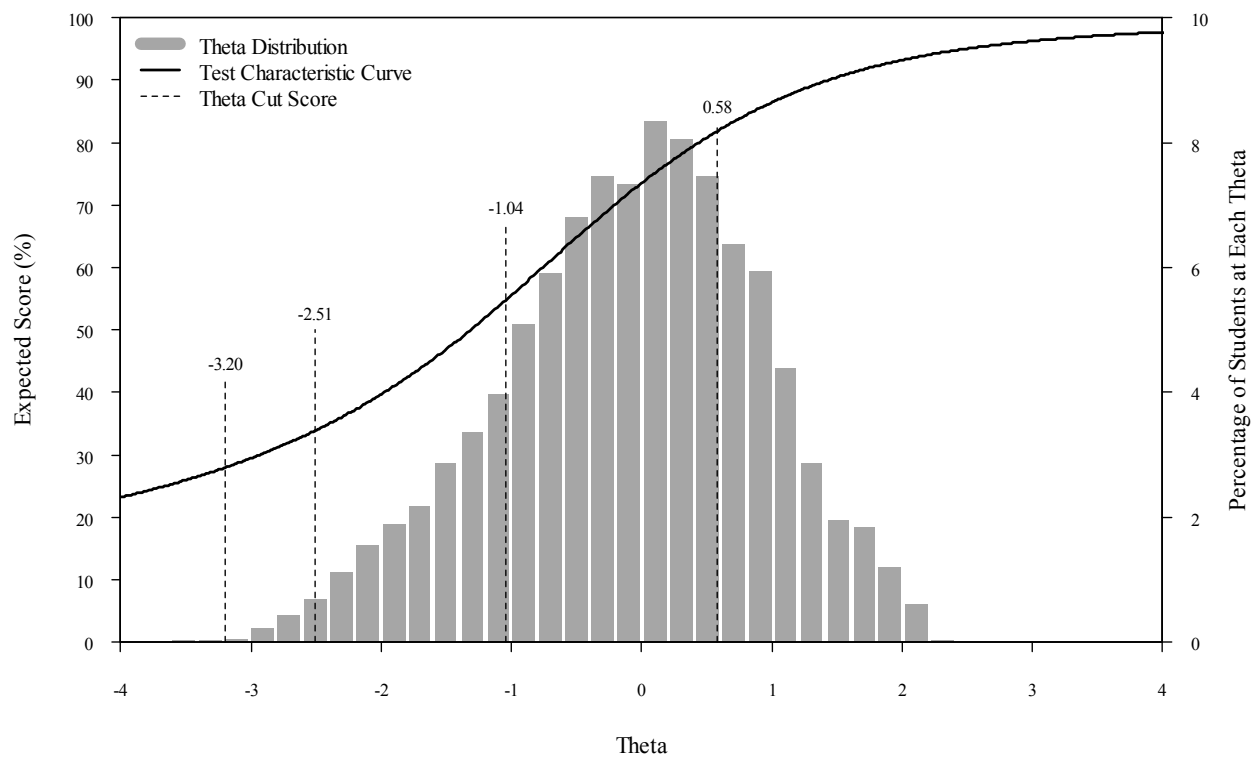


Figure 7.12 Test Characteristic Curve and Distribution of Thetas for Mathematics: Junior Division (French)

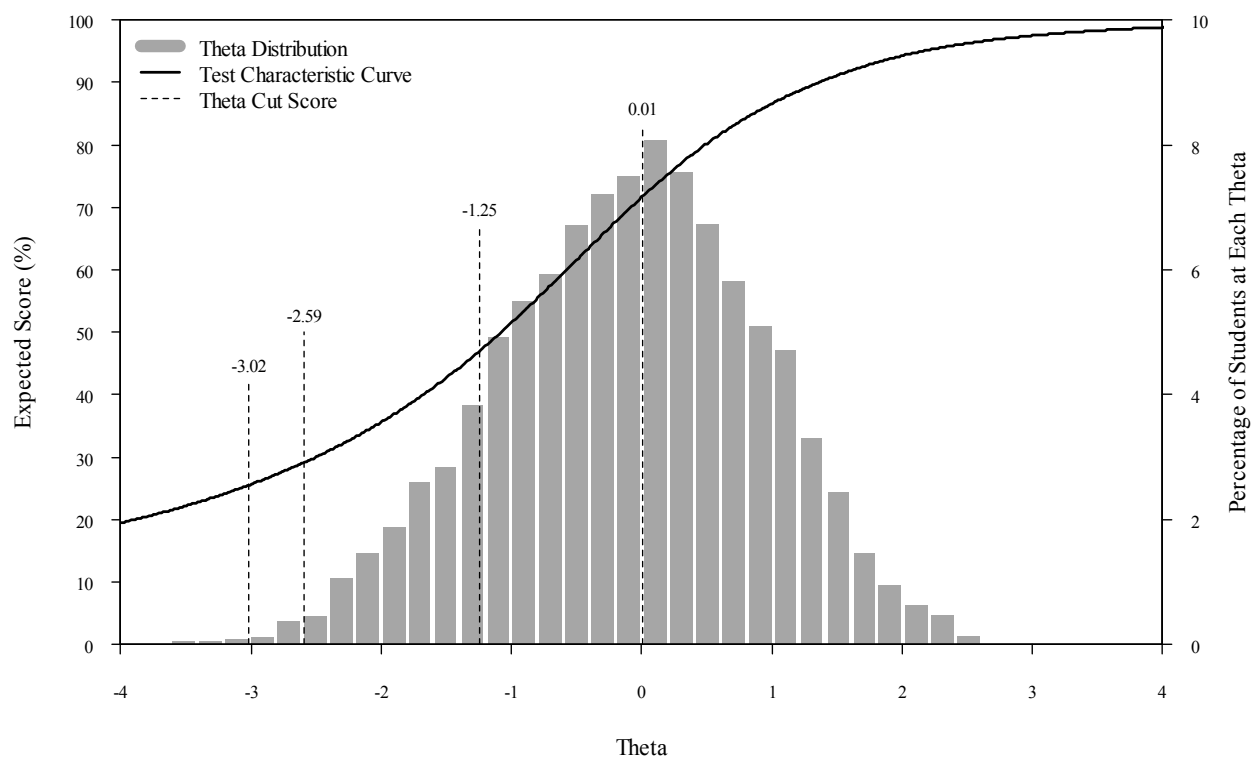


Figure 7.13 Test Information Function for Reading: Primary Division (English)

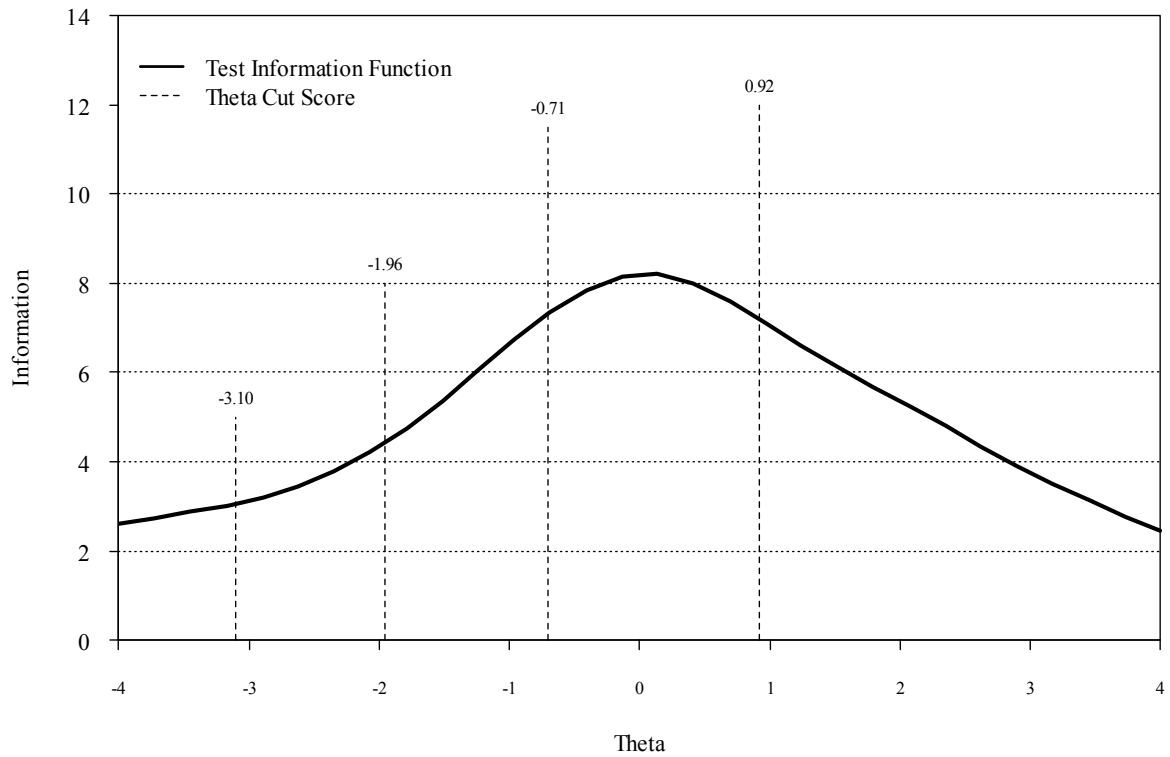


Figure 7.14 Test Information Function for Reading: Junior Division (English)

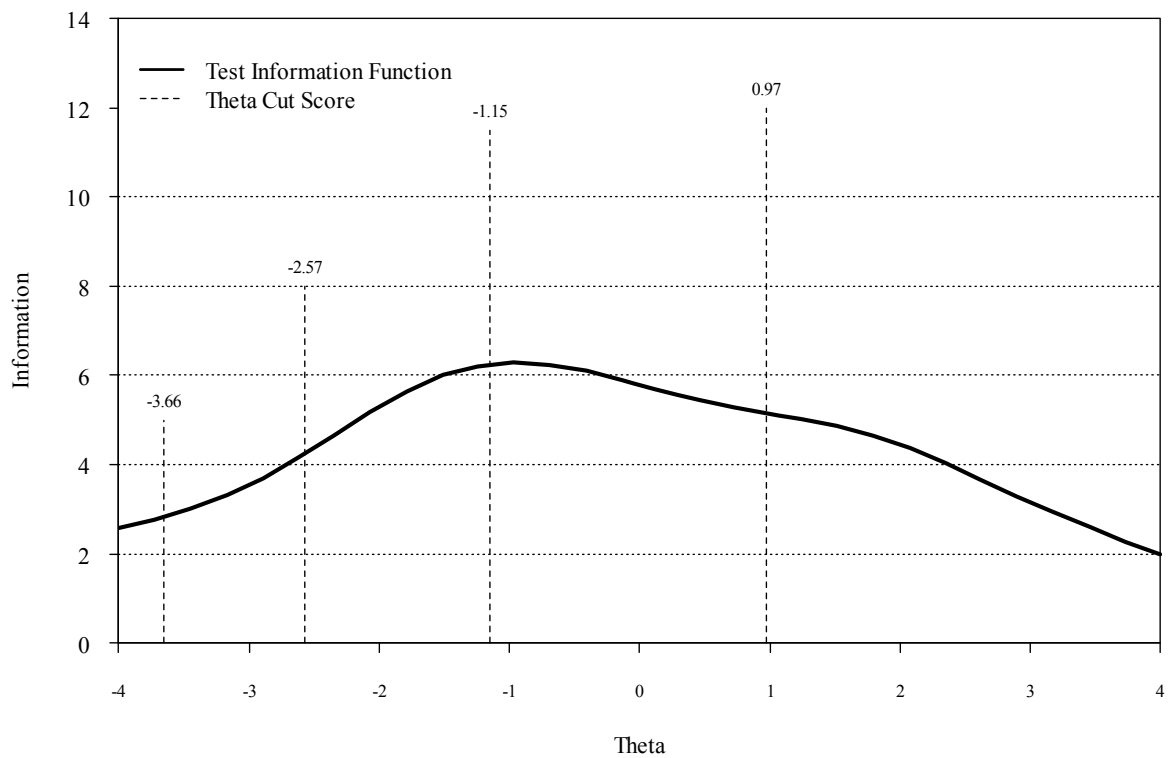


Figure 7.15 Test Information Function for Reading: Primary Division (French)

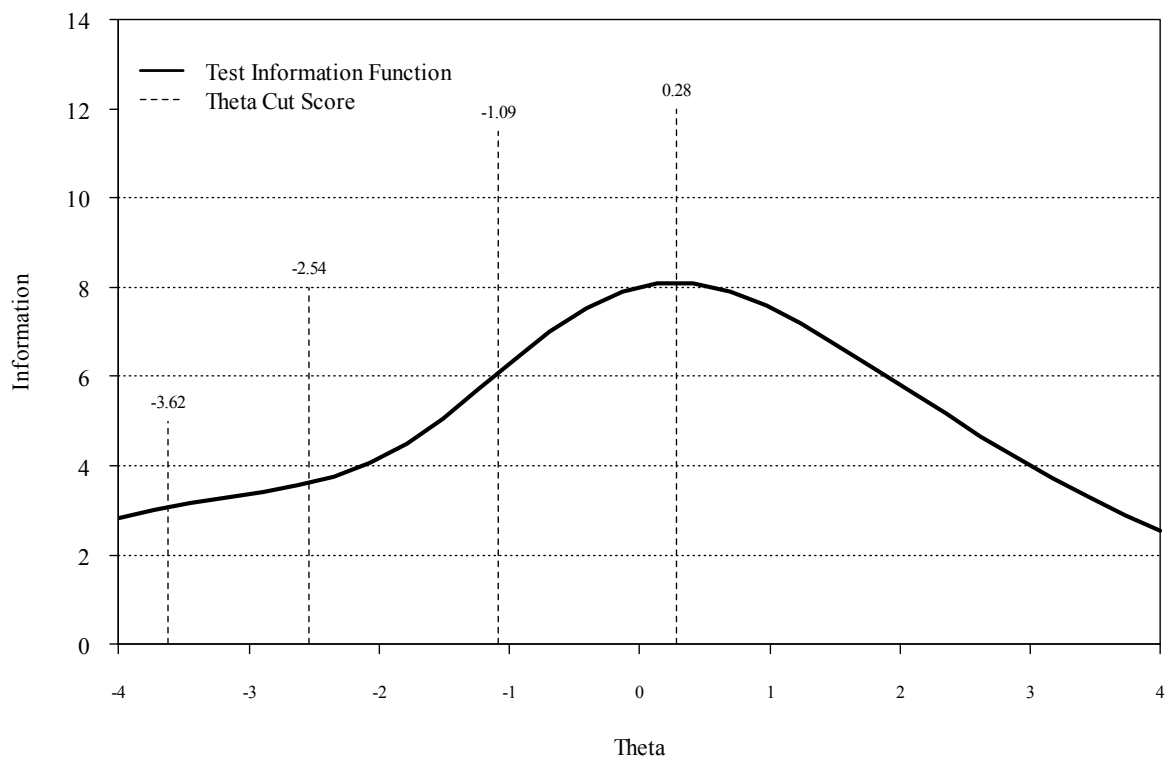


Figure 7.16 Test Information Function for Reading: Junior Division (French)

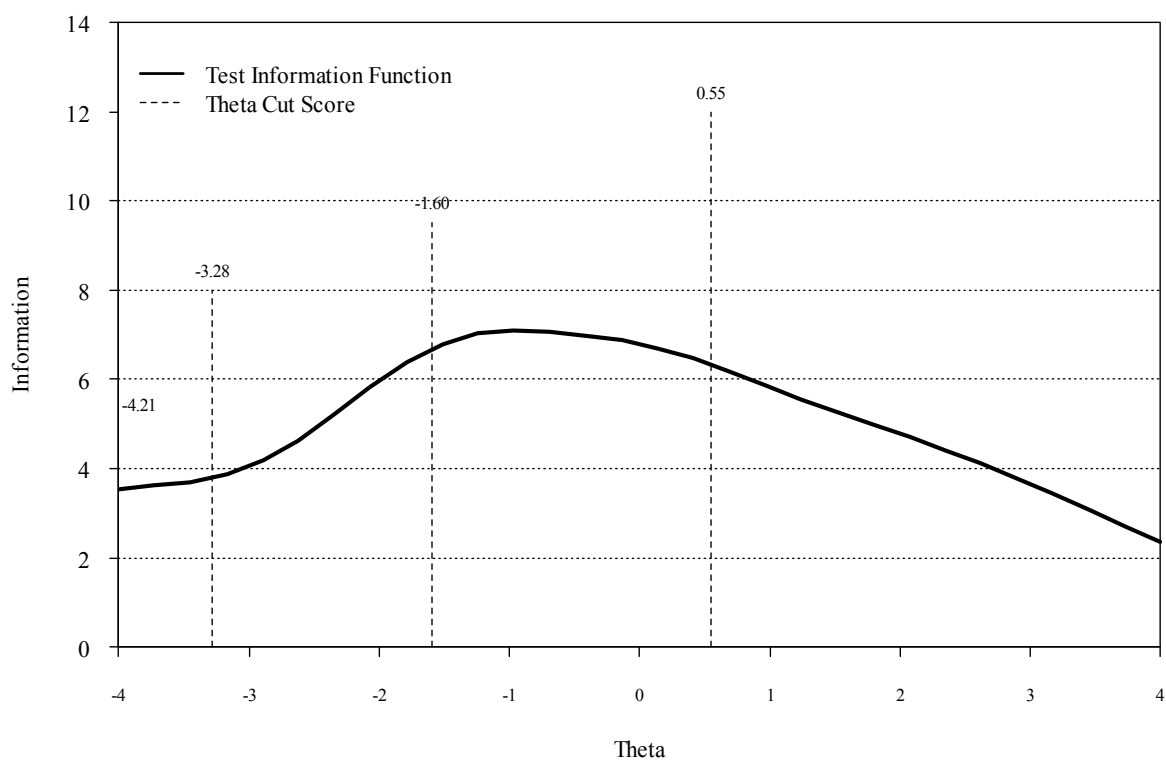


Figure 7.17 Test Information Function for Writing: Primary Division (English)

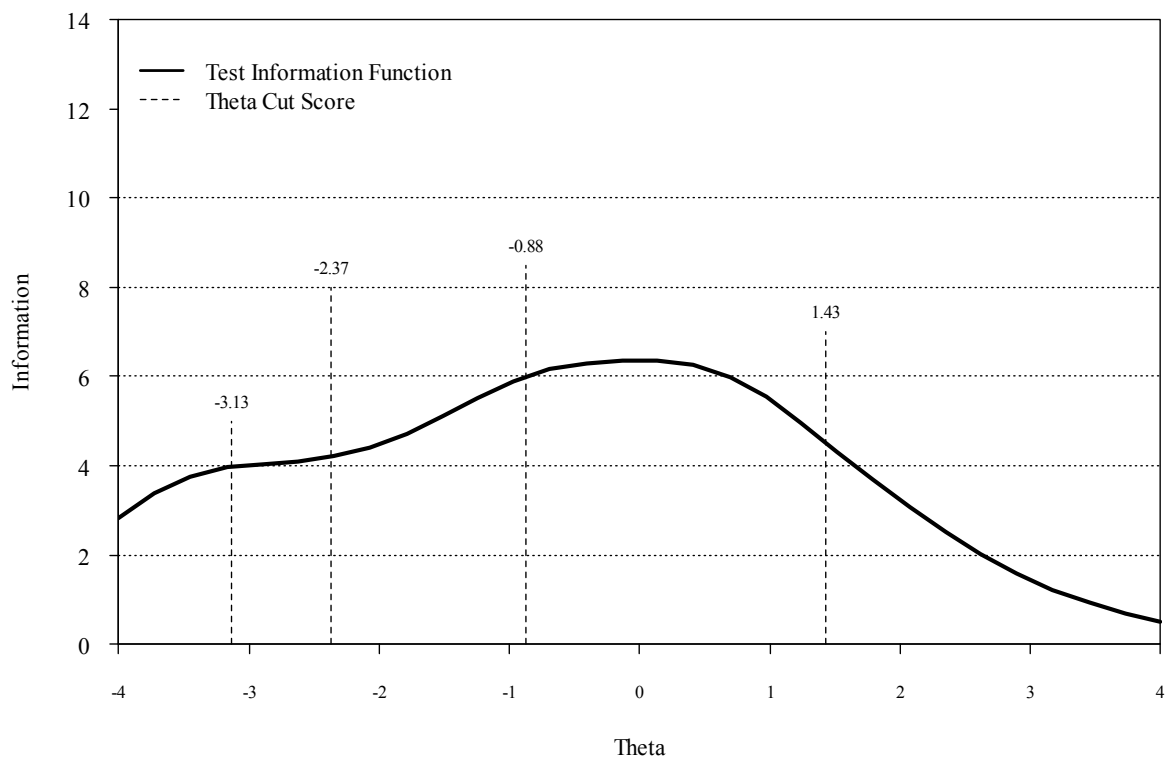


Figure 7.18 Test Information Function for Writing: Junior Division (English)

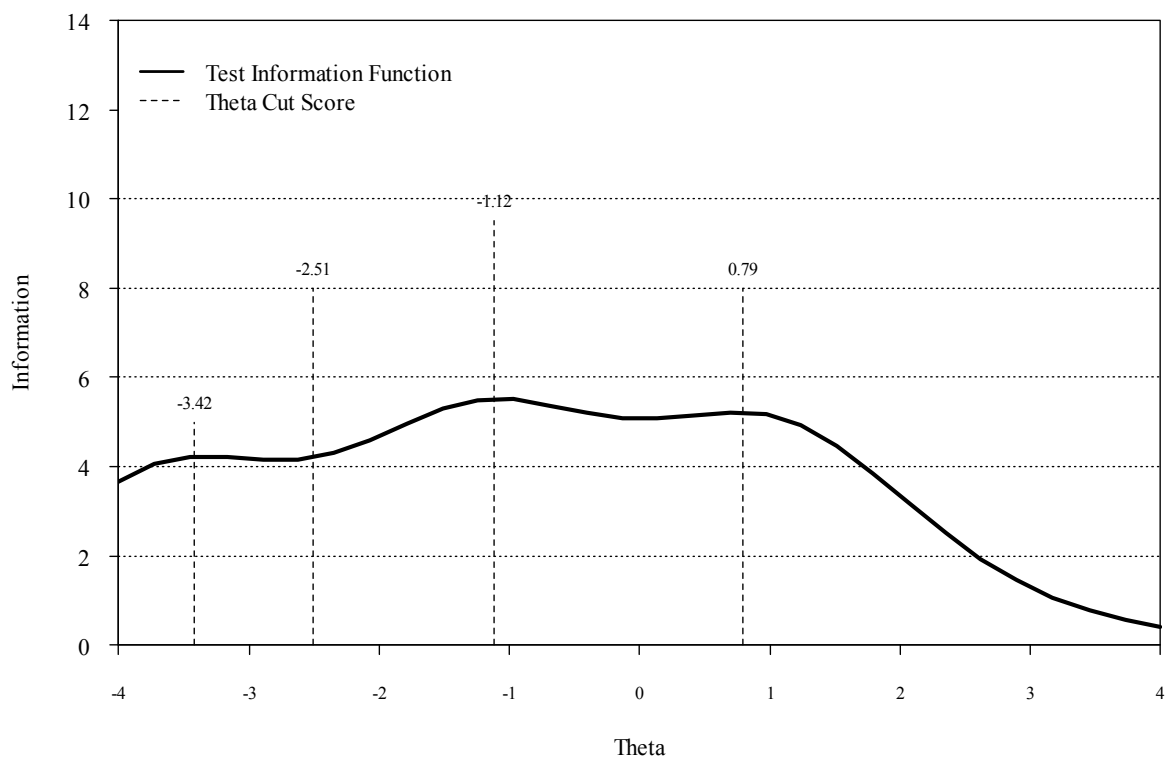


Figure 7.19 Test Information Function for Writing: Primary Division (French)

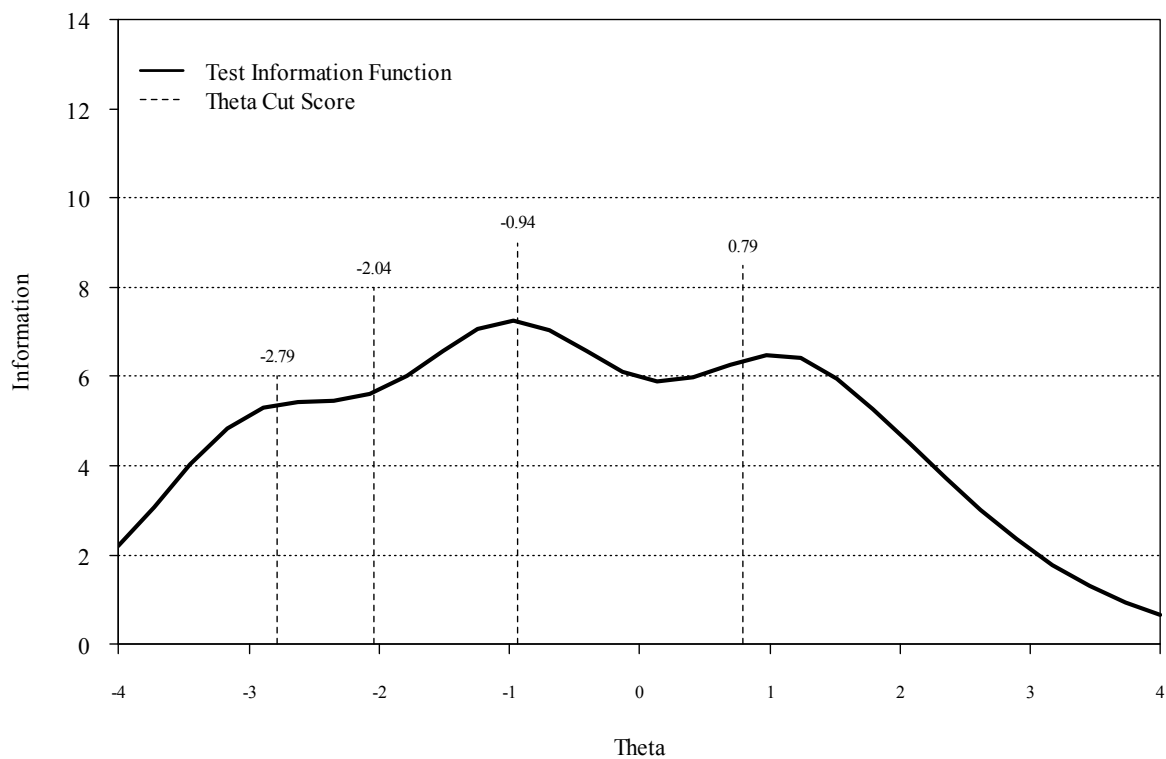


Figure 7.20 Test Information Function for Writing: Junior Division (French)

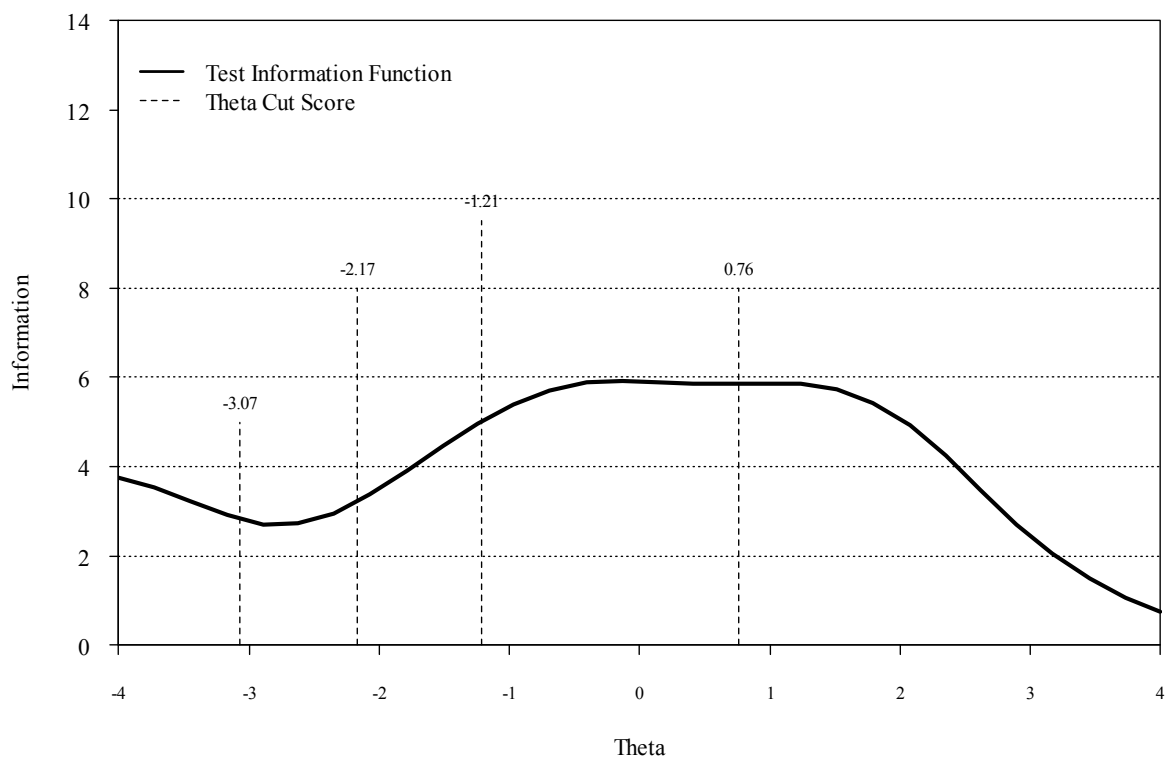


Figure 7.21 Test Information Function for Mathematics: Primary Division (English)

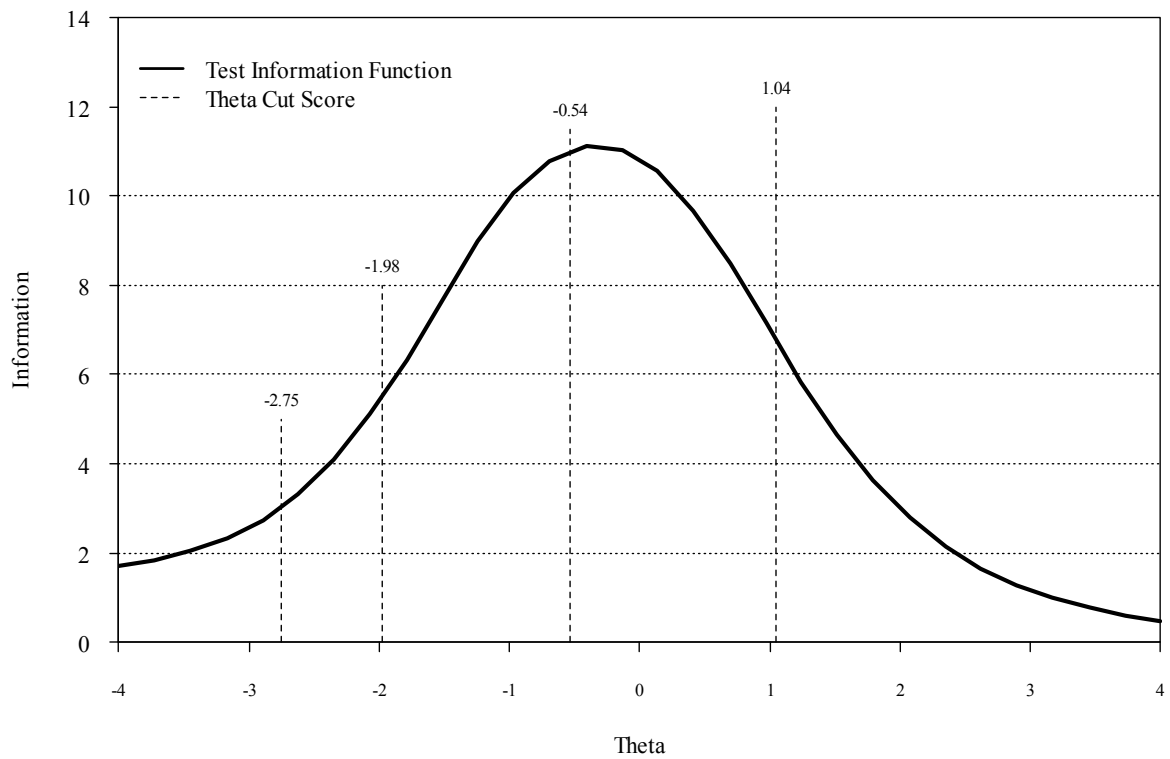


Figure 7.22 Test Information Function for Mathematics: Junior Division (English)

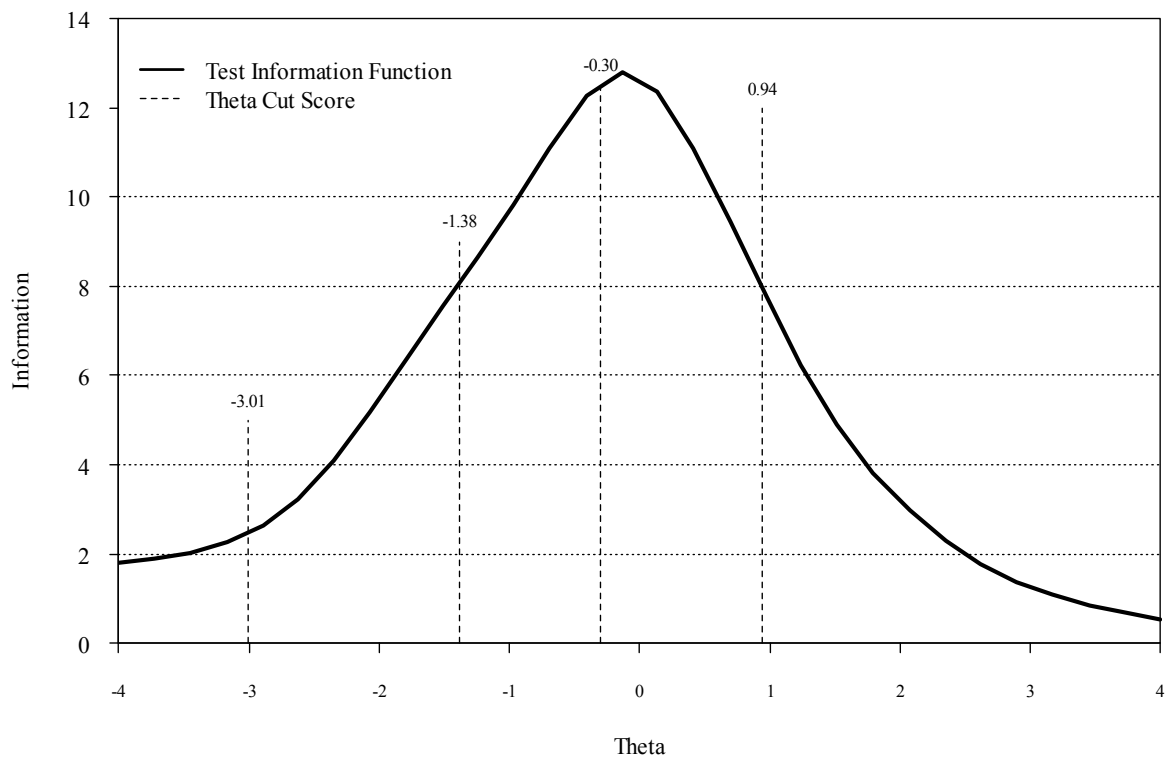


Figure 7.23 Test Information Function for Mathematics: Primary Division (French)

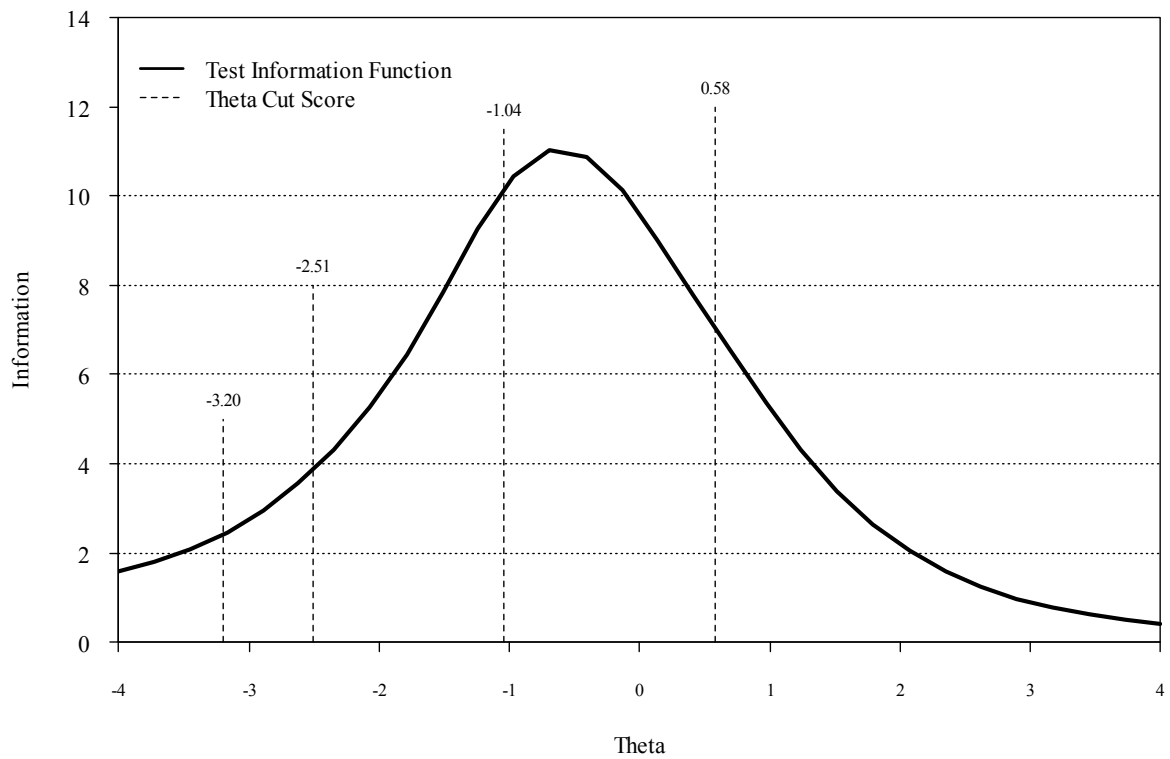
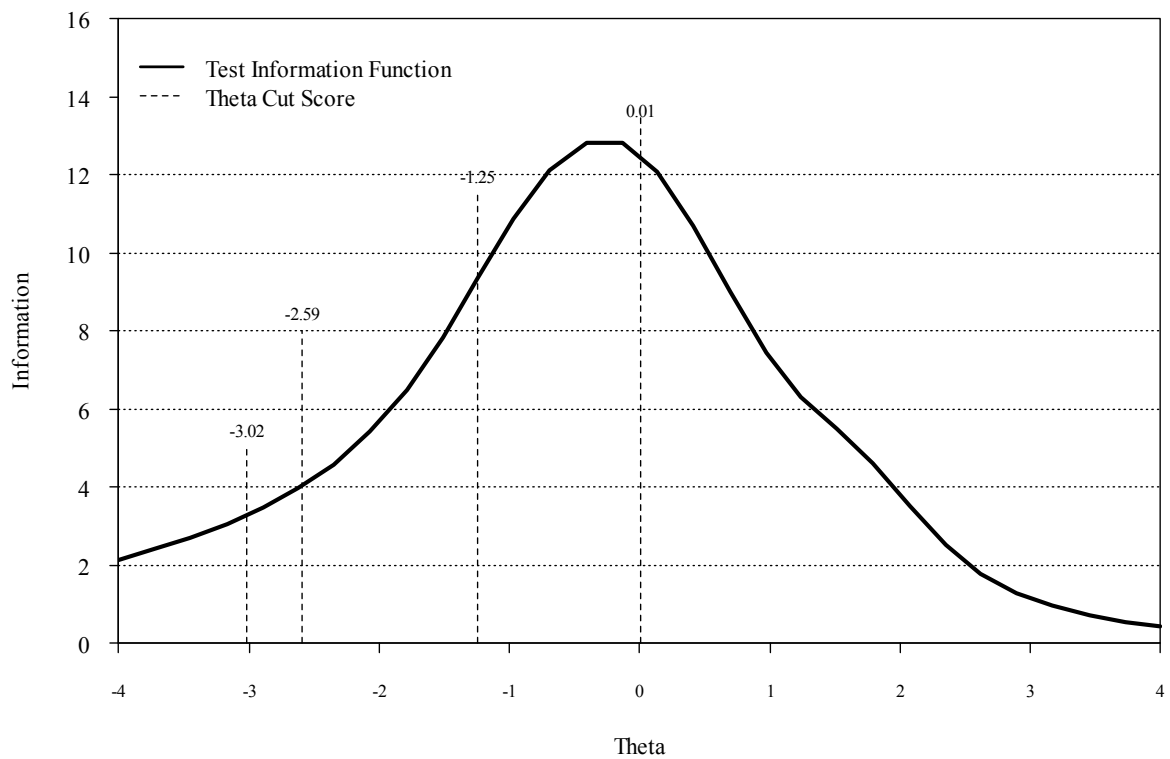


Figure 7.24 Test Information Function for Mathematics: Junior Division (French)



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

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

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

Assessment	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*	
			Item Difficulty	Item-Total Correlation	Slope	Location
Primary Reading (English)	36	Min.	0.36	0.13	0.21	-1.75
		Max.	0.85	0.49	1.19	1.25
		Mean	0.62	0.37 [†]	0.60	-0.37
		SD	0.13	0.09	0.20	0.65
Junior Reading (English)	36	Min.	0.29	0.16	0.31	-2.87
		Max.	0.89	0.47	0.99	1.58
		Mean	0.67	0.33	0.56	-0.82
		SD	0.16	0.08	0.18	0.99
Primary Reading (French)	36	Min.	0.42	0.04	0.06	-1.40
		Max.	0.84	0.55	1.06	0.97
		Mean	0.58	0.37	0.63	-0.16
		SD	0.12	0.09	0.20	0.64
Junior Reading (French)	36	Min.	0.37	0.19	0.31	-2.47
		Max.	0.96	0.53	1.10	1.54
		Mean	0.68	0.36	0.63	-0.88
		SD	0.15	0.09	0.20	0.88

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.	0.42	0.28	0.47	-1.67
		Max.	0.82	0.61	1.07	0.78
		Mean	0.63	0.46 [†]	0.73	-0.64
		SD	0.09	0.12	0.17	0.62
Junior Writing (English)	14	Min.	0.46	0.24	0.29	-2.93
		Max.	0.94	0.61	1.08	0.62
		Mean	0.67	0.45	0.70	-1.05
		SD	0.12	0.18	0.23	0.81
Primary Writing (French)	14	Min.	0.43	0.27	0.51	-2.23
		Max.	0.91	0.62	1.20	0.58
		Mean	0.65	0.46	0.82	-0.77
		SD	0.16	0.14	0.24	0.88
Junior Writing (French)	14	Min.	0.49	0.30	0.45	-1.77
		Max.	0.87	0.62	1.05	0.27
		Mean	0.63	0.46	0.79	-0.79
		SD	0.11	0.12	0.19	0.54
Primary Mathematics (English)	36	Min.	0.38	0.24	0.38	-2.54
		Max.	0.93	0.66	1.13	0.91
		Mean	0.67	0.41	0.70	-0.87
		SD	0.12	0.13	0.21	0.87
Junior Mathematics (English)	36	Min.	0.35	0.23	0.35	-2.73
		Max.	0.89	0.66	1.55	1.24
		Mean	0.65	0.42	0.74	-0.72
		SD	0.12	0.14	0.26	0.83
Primary Mathematics (French)	36	Min.	0.47	0.27	0.20	-3.93
		Max.	0.94	0.54	1.21	0.51
		Mean	0.73	0.41	0.72	-1.16
		SD	0.11	0.09	0.25	0.91
Junior Mathematics (French)	36	Min.	0.19	0.17	0.25	-3.62
		Max.	0.83	0.68	1.66	1.63
		Mean	0.66	0.42	0.76	-0.76
		SD	0.13	0.15	0.30	0.98

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 2014–2015 the mean percentages ranged from 56.19% (English, spring) to 61.96% (French, spring) for applied mathematics. The mean percentages ranged from 65.81% (French, spring) to 71.06% (French, winter) for academic mathematics.

Cronbach's alpha estimates ranged from 0.85 to 0.86 for applied mathematics and from 0.86 to 0.88 for academic mathematics. The corresponding SEMs were likewise similar, ranging from 6.58% to 7.00% of the possible maximum score for applied mathematics and from 6.33% to 6.85% 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	16 210	0	52	29.60	9.40	0.85	3.62
Applied, Spring (English)	31	24	7	52	16 491	1	52	29.22	9.52	0.85	3.64
Academic, Winter (English)	31	24	7	52	41 232	0	52	35.71	9.68	0.86	3.56
Academic, Spring (English)	31	24	7	52	45 776	1	52	36.95	9.52	0.88	3.29
Applied, Winter (French)	31	24	7	52	264	5	50	29.70	9.34	0.86	3.51
Applied, Spring (French)	31	24	7	52	1 028	9	51	32.22	9.01	0.86	3.42
Academic, Winter (French)	31	24	7	52	910	10	52	34.56	9.41	0.86	3.46
Academic, Spring (French)	31	24	7	52	3 052	8	52	34.22	9.45	0.87	3.47

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 -0.09 to -0.03 for the applied version of the mathematics assessment and from -0.05 to 0.01 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)	16 210	-3.51	2.77	-0.08	0.97
Applied, Spring (English)	16 491	-3.40	2.81	-0.05	0.98
Academic, Winter (English)	41 232	-3.60	2.17	-0.05	0.95
Academic, Spring (English)	45 776	-3.69	2.13	0.01	0.95
Applied, Winter (French)	264	-3.13	2.22	-0.09	0.98
Applied, Spring (French)	1 028	-2.65	2.51	-0.03	0.95
Academic, Winter (French)	910	-2.89	2.33	-0.04	0.95
Academic, Spring (French)	3 052	-3.29	2.34	-0.02	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 the winter and spring administrations, which are displayed in Figures 7.25–7.28, are very similar 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 each assessment 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)

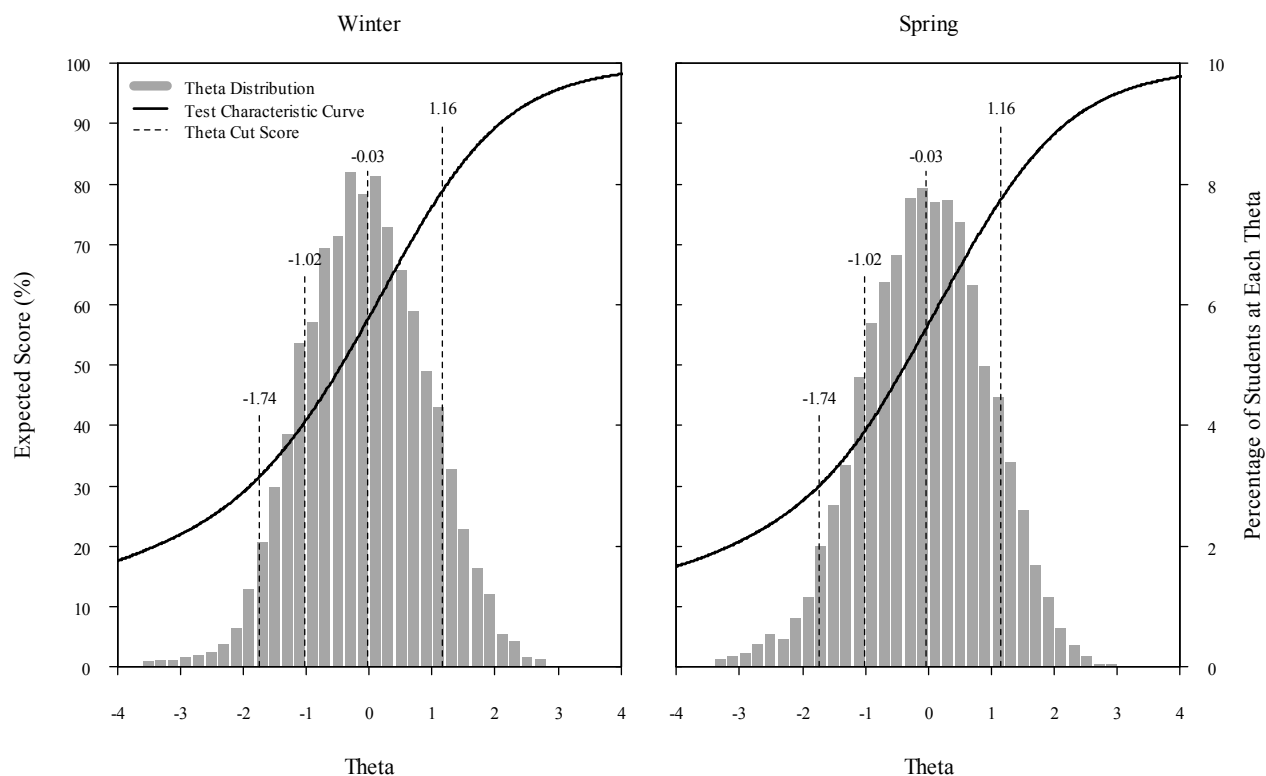


Figure 7.26 Test Characteristic Curves and Distributions of Thetas: Grade 9 Academic Math (English)

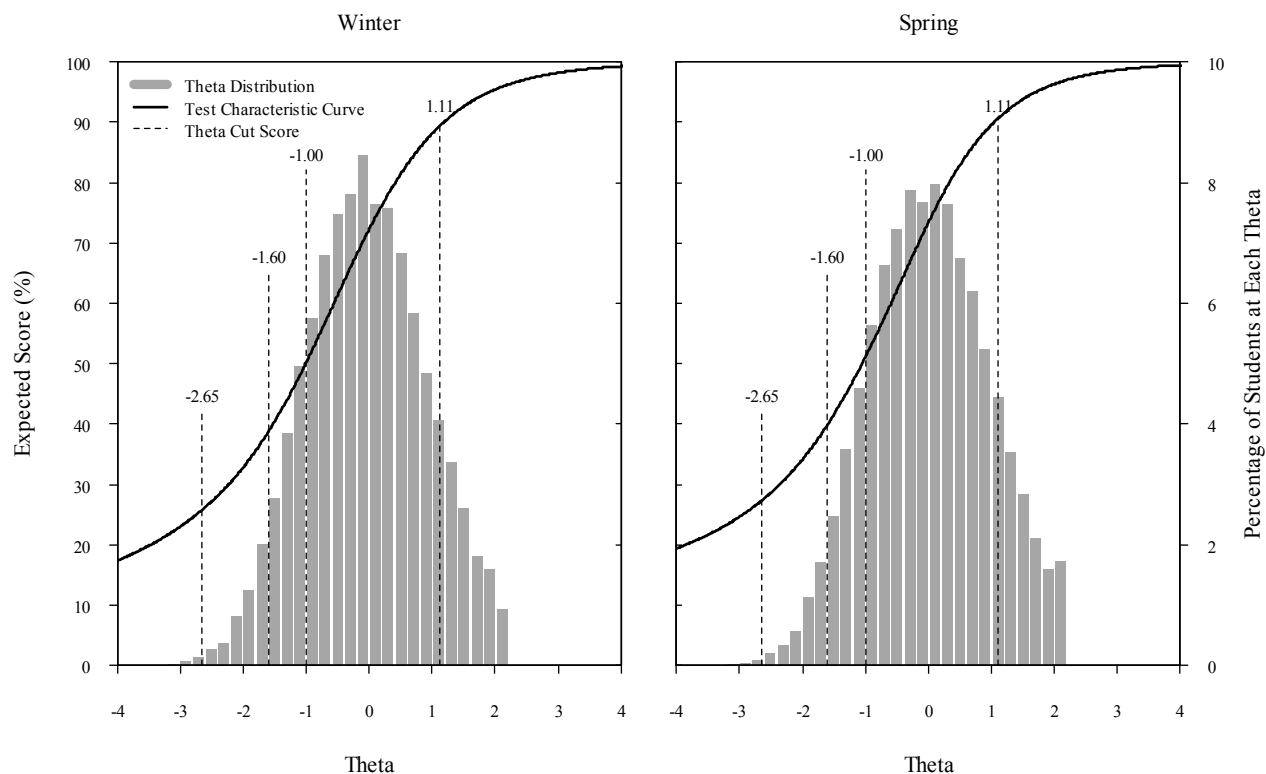


Figure 7.27 Test Characteristic Curves and Distributions of Thetas: Grade 9 Applied Math (French)

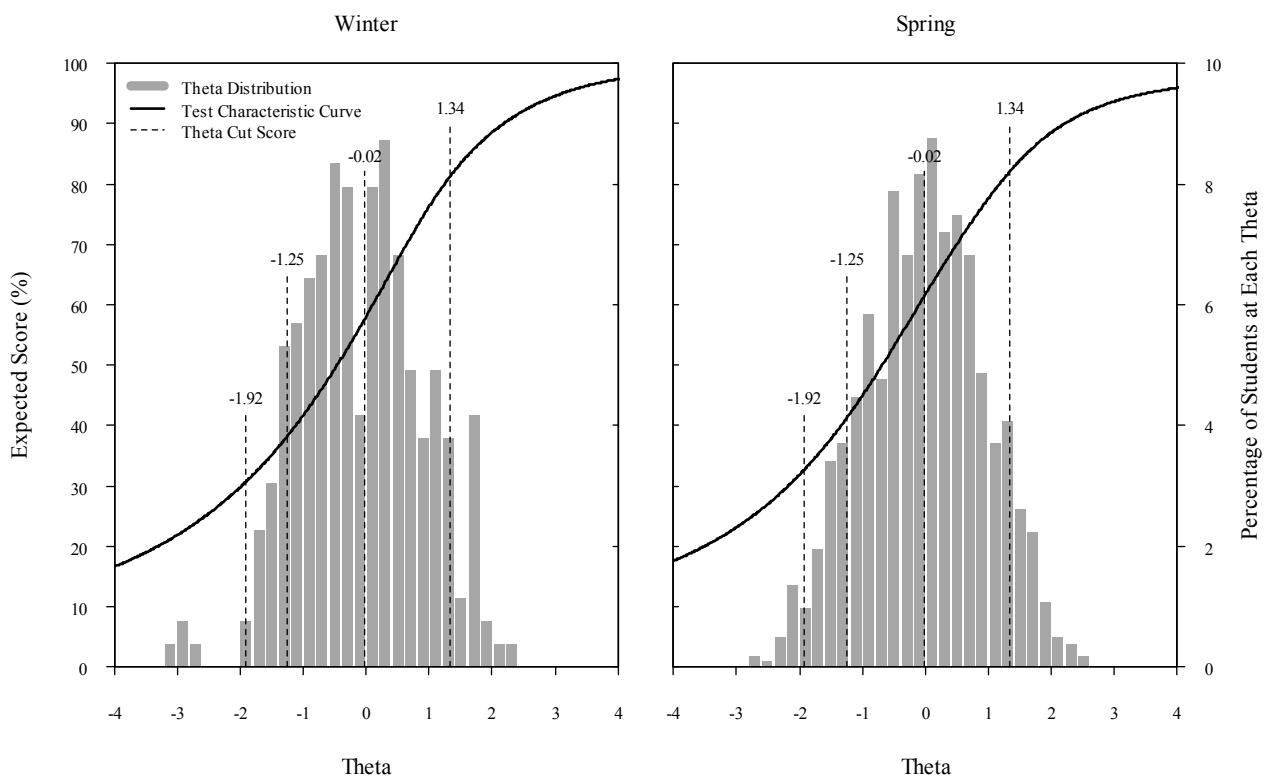


Figure 7.28 Test Characteristic Curves and Distributions of Thetas: Grade 9 Academic Math (French)

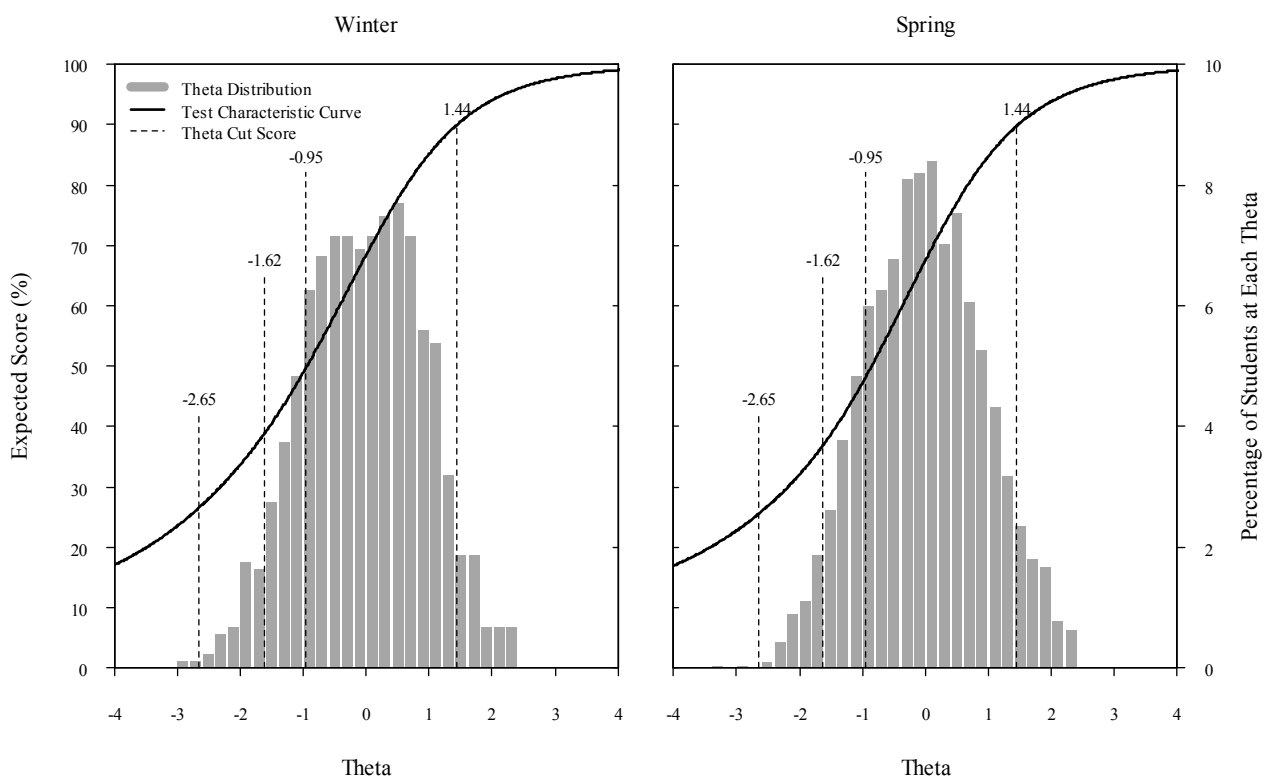


Figure 7.29 Test Information Functions: Grade 9 Applied Math (English)

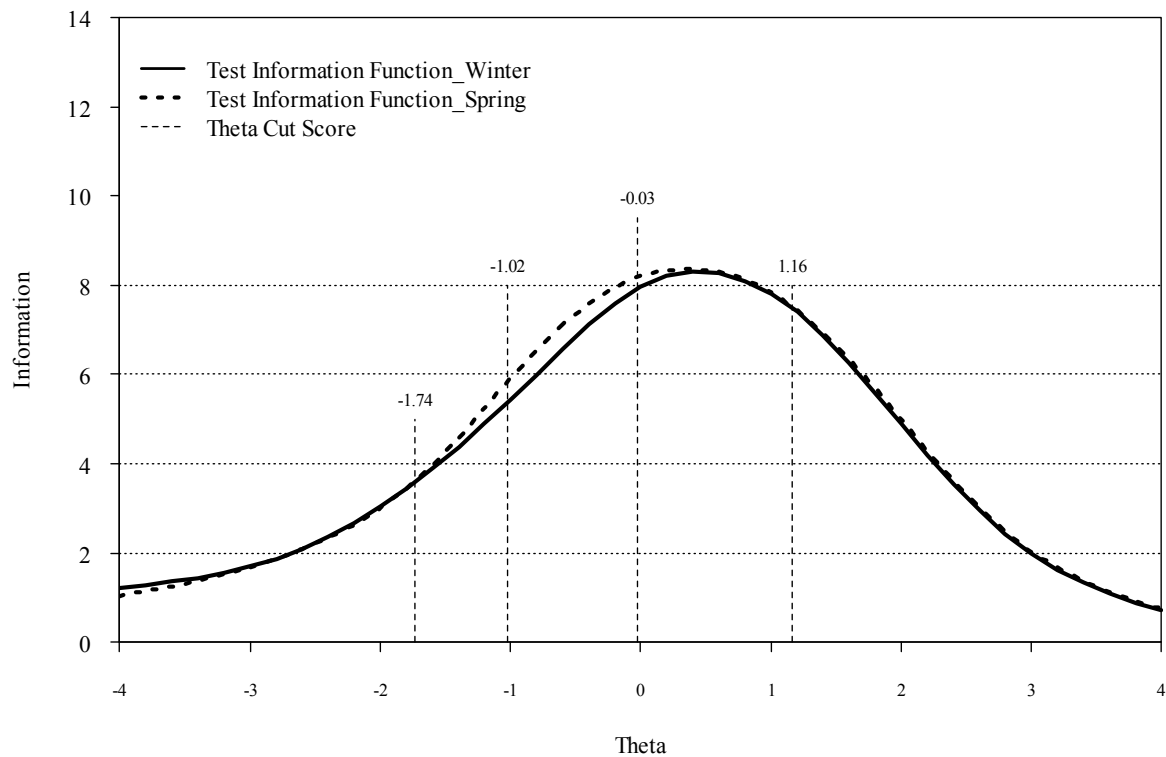


Figure 7.30 Test Information Functions: Grade 9 Academic Math (English)

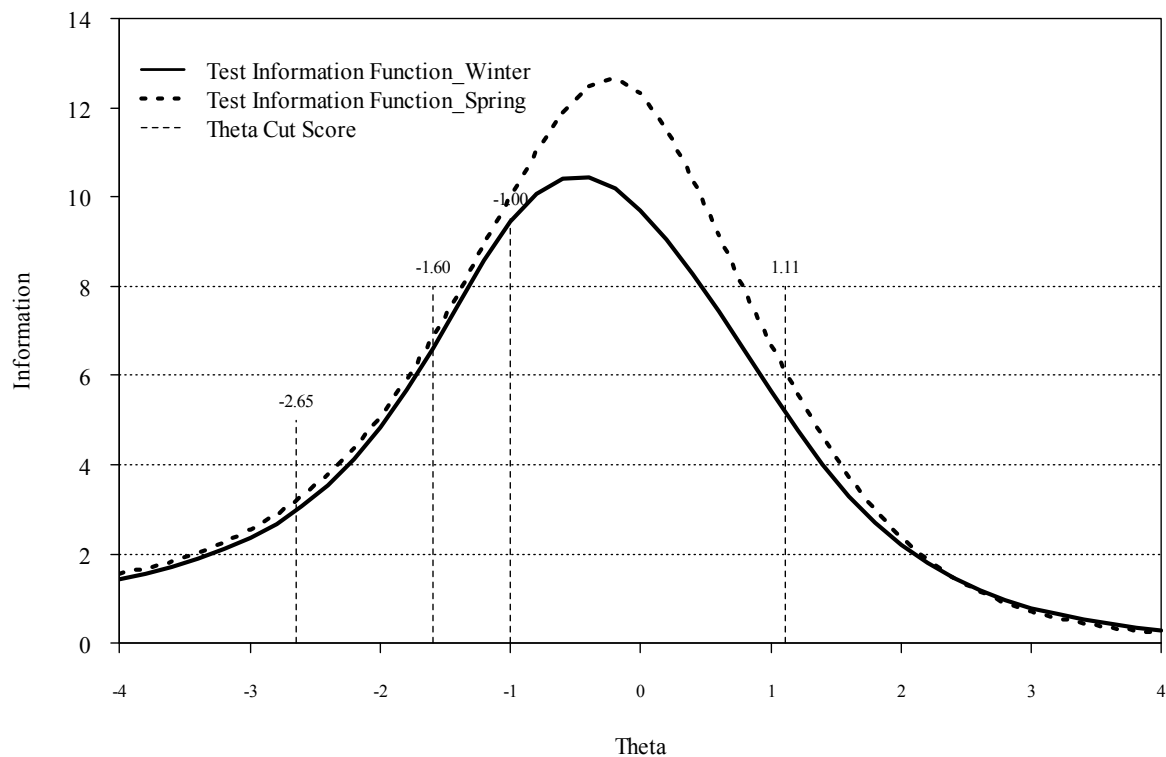


Figure 7.31 Test Information Functions: Grade 9 Applied Math (French)

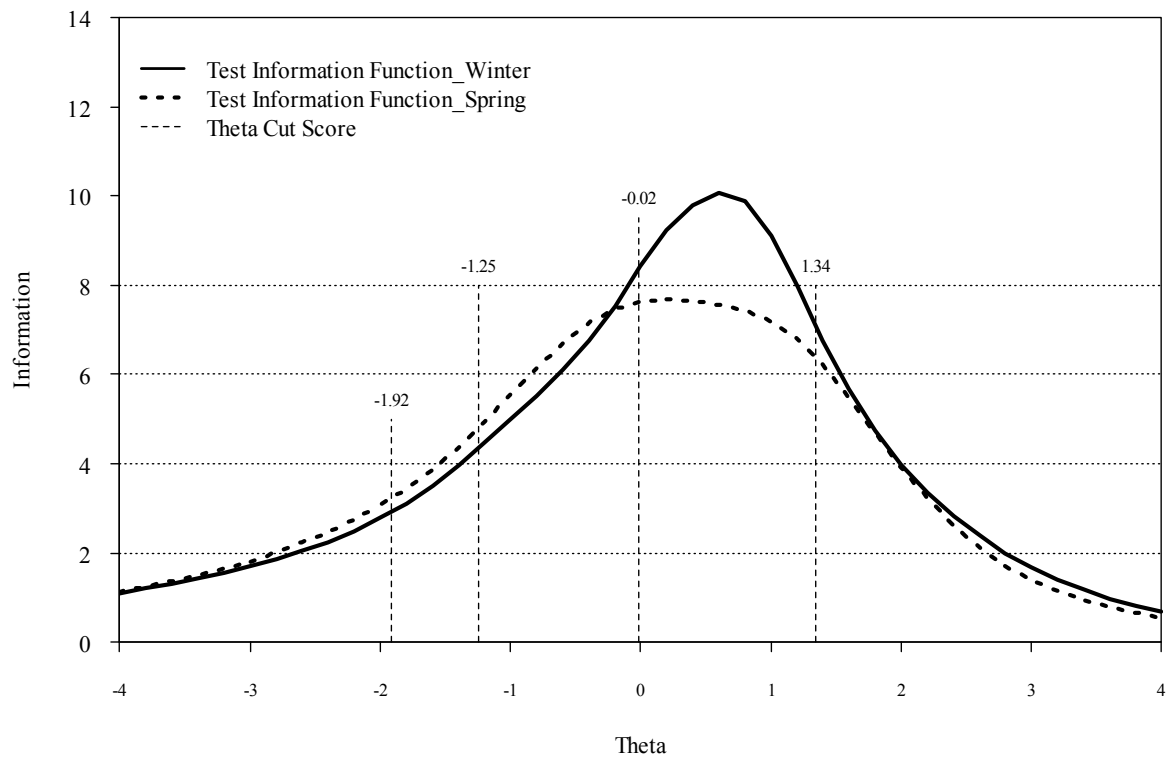
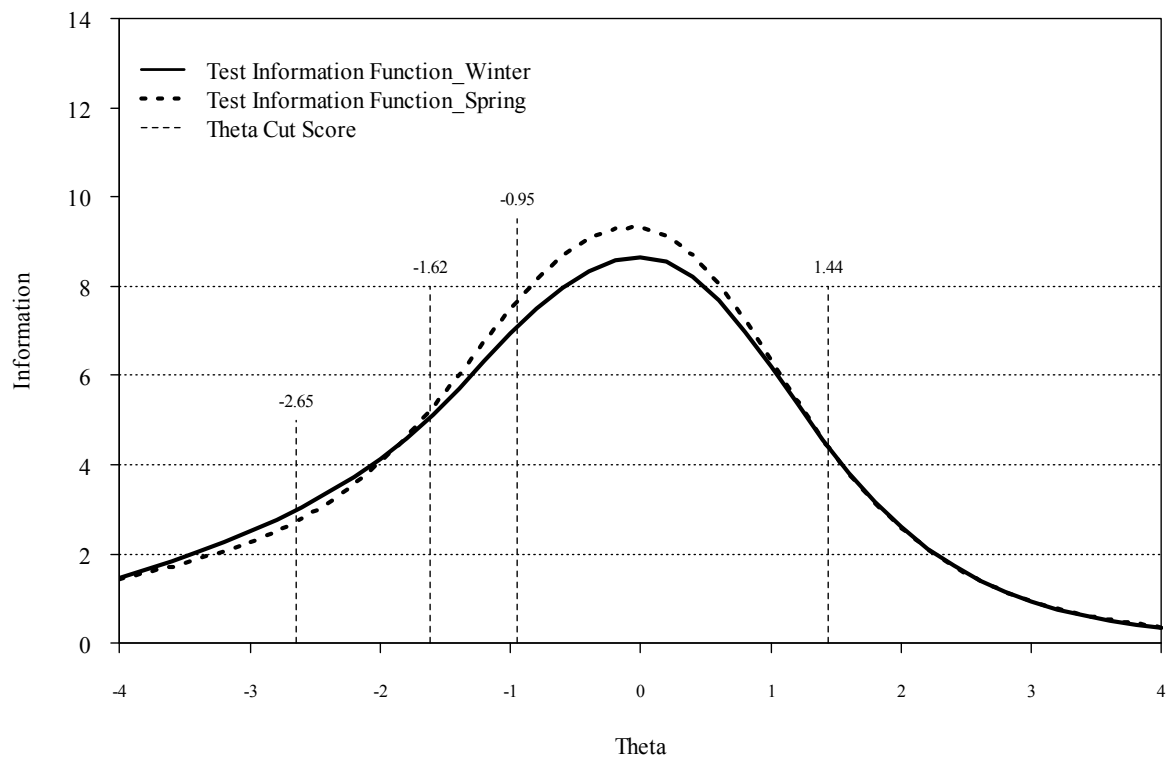


Figure 7.32 Test Information Functions: Grade 9 Academic Math (French)



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.	31.15	0.16	-2.32	0.27
	31	Max.	86.56	0.66	1.63	1.01
		Mean	56.46	0.38 [†]	-0.10	0.63
		SD	14.97	0.14	1.06	0.22
Applied, Spring (English)		Min.	30.40	0.16	-2.32	0.27
	31	Max.	81.28	0.70	1.65	1.09
		Mean	56.99	0.38	-0.10	0.64
		SD	14.22	0.14	0.99	0.24
Academic, Winter (English)		Min.	46.40	0.23	-2.40	0.38
	31	Max.	89.32	0.72	0.55	1.14
		Mean	69.46	0.42	-0.80	0.65
		SD	10.12	0.13	0.69	0.20
Academic, Spring (English)		Min.	36.61	0.24	-2.84	0.38
	31	Max.	91.72	0.86	1.11	1.22
		Mean	70.23	0.45	-0.80	0.75
		SD	10.82	0.13	0.84	0.21
Applied, Winter (French)		Min.	28.29	0.10	-2.45	0.22
	31	Max.	88.29	0.67	2.33	1.94
		Mean	58.88	0.38	-0.12	0.69
		SD	15.51	0.16	1.10	0.40
Applied, Spring (French)		Min.	31.06	0.26	-2.45	0.31
	31	Max.	88.05	0.61	1.23	1.29
		Mean	60.74	0.40	-0.32	0.68
		SD	14.10	0.10	0.96	0.26
Academic, Winter (French)		Min.	39.44	0.22	-2.44	0.32
	31	Max.	86.74	0.64	0.97	1.27
		Mean	65.55	0.42	-0.58	0.66
		SD	11.78	0.12	0.85	0.22
Academic, Spring (French)		Min.	45.50	0.22	-2.13	0.35
	31	Max.	91.98	0.68	0.83	1.27
		Mean	65.91	0.43	-0.54	0.68
		SD	11.02	0.13	0.72	0.22

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 77.5% for English-language students and 75.4% for French-language students.

Cronbach's alpha estimates are 0.89 and 0.88 for English and French, respectively. The corresponding SEMs are 4.0% and 4.1% of the possible maximum score. The obtained reliability coefficients and SEMs are acceptable and indicate that test scores from these assessments are at a satisfactory level of precision.

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

Language	No. of Items	Item Type				Possible Max. Score	No. of Students	Min.	Max.	Mean	SD	R	SEM
		MC	OR	SW	LW								
English	47	39	4	2	2	81	127 867	2.0	81.0	62.8	9.67	0.89	3.26
French	47	39	4	2	2	81	5 313	23.0	81.0	61.1	9.53	0.88	3.32

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	127 867	-3.87	2.83	0.12	0.90
French	5 313	-3.21	2.95	0.06	0.90

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)

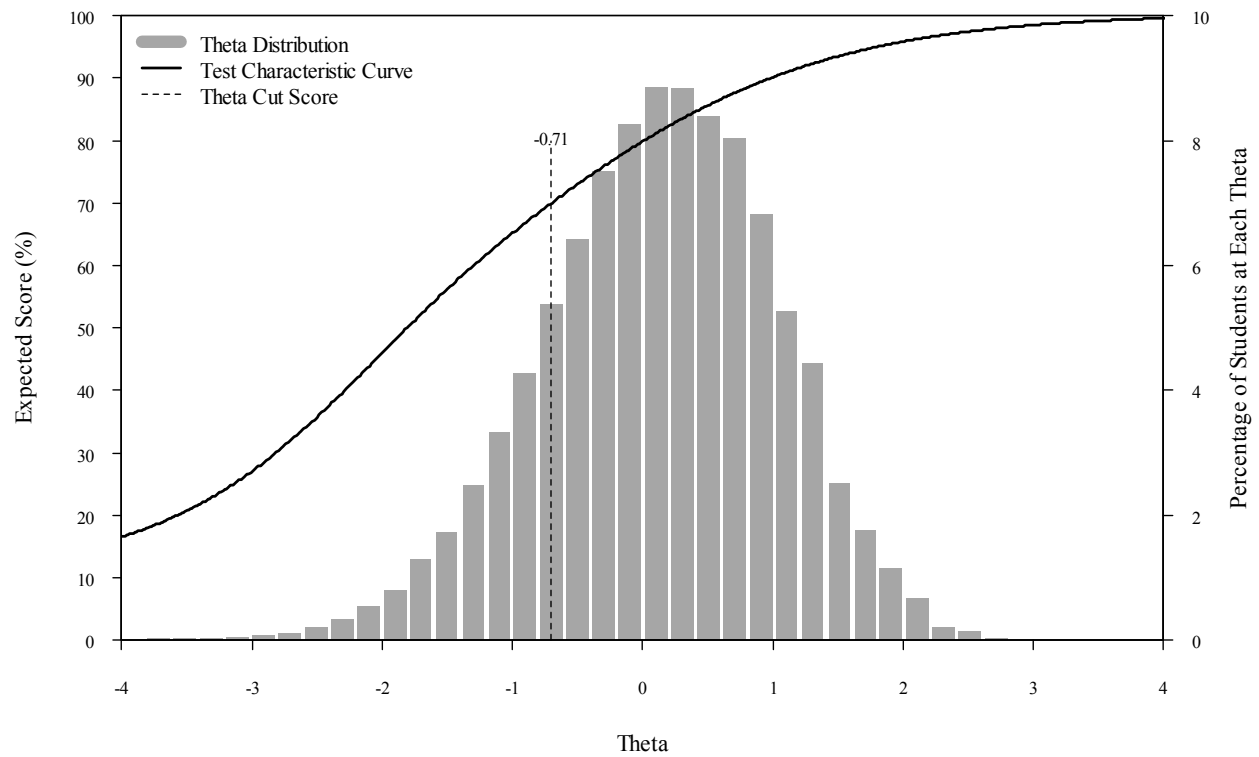


Figure 7.34 Test Characteristic Curve and Distribution of Theta: OSSLT (French)

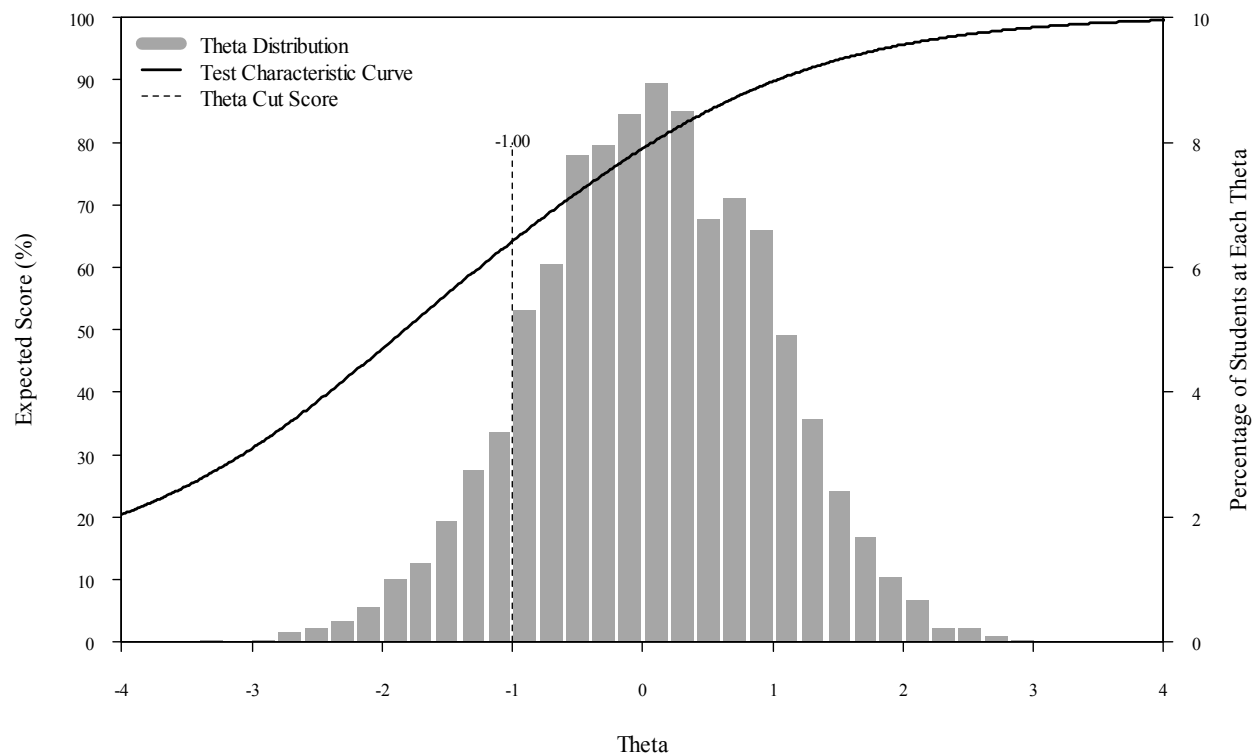


Figure 7.35 Test Information Function: OSSLT (English)

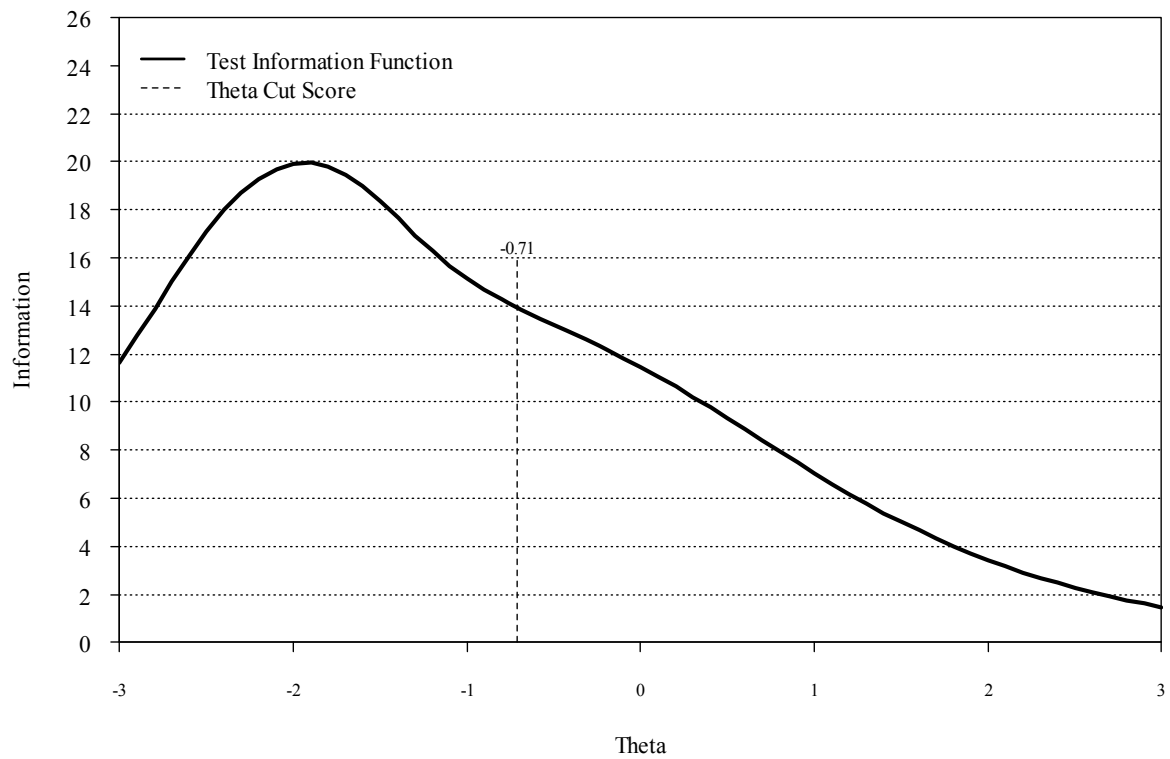
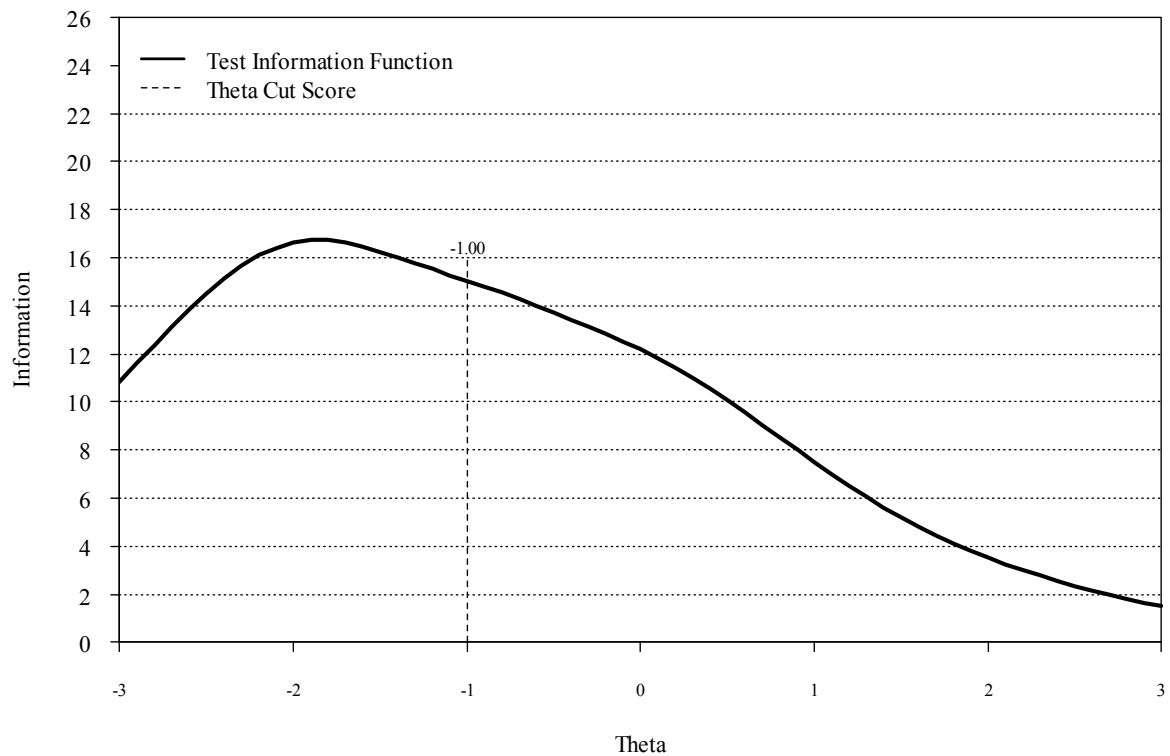


Figure 7.36 Test Information Function: OSSLT (French)



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	47	Min.	0.36	0.16	-2.69
		Max.	0.96	0.57	1.55
		Mean	0.76	0.34 [†]	-1.07
		SD	0.11	0.09	0.84
French	47	Min.	0.34	0.10	-2.79
		Max.	0.94	0.56	1.74
		Mean	0.75	0.33 [†]	-1.17
		SD	0.13	0.11	1.00

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 2014–2015 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, $m_{Fk} = 1/[n_{Fk}(\sum_t y_{tk} \cdot n_{Ftk})]$ is the mean item score of the focal group members at the k^{th} level and m_{Rk} is the analogous value for the reference group. The SMD is divided by the item standard deviation of the total group to obtain an effect size value for the SMD, and these effect sizes, in conjunction with Mantel's (1963) extension of the MH chi-square ($MH \chi^2$), are used to classify OR items into three categories of DIF magnitude, as shown in Table 7.12.

Table 7.12 DIF Classification Rules for Open-Response Items

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

For each assessment, except for the French-language Grade 9 version, two random samples of 2000 examinees were selected from the 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. Negative estimates of Δ and effect size indicate that the girls outperformed the boys or that the SLLs outperformed the non-SLLs; positive Δ and effect-size estimates indicate that the boys outperformed the girls or that the non-SLLs outperformed the SLLs.

The Primary- and Junior-Division Assessments

For the reading, writing and mathematics components of the 2014–2015 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 <i>k</i> = 36	1 (MC) 0 (OR)	0 (MC) 0 (OR)	2 (MC) 0 (OR)	0 (MC) 1 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)
Writing <i>k</i> = 14	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)
Math <i>k</i> = 36	2 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	2 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (MC) 0 (OR)	0 (MC) 0 (OR)

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

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

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

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

Of the 344 items comprising the primary- and junior-division assessments, 14 items were found to have gender-based DIF. The majority of these items had B-level DIF (11), and only three items had C-level DIF. Of the 13 multiple-choice items showing B- or C-level DIF, 12 favoured the boys, while one favoured the girls. One open-response item was found to favour girls. Overall, more items were found to favour boys than girls. The reading component of the English-language junior-division assessment had the largest number of gender-based DIF items (four).

The summaries for SLL-based DIF are reported in Tables 7.15 and 7.16. Twelve out of the 344 items across all the assessments showed SLL-based DIF; 10 had B-level DIF, and two had C-level DIF. All of the eleven multiple-choice DIF items favoured non-SLL students, and the one open-response item also favoured non-SLL students. The reading component of the English-language junior-division assessment was found to have the largest number of SLL-based DIF items (four).

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	3 (MC) 0 (OR)	0 (MC) 0 (OR)	2 (MC) 0 (OR)	0 (MC) 0 (OR)	1 (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)	2 (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)	1 (MC) 0 (OR)	0 (MC) 1 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)

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

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

Component	Primary English		Junior English		Primary French		Junior French	
	Non-SLLs	SLLs	Non-SLLs	SLLs	Non-SLLs	SLLs	Non-SLLs	SLLs
Reading k = 36	0 (MC) 0 (OR)	0 (MC) 0 (OR)	2 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Writing k = 14	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)
Math k = 36	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)	0 (MC) 0 (OR)

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

All items identified as having B-level or C-level DIF on the primary- and junior-division assessments were reviewed by the assessment team. Since the reviewers did not identify any apparent bias in the content of these items, they were not removed from the 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)	0 (MC)	5 (MC)
	1 (OR)	1 (OR)	1 (OR)	0 (OR)
Applied, Spring k = 31	0 (MC)	0 (MC)	2 (MC)	1 (MC)
	0 (OR)	0 (OR)	1 (OR)	1 (OR)
Academic, Winter k = 31	2 (MC)	0 (MC)	2 (MC)	2 (MC)
	0 (OR)	0 (OR)	1 (OR)	1 (OR)
Academic, Spring k = 31	1 (MC)	0 (MC)	2 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	1 (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)	1 (MC)	1 (MC)
	0 (OR)	0 (OR)	1 (OR)	0 (OR)
Applied, Spring k = 31	0 (MC)	0 (MC)	0 (MC)	1 (MC)
	0 (OR)	0 (OR)	1 (OR)	0 (OR)
Academic, Winter k = 31	0 (MC)	0 (MC)	0 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Academic, Spring k = 31	0 (MC)	0 (MC)	0 (MC)	0 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)

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

Of the 248 items across eight Grade 9 math assessments, 18 multiple-choice items showed B-level gender-based DIF. Of the 18 B-level gender-based multiple-choice DIF items, 10 favoured boys and 8 favoured girls. Eight open-response items showed B-level gender-based DIF, with four favouring boys and four favouring girls. Three multiple-choice items showed C-level gender-based DIF: one favoured boys and two favoured girls. Two open-response items showed C-level gender-based DIF, which favoured boys.

Across the Grade 9 math assessments, 8–11 multiple-choice and one open-response items were used in both versions of the winter and spring administrations. Overall, eight items that were repeated showed gender-based DIF. Of the eight items, four showed DIF in the winter administration, four showed DIF in the spring administration, and four were found having DIF across both administrations. One open-response item that was used in both winter and spring administrations showed gender-based DIF in spring administration.

SLL-based DIF was conducted only for the English-language courses. There are eleven multiple-choice B-level SLL-based DIF items. Of these items, six items favoured SLL students, and five favoured non-SLL students. Three open-response items showed B-level SLL-based DIF. Two favoured SLL students, and one 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.

Three repeated items showed SLL-based DIF; all showed DIF in both winter and spring administrations. Overall, more SLL-based DIF items were found to favour SLL students. More SLL-based DIF items were found for the English-language applied 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)	0 (MC)	2 (MC)
	0 (OR)	0 (OR)	0 (OR)	0 (OR)
Spring k = 31	3 (MC)	2 (MC)	1 (MC)	1 (MC)
	1 (OR)	1 (OR)	0 (OR)	1 (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)	2 (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 = 51)		French (k = 51)	
Males	Females	Males	Females
2 (MC)	1 (LW)	4 (MC)	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 = 51)		French (k = 51)	
Males	Females	Males	Females
3 (MC)	0	1 (MC)	0

Note. MC = multiple choice.

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

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

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

One multiple-choice item 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 = 51)			
B-Level DIF		C-Level DIF	
Non-SLLs	SLLs	Non-SLLs	SLLs
4 (MC)	1 (MC) 2 (SW) 1 (LW)	1 (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	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	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	0.86
Junior Reading (English)	0.85	0.79	0.92	0.88
Primary Reading (French)	0.82	0.75	0.92	0.88
Junior Reading (French)	0.86	0.81	0.95	0.92
Primary Writing (English)	0.85	0.79	0.89	0.84
Junior Writing (English)	0.81	0.73	0.90	0.86
Primary Writing (French)	0.80	0.72	0.89	0.84
Junior Writing (French)	0.81	0.74	0.91	0.87
Primary Mathematics (English)	0.81	0.74	0.90	0.86
Junior Mathematics (English)	0.79	0.70	0.91	0.87
Primary Mathematics (French)	0.83	0.76	0.91	0.89
Junior Mathematics (French)	0.85	0.78	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	0.88	0.83
Applied, Spring (English)	0.70	0.59	0.88	0.84
Academic, Winter (English)	0.83	0.76	0.94	0.91
Academic, Spring (English)	0.84	0.78	0.95	0.93
Applied, Winter (French)	0.74	0.64	0.87	0.84
Applied, Spring (French)	0.76	0.66	0.88	0.84
Academic, Winter (French)	0.85	0.79	0.93	0.90
Academic, Spring (French)	0.85	0.80	0.93	0.90

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	0.91

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:

1. To provide achievement data to evaluate the quality of the Ontario educational system for accountability purposes at the school, board and provincial levels, including monitoring changes in achievement across years.
2. To provide information to students and parents on students' achievement of the curriculum expectations in reading, writing and mathematics at selected grade levels.
3. To provide information to be used for school improvement planning.

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 (Grades 3 and 6, respectively); 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 but one item in the mathematics component and for all items in the reading and writing components. The aggregate exact-plus-adjacent validity estimates for the items in each component ranged from 98 to 100%. For Grade 9 mathematics, the EQAO target of 95% exact-plus-adjacent agreement was met for all but one item in the English-language academic assessment, and the aggregate validity estimates ranged from 98.5 to 100%. For the OSSLT, the EQAO target of 95% exact-plus-adjacent agreement was met for all items, and the aggregate validity estimates ranged from 96.5 to 99.4% (see Appendix 4.1).

In addition, for the paper-based scoring process, student responses to multiple-choice items are captured by optical-scan forms. 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, 0.85 to 0.88 for Grade 9 mathematics, 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 0.87 to 0.95, 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 (Radwan & Xu, 2012).

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.

Conclusion

This chapter follows the argument-based approach to validation, as specified in the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 1999) and by Kane (2006, 2013). With this approach, the claims about proposed interpretations or uses are stated, and then these claims are evaluated. Three purposes of the EQAO assessments are clearly given at the beginning of this chapter, and various sources of evidence are summarized in the previous sections to evaluate these purposes.

In order to provide data for accountability purposes, for informing individual students of their achievement and for school improvement planning, the assessments must be carefully constructed and closely aligned with curriculum expectations, the results must be reliable and based on accurate scaling, equating and standard setting, and there should be a convergent relationship between EQAO assessments and other assessments that measure a similar construct. The types of validity evidence presented in this chapter and throughout this technical report support these claims. It is always challenging to collect evidence based on consequences of testing, but several research projects have been proposed at EQAO to address the intended and unintended outcomes and the positive and negative systemic effects from the interpretations and uses of EQAO assessment results.

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
- Radwan, N., Xu, Y. (2012). *Comparison of the Performance of Ontario Students on the OSSLT/TPCL and the PISA 2009 Reading Assessment*. Retrieved February 10, 2015, from http://www.eqao.com/en/research_data/Research_Reports/DMA-docs/comparison-OSSLT-PISA-2009.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 FOR ALL ASSESSMENTS AND INTERRATER RELIABILITY FOR OSSLT

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

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
22981	NR	NR	4 496	99.1	75.8	23.3	16.6	6.7	0.9
22983	NR	NR	4 013	97.5	71.6	25.9	12.4	13.5	2.5
22951	NR	NR	2 511	100.0	88.1	11.9	8.3	3.6	0.0
22953	NR	NR	2 694	98.7	81.5	17.1	11.3	5.8	1.3
22859	NR	NR	2 683	99.7	89.1	10.6	6.2	4.4	0.3
22860	NR	NR	3 072	99.2	79.8	19.4	10.5	8.9	0.8
23085	1 (C)	11	3 152	99.4	91.8	7.6	3.3	4.3	0.6
23088	1 (C)	12	2 286	99.8	86.1	13.7	9.4	4.4	0.2
20045	1 (D)	5	2 146	99.7	83.7	15.9	11.6	4.4	0.3
20043	1 (D)	6	2 946	99.5	83.2	16.4	12.1	4.3	0.5
Aggregate			29 999	99.2	82.1	17.0	10.6	6.5	0.8

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
23033	NR	NR	681	99.6	79.4	20.1	9.8	10.3	0.4
23032	NR	NR	695	99.0	68.2	30.8	14.0	16.8	1.0
19993	NR	NR	806	98.5	75.7	22.8	11.5	11.3	1.5
19995	NR	NR	807	99.0	80.8	18.2	10.5	7.7	1.0
23117	NR	NR	635	99.4	83.6	15.7	4.3	11.5	0.6
23116	NR	NR	712	98.9	83.7	15.2	10.0	5.2	1.1
22660	1 (C)	11	778	100.0	79.8	20.2	6.3	13.9	0.0
22659	1 (C)	12	1241	99.0	77.0	22.0	16.1	5.9	1.0
23128	1 (D)	5	577	99.1	85.8	13.3	10.2	3.1	0.9
23139	1 (D)	6	708	96.8	72.6	24.2	18.2	5.9	3.2
Aggregate			7640	98.9	78.4	20.5	11.5	9.0	1.1

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
22752	NR	NR	360	96.9	87.2	9.7	3.1	6.7	3.1
22753	NR	NR	234	97.9	77.8	20.1	10.7	9.4	2.1
17633	NR	NR	217	100.0	94.0	6.0	4.6	1.4	0.0
17634	NR	NR	142	100.0	97.2	2.8	2.1	0.7	0.0
22777	NR	NR	410	97.6	92.0	5.6	4.9	0.7	2.4
22778	NR	NR	241	98.8	94.2	4.6	2.1	2.5	1.2
15758	1 (C)	11	277	97.8	85.6	12.3	10.5	1.8	2.2
15757	1 (C)	12	231	100.0	95.2	4.8	2.2	2.6	0.0
26320	1 (D)	5	226	100.0	94.7	5.3	1.8	3.5	0.0
26319	1 (D)	6	393	99.7	88.3	11.5	7.6	3.8	0.3
Aggregate			2731	98.7	90.1	8.6	5.2	3.4	1.3

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
23211	NR	NR	242	100.0	78.9	21.1	12.4	8.7	0.0
23210	NR	NR	204	100.0	82.8	17.2	10.3	6.9	0.0
23201	NR	NR	231	98.7	87.0	11.7	8.2	3.5	1.3
23200	NR	NR	351	100.0	81.2	18.8	3.4	15.4	0.0
23173	NR	NR	253	100.0	87.7	12.3	5.5	6.7	0.0
23174	NR	NR	247	99.2	88.7	10.5	6.5	4.0	0.8
23172	1 (C)	11	264	99.6	93.6	6.1	2.3	3.8	0.4
23171	1 (C)	12	631	99.7	91.1	8.6	4.4	4.1	0.3
23220	1 (D)	5	231	100.0	90.9	9.1	2.6	6.5	0.0
23390	1 (D)	6	215	98.1	79.5	18.6	13.0	5.6	1.9
Aggregate			2869	99.6	86.8	12.8	6.3	6.5	0.4

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
25981 T	D	7	1 987	98.4	67.5	30.9	17.5	13.4	1.6
25981 V	D	7	1 987	99.2	68.0	31.2	11.4	19.8	0.8
Aggregate Long Writing			3 974	98.8	67.7	31.1	14.4	16.6	1.2
22940 T	NR	NR	2 420	98.1	73.4	24.7	11.0	13.7	1.9
22940 V	NR	NR	2 420	98.7	74.9	23.8	14.4	9.4	1.3
22804 T	C	13	1 628	99.9	82.3	17.6	8.3	9.3	0.1
22804 V	C	13	1 628	99.3	84.8	14.5	9.5	5.0	0.7
Aggregate Short Writing			8 096	99.0	78.9	20.1	10.8	9.3	1.0
Aggregate All Items			12 070	98.9	73.3	25.6	12.6	13.0	1.1

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
25991 T	D	7	2084	98.3	65.2	33.1	10.1	23.0	1.7
25991 V	D	7	2084	98.9	69.8	29.2	12.4	16.8	1.1
Aggregate Long Writing			4168	98.6	67.5	31.1	11.3	19.9	1.4
19774 T	NR	NR	910	97.4	64.2	33.2	15.3	17.9	2.6
19774 V	NR	NR	910	99.1	69.9	29.2	11.0	18.2	0.9
22685 T	C	13	1829	95.2	51.9	43.4	15.0	28.4	4.8
22685 V	C	13	1829	97.9	61.3	36.6	11.6	24.9	2.1
Aggregate Short Writing			5478	97.4	61.8	35.6	13.2	22.4	2.6
Aggregate All Items			9646	98.0	64.7	33.4	12.2	21.1	2.0

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
25919 T	D	7	705	99.1	71.8	27.4	10.8	16.6	0.9
25919 V	D	7	705	99.6	69.4	30.2	10.8	19.4	0.4
Aggregate Long Writing			1410	99.4	70.6	28.8	10.8	18.0	0.6
22814 T	NR	NR	628	99.4	83.8	15.6	7.8	7.8	0.6
22814 V	NR	NR	628	97.3	76.3	21.0	2.5	18.5	2.7
22811 T	C	13	650	99.4	75.4	24.0	11.5	12.5	0.6
22811 V	C	13	650	99.5	74.5	25.1	9.8	15.2	0.5
Aggregate Short Writing			2556	98.9	77.5	21.4	7.9	13.5	1.1
Aggregate All Items			3966	99.1	74.0	25.1	9.4	15.8	0.9

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
23297 T	D	7	299	99.7	71.6	28.1	15.1	13.0	0.3
23297 V	D	7	299	99.3	70.9	28.4	11.4	17.1	0.7
Aggregate Long Writing			598	99.5	71.2	28.3	13.2	15.1	0.5
26299 T	NR	NR	339	100.0	80.2	19.8	6.8	13.0	0.0
26299 V	NR	NR	339	100.0	73.2	26.8	15.6	11.2	0.0
23555 T	C	13	372	99.7	78.2	21.5	17.7	3.8	0.3
23555 V	C	13	372	99.2	67.7	31.5	8.9	22.6	0.8
Aggregate Short Writing			1422	99.7	74.8	24.9	12.3	12.6	0.3
Aggregate All Items			2020	99.6	73.0	26.6	12.7	13.8	0.4

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
22301	3(1)	8	936	98.0	92.3	5.7	1.1	4.6	2.0
22321	3(1)	9	930	99.9	98.3	1.6	0.6	1.0	0.1
22644	NR	NR	587	100.0	92.5	7.5	2.0	5.5	0.0
22254	NR	NR	965	100.0	98.9	1.1	0.4	0.7	0.0
15096	3(2)	10	789	98.9	94.2	4.7	3.8	0.9	1.1
19252	3(2)	11	832	99.4	88.5	10.9	6.6	4.3	0.6
26475	NR	NR	678	99.3	88.2	11.1	8.1	2.9	0.7
19269	NR	NR	672	100.0	97.6	2.4	0.6	1.8	0.0
Aggregate			6389	99.4	94.0	5.4	2.8	2.6	0.6

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
22386	3(1)	8	902	99.8	97.9	1.9	1.8	0.1	0.2
22276	NR	NR	1237	99.1	94.7	4.4	3.2	1.1	0.9
22343	NR	NR	858	100.0	94.9	5.1	1.2	4.0	0.0
23606	NR	NR	988	100.0	94.7	5.3	0.8	4.5	0.0
20635	3(2)	9	808	99.9	97.5	2.4	1.1	1.2	0.1
20469	3(2)	10	1105	100.0	99.6	0.4	0.2	0.2	0.0
22384	3(2)	11	920	100.0	99.8	0.2	0.0	0.2	0.0
20529	NR	NR	826	99.3	93.1	6.2	4.0	2.2	0.7
Aggregate			7644	99.7	96.6	3.2	1.5	1.6	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
19784	3(1)	8	54	96.3	90.7	5.6	0.0	5.6	3.7
14654	3(1)	11	82	97.6	75.6	22.0	1.2	20.7	2.4
22185	NR	NR	35	100.0	91.4	8.6	5.7	2.9	0.0
19837	NR	NR	142	100.0	95.1	4.9	0.0	4.9	0.0
16385	3(2)	9	33	100.0	97.0	3.0	0.0	3.0	0.0
19705	3(2)	10	117	94.9	92.3	2.6	1.7	0.9	5.1
14598	NR	NR	137	99.3	94.2	5.1	4.4	0.7	0.7
22193	NR	NR	76	98.7	86.8	11.8	10.5	1.3	1.3
Aggregate			676	98.2	90.7	7.5	2.8	4.7	1.8

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
22491	3(1)	8	230	100.0	100.0	0.0	0.0	0.0	0.0
23315	3(1)	9	200	100.0	97.0	3.0	3.0	0.0	0.0
11577	NR	NR	80	100.0	86.3	13.8	11.3	2.5	0.0
20198	NR	NR	117	100.0	91.5	8.5	8.5	0.0	0.0
22490	3(2)	10	53	100.0	96.2	3.8	1.9	1.9	0.0
23316	3(2)	11	28	100.0	96.4	3.6	0.0	3.6	0.0
20150	NR	NR	99	100.0	96.0	4.0	2.0	2.0	0.0
20167	NR	NR	107	100.0	97.2	2.8	2.8	0.0	0.0
Aggregate			914	100.0	96.0	4.0	3.4	0.7	0.0

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	21586	20	146	99.3	93.2	6.2	1.4	4.8	0.7
	19625	22	162	99.4	87.0	12.3	5.6	6.8	0.6
	21588	24	171	100.0	81.3	18.7	2.9	15.8	0.0
	21568	NR	164	98.2	93.9	4.3	3.7	0.6	1.8
	21570	NR	135	95.6	88.9	6.7	3.0	3.7	4.4
	14845	NR	200	98.5	89.5	9.0	5.5	3.5	1.5
	21534	NR	102	100.0	91.2	8.8	1.0	7.8	0.0
Aggregate			1080	98.7	89.1	9.6	3.5	6.1	1.3
Spring	21585	19	208	97.1	82.2	14.9	5.8	9.1	0.0
	19624	21	128	98.4	85.2	13.3	3.1	10.2	0.0
	21551	23	200	100.0	93.0	7.0	6.5	0.5	0.0
	21572	25	90	100.0	93.3	6.7	1.1	5.6	0.0
	21530	NR	171	99.4	95.9	3.5	2.9	0.6	0.0
	21531	NR	111	100.0	83.8	16.2	5.4	10.8	0.0
	19627	NR	153	100.0	98.0	2.0	0.7	1.3	0.0
Aggregate			1061	99.2	90.2	9.0	4.0	5.0	0.0
Aggregate Across Administrations			2141	98.9	89.6	9.3	3.7	5.6	0.6

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	15680	12	486	99.6	91.6	8.0	4.5	3.5	0.4
	19587	13	300	97.0	80.0	17.0	10.0	7.0	3.0
	21661	15	295	100.0	90.2	9.8	7.1	2.7	0.0
	26861	18	217	100.0	92.2	7.8	4.6	3.2	0.0
	21642	NR	490	99.4	91.6	7.8	4.3	3.5	0.6
	15702	NR	306	92.5	88.2	4.2	3.6	0.7	7.5
	21644	NR	480	99.8	95.0	4.8	2.1	2.7	0.2
Aggregate			2574	98.5	90.4	8.2	4.9	3.3	1.5
Spring	26868	14	789	100.0	97.0	3.0	2.4	0.6	0.0
	14943	16	275	100.0	95.6	4.4	2.2	2.2	0.0
	19608	17	593	100.0	83.0	17.0	14.7	2.4	0.0
	19567	NR	156	100.0	98.7	1.3	0.6	0.6	0.0
	26865	NR	704	99.6	83.7	15.9	8.0	8.0	0.4
	21624	NR	407	99.5	85.3	14.3	10.8	3.4	0.0
	19591	NR	259	99.2	90.0	9.3	1.2	8.1	0.8
Aggregate			3183	99.8	89.3	10.5	6.8	3.7	0.2
Aggregate Across Administrations			5757	99.2	89.8	9.3	5.8	3.5	0.8

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	22012	12	14	100.0	100.0	0.0	0.0	0.0	0.0
	20429	15	18	100.0	100.0	0.0	0.0	0.0	0.0
	22019	16	15	100.0	100.0	0.0	0.0	0.0	0.0
	20366	17	15	100.0	100.0	0.0	0.0	0.0	0.0
	20391	NR	18	100.0	100.0	0.0	0.0	0.0	0.0
	20426	NR	19	100.0	94.7	5.3	0.0	5.3	0.0
	18496	NR	18	100.0	94.4	5.6	0.0	5.6	0.0
Aggregate			117	100.0	98.3	1.7	0.0	1.7	0.0
Spring	22016	13	56	98.2	82.1	16.1	5.4	10.7	0.0
	20369	14	52	100.0	92.3	7.7	5.8	1.9	0.0
	20429	15	48	100.0	97.9	0.0	0.0	2.1	0.0
	21684	18	58	100.0	86.2	13.8	3.4	10.3	0.0
	15307	NR	57	100.0	93.0	7.0	3.5	3.5	0.0
	20448	NR	58	100.0	91.4	8.6	6.9	1.7	0.0
	21787	NR	52	100.0	80.8	19.2	11.5	7.7	0.0
Aggregate			381	99.7	89.0	10.5	5.2	5.5	0.0
Aggregate Across Administrations			498	99.9	93.6	6.1	2.6	3.6	0.0

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	20269	11	24	100.0	95.8	4.2	4.2	0.0	0.0
	18490	12	25	100.0	100.0	0.0	0.0	0.0	0.0
	22714	14	30	100.0	83.3	16.7	13.3	3.3	0.0
	15441	16	28	100.0	96.4	3.6	0.0	3.6	0.0
	20330	NR	26	96.2	96.2	0.0	0.0	0.0	3.8
	20307	NR	26	100.0	96.2	3.8	0.0	3.8	0.0
	22030	NR	27	100.0	100.0	0.0	0.0	0.0	0.0
Aggregate			186	99.5	95.2	4.3	2.7	1.6	0.5
Spring	20331	10	50	100.0	96.0	4.0	2.0	2.0	0.0
	18490	12	40	100.0	90.0	0.0	10.0	0.0	0.0
	20346	13	63	100.0	95.2	0.0	4.8	0.0	0.0
	21968	15	63	98.4	92.1	6.3	4.8	1.6	0.0
	20289	NR	59	100.0	98.3	0.0	1.7	0.0	0.0
	20287	NR	93	100.0	96.8	0.0	0.0	3.2	0.0
	15399	NR	70	98.6	91.4	7.1	5.7	1.4	0.0
Aggregate			438	99.5	94.5	2.5	3.7	1.4	0.0
Aggregate Across Administrations			624	99.5	94.8	3.4	3.2	1.5	0.3

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
21197_555	I	6	7 830	99.4	91.2	8.2	3.7	4.5	0.6
17429_303	V	7	8 401	99.6	91.2	8.5	3.6	4.8	0.4
21321_567	NR	NR	8 445	98.8	82.8	16.0	9.1	6.8	1.2
21322_567	NR	NR	7 541	99.9	88.5	11.4	5.4	6.0	0.1
Aggregate			32 217	99.4	88.4	11.0	5.5	5.5	0.6

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
26731 T	IV	1	17 636	98.8	81.4	17.4	9.9	7.5	1.2
26731 V	IV	1	11 635	99.4	84.5	14.8	7.5	7.3	0.6
23367 T	NR	NR	23 674	98.6	78.2	20.4	11.7	8.7	1.4
23367 V	NR	NR	17 212	98.4	75.8	22.7	12.1	10.6	1.6
Aggregate Long Writing			70 157	98.7	79.5	19.3	10.6	8.6	1.3
26495 T & V	III	1	10 010	96.8	70.4	26.4	17.7	8.7	3.2
23689 T & V	NR	NR	10 146	96.5	73.1	23.4	12.5	10.9	3.5
Aggregate Short Writing			20 156	96.6	71.7	24.9	15.1	9.8	3.4
Aggregate All Items			90 313	98.3	77.7	20.5	11.6	8.9	1.7

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
21119 581	I	6	249	99.2	92.8	6.4	3.2	3.2	0.8
21828 582	V	7	208	99.0	95.2	3.8	1.9	1.9	1.0
21106 580	NR	NR	199	100.0	97.0	3.0	1.5	1.5	0.0
23729 580	NR	NR	167	99.4	94.0	5.4	0.0	5.4	0.6
Aggregate			823	99.4	94.7	4.7	1.8	2.9	0.6

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
26724 T	IV	1	523	97.3	73.8	23.5	8.4	15.1	2.7
26724 V	IV	1	523	99.4	74.4	25.0	9.6	15.5	0.6
41028 T	NR	NR	806	98.5	73.6	24.9	11.8	13.2	1.5
41028 V	NR	NR	806	98.0	61.5	36.5	20.7	15.8	2.0
Aggregate Long Writing			2 658	98.3	70.1	28.2	13.4	14.8	1.7
21121 T & V	III	1	354	96.0	70.3	25.7	18.6	7.1	4.0
21154 T & V	NR	NR	310	97.1	72.6	24.5	6.8	17.7	2.9
Aggregate Short Writing			664	96.5	71.4	25.2	13.1	12.0	3.5
Aggregate All Items			3 322	98.0	70.4	27.6	13.3	14.2	2.0

Note. NR = not released.

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

Table 4.1.21 Interrater Reliability Estimates for Reading: OSSLT (English)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
21197 555	I	6	171 743	97.4	58.0	39.4	2.6
17429 303	V	7	171 742	97.6	64.0	33.6	2.4
21321 567	NR	NR	171 741	97.6	58.5	39.1	2.4
21322 567	NR	NR	171 740	98.0	64.8	33.2	2.0
Aggregate			686 966	97.6	61.3	36.3	2.4

Note. NR = not released.

Table 4.1.22 Interrater Reliability Estimates for Writing: OSSLT (English)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
26731 T	IV	1	171 742	91.6	47.2	44.4	8.4
26731 V	IV	1	171 742	97.0	59.3	37.6	3.0
23367 T	NR	NR	171 737	91.2	45.8	45.4	8.8
23367 V	NR	NR	171 737	96.0	58.4	37.6	4.0
Aggregate Long Writing			686 958	94.0	52.7	41.2	6.0
26495 T & V	III	1	171 743	92.2	51.5	40.7	7.8
23689 T & V	NR	NR	171 739	92.0	51.0	41.1	8.0
Aggregate Short Writing			343 482	92.1	51.2	40.9	7.9
Aggregate All Items			1 030 440	93.3	52.2	41.1	6.7

Note. NR = not released.

Table 4.1.23 Interrater Reliability Estimates for Reading: OSSLT (French)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
21119 581	I	6	6 143	98.9	69.7	29.2	1.1
21828 582	V	7	6 143	94.5	69.4	25.1	5.5
21106 580	NR	NR	6 143	99.2	76.5	22.7	0.8
23729 580	NR	NR	6 143	98.1	71.6	26.5	1.9
Aggregate			24 572	97.7	71.8	25.9	2.3

Note. NR = not released.

Table 4.1.24 Interrater Reliability Estimates for Writing: OSSLT (French)

Item Code	Section	Sequence	No. of Pairs	% Exact-Plus-Adjacent	% Exact	% Adjacent	% Non-Adjacent
26724_T	IV	1	6 143	94.0	56.4	37.5	6.0
26724_V	IV	1	6 143	98.2	63.4	34.9	1.8
41028_T	NR	NR	6 143	91.3	46.7	44.7	8.7
41028_V	NR	NR	6 143	95.7	51.6	44.1	4.3
Aggregate Long Writing			24 572	94.8	54.5	40.3	5.2
21121_T & V	III	1	6 143	90.8	50.9	39.9	9.2
21154_T & V	NR	NR	6 143	90.8	50.0	40.8	9.2
Aggregate Short Writing			12 286	90.8	50.4	40.3	9.2
Aggregate All Items			36 858	93.5	53.2	40.3	6.5

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
22980	NR	NR	R3.0	C	4	0.70	0.38	-0.63	0.74
22973	NR	NR	R2.0	I	2	0.61	0.23	-0.40	0.31
22970	NR	NR	R1.0	C	3	0.54	0.37	0.16	0.78
22978	NR	NR	R3.0	I	1	0.55	0.44	0.08	1.19
22981	NR	NR	R1.0	C	4*	0.55 (2.19)	0.48	-0.62	0.56
22983	NR	NR	R1.0	C	4*	0.43 (1.70)	0.50	-0.09	0.53
26387	NR	NR	R2.0	C	1	0.63	0.32	-0.40	0.53
22948	NR	NR	R3.0	I	3	0.53	0.34	0.29	0.67
22934	NR	NR	R1.0	I	1	0.46	0.27	0.84	0.58
22946	NR	NR	R3.0	E	2	0.61	0.34	-0.21	0.55
22951	NR	NR	R1.0	C	4*	0.50 (2.01)	0.47	-0.28	0.42
22953	NR	NR	R2.0	I	4*	0.41 (1.64)	0.44	0.16	0.43
22855	NR	NR	R2.0	C	1	0.82	0.32	-1.75	0.58
22857	NR	NR	R3.0	E	4	0.55	0.38	0.09	0.78
22854	NR	NR	R1.0	I	3	0.47	0.22	0.96	0.40
22858	NR	NR	R3.0	C	4	0.62	0.39	-0.24	0.78
22859	NR	NR	R1.0	E	4*	0.54 (2.16)	0.49	-0.28	0.54
22860	NR	NR	R1.0	C	4*	0.54 (2.16)	0.51	-0.59	0.54
23069	1 (C)	1	R1.0	E	4	0.80	0.39	-1.26	0.75
23083	1 (C)	2	R3.0	C	4	0.83	0.38	-1.44	0.77
23082	1 (C)	3	R3.0	I	1	0.72	0.41	-0.86	0.76
23073	1 (C)	4	R1.0	I	4	0.35	0.30	1.25	0.83
23077	1 (C)	5	R2.0	I	1	0.81	0.32	-1.68	0.56
23076	1 (C)	6	R1.0	C	3	0.64	0.15	-0.64	0.21
23078	1 (C)	7	R2.0	C	1	0.65	0.26	-0.67	0.38
23084	1 (C)	8	R3.0	I	3	0.73	0.45	-0.80	0.93
23075	1 (C)	9	R1.0	I	1	0.71	0.27	-1.08	0.42
23079	1 (C)	10	R2.0	C	2	0.68	0.38	-0.58	0.65
23085	1 (C)	11	R1.0	C	4*	0.46 (1.82)	0.50	-0.10	0.55
23088	1 (C)	12	R2.0	I	4*	0.50 (2.00)	0.47	-0.04	0.52
20050	1 (D)	1	R2.0	I	1	0.60	0.27	-0.28	0.41
20047	1 (D)	2	R1.0	I	4	0.67	0.44	-0.48	0.92
20046	1 (D)	3	R1.0	I	2	0.67	0.36	-0.57	0.64
20051	1 (D)	4	R3.0	I	2	0.66	0.31	-0.50	0.56
20045	1 (D)	5	R2.0	C	4*	0.42 (1.67)	0.45	-0.15	0.48
20043	1 (D)	6	R1.0	C	4*	0.58 (2.31)	0.45	-0.65	0.47

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.

*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
22981	NR	NR	% of Students	0.52	0.52	11.05	57.68	27.81	2.41
			Parameters			-4.41	-2.40	0.84	3.49
22983	NR	NR	% of Students	1.32	1.08	42.79	40.12	11.51	3.17
			Parameters			-4.47	-0.30	1.60	2.82
22951	NR	NR	% of Students	1.17	1.05	31.42	32.42	31.01	2.93
			Parameters			-5.05	-0.88	0.69	4.10
22953	NR	NR	% of Students	2.74	1.96	41.69	42.10	7.91	3.60
			Parameters			-4.02	-0.21	2.01	2.92
22859	NR	NR	% of Students	0.70	0.86	15.53	49.36	32.19	1.35
			Parameters			-4.33	-1.82	0.70	4.34
22860	NR	NR	% of Students	0.97	0.67	16.45	51.49	25.58	4.83
			Parameters			-4.22	-1.80	0.84	2.82
23085	1 (C)	11	% of Students	0.94	0.70	36.97	40.94	18.88	1.57
			Parameters			-4.88	-0.56	1.26	3.77
23088	1 (C)	12	% of Students	1.40	0.57	21.99	51.44	23.28	1.33
			Parameters			-4.23	-1.38	1.22	4.25
20045	1 (D)	5	% of Students	1.14	0.58	45.65	38.88	11.03	2.72
			Parameters			-5.34	-0.18	1.80	3.13
20043	1 (D)	6	% of Students	1.03	0.32	11.84	44.36	39.60	2.85
			Parameters			-4.72	-2.27	0.31	4.09

Note. The total number of students is 34 971. 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
23025	NR	NR	R3.0	I	1	0.84	0.36	-1.80	0.66
23028	NR	NR	R2.0	C	4	0.74	0.28	-1.64	0.36
23031	NR	NR	R1.0	E	2	0.61	0.27	-0.37	0.38
23029	NR	NR	R1.0	I	1	0.50	0.30	0.49	0.55
23033	NR	NR	R2.0	I	4*	0.58 (2.30)	0.48	-1.01	0.45
23032	NR	NR	R1.0	C	4*	0.62 (2.48)	0.47	-1.29	0.46
19996	NR	NR	R1.0	E	2	0.83	0.20	-2.87	0.31
19999	NR	NR	R3.0	C	3	0.60	0.28	-0.21	0.44
19998	NR	NR	R2.0	C	1	0.80	0.37	-1.48	0.65
19997	NR	NR	R1.0	I	3	0.81	0.41	-1.55	0.71
19993	NR	NR	R1.0	C	4*	0.56 (2.24)	0.48	-1.09	0.47
19995	NR	NR	R1.0	C	4*	0.55 (2.20)	0.46	-1.11	0.39
23106	NR	NR	R1.0	I	1	0.85	0.33	-1.88	0.63
23109	NR	NR	R2.0	I	4	0.67	0.28	-0.74	0.44
23108	NR	NR	R1.0	C	2	0.39	0.28	1.09	0.68
23111	NR	NR	R3.0	I	3	0.80	0.39	-1.35	0.75
23117	NR	NR	R1.0	C	4*	0.55 (2.19)	0.43	-1.32	0.35
23116	NR	NR	R1.0	C	4*	0.49 (1.97)	0.46	-0.33	0.48
22651	1 (C)	1	R2.0	I	1	0.62	0.29	-0.36	0.46
22654	1 (C)	2	R3.0	C	3	0.82	0.37	-1.43	0.74
22649	1 (C)	3	R1.0	I	4	0.29	0.24	1.58	0.99
22650	1 (C)	4	R1.0	I	2	0.56	0.30	0.05	0.47
22652	1 (C)	5	R2.0	I	2	0.52	0.26	0.39	0.39
22653	1 (C)	6	R3.0	E	1	0.85	0.33	-2.03	0.59
22646	1 (C)	7	R1.0	E	1	0.83	0.37	-1.63	0.73
22655	1 (C)	8	R1.0	I	4	0.65	0.45	-0.41	0.95
22647	1 (C)	9	R1.0	I	3	0.86	0.39	-1.60	0.94
22657	1 (C)	10	R1.0	E	1	0.86	0.34	-1.92	0.68
22660	1 (C)	11	R2.0	I	4*	0.63 (2.53)	0.49	-1.29	0.48
22659	1 (C)	12	R1.0	I	4*	0.55 (2.21)	0.49	-0.81	0.55
23124	1 (D)	1	R3.0	C	3	0.78	0.36	-1.46	0.58
23125	1 (D)	2	R1.0	I	4	0.46	0.19	1.21	0.34
23121	1 (D)	3	R2.0	I	1	0.49	0.35	0.45	0.71
23119	1 (D)	4	R1.0	E	4	0.72	0.37	-0.97	0.61
23128	1 (D)	5	R2.0	I	4*	0.43 (1.72)	0.46	-0.15	0.48
23139	1 (D)	6	R1.0	C	4*	0.46 (1.83)	0.39	-0.61	0.36

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
23033	NR	NR	% of Students	0.55	0.60	15.24	41.42	36.81	5.38
			Parameters			-5.31	-2.11	0.17	3.20
23032	NR	NR	% of Students	0.59	0.18	6.76	42.29	44.13	6.04
			Parameters			-4.94	-3.15	-0.22	3.14
19993	NR	NR	% of Students	0.45	0.32	15.42	48.99	29.10	5.73
			Parameters			-5.54	-2.29	0.58	2.88
19995	NR	NR	% of Students	0.72	0.46	19.53	43.54	29.62	6.13
			Parameters			-6.14	-2.08	0.58	3.21
23117	NR	NR	% of Students	0.31	0.33	25.03	33.58	35.92	4.83
			Parameters			-8.05	-1.60	0.34	4.01
23116	NR	NR	% of Students	0.74	1.34	20.70	58.29	16.02	2.91
			Parameters			-4.42	-1.74	1.57	3.29
22660	1 (C)	11	% of Students	0.60	0.50	7.89	37.67	43.10	10.24
			Parameters			-4.52	-2.70	-0.32	2.36
22659	1 (C)	12	% of Students	0.68	0.35	13.22	53.68	27.89	4.16
			Parameters			-4.73	-2.16	0.72	2.93
23128	1 (D)	5	% of Students	0.83	0.76	39.41	45.77	11.44	1.78
			Parameters			-5.44	-0.64	1.90	3.59
23139	1 (D)	6	% of Students	0.78	0.22	38.04	40.19	18.18	2.59
			Parameters			-7.47	-0.79	1.62	4.20

Note. The total number of students is 37 020. 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
22748	NR	NR	A	L	1	0.79	0.41	-1.11	0.87
22750	NR	NR	C	I	3	0.48	0.35	0.56	0.80
22749	NR	NR	B	I	2	0.56	0.31	0.09	0.52
22751	NR	NR	C	L	4	0.78	0.37	-1.09	0.79
22752	NR	NR	A	L	4*	0.53 (2.12)	0.41	-0.40	0.49
22753	NR	NR	A	L	4*	0.51 (2.02)	0.33	-0.41	0.32
17629	NR	NR	A	I	2	0.55	0.37	0.22	0.76
17630	NR	NR	B	L	3	0.43	0.30	0.93	0.64
17631	NR	NR	C	E	4	0.60	0.28	-0.16	0.46
17632	NR	NR	C	I	3	0.71	0.47	-0.60	1.06
17633	NR	NR	A	I	4*	0.53 (2.10)	0.49	-0.44	0.71
17634	NR	NR	A	L	4*	0.57 (2.27)	0.48	-0.70	0.51
22773	NR	NR	C	E	2	0.55	0.33	0.18	0.62
22775	NR	NR	C	L	1	0.57	0.36	-0.01	0.62
22772	NR	NR	B	L	4	0.82	0.39	-1.36	0.83
22771	NR	NR	A	I	1	0.47	0.28	0.74	0.53
22777	NR	NR	A	I	4*	0.49 (1.95)	0.47	-0.18	0.57
22778	NR	NR	A	L	4*	0.57 (2.28)	0.44	-0.62	0.42
15750	1 (C)	1	B	L	2	0.59	0.41	-0.09	0.80
15749	1 (C)	2	B	I	4	0.56	0.30	0.22	0.56
15756	1 (C)	3	C	L	1	0.43	0.06	0.00	0.06
15754	1 (C)	4	C	I	2	0.57	0.42	0.02	0.88
15751	1 (C)	5	C	I	3	0.53	0.39	0.24	0.84
15753	1 (C)	6	C	L	2	0.76	0.33	-1.05	0.63
15755	1 (C)	7	C	L	4	0.65	0.43	-0.36	0.87
15748	1 (C)	8	A	I	3	0.41	0.30	0.97	0.74
15752	1 (C)	9	C	I	1	0.47	0.35	0.58	0.85
15747	1 (C)	10	A	E	1	0.74	0.37	-0.92	0.69
15758	1 (C)	11	A	I	4*	0.45 (1.78)	0.45	0.20	0.47
15757	1 (C)	12	A	L	4*	0.48 (1.91)	0.56	-0.29	0.65
26317	1 (D)	1	B	I	2	0.48	0.28	0.62	0.51
26316	1 (D)	2	A	I	1	0.47	0.35	0.60	0.81
26315	1 (D)	3	A	I	4	0.74	0.23	-1.40	0.37
26318	1 (D)	4	C	I	4	0.74	0.38	-0.86	0.73
26320	1 (D)	5	B	L	4*	0.43 (1.70)	0.38	0.38	0.43
26319	1 (D)	6	A	I	4*	0.48 (1.92)	0.38	-0.14	0.44

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
22752	NR	NR	% of Students	0.39	1.11	7.92	70.21	17.88	2.48
			Parameters			-3.89	-2.70	1.53	3.45
22753	NR	NR	% of Students	0.57	0.78	20.15	55.86	20.29	2.36
			Parameters			-6.17	-2.15	1.82	4.87
17633	NR	NR	% of Students	0.41	0.19	9.88	70.25	17.40	1.88
			Parameters			-3.96	-2.20	1.37	3.03
17634	NR	NR	% of Students	0.81	0.49	12.40	50.49	29.60	6.19
			Parameters			-4.12	-2.04	0.62	2.74
22777	NR	NR	% of Students	0.62	0.36	23.85	55.69	18.40	1.09
			Parameters			-4.64	-1.44	1.41	3.96
22778	NR	NR	% of Students	0.62	0.77	12.03	47.51	35.49	3.59
			Parameters			-4.73	-2.33	0.50	4.08
15758	1 (C)	11	% of Students	1.79	1.62	32.29	48.79	13.59	1.92
			Parameters			-3.82	-0.75	1.76	3.63
15757	1 (C)	12	% of Students	0.88	0.93	31.63	45.29	16.78	4.50
			Parameters			-3.95	-0.79	1.13	2.45
26320	1 (D)	5	% of Students	0.78	0.62	35.85	54.20	8.16	0.41
			Parameters			-5.62	-0.85	2.76	5.22
26319	1 (D)	6	% of Students	0.72	0.37	23.10	60.12	14.49	1.21
			Parameters			-5.44	-1.51	2.04	4.36

Note. The total number of students is 8107. 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
23393	NR	NR	C	I	3	0.78	0.32	-1.39	0.54
23409	NR	NR	B	L	4	0.62	0.25	-0.36	0.38
23207	NR	NR	A	I	1	0.90	0.30	-2.44	0.63
23477	NR	NR	C	I	4	0.69	0.23	-0.91	0.38
23211	NR	NR	A	L	4*	0.53 (2.12)	0.46	-1.16	0.49
23210	NR	NR	A	I	4*	0.48 (1.93)	0.56	-0.46	0.69
23487	NR	NR	A	I	1	0.75	0.41	-1.01	0.77
23407	NR	NR	B	L	3	0.45	0.25	0.99	0.41
23202	NR	NR	A	E	3	0.90	0.38	-1.97	0.94
24899	NR	NR	C	L	2	0.35	0.23	1.54	0.60
23201	NR	NR	A	L	4*	0.50 (2.00)	0.45	-0.41	0.51
23200	NR	NR	A	L	4*	0.52 (2.06)	0.39	-0.90	0.40
23154	NR	NR	A	E	2	0.86	0.30	-2.23	0.53
23482	NR	NR	B	I	3	0.75	0.31	-1.31	0.49
23177	NR	NR	C	I	4	0.63	0.40	-0.31	0.72
23175	NR	NR	A	I	1	0.55	0.45	0.07	0.99
23173	NR	NR	A	I	4*	0.58 (2.30)	0.48	-1.72	0.46
23174	NR	NR	A	L	4*	0.59 (2.34)	0.52	-1.10	0.52
23162	1 (C)	1	A	I	3	0.75	0.36	-1.06	0.63
23761	1 (C)	2	C	L	3	0.72	0.40	-0.80	0.70
23167	1 (C)	3	C	I	1	0.85	0.31	-1.88	0.61
23160	1 (C)	4	A	E	4	0.88	0.45	-1.62	1.10
23165	1 (C)	5	B	I	2	0.94	0.35	-2.47	0.97
23166	1 (C)	6	C	I	2	0.67	0.24	-0.62	0.42
23159	1 (C)	7	A	E	4	0.87	0.42	-1.52	1.01
23164	1 (C)	8	A	I	2	0.61	0.45	-0.26	0.85
23163	1 (C)	9	A	I	2	0.76	0.37	-1.12	0.65
23161	1 (C)	10	A	I	3	0.79	0.40	-1.25	0.79
23172	1 (C)	11	A	I	4*	0.54 (2.17)	0.50	-0.72	0.53
23171	1 (C)	12	A	L	4*	0.57 (2.26)	0.43	-0.99	0.57
23226	1 (D)	1	A	L	3	0.69	0.22	-1.13	0.31
23228	1 (D)	2	B	I	3	0.48	0.36	0.44	0.81
23227	1 (D)	3	A	I	2	0.60	0.31	-0.15	0.52
23229	1 (D)	4	C	L	1	0.59	0.33	-0.09	0.52
23220	1 (D)	5	A	I	4*	0.60 (2.41)	0.49	-1.28	0.63
23390	1 (D)	6	B	L	4*	0.49 (1.96)	0.45	-0.13	0.47

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. A = comprehension; B = organization; C = vocabulary and linguistic conventions; L = connections; E = explicit; I = implicit; 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
23211	NR	NR	% of Students	0.13	0.22	19.20	51.86	25.38	3.22
			Parameters			-6.90	-2.03	0.92	3.36
23210	NR	NR	% of Students	0.33	0.67	28.05	50.25	18.28	2.42
			Parameters			-4.80	-1.04	1.16	2.84
23201	NR	NR	% of Students	0.36	0.20	16.74	65.64	16.09	0.96
			Parameters			-5.48	-2.15	1.75	4.23
23200	NR	NR	% of Students	0.52	0.22	15.98	63.38	15.99	3.92
			Parameters			-6.10	-2.52	1.65	3.40
23173	NR	NR	% of Students	0.13	0.15	12.78	50.68	29.48	6.78
			Parameters			-7.45	-2.65	0.52	2.69
23174	NR	NR	% of Students	0.30	0.45	14.54	39.50	40.28	4.93
			Parameters			-5.59	-1.91	0.07	3.04
23172	1 (C)	11	% of Students	0.80	0.23	16.74	49.39	29.50	3.34
			Parameters			-4.91	-1.84	0.64	3.22
23171	1 (C)	12	% of Students	0.51	0.04	2.51	69.62	24.96	2.37
			Parameters			-4.57	-3.85	1.03	3.41
23220	1 (D)	5	% of Students	0.52	0.07	3.02	56.02	35.26	5.11
			Parameters			-4.52	-3.49	0.25	2.63
23390	1 (D)	6	% of Students	1.61	0.49	21.01	57.04	18.15	1.70
			Parameters			-4.41	-1.65	1.54	4.03

Note. The total number of students is 6903. NR = not released.

Table 7.1.9 Item Statistics: Primary Writing (English)

Item Code	Section	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
22940_T	NR	NR	W2.0	T	4*	0.52 (2.07)	0.52	-0.20	0.66
22940_V	NR	NR	W3.0	V	3*	0.68 (2.05)	0.58	-1.22	0.91
22984	NR	NR	W1.0	T	3	0.69	0.39	-0.65	0.71
22979	NR	NR	W3.0	V	4	0.58	0.35	-0.03	0.65
22986	NR	NR	W3.0	V	3	0.83	0.35	-1.67	0.66
22925	NR	NR	W1.0	T	2	0.65	0.35	-0.43	0.60
22804_T	C	13	W2.0	T	4*	0.58 (2.30)	0.60	-0.59	0.87
22804_V	C	13	W3.0	V	3*	0.66 (1.99)	0.61	-1.25	1.07
22969	C	14	W2.0	T	4	0.64	0.39	-0.41	0.68
22913	C	15	W3.0	V	1	0.72	0.28	-1.07	0.47
22971	C	16	W2.0	T	3	0.66	0.32	-0.59	0.51
22952	C	17	W1.0	T	4	0.42	0.35	0.78	0.79
25981_T	D	7	W2.0	T	4*	0.55 (2.19)	0.59	-0.44	0.74
25981_V	D	7	W3.0	V	3*	0.67 (2.00)	0.59	-1.27	0.98

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	Section	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
22940_T	NR	NR	% of Students	1.04	1.53	16.46	56.13	21.02	3.82
			Parameters			-2.88	-1.62	1.16	2.54
22940_V	NR	NR	% of Students	1.04	0.36	25.06	41.17	32.37	
			Parameters			-3.34	-0.85	0.52	
22804_T	C	13	% of Students	0.65	0.96	12.85	45.78	32.98	6.78
			Parameters			-3.01	-1.64	0.30	1.98
22804_V	C	13	% of Students	0.65	0.16	27.56	43.15	28.47	
			Parameters			-3.63	-0.78	0.66	
25981_T	D	7	% of Students	0.69	2.52	18.78	40.92	29.78	7.31
			Parameters			-2.85	-1.26	0.37	1.98
25981_V	D	7	% of Students	0.69	0.38	23.51	49.57	25.85	
			Parameters			-3.65	-1.00	0.84	

Note. The total number of students is 35 216. NR = not released.

Table 7.1.11 Item Statistics: Junior Writing (English)

Item Code	Section	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
16520	NR	NR	W1.0	T	3	0.62	0.25	-0.45	0.32
22719	NR	NR	W1.0	T	4	0.74	0.26	-1.38	0.40
22723	NR	NR	W3.0	V	2	0.79	0.41	-1.36	0.77
22725	NR	NR	W3.0	V	4	0.84	0.38	-1.78	0.74
19774_T	NR	NR	W2.0	T	4*	0.59 (2.36)	0.60	-0.81	0.81
19774_V	NR	NR	W3.0	V	3*	0.66 (1.97)	0.59	-1.25	1.08
17940	NR	NR	W3.0	V	3	0.94	0.32	-2.93	0.78
19802	C	16	W2.0	T	2	0.59	0.24	-0.26	0.29
22699	C	15	W3.0	V	2	0.72	0.39	-0.98	0.63
22736	C	14	W1.0	T	1	0.46	0.33	0.62	0.63
22685_T	C	13	W2.0	T	4*	0.56 (2.24)	0.56	-0.72	0.71
22685_V	C	13	W3.0	V	3*	0.68 (2.04)	0.58	-1.46	0.84
25991_T	D	7	W2.0	T	4*	0.56 (2.22)	0.61	-0.78	0.84
25991_V	D	7	W3.0	V	3*	0.62 (1.86)	0.60	-1.20	1.02

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	Section	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
19774_T	NR	NR	% of Students	0.59	0.86	13.56	41.8	33.43	9.75
			Parameters			-3.28	-1.74	0.13	1.65
19774_V	NR	NR	% of Students	0.59	0.16	19.01	62.76	17.48	
			Parameters			-3.59	-1.31	1.15	
22685_T	C	13	% of Students	0.77	0.53	15.66	47.94	28.34	6.76
			Parameters			-3.74	-1.73	0.54	2.05
22685_V	C	13	% of Students	0.77	0.32	21.83	49.18	27.89	
			Parameters			-3.79	-1.28	0.68	
25991_T	D	7	% of Students	0.63	0.29	20.22	42.08	29.45	7.32
			Parameters			-4.00	-1.31	0.34	1.86
25991_V	D	7	% of Students	0.63	0.18	28.7	53.71	16.78	
			Parameters			-3.86	-0.91	1.16	

Note. The total number of students is 37 192. NR = not released.

Table 7.1.13 Item Statistics: Primary Writing (French)

Item Code	Section	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
22815	NR	NR	A	T	2	0.76	0.36	-1.07	0.64
22904	NR	NR	A	T	3	0.81	0.37	-1.38	0.68
22916	NR	NR	C	V	1	0.87	0.27	-2.23	0.55
23403	NR	NR	B	T	4	0.62	0.36	-0.21	0.62
22814_T	NR	NR	A	T	4*	0.47 (1.88)	0.59	0.00	0.93
22814_V	NR	NR	C	V	3*	0.64 (1.91)	0.60	-0.84	1.20
22819	C	14	C	V	1	0.91	0.31	-2.21	0.77
22880	C	17	C	V	2	0.50	0.29	0.58	0.53
23418	C	16	B	T	4	0.84	0.39	-1.46	0.83
23447	C	15	A	T	3	0.73	0.32	-1.05	0.51
22811_T	C	13	A	T	4*	0.43 (1.72)	0.61	0.32	0.99
22811_V	C	13	C	V	3*	0.58 (1.75)	0.60	-0.78	1.17
25919_T	D	7	A	T	4*	0.43 (1.73)	0.62	0.22	0.94
25919_V	D	7	C	V	3*	0.57 (1.71)	0.59	-0.69	1.08

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	Section	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
22814_T	NR	NR	% of Students	0.88	1.45	28.09	51.39	15.69	2.50
			Parameters			-2.89	-0.78	1.19	2.48
22814_V	NR	NR	% of Students	0.88	0.95	21.83	59.86	16.48	
			Parameters			-2.77	-1.00	1.26	
22811_T	C	13	% of Students	0.55	3.08	37.85	42.42	14.68	1.41
			Parameters			-2.63	-0.26	1.27	2.90
22811_V	C	13	% of Students	0.55	0.31	35.89	50.82	12.43	
			Parameters			-3.34	-0.49	1.49	
25919_T	D	7	% of Students	0.62	4.19	37.5	40.54	14.49	2.66
			Parameters			-2.36	-0.33	1.20	2.36
25919_V	D	7	% of Students	0.62	0.48	38.01	49.19	11.70	
			Parameters			-3.27	-0.39	1.59	

Note. The total number of students is 8122. NR = not released.

Table 7.1.15 Item Statistics: Junior Writing (French)

Item Code	Section	Sequence	Expectation	Cognitive Skill	Answer Key/ Max. Score	CTT Item Statistics		IRT Item Parameters	
						Difficulty	Item-Total Correlation	Location	Slope
12460	NR	NR	B	T	3	0.87	0.37	-1.77	0.83
23292	NR	NR	A	T	1	0.58	0.33	-0.02	0.60
23559	NR	NR	A	T	2	0.75	0.37	-1.05	0.67
23561	NR	NR	C	V	3	0.51	0.39	0.27	0.86
26299_T	NR	NR	A	T	4*	0.49 (1.95)	0.54	-0.26	0.82
26299_V	NR	NR	C	V	3*	0.56 (1.69)	0.59	-0.85	1.02
20708	C	16	C	V	4	0.70	0.39	-0.70	0.73
23287	C	17	A	T	2	0.68	0.38	-0.60	0.69
23298	C	14	A	T	1	0.77	0.34	-1.36	0.55
26312	C	15	B	T	2	0.71	0.30	-0.99	0.45
23555_T	C	13	A	T	4*	0.52 (2.08)	0.61	-0.59	1.03
23555_V	C	13	C	V	3*	0.58 (1.74)	0.59	-1.17	1.05
23297_T	D	7	A	T	4*	0.58 (2.31)	0.62	-0.85	0.97
23297_V	D	7	C	V	3*	0.56 (1.68)	0.55	-1.19	0.84

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	Section	Sequence		Score Points					
				Missing	Illegible	10	20	30	40
26299_T	NR	NR	% of Students	0.52	0.29	23.33	57.21	16.87	1.77
			Parameters			-3.92	-1.20	1.23	2.85
26299_V	NR	NR	% of Students	0.52	0.20	39.04	50.62	9.62	
			Parameters			-3.77	-0.49	1.71	
23555_T	C	13	% of Students	0.55	0.22	18.34	56.20	21.68	3.01
			Parameters			-4.30	-1.29	0.89	2.34
23555_V	C	13	% of Students	0.55	0.06	37.17	49.61	12.62	
			Parameters			-4.59	-0.45	1.53	
23297_T	D	7	% of Students	0.62	0.25	12.22	47.91	33.10	5.90
			Parameters			-3.83	-1.79	0.26	1.96
23297_V	D	7	% of Students	0.62	0.09	41.85	46.38	11.07	
			Parameters			-4.91	-0.34	1.68	

Note. The total number of students is 6923. 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
22247	3(1)	1	PV1	KU	P	1	79.01	0.24	-1.75	0.42
22295	3(1)	2	PV2	TH	P	3	66.40	0.40	-0.47	0.77
12568	3(1)	3	NV2	AP	N	3	64.93	0.37	-0.47	0.64
16784	3(1)	4	GV2	KU	G	1	92.84	0.29	-2.54	0.74
19247	3(1)	5	MV2	AP	M	3	52.29	0.34	0.30	0.78
16839	3(1)	7	MV1	KU	M	4	56.86	0.38	0.07	1.11
22301	3(1)	8	DV1	AP	D	4*	80.50(3.22)	0.54	-2.29	0.40
22321	3(1)	9	MV2	TH	M	4*	81.00(3.24)	0.58	-2.41	0.50
10728	3(1)	13	PV2	AP	P	2	72.69	0.38	-0.88	0.71
22289	3(1)	16	MV1	AP	M	1	56.86	0.35	0.04	0.69
22352	NR	NR	NV3	TH	N	1	48.79	0.30	0.50	0.91
22286	NR	NR	NV3	AP	N	4	73.00	0.33	-0.96	0.61
16666	NR	NR	NV3	KU	N	1	72.85	0.36	-0.98	0.63
22644	NR	NR	NV1	AP	N	4*	65.50(2.62)	0.61	-1.64	0.48
13144	NR	NR	GV3	AP	G	3	63.87	0.29	-0.47	0.45
22254	NR	NR	GV1	TH	G	4*	55.25(2.21)	0.49	-1.19	0.44
22358	NR	NR	PV1	KU	P	2	86.04	0.33	-2.18	0.59
19226	NR	NR	DV2	AP	D	2	38.36	0.33	0.91	0.93
19259	3(2)	6	DV3	AP	D	1	71.26	0.34	-0.78	0.69
15096	3(2)	10	NV3	TH	N	4*	71.00(2.84)	0.57	-2.00	0.38
19252	3(2)	11	GV3	AP	G	4*	61.75(2.47)	0.47	-1.46	0.38
17235	3(2)	12	MV2	KU	M	4	81.22	0.35	-1.43	0.69
16846	3(2)	14	GV1	AP	G	4	50.59	0.40	0.26	0.94
22349	3(2)	15	NV1	TH	N	1	48.56	0.39	0.39	1.04
22250	3(2)	17	DV1	KU	D	3	80.63	0.35	-1.34	0.76
22235	3(2)	18	NV1	KU	N	2	78.31	0.48	-1.04	1.13
20966	NR	NR	NV3	AP	N	2	79.91	0.43	-1.22	0.88
22311	NR	NR	MV1	AP	M	3	70.89	0.41	-0.69	0.79
15099	NR	NR	MV1	KU	M	2	53.01	0.29	0.30	0.57
15107	NR	NR	MV2	TH	M	2	74.90	0.46	-0.85	0.99
11993	NR	NR	MV2	KU	M	1	66.48	0.38	-0.45	0.73
17424	NR	NR	GV1	KU	G	3	64.53	0.38	-0.44	0.73
22316	NR	NR	PV1	AP	P	1	75.63	0.42	-0.97	0.89
19254	NR	NR	PV2	KU	P	4	71.87	0.33	-0.90	0.58
26475	NR	NR	PV1	AP	P	4*	59.00(2.36)	0.55	-1.30	0.48
19269	NR	NR	DV2	TH	D	4*	61.25(2.45)	0.66	-1.10	0.67

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
22301	3(1)	8	% of Students	0.66	0.54	11.63	9.77	19.03	58.38
			Parameters			-5.19	-0.83	-1.50	-1.63
22321	3(1)	9	% of Students	0.29	0.1	9.38	12.02	22.25	55.95
			Parameters			-5.76	-1.40	-1.35	-1.11
22644	NR	NR	% of Students	0.78	0.14	27.32	17.24	17.74	36.78
			Parameters			-5.62	-0.25	-0.15	-0.53
22254	NR	NR	% of Students	0.51	0.28	25.94	36.80	24.26	12.21
			Parameters			-5.92	-0.94	0.61	1.50
15096	3(2)	10	% of Students	0.97	0.25	22.79	17.36	7.96	50.68
			Parameters			-5.96	-0.40	0.95	-2.59
19252	3(2)	11	% of Students	0.76	0.48	12.98	36.16	37.21	12.41
			Parameters			-5.40	-2.41	-0.14	2.13
26475	NR	NR	% of Students	0.71	0.18	23.36	30.16	29.80	15.79
			Parameters			-5.50	-1.04	0.02	1.33
19269	NR	NR	% of Students	0.91	0.23	25.96	24.34	24.06	24.51
			Parameters			-4.25	-0.68	-0.03	0.55

Note. The total number of students is 35 687. 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
22217	3(1)	1	MV2	AP	M	3	57.09	0.32	0.04	0.70
23597	3(1)	2	PV1	AP	P	1	89.34	0.40	-1.74	1.22
20522	3(1)	3	MV2	TH	M	2	71.09	0.38	-0.77	0.75
22223	3(1)	4	PV1	TH	P	1	63.61	0.41	-0.32	0.93
22219	3(1)	5	GV1	KU	G	2	57.61	0.35	-0.03	0.71
22379	3(1)	6	DV2	AP	D	2	61.47	0.33	-0.27	0.59
22470	3(1)	7	NV2	AP	N	2	88.03	0.40	-1.77	1.04
22386	3(1)	8	DV3	AP	D	4*	62.00(2.48)	0.66	-1.21	0.62
15035	3(1)	12	PV1	AP	P	4	76.40	0.33	-1.25	0.59
22369	3(1)	13	MV2	KU	M	3	52.38	0.29	0.31	0.62
22365	NR	NR	NV2	KU	N	4	72.48	0.37	-0.91	0.68
22276	NR	NR	NV1	AP	N	4*	54.25(2.17)	0.50	-0.87	0.35
22218	NR	NR	MV2	TH	M	3	34.94	0.23	1.24	0.91
15016	NR	NR	GV1	TH	G	3	72.50	0.38	-0.87	0.72
22343	NR	NR	GV3	TH	G	4*	60.75(2.43)	0.59	-1.34	0.50
12663	NR	NR	PV2	KU	P	1	75.90	0.29	-1.31	0.52
23606	NR	NR	PV1	TH	P	4*	80.75(3.23)	0.52	-2.73	0.39
12720	NR	NR	DV3	AP	D	2	62.09	0.52	-0.25	1.55
20635	3(2)	9	NV2	TH	N	4*	73.00(2.92)	0.54	-2.33	0.39
20469	3(2)	10	MV2	AP	M	4*	59.25(2.37)	0.65	-1.03	0.56
22384	3(2)	11	GV1	AP	G	4*	58.75(2.35)	0.50	-1.09	0.48
22225	3(2)	14	DV1	KU	D	2	66.55	0.32	-0.64	0.52
22260	3(2)	15	NV2	TH	N	3	60.77	0.44	-0.19	1.08
20492	3(2)	16	DV3	TH	D	2	56.05	0.33	0.13	0.92
15020	3(2)	17	GV3	AP	G	2	68.33	0.36	-0.68	0.66
22274	3(2)	18	NV1	KU	N	1	57.02	0.46	-0.07	1.02
22325	NR	NR	NV1	TH	N	3	78.43	0.41	-1.25	0.80
22214	NR	NR	NV3	AP	N	2	51.35	0.41	0.21	1.06
22330	NR	NR	MV1	AP	M	3	60.95	0.34	-0.27	0.58
20484	NR	NR	MV2	AP	M	3	44.85	0.25	0.88	0.55
22216	NR	NR	MV2	KU	M	3	63.81	0.40	-0.37	0.79
20500	NR	NR	GV3	TH	G	3	58.52	0.39	-0.08	0.85
23599	NR	NR	PV1	AP	P	3	76.62	0.33	-1.58	0.47
17141	NR	NR	PV2	KU	P	1	82.66	0.41	-1.40	0.90
22270	NR	NR	DV2	AP	D	4	81.31	0.45	-1.30	1.00
20529	NR	NR	DV2	TH	D	4*	55.00(2.20)	0.62	-0.84	0.54

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
22386	3(1)	8	% of Students	0.73	0.36	29.45	19.88	19.75	29.82
			Parameters			-4.44	-0.32	-0.13	0.07
22276	NR	NR	% of Students	2.35	0.92	33.27	22.15	25.52	15.79
			Parameters			-4.74	0.13	-0.19	1.32
22343	NR	NR	% of Students	0.78	0.32	23.61	30.76	20.64	23.89
			Parameters			-5.04	-1.03	0.39	0.30
23606	NR	NR	% of Students	0.27	0.13	12.24	12.34	14.13	60.89
			Parameters			-6.79	-1.18	-0.70	-2.24
20635	3(2)	9	% of Students	0.62	0.08	11.67	23.63	22.56	41.45
			Parameters			-5.98	-2.25	-0.33	-0.75
20469	3(2)	10	% of Students	1.56	0.51	28.90	25.39	16.87	26.78
			Parameters			-3.91	-0.59	0.37	0.02
22384	3(2)	11	% of Students	0.69	0.12	13.50	41.36	38.13	6.19
			Parameters			-5.09	-2.13	-0.01	2.86
20529	NR	NR	% of Students	2.11	0.56	33.54	26.54	16.06	21.18
			Parameters			-3.92	-0.35	0.60	0.32

Note. The total number of students in writing the Mathematics component of the English-language junior-division assessment is 36 989. 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
17780	3(1)	1	AA3	CC	A	2	86.93	0.33	-1.75	0.76
26538	3(1)	2	NA4	MA	N	2	82.71	0.44	-1.35	0.99
16440	3(1)	3	AA2	HP	A	3	77.54	0.36	-1.09	0.74
22119	3(1)	4	NA1	MA	N	4	74.03	0.47	-0.81	1.01
14619	3(1)	7	GA1	CC	G	2	60.59	0.34	-0.15	0.71
19784	3(1)	8	NA4	HP	N	4*	72.50(2.90)	0.52	-2.26	0.35
14654	3(1)	11	MA3	HP	M	4*	69.50(2.78)	0.54	-1.74	0.42
19721	3(1)	12	TA2	MA	T	2	53.58	0.33	0.24	0.80
19672	3(1)	14	GA2	MA	G	4	64.62	0.41	-0.29	0.86
23415	3(1)	15	MA2	MA	M	1	71.39	0.32	-0.81	0.59
12613	3(1)	16	AA1	CC	A	4	80.50	0.31	-1.54	0.54
22178	NR	NR	NA1	CC	N	1	87.22	0.32	-1.99	0.66
26537	NR	NR	NA4	CC	N	1	75.78	0.38	-0.96	0.76
14637	NR	NR	MA1	MA	M	3	73.90	0.49	-0.71	1.21
16422	NR	NR	MA4	CC	M	4	71.33	0.40	-0.64	0.85
22185	NR	NR	GA1	MA	G	4*	71.25(2.85)	0.52	-1.86	0.43
19782	NR	NR	TA1	HP	T	3	79.63	0.50	-1.03	1.18
19837	NR	NR	TA1	HP	T	4*	62.75(2.51)	0.53	-1.54	0.44
22124	3(2)	5	MA3	CC	M	2	74.52	0.41	-0.83	0.80
22191	3(2)	6	TA1	CC	T	1	92.21	0.27	-2.98	0.55
16385	3(2)	9	TA1	MA	T	4*	81.75(3.27)	0.35	-3.93	0.20
19705	3(2)	10	GA1	HP	G	4*	53.75(2.15)	0.41	0.13	0.32
14639	3(2)	13	MA1	HP	M	1	47.14	0.36	0.51	0.87
22156	3(2)	17	NA2	MA	N	3	81.19	0.39	-1.38	0.77
22158	3(2)	18	NA3	HP	N	3	67.80	0.42	-0.50	0.83
23458	NR	NR	NA2	HP	N	4	70.73	0.38	-0.79	0.63
19791	NR	NR	NA3	CC	N	4	94.17	0.33	-2.39	0.96
16322	NR	NR	NA3	MA	N	3	71.96	0.45	-0.70	0.94
14598	NR	NR	NA2	MA	N	4*	54.50(2.18)	0.50	-1.40	0.36
16380	NR	NR	MA2	CC	M	2	91.20	0.31	-2.40	0.70
12646	NR	NR	GA1	MA	G	3	74.74	0.37	-0.92	0.74
22064	NR	NR	GA1	HP	G	3	70.70	0.36	-0.66	0.74
20833	NR	NR	AA2	CC	A	2	75.10	0.45	-0.84	0.97
16331	NR	NR	AA3	MA	A	2	64.35	0.39	-0.26	0.88
22193	NR	NR	AA2	HP	A	4*	58.25(2.33)	0.47	-1.46	0.37
19816	NR	NR	TA1	MA	T	4	75.80	0.44	-0.83	1.01

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
19784	3(1)	8	% of Students	0.73	0.01	16.93	21.77	12.70	47.86
			Parameters			-6.43	-1.16	0.60	-2.04
14654	3(1)	11	% of Students	0.58	0.3	14.88	19.46	35.06	29.72
			Parameters			-5.22	-1.28	-1.02	0.57
22185	NR	NR	% of Students	0.78	0.11	9.28	30.16	23.05	36.62
			Parameters			-4.56	-2.54	0.03	-0.37
19837	NR	NR	% of Students	0.55	0.2	18.24	34.74	21.67	24.60
			Parameters			-5.53	-1.53	0.56	0.36
16385	3(2)	9	% of Students	0.54	0.07	14.16	3.24	21.40	60.59
			Parameters			-10.15	3.69	-6.19	-3.08
19705	3(2)	10	% of Students	1.28	3.43	17.43	38.89	36.46	2.50
			Parameters			-3.39	-1.82	0.21	5.53
14598	NR	NR	% of Students	0.92	0.5	43.24	18.64	9.69	27.01
			Parameters			-6.59	1.00	1.17	-1.18
22193	NR	NR	% of Students	0.66	0.34	15.40	50.09	16.82	16.68
			Parameters			-5.50	-2.65	1.83	0.49

Note. The total number of students is 8123. 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
12817	3(1)	1	MA1	CC	M	2	67.32	0.32	-0.63	0.56
20156	3(1)	3	GA2	CC	G	4	76.97	0.41	-1.11	0.78
20161	3(1)	4	TA2	HP	T	1	19.33	0.20	1.63	1.66
22486	3(1)	5	MA2	HP	M	1	72.20	0.44	-0.71	0.97
22478	3(1)	6	AA2	MA	A	4	78.10	0.32	-1.30	0.57
22491	3(1)	8	GA2	HP	G	4*	67.00(2.68)	0.50	-1.29	0.38
23315	3(1)	9	MA2	MA	M	4*	70.75(2.83)	0.57	-1.59	0.51
16293	3(1)	14	TA2	CC	T	1	66.96	0.40	-0.51	0.75
14675	NR	NR	NA1	MA	N	4	61.71	0.33	-0.32	0.53
22409	NR	NR	NA2	CC	N	3	80.89	0.25	-1.94	0.45
20116	NR	NR	NA3	MA	N	2	76.90	0.43	-0.95	0.94
11577	NR	NR	NA3	MA	N	4*	64.50(2.58)	0.54	-1.82	0.45
23663	NR	NR	MA1	MA	M	4	65.90	0.45	-0.35	1.02
13334	NR	NR	MA1	HP	M	3	57.27	0.42	0.04	1.16
20202	NR	NR	GA1	MA	G	3	74.25	0.37	-1.15	0.85
12795	NR	NR	AA1	CC	A	1	80.67	0.42	-1.23	0.89
20198	NR	NR	AA1	HP	A	4*	74.00(2.96)	0.42	-3.62	0.25
22405	NR	NR	TA1	HP	T	2	34.25	0.17	1.61	0.67
22441	3(2)	2	NA2	MA	N	1	73.70	0.52	-0.75	1.31
26478	3(2)	7	NA1	CC	N	1	56.97	0.43	0.02	1.03
22490	3(2)	10	AA2	MA	A	4*	73.00(2.92)	0.62	-1.58	0.56
23316	3(2)	11	NA1	HP	N	4*	63.50(2.54)	0.68	-1.23	0.72
11657	3(2)	12	TA1	MA	T	3	65.25	0.35	-0.46	0.63
18016	3(2)	13	AA1	MA	A	2	72.23	0.32	-0.89	0.59
14716	3(2)	15	GA2	HP	G	3	54.53	0.27	0.23	0.50
15927	3(2)	16	NA3	HP	N	4	46.71	0.39	0.46	1.06
11517	3(2)	17	MA3	MA	M	2	52.38	0.40	0.23	1.05
16340	3(2)	18	GA1	CC	G	4	70.94	0.47	-0.66	1.01
16353	NR	NR	MA2	CC	M	1	68.62	0.31	-0.71	0.51
22400	NR	NR	MA3	CC	M	3	79.04	0.28	-1.76	0.44
18314	NR	NR	GA1	HP	G	1	52.84	0.41	0.22	0.91
14714	NR	NR	GA2	CC	G	3	74.24	0.42	-0.93	0.80
20150	NR	NR	GA1	MA	G	4*	71.50(2.86)	0.64	-1.40	0.52
20104	NR	NR	AA2	HP	A	4	56.21	0.48	0.01	1.13
22508	NR	NR	TA2	MA	T	4	82.87	0.40	-1.46	0.81
20167	NR	NR	TA1	HP	T	4*	70.25(2.81)	0.57	-1.34	0.52

Note. The guessing parameter was set at a constant of 0.2 for multiple-choice items. CC = knowledge and understanding; MA = application; HP = thinking; 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
22491	3(1)	8	% of Students	0.98	1.24	16.12	8.13	58.00	15.52
			Parameters			-4.59	0.29	-3.31	2.46
23315	3(1)	9	% of Students	0.77	0.13	13.43	14.84	43.66	27.17
			Parameters			-4.48	-1.12	-1.57	0.80
11577	NR	NR	% of Students	0.16	0.26	13.86	38.27	22.38	25.07
			Parameters			-6.08	-2.11	0.58	0.34
20198	NR	NR	% of Students	0.25	0.13	17.82	18.30	12.38	51.11
			Parameters			-11.05	-0.83	0.82	-3.43
22490	3(2)	10	% of Students	1.11	0.22	6.28	33.52	16.36	42.51
			Parameters			-3.05	-2.88	0.40	-0.79
23316	3(2)	11	% of Students	0.59	0.12	23.89	27.39	17.09	30.93
			Parameters			-4.28	-0.93	0.28	0.02
20150	NR	NR	% of Students	1.37	1.06	20.88	11.83	17.58	47.28
			Parameters			-3.69	-0.22	-0.78	-0.90
20167	NR	NR	% of Students	1.78	0.38	6.28	28.31	35.36	27.89
			Parameters			-2.66	-2.70	-0.59	0.59

Note. The total number of students is 6912. 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. For gender-based DIF, negative values of Δ for multiple-choice items and negative effect-size values for open-response items indicate that the girls outperformed the boys; positive values of Δ for multiple-choice items and positive effect sizes for open-response items indicate that the boys outperformed the girls. For SLL-based DIF, negative values of Δ for multiple-choice items and negative effect sizes for open-response items indicate that the SLLs outperformed the non-SLLs; positive values of Δ for multiple-choice items and positive effect sizes for open-response items indicate that the non-SLLs outperformed the SLLs.

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

Item Code	Booklet (Section)	Sequence	Sample 1				Sample 2			
			Δ	Lower Limit	Upper Limit	DIF Level	Δ	Lower Limit	Upper Limit	DIF Level
22980	NR	NR	0.27	-0.08	0.63		0.29	-0.07	0.64	
22973	NR	NR	0.20	-0.11	0.52		0.12	-0.19	0.43	
22970	NR	NR	-0.27	-0.59	0.05		-0.61	-0.93	-0.29	
22978	NR	NR	0.19	-0.16	0.53		0.10	-0.24	0.44	
26387	NR	NR	-0.35	-0.68	-0.02		-0.01	-0.34	0.32	
22948	NR	NR	0.29	-0.03	0.61		0.24	-0.08	0.56	
22934	NR	NR	0.14	-0.17	0.46		0.23	-0.08	0.55	
22946	NR	NR	0.41	0.08	0.73		0.27	-0.06	0.60	
22855	NR	NR	0.64	0.23	1.05		0.87	0.47	1.28	
22857	NR	NR	0.63	0.30	0.96		0.78	0.45	1.12	
22854	NR	NR	-0.10	-0.41	0.21		-0.03	-0.33	0.28	
22858	NR	NR	-0.05	-0.39	0.30		0.19	-0.15	0.52	
23069	1 (C)	1	0.49	0.09	0.90		-0.20	-0.60	0.21	
23083	1 (C)	2	0.38	-0.05	0.81		0.07	-0.36	0.51	
23082	1 (C)	3	0.09	-0.27	0.46		-0.07	-0.44	0.30	
23073	1 (C)	4	-0.34	-0.67	0.00		-0.65	-0.99	-0.32	
23077	1 (C)	5	-0.06	-0.46	0.34		-0.37	-0.77	0.03	
23076	1 (C)	6	1.11	0.79	1.42	B+	1.21	0.89	1.52	B+
23078	1 (C)	7	-0.13	-0.45	0.19		-0.34	-0.66	-0.01	
23084	1 (C)	8	-0.21	-0.59	0.17		0.12	-0.26	0.51	
23075	1 (C)	9	0.03	-0.31	0.37		0.33	-0.01	0.68	
23079	1 (C)	10	0.25	-0.10	0.60		0.27	-0.08	0.61	
20050	1 (D)	1	-0.47	-0.79	-0.16		-0.12	-0.44	0.20	
20047	1 (D)	2	0.36	0.01	0.72		0.23	-0.13	0.58	
20046	1 (D)	3	0.54	0.19	0.88		0.55	0.21	0.89	
20051	1 (D)	4	0.45	0.12	0.78		0.17	-0.16	0.50	

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

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
22981	NR	NR	-0.16	0.00		-0.14	0.00	
22983	NR	NR	-0.10	0.01		-0.05	0.38	
22951	NR	NR	-0.03	0.57		-0.03	0.13	
22953	NR	NR	0.04	0.05		0.01	0.07	
22859	NR	NR	-0.08	0.05		-0.10	0.00	
22860	NR	NR	-0.04	0.14		-0.05	0.02	
23085	1 (C)	11	-0.07	0.00		-0.09	0.00	
23088	1 (C)	12	-0.02	0.19		-0.03	0.10	
20045	1 (D)	5	0.01	0.35		-0.04	0.36	
20043	1 (D)	6	-0.05	0.41		-0.03	0.41	

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
23025	NR	NR	-0.01	-0.43	0.42		-0.19	-0.62	0.25	
23028	NR	NR	-0.67	-1.03	-0.32		-0.90	-1.24	-0.55	
23031	NR	NR	-0.28	-0.60	0.04		-0.39	-0.70	-0.07	
23029	NR	NR	0.38	0.06	0.70		0.33	0.02	0.64	
19996	NR	NR	0.51	0.11	0.90		0.07	-0.32	0.47	
19999	NR	NR	0.88	0.56	1.20		0.51	0.20	0.83	
19998	NR	NR	0.71	0.31	1.11		0.43	0.03	0.82	
19997	NR	NR	0.00	-0.41	0.42		-0.56	-0.98	-0.15	
23106	NR	NR	-0.52	-0.96	-0.08		-0.77	-1.21	-0.32	
23109	NR	NR	-0.27	-0.60	0.05		-0.04	-0.37	0.29	
23108	NR	NR	0.56	0.24	0.88		0.66	0.34	0.98	
23111	NR	NR	-0.01	-0.42	0.40		0.11	-0.30	0.51	
22651	1 (C)	1	0.43	0.11	0.75		0.21	-0.12	0.53	
22654	1 (C)	2	1.72	1.29	2.15	C+	1.75	1.32	2.18	C+
22649	1 (C)	3	0.90	0.55	1.25		0.96	0.62	1.30	
22650	1 (C)	4	0.52	0.20	0.83		0.44	0.13	0.76	
22652	1 (C)	5	-0.02	-0.33	0.29		-0.04	-0.34	0.27	
22653	1 (C)	6	0.46	0.02	0.90		0.71	0.27	1.15	
22646	1 (C)	7	0.65	0.22	1.08		0.50	0.06	0.93	
22655	1 (C)	8	1.44	1.08	1.81	B+	1.61	1.24	1.98	C+
22647	1 (C)	9	-0.39	-0.86	0.08		-0.62	-1.09	-0.15	
22657	1 (C)	10	0.82	0.35	1.28		0.46	-0.01	0.92	
23124	1 (D)	1	0.68	0.30	1.07		0.58	0.20	0.96	
23125	1 (D)	2	0.21	-0.10	0.51		0.43	0.12	0.73	
23121	1 (D)	3	0.88	0.56	1.20		1.02	0.70	1.34	B+
23119	1 (D)	4	1.42	1.04	1.79	B+	1.49	1.12	1.86	B+

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
23033	NR	NR	-0.20	0.00	B-	-0.20	0.00	B-
23032	NR	NR	-0.15	0.00		-0.12	0.00	
19993	NR	NR	-0.14	0.00		-0.18	0.00	B-
19995	NR	NR	-0.09	0.00		-0.10	0.00	
23117	NR	NR	-0.17	0.00		-0.14	0.00	
23116	NR	NR	-0.03	0.51		-0.04	0.20	
22660	1 (C)	11	-0.07	0.05		-0.06	0.16	
22659	1 (C)	12	-0.08	0.00		-0.04	0.00	
23128	1 (D)	5	-0.10	0.00		-0.05	0.03	
23139	1 (D)	6	0.00	0.86		0.02	0.64	

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
22748	NR	NR	-0.51	-0.92	-0.11		-0.41	-0.81	-0.01	
22750	NR	NR	-0.01	-0.33	0.31		-0.02	-0.34	0.30	
22749	NR	NR	-0.05	-0.36	0.27		-0.18	-0.49	0.14	
22751	NR	NR	0.12	-0.28	0.51		-0.14	-0.53	0.25	
17629	NR	NR	0.03	-0.29	0.35		0.12	-0.21	0.44	
17630	NR	NR	0.09	-0.23	0.40		0.18	-0.14	0.50	
17631	NR	NR	0.44	0.12	0.76		0.56	0.24	0.88	
17632	NR	NR	0.07	-0.31	0.45		-0.13	-0.51	0.25	
22773	NR	NR	0.59	0.27	0.91		0.45	0.13	0.77	
22775	NR	NR	-0.28	-0.60	0.05		-0.11	-0.44	0.21	
22772	NR	NR	-0.53	-0.96	-0.09		-0.43	-0.86	0.00	
22771	NR	NR	0.13	-0.18	0.44		0.03	-0.28	0.34	
15750	1 (C)	1	-0.05	-0.38	0.29		-0.07	-0.40	0.27	
15749	1 (C)	2	0.07	-0.25	0.39		0.00	-0.32	0.31	
15756	1 (C)	3	0.32	0.02	0.62		0.65	0.35	0.95	
15754	1 (C)	4	0.33	-0.01	0.66		0.21	-0.12	0.55	
15751	1 (C)	5	0.40	0.07	0.73		0.29	-0.04	0.63	
15753	1 (C)	6	-0.44	-0.81	-0.07		-0.35	-0.72	0.02	
15755	1 (C)	7	-0.35	-0.70	0.00		-0.25	-0.60	0.09	
15748	1 (C)	8	-0.11	-0.43	0.21		0.09	-0.22	0.41	
15752	1 (C)	9	0.51	0.19	0.83		0.50	0.18	0.83	
15747	1 (C)	10	0.09	-0.28	0.45		-0.11	-0.48	0.25	
26317	1 (D)	1	-0.16	-0.47	0.15		-0.13	-0.44	0.18	
26316	1 (D)	2	0.37	0.05	0.69		0.43	0.11	0.75	
26315	1 (D)	3	-0.22	-0.57	0.12		-0.19	-0.53	0.16	
26318	1 (D)	4	-0.14	-0.51	0.23		-0.27	-0.63	0.10	

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
22752	NR	NR	-0.10	0.01		-0.09	0.02	
22753	NR	NR	-0.07	0.08		-0.08	0.02	
17633	NR	NR	-0.08	0.00		-0.07	0.00	
17634	NR	NR	-0.01	0.00		-0.03	0.05	
22777	NR	NR	-0.06	0.04		-0.05	0.02	
22778	NR	NR	-0.09	0.01		-0.11	0.00	
15758	1 (C)	11	0.01	0.71		-0.02	0.02	
15757	1 (C)	12	-0.01	0.10		-0.01	0.87	
26320	1 (D)	5	0.07	0.10		0.07	0.04	
26319	1 (D)	6	-0.03	0.73		-0.05	0.25	

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
23393	NR	NR	-0.21	-0.58	0.17		0.05	-0.33	0.43	
23409	NR	NR	0.80	0.48	1.12		0.65	0.33	0.97	
23207	NR	NR	0.96	0.44	1.49		0.71	0.17	1.24	
23477	NR	NR	-0.07	-0.41	0.26		-0.06	-0.39	0.28	
23487	NR	NR	-0.06	-0.44	0.33		0.05	-0.34	0.43	
23407	NR	NR	-0.33	-0.65	-0.02		-0.26	-0.57	0.05	
23202	NR	NR	-0.11	-0.67	0.44		-0.36	-0.92	0.19	
24899	NR	NR	0.91	0.58	1.23		1.04	0.70	1.37	B+
23154	NR	NR	0.17	-0.28	0.63		0.38	-0.08	0.83	
23482	NR	NR	-0.59	-0.96	-0.22		-0.49	-0.86	-0.12	
23177	NR	NR	0.32	-0.02	0.66		0.24	-0.10	0.58	
23175	NR	NR	-0.13	-0.47	0.21		-0.08	-0.42	0.27	
23162	1 (C)	1	-0.39	-0.76	-0.02		-0.28	-0.65	0.09	
23761	1 (C)	2	-0.07	-0.44	0.29		0.01	-0.35	0.38	
23167	1 (C)	3	0.85	0.40	1.30		1.10	0.64	1.55	B+
23160	1 (C)	4	0.79	0.26	1.32		0.46	-0.06	0.99	
23165	1 (C)	5	-0.67	-1.36	0.03		-0.60	-1.31	0.11	
23166	1 (C)	6	0.46	0.13	0.79		0.55	0.22	0.89	
23159	1 (C)	7	-0.05	-0.54	0.44		0.11	-0.37	0.60	
23164	1 (C)	8	0.32	-0.03	0.67		0.23	-0.13	0.58	
23163	1 (C)	9	-0.31	-0.68	0.07		0.01	-0.37	0.39	
23161	1 (C)	10	-0.09	-0.50	0.31		0.07	-0.35	0.48	
23226	1 (D)	1	-0.98	-1.32	-0.65		-1.02	-1.35	-0.68	B-
23228	1 (D)	2	0.13	-0.20	0.46		0.25	-0.08	0.58	
23227	1 (D)	3	0.29	-0.03	0.61		0.45	0.13	0.78	
23229	1 (D)	4	1.34	1.01	1.67	B+	1.26	0.92	1.59	B+

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
23211	NR	NR	-0.12	0.00		-0.13	0.00	
23210	NR	NR	-0.04	0.08		-0.06	0.02	
23201	NR	NR	-0.18	0.00	B-	-0.16	0.00	
23200	NR	NR	-0.16	0.00		-0.17	0.00	
23173	NR	NR	-0.08	0.00		-0.05	0.22	
23174	NR	NR	-0.02	0.58		-0.05	0.27	
23172	1 (C)	11	-0.04	0.00		-0.07	0.03	
23171	1 (C)	12	-0.02	0.16		0.02	0.44	
23220	1 (D)	5	-0.09	0.00		-0.10	0.00	
23390	1 (D)	6	-0.04	0.06		-0.06	0.29	

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
22980	NR	NR	0.70	0.36	1.05		0.61	0.27	0.95	
22973	NR	NR	0.05	-0.26	0.36		0.32	0.01	0.62	
22970	NR	NR	1.03	0.71	1.35	B+	1.10	0.78	1.42	B+
22978	NR	NR	0.27	-0.07	0.61		0.19	-0.15	0.52	
26387	NR	NR	0.27	-0.05	0.59		0.30	-0.03	0.62	
22948	NR	NR	0.18	-0.14	0.49		0.08	-0.24	0.39	
22934	NR	NR	0.07	-0.25	0.38		0.19	-0.12	0.50	
22946	NR	NR	0.30	-0.02	0.61		0.00	-0.32	0.31	
22855	NR	NR	0.62	0.23	1.00		0.61	0.22	1.00	
22857	NR	NR	1.18	0.86	1.50	B+	1.10	0.77	1.42	B+
22854	NR	NR	0.02	-0.28	0.33		-0.09	-0.39	0.22	
22858	NR	NR	-0.10	-0.43	0.23		-0.35	-0.69	-0.02	
23069	1 (C)	1	0.52	0.14	0.91		0.62	0.24	1.00	
23083	1 (C)	2	0.55	0.15	0.95		0.44	0.04	0.84	
23082	1 (C)	3	0.83	0.47	1.18		0.34	-0.02	0.69	
23073	1 (C)	4	0.38	0.04	0.72		0.49	0.15	0.82	
23077	1 (C)	5	0.67	0.30	1.05		0.92	0.53	1.31	
23076	1 (C)	6	0.92	0.61	1.22		0.98	0.67	1.28	
23078	1 (C)	7	-0.03	-0.36	0.29		0.02	-0.30	0.34	
23084	1 (C)	8	0.61	0.25	0.97		0.57	0.20	0.94	
23075	1 (C)	9	-0.12	-0.46	0.21		0.17	-0.17	0.50	
23079	1 (C)	10	0.14	-0.21	0.48		-0.04	-0.38	0.31	
20050	1 (D)	1	0.36	0.05	0.67		0.44	0.13	0.76	
20047	1 (D)	2	1.13	0.78	1.48	B+	1.09	0.74	1.44	B+
20046	1 (D)	3	0.52	0.19	0.86		0.78	0.45	1.12	
20051	1 (D)	4	0.22	-0.11	0.55		0.15	-0.17	0.48	

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
22981	NR	NR	-0.01	0.12		-0.03	0.40	
22983	NR	NR	-0.02	0.06		-0.02	0.77	
22951	NR	NR	-0.15	0.00		-0.15	0.00	
22953	NR	NR	-0.08	0.03		-0.08	0.00	
22859	NR	NR	-0.07	0.00		-0.03	0.00	
22860	NR	NR	-0.13	0.00		-0.10	0.00	
23085	1 (C)	11	-0.10	0.00		-0.07	0.00	
23088	1 (C)	12	-0.13	0.00		-0.16	0.00	
20045	1 (D)	5	-0.12	0.00		-0.13	0.00	
20043	1 (D)	6	-0.01	0.08		0.02	0.39	

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
23025	NR	NR	0.39	-0.03	0.82		0.48	0.05	0.90	
23028	NR	NR	0.30	-0.06	0.66		-0.05	-0.41	0.30	
23031	NR	NR	0.52	0.19	0.85		0.64	0.31	0.97	
23029	NR	NR	0.36	0.03	0.69		0.44	0.11	0.77	
19996	NR	NR	-0.24	-0.65	0.17		-0.26	-0.67	0.16	
19999	NR	NR	0.30	-0.03	0.63		0.56	0.23	0.89	
19998	NR	NR	1.44	1.05	1.83	B+	1.35	0.96	1.74	B+
19997	NR	NR	0.84	0.43	1.25		0.84	0.43	1.26	
23106	NR	NR	0.23	-0.19	0.66		0.06	-0.36	0.49	
23109	NR	NR	0.42	0.08	0.76		0.23	-0.10	0.57	
23108	NR	NR	1.07	0.72	1.41	B+	0.80	0.45	1.15	
23111	NR	NR	1.80	1.41	2.18	C+	1.99	1.60	2.38	C+
22651	1 (C)	1	-0.35	-0.69	-0.02		-0.44	-0.78	-0.10	
22654	1 (C)	2	2.34	1.94	2.73	C+	2.40	2.00	2.79	C+
22649	1 (C)	3	0.93	0.54	1.31		0.86	0.48	1.24	
22650	1 (C)	4	0.53	0.21	0.86		0.45	0.12	0.77	
22652	1 (C)	5	-0.16	-0.48	0.17		0.09	-0.23	0.42	
22653	1 (C)	6	0.23	-0.20	0.66		0.11	-0.32	0.54	
22646	1 (C)	7	0.85	0.45	1.26		1.17	0.75	1.58	B+
22655	1 (C)	8	-0.21	-0.58	0.15		0.16	-0.21	0.52	
22647	1 (C)	9	1.27	0.84	1.70	B+	1.70	1.25	2.14	C+
22657	1 (C)	10	0.08	-0.38	0.54		-0.27	-0.72	0.18	
23124	1 (D)	1	-0.06	-0.44	0.31		0.26	-0.12	0.65	
23125	1 (D)	2	0.17	-0.15	0.48		0.22	-0.10	0.54	
23121	1 (D)	3	-0.53	-0.87	-0.19		-0.78	-1.12	-0.44	
23119	1 (D)	4	-0.55	-0.93	-0.18		-0.30	-0.68	0.07	

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
23033	NR	NR	-0.05	0.60		-0.06	0.15	
23032	NR	NR	-0.10	0.00		-0.11	0.00	
19993	NR	NR	-0.12	0.00		-0.06	0.23	
19995	NR	NR	-0.02	0.08		0.00	0.10	
23117	NR	NR	-0.14	0.00		-0.14	0.00	
23116	NR	NR	-0.05	0.42		-0.01	0.07	
22660	1 (C)	11	-0.04	0.06		-0.03	0.30	
22659	1 (C)	12	-0.15	0.00		-0.09	0.00	
23128	1 (D)	5	-0.14	0.00		-0.11	0.00	
23139	1 (D)	6	-0.15	0.00		-0.15	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
22748	NR	NR	0.06	-0.35	0.48		0.32	-0.10	0.74	
22750	NR	NR	1.09	0.75	1.42	B+	1.20	0.86	1.53	B+
22749	NR	NR	0.02	-0.31	0.35		-0.23	-0.57	0.10	
22751	NR	NR	0.11	-0.29	0.51		0.19	-0.21	0.58	
17629	NR	NR	0.46	0.13	0.80		0.37	0.04	0.71	
17630	NR	NR	-0.06	-0.39	0.27		-0.39	-0.73	-0.06	
17631	NR	NR	0.15	-0.18	0.47		0.32	-0.01	0.65	
17632	NR	NR	0.30	-0.08	0.69		0.43	0.05	0.82	
22773	NR	NR	-0.03	-0.36	0.30		-0.15	-0.48	0.18	
22775	NR	NR	0.00	-0.34	0.34		0.02	-0.32	0.36	
22772	NR	NR	0.11	-0.33	0.54		0.04	-0.40	0.48	
22771	NR	NR	0.12	-0.20	0.45		-0.04	-0.36	0.29	
15750	1 (C)	1	0.30	-0.04	0.65		0.30	-0.05	0.65	
15749	1 (C)	2	0.15	-0.18	0.48		0.36	0.03	0.69	
15756	1 (C)	3	-0.20	-0.52	0.11		-0.29	-0.60	0.02	
15754	1 (C)	4	-0.14	-0.49	0.20		-0.17	-0.52	0.18	
15751	1 (C)	5	0.11	-0.23	0.45		-0.05	-0.39	0.29	
15753	1 (C)	6	0.10	-0.27	0.48		0.07	-0.31	0.45	
15755	1 (C)	7	0.18	-0.18	0.54		0.13	-0.23	0.49	
15748	1 (C)	8	-0.33	-0.66	0.00		-0.41	-0.75	-0.07	
15752	1 (C)	9	0.36	0.03	0.70		0.11	-0.22	0.45	
15747	1 (C)	10	0.06	-0.31	0.44		-0.08	-0.46	0.30	
26317	1 (D)	1	0.02	-0.30	0.35		0.07	-0.26	0.40	
26316	1 (D)	2	0.05	-0.29	0.38		0.08	-0.25	0.42	
26315	1 (D)	3	0.05	-0.30	0.41		0.09	-0.27	0.45	
26318	1 (D)	4	0.20	-0.17	0.57		0.47	0.09	0.84	

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
22752	NR	NR	0.02	0.34		0.03	0.84	
22753	NR	NR	0.03	0.78		0.03	0.33	
17633	NR	NR	0.01	0.09		-0.03	0.07	
17634	NR	NR	-0.02	0.95		-0.01	0.86	
22777	NR	NR	-0.09	0.01		-0.10	0.00	
22778	NR	NR	-0.12	0.00		-0.09	0.01	
15758	1 (C)	11	0.02	0.53		0.00	0.83	
15757	1 (C)	12	-0.01	0.68		0.01	0.56	
26320	1 (D)	5	-0.03	0.45		0.00	0.34	
26319	1 (D)	6	0.06	0.09		0.07	0.05	

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
23393	NR	NR	-0.13	-0.67	0.41		0.09	-0.45	0.64	
23409	NR	NR	-0.24	-0.70	0.21		-0.44	-0.89	0.02	
23207	NR	NR	0.77	0.05	1.49		0.87	0.16	1.58	
23477	NR	NR	-0.59	-1.07	-0.12		-0.61	-1.08	-0.13	
23487	NR	NR	0.25	-0.28	0.78		0.47	-0.06	1.01	
23407	NR	NR	-0.44	-0.89	0.02		-0.39	-0.84	0.07	
23202	NR	NR	-0.17	-0.91	0.57		0.48	-0.30	1.26	
24899	NR	NR	0.49	0.02	0.97		0.39	-0.09	0.86	
23154	NR	NR	0.64	0.00	1.28		0.27	-0.35	0.88	
23482	NR	NR	0.02	-0.49	0.54		0.09	-0.42	0.60	
23177	NR	NR	-0.04	-0.53	0.45		0.00	-0.49	0.48	
23175	NR	NR	0.19	-0.31	0.69		0.42	-0.07	0.91	
23162	1 (C)	1	0.27	-0.25	0.79		0.11	-0.41	0.63	
23761	1 (C)	2	0.11	-0.40	0.62		0.40	-0.10	0.91	
23167	1 (C)	3	0.10	-0.52	0.72		-0.10	-0.71	0.52	
23160	1 (C)	4	0.05	-0.66	0.76		0.10	-0.62	0.83	
23165	1 (C)	5	0.68	-0.29	1.65		0.32	-0.61	1.26	
23166	1 (C)	6	-0.26	-0.72	0.20		0.03	-0.43	0.49	
23159	1 (C)	7	0.64	-0.05	1.32		-0.01	-0.67	0.65	
23164	1 (C)	8	0.40	-0.10	0.90		0.28	-0.22	0.78	
23163	1 (C)	9	0.32	-0.20	0.85		0.40	-0.13	0.93	
23161	1 (C)	10	0.35	-0.21	0.90		0.78	0.22	1.34	
23226	1 (D)	1	0.85	0.39	1.32		0.99	0.53	1.46	
23228	1 (D)	2	0.10	-0.36	0.57		0.26	-0.20	0.73	
23227	1 (D)	3	0.02	-0.44	0.49		-0.12	-0.58	0.34	
23229	1 (D)	4	-0.04	-0.51	0.42		-0.01	-0.48	0.46	

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
23211	NR	NR	-0.01	0.25		-0.02	0.33	
23210	NR	NR	0.04	0.37		0.03	0.88	
23201	NR	NR	0.08	0.04		0.07	0.32	
23200	NR	NR	-0.01	0.54		0.00	0.66	
23173	NR	NR	-0.06	0.01		-0.09	0.02	
23174	NR	NR	-0.04	0.57		-0.03	0.39	
23172	1 (C)	11	0.05	0.33		0.07	0.12	
23171	1 (C)	12	0.06	0.33		0.06	0.33	
23220	1 (D)	5	-0.01	0.06		-0.03	0.35	
23390	1 (D)	6	-0.03	0.21		-0.02	0.73	

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
22925	NR	NR	-0.49	-0.84	-0.14		-0.59	-0.93	-0.24	
22979	NR	NR	-0.48	-0.81	-0.14		-0.03	-0.36	0.31	
22984	NR	NR	0.27	-0.11	0.64		0.15	-0.21	0.51	
22986	NR	NR	-0.29	-0.73	0.15		-0.01	-0.43	0.42	
22969	1(C)	14	-0.25	-0.61	0.10		-0.36	-0.71	0.00	
22913	1(C)	15	-0.07	-0.43	0.28		-0.09	-0.44	0.27	
22971	1(C)	16	-0.15	-0.49	0.19		-0.35	-0.69	-0.01	
22952	1(C)	17	-0.33	-0.66	0.01		-0.34	-0.68	0.00	

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

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
22940 T	NR	NR	0.04	0.08		0.00	0.97	
22940 V	NR	NR	-0.01	0.61		-0.02	0.24	
22804 T	1(C)	13	-0.07	0.03		-0.08	0.00	
22804 V	1(C)	13	0.01	0.08		-0.02	0.79	
25981 T	1(D)	7	-0.06	0.04		-0.04	0.29	
25981 V	1(D)	7	-0.02	0.19		-0.01	0.12	

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
16520	NR	NR	-0.65	-0.99	-0.32		-0.90	-1.24	-0.57	
22719	NR	NR	-0.48	-0.85	-0.11		-0.55	-0.92	-0.19	
22723	NR	NR	0.47	0.05	0.89		0.58	0.15	1.01	
22725	NR	NR	0.08	-0.40	0.56		-0.34	-0.81	0.12	
22736	1(C)	14	-0.92	-1.26	-0.58		-0.80	-1.13	-0.46	
22699	1(C)	15	-0.70	-1.09	-0.31		-0.62	-1.00	-0.24	
19802	1(C)	16	-0.08	-0.41	0.24		-0.29	-0.62	0.04	
17940	1(C)	17	-0.50	-1.22	0.22		-0.32	-1.04	0.40	

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

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
19774 T	NR	NR	0.01	0.23		0.00	0.10	
19774 V	NR	NR	-0.08	0.00		-0.04	0.05	
22685 T	1(C)	13	-0.01	0.28		-0.04	0.42	
22685 V	1(C)	13	-0.08	0.00		-0.12	0.00	
25991 T	1(D)	7	-0.01	0.11		-0.01	0.63	
25991 V	1(D)	7	-0.09	0.00		-0.13	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
22815	NR	NR	0.11	-0.46	0.69		-0.47	-1.03	0.08	
22904	NR	NR	-0.24	-0.86	0.38		-0.15	-0.76	0.46	
22916	NR	NR	-0.24	-0.92	0.43		-0.53	-1.24	0.18	
23403	NR	NR	-0.32	-0.82	0.17		0.08	-0.40	0.57	
22819	1(C)	14	-0.44	-1.25	0.36		-0.33	-1.16	0.51	
23447	1(C)	15	-0.02	-0.55	0.50		0.52	0.01	1.03	
23418	1(C)	16	0.18	-0.49	0.86		0.29	-0.39	0.96	
22880	1(C)	17	0.17	-0.29	0.63		0.00	-0.46	0.46	

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

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
22814 T	NR	NR	-0.03	0.17		0.07	0.21	
22814 V	NR	NR	-0.02	0.61		-0.04	0.51	
22811 T	1(C)	13	0.02	0.16		-0.02	0.45	
22811 V	1(C)	13	-0.03	0.85		-0.04	0.23	
25919 T	1(D)	7	-0.07	0.10		-0.05	0.04	
25919 V	1(D)	7	0.06	0.07		0.09	0.05	

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
12460	NR	NR	-0.67	-1.39	0.05		-0.33	-1.06	0.41	
23292	NR	NR	-0.70	-1.19	-0.21		-1.35	-1.84	-0.86	B-
23559	NR	NR	0.21	-0.35	0.77		-0.31	-0.85	0.24	
23561	NR	NR	-0.65	-1.14	-0.15		-0.47	-0.97	0.02	
23298	1(C)	14	-1.26	-1.84	-0.68	B-	-1.53	-2.12	-0.95	C-
26312	1(C)	15	-1.13	-1.65	-0.61	B-	-0.67	-1.18	-0.17	
20708	1(C)	16	0.96	0.43	1.48		0.82	0.3	1.33	
23287	1(C)	17	-0.62	-1.14	-0.10		-1.20	-1.72	-0.68	B-

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

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
26299 T	NR	NR	-0.03	0.77		0.03	0.04	
26299 V	NR	NR	-0.09	0.00		-0.13	0.00	
23555 T	C	13	-0.10	0.04		-0.16	0.00	
23555 V	C	13	-0.08	0.01		-0.19	0.00	B+
23297 T	D	7	-0.02	0.78		-0.07	0.08	
23297 V	D	7	0.01	0.08		-0.04	0.33	

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
22925	NR	NR	0.26	-0.25	0.77		0.18	-0.32	0.68	
22979	NR	NR	-0.06	-0.56	0.43		-0.12	-0.60	0.37	
22984	NR	NR	0.26	-0.27	0.79		0.47	-0.04	0.98	
22986	NR	NR	0.74	0.12	1.35		0.61	0.02	1.20	
22969	1(C)	14	0.18	-0.34	0.69		0.18	-0.33	0.70	
22913	1(C)	15	-0.42	-0.93	0.09		-0.37	-0.88	0.15	
22971	1(C)	16	0.32	-0.19	0.82		0.13	-0.37	0.64	
22952	1(C)	17	0.23	-0.27	0.72		-0.23	-0.74	0.27	

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

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
22940 T	NR	NR	-0.01	0.36		0.02	0.66	
22940 V	NR	NR	-0.04	0.48		0.00	0.99	
22804 T	C	13	0.00	0.98		0.00	0.79	
22804 V	C	13	-0.11	0.01		-0.05	0.45	
25981 T	D	7	0.05	0.04		0.06	0.23	
25981 V	D	7	-0.05	0.11		-0.08	0.02	

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
16520	NR	NR	0.76	0.29	1.24		0.14	-0.34	0.62	
22719	NR	NR	0.66	0.15	1.17		0.40	-0.12	0.92	
22723	NR	NR	1.07	0.48	1.66	B+	1.07	0.50	1.65	B+
22725	NR	NR	0.31	-0.34	0.96		-0.25	-0.91	0.40	
22736	1(C)	14	0.92	0.42	1.41		-0.12	-0.61	0.37	
22699	1(C)	15	0.77	0.23	1.31		0.11	-0.43	0.66	
19802	1(C)	16	0.60	0.13	1.07		-0.18	-0.66	0.31	
17940	1(C)	17	1.31	0.36	2.26	B+	1.20	0.33	2.07	B+

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

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
19774 T	NR	NR	-0.05	0.15		0.03	0.17	
19774 V	NR	NR	-0.07	0.11		-0.06	0.14	
22685 T	C	13	-0.11	0.01		0.04	0.32	
22685 V	C	13	-0.07	0.11		-0.04	0.33	
25991 T	D	7	-0.08	0.14		-0.04	0.48	
25991 V	D	7	-0.11	0.03		-0.05	0.07	

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
22815	NR	NR	-0.05	-0.53	0.43		-0.54	-1.00	-0.08	
22904	NR	NR	0.04	-0.49	0.56		-0.39	-0.89	0.11	
22916	NR	NR	0.47	-0.11	1.06		0.97	0.37	1.57	
23403	NR	NR	-0.07	-0.48	0.35		0.21	-0.21	0.63	
22819	1(C)	14	-0.73	-1.41	-0.05		0.30	-0.43	1.03	
23447	1(C)	15	0.01	-0.43	0.46		-0.42	-0.86	0.02	
23418	1(C)	16	0.11	-0.45	0.67		0.56	0.00	1.13	
22880	1(C)	17	-0.05	-0.45	0.35		-0.03	-0.42	0.37	

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

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
22814_T	NR	NR	0.01	0.55		0.03	0.49	
22814_V	NR	NR	0.01	0.64		0.02	0.30	
22811_T	C	13	0.02	0.23		0.05	0.01	
22811_V	C	13	-0.05	0.08		0.01	0.46	
25919_T	D	7	0.04	0.19		-0.01	0.58	
25919_V	D	7	-0.04	0.28		-0.06	0.17	

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
12460	NR	NR	-0.37	-1.05	0.31		-0.13	-0.80	0.54	
23292	NR	NR	0.27	-0.19	0.73		-0.01	-0.46	0.45	
23559	NR	NR	0.52	-0.01	1.05		-0.04	-0.56	0.47	
23561	NR	NR	-0.02	-0.49	0.45		-0.50	-0.97	-0.03	
23298	1(C)	14	0.26	-0.27	0.79		0.65	0.12	1.18	
26312	1(C)	15	0.20	-0.28	0.68		0.14	-0.33	0.60	
20708	1(C)	16	-0.18	-0.69	0.33		-0.24	-0.74	0.26	
23287	1(C)	17	0.42	-0.07	0.90		0.35	-0.13	0.83	

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

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
26299_T	NR	NR	0.00	0.94		0.03	0.44	
26299_V	NR	NR	0.02	0.63		0.06	0.06	
23555_T	C	13	-0.11	0.02		-0.05	0.54	
23555_V	C	13	-0.02	0.50		0.01	0.74	
23297_T	D	7	-0.03	0.80		-0.06	0.18	
23297_V	D	7	-0.01	0.11		0.00	0.73	

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
22247	3(1)	1	0.48	0.10	0.86		0.21	-0.16	0.59	
22295	3(1)	2	0.39	0.04	0.73		0.82	0.47	1.18	
12568	3(1)	3	0.06	-0.28	0.40		-0.11	-0.45	0.23	
16784	3(1)	4	-0.20	-0.80	0.40		-0.40	-1.01	0.21	
19247	3(1)	5	1.51	1.18	1.84	C+	1.47	1.14	1.79	B+
16839	3(1)	7	1.13	0.79	1.47	B+	1.06	0.72	1.40	B+
10728	3(1)	13	-0.66	-1.03	-0.29		-0.57	-0.93	-0.21	
22289	3(1)	16	-0.18	-0.50	0.15		-0.08	-0.40	0.24	
22352	NR	NR	-0.17	-0.49	0.15		0.29	-0.03	0.61	
22286	NR	NR	-0.64	-1.00	-0.28		-0.87	-1.22	-0.51	
16666	NR	NR	0.40	0.03	0.77		0.41	0.06	0.77	
13144	NR	NR	0.13	-0.19	0.45		-0.19	-0.52	0.13	
22358	NR	NR	-0.47	-0.94	0.00		-0.54	-0.99	-0.09	
19226	NR	NR	0.05	-0.27	0.37		0.13	-0.20	0.46	
19259	3(2)	6	0.09	-0.27	0.44		-0.20	-0.55	0.15	
17235	3(2)	12	0.74	0.32	1.16		0.36	-0.05	0.77	
16846	3(2)	14	-0.03	-0.36	0.31		0.01	-0.32	0.34	
22349	3(2)	15	0.96	0.62	1.29		1.13	0.79	1.46	B+
22250	3(2)	17	-0.36	-0.76	0.04		-0.27	-0.68	0.14	
22235	3(2)	18	-0.40	-0.83	0.03		-0.12	-0.53	0.30	
20966	NR	NR	0.02	-0.41	0.44		-0.34	-0.77	0.08	
22311	NR	NR	0.30	-0.07	0.67		0.07	-0.28	0.43	
15099	NR	NR	0.64	0.32	0.95		0.37	0.06	0.69	
15107	NR	NR	-0.38	-0.79	0.02		-0.13	-0.53	0.26	
11993	NR	NR	0.14	-0.21	0.49		-0.03	-0.37	0.31	
17424	NR	NR	-0.19	-0.53	0.15		-0.04	-0.38	0.29	
22316	NR	NR	-0.54	-0.93	-0.14		-0.84	-1.22	-0.46	
19254	NR	NR	0.21	-0.15	0.56		0.17	-0.18	0.52	

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
22301	3(1)	8	-0.09	0.00		-0.07	0.09	
22321	3(1)	9	-0.08	0.00		-0.06	0.06	
22644	NR	NR	0.07	0.00		0.04	0.15	
22254	NR	NR	-0.10	0.00		-0.06	0.03	
15096	3(2)	10	-0.02	0.15		-0.01	0.37	
19252	3(2)	11	-0.13	0.00		-0.14	0.00	
26475	NR	NR	-0.02	0.58		0.00	0.97	
19269	NR	NR	0.04	0.00		0.05	0.01	

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
22217	3(1)	1	-0.34	-0.66	-0.02		-0.12	-0.44	0.20	
23597	3(1)	2	0.05	-0.51	0.61		0.26	-0.30	0.81	
20522	3(1)	3	0.36	0.00	0.72		0.21	-0.15	0.58	
22223	3(1)	4	-0.24	-0.59	0.11		-0.02	-0.37	0.32	
22219	3(1)	5	-0.17	-0.50	0.15		-0.32	-0.65	0.01	
22379	3(1)	6	0.26	-0.07	0.58		0.44	0.10	0.77	
22470	3(1)	7	-0.37	-0.89	0.14		-0.36	-0.90	0.19	
15035	3(1)	12	0.00	-0.38	0.37		-0.10	-0.46	0.27	
22369	3(1)	13	0.35	0.04	0.67		0.47	0.15	0.79	
22365	NR	NR	-0.51	-0.86	-0.15		-0.13	-0.49	0.23	
22218	NR	NR	0.57	0.24	0.90		0.68	0.35	1.01	
15016	NR	NR	0.42	0.06	0.79		0.44	0.07	0.81	
12663	NR	NR	-0.37	-0.73	0.00		-0.91	-1.27	-0.55	
12720	NR	NR	0.48	0.09	0.86		0.72	0.34	1.10	
22225	3(2)	14	-0.50	-0.84	-0.17		-0.41	-0.74	-0.07	
22260	3(2)	15	0.33	-0.03	0.69		-0.23	-0.58	0.12	
20492	3(2)	16	-0.01	-0.34	0.31		0.30	-0.02	0.62	
15020	3(2)	17	-0.90	-1.25	-0.55		-0.72	-1.07	-0.37	
22274	3(2)	18	1.34	0.97	1.70	B+	0.96	0.60	1.31	
22325	NR	NR	-0.39	-0.80	0.01		-0.26	-0.68	0.15	
22214	NR	NR	0.04	-0.30	0.38		0.19	-0.14	0.53	
22330	NR	NR	-0.94	-1.27	-0.61		-0.64	-0.97	-0.32	
20484	NR	NR	-0.24	-0.55	0.07		-0.28	-0.59	0.03	
22216	NR	NR	0.12	-0.23	0.46		0.18	-0.17	0.52	
20500	NR	NR	0.21	-0.12	0.55		-0.13	-0.46	0.21	
23599	NR	NR	-0.17	-0.55	0.21		-0.33	-0.72	0.05	
17141	NR	NR	-0.18	-0.62	0.26		-0.27	-0.72	0.17	
22270	NR	NR	-0.63	-1.07	-0.19		-0.08	-0.53	0.38	

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
22386	3(1)	8	0.05	0.00		0.04	0.07	
22276	NR	NR	-0.08	0.00		-0.08	0.01	
22343	NR	NR	0.02	0.18		0.09	0.00	
23606	NR	NR	0.09	0.00		0.07	0.03	
20635	3(2)	9	0.01	0.01		-0.04	0.24	
20469	3(2)	10	-0.03	0.25		-0.03	0.50	
22384	3(2)	11	-0.07	0.03		-0.03	0.59	
20529	NR	NR	-0.04	0.00		-0.06	0.09	

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
17780	3(1)	1	0.58	0.11	1.05		0.76	0.29	1.23	
26538	3(1)	2	-0.55	-1.01	-0.09		-0.54	-0.99	-0.09	
16440	3(1)	3	-0.09	-0.48	0.30		-0.09	-0.47	0.29	
22119	3(1)	4	0.42	0.01	0.83		0.70	0.29	1.10	
14619	3(1)	7	-0.39	-0.72	-0.06		-0.05	-0.38	0.27	
19721	3(1)	12	0.29	-0.04	0.61		0.12	-0.20	0.44	
19672	3(1)	14	-0.17	-0.52	0.17		0.07	-0.28	0.41	
23415	3(1)	15	2.12	1.76	2.48	C+	1.74	1.38	2.10	C+
12613	3(1)	16	-0.39	-0.78	0.00		-0.81	-1.20	-0.42	
22178	NR	NR	-0.36	-0.83	0.11		0.07	-0.40	0.54	
26537	NR	NR	0.19	-0.20	0.57		0.28	-0.10	0.66	
14637	NR	NR	1.36	0.95	1.76	B+	1.73	1.32	2.14	C+
16422	NR	NR	0.41	0.04	0.78		0.70	0.33	1.06	
19782	NR	NR	-0.35	-0.80	0.11		-0.17	-0.62	0.27	
22124	3(2)	5	-0.40	-0.78	-0.03		-0.46	-0.84	-0.09	
22191	3(2)	6	-0.59	-1.16	-0.02		-0.77	-1.36	-0.17	
14639	3(2)	13	0.59	0.26	0.92		0.61	0.28	0.94	
22156	3(2)	17	0.50	0.07	0.92		0.79	0.37	1.21	
22158	3(2)	18	-0.02	-0.38	0.34		0.24	-0.11	0.60	
23458	NR	NR	0.88	0.52	1.24		1.08	0.71	1.44	B+
19791	NR	NR	0.67	-0.02	1.35		0.88	0.19	1.58	
16322	NR	NR	0.80	0.42	1.18		0.81	0.43	1.19	
16380	NR	NR	1.06	0.49	1.62	B+	0.78	0.24	1.33	
12646	NR	NR	-0.75	-1.12	-0.37		-0.37	-0.74	0.00	
22064	NR	NR	-0.85	-1.20	-0.50		-0.46	-0.81	-0.10	
20833	NR	NR	1.18	0.78	1.58	B+	1.05	0.66	1.45	B+
16331	NR	NR	0.54	0.19	0.88		0.68	0.34	1.03	
19816	NR	NR	-0.04	-0.43	0.35		-0.14	-0.53	0.26	

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
19784	3(1)	8	-0.09	0.01		-0.07	0.07	
14654	3(1)	11	-0.08	0.02		-0.07	0.03	
22185	NR	NR	-0.02	0.01		-0.05	0.00	
19837	NR	NR	0.00	0.61		-0.02	0.84	
16385	3(2)	9	-0.08	0.03		-0.05	0.06	
19705	3(2)	10	-0.12	0.00		-0.15	0.00	
14598	NR	NR	0.00	0.16		-0.05	0.00	
22193	NR	NR	-0.01	0.01		-0.04	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
12817	3(1)	1	0.69	0.35	1.03		0.65	0.31	0.99	
20156	3(1)	3	-0.44	-0.84	-0.04		-0.36	-0.76	0.03	
20161	3(1)	4	0.82	0.43	1.22		1.08	0.68	1.48	B+
22486	3(1)	5	-0.38	-0.76	0.00		-0.22	-0.59	0.16	
22478	3(1)	6	-0.18	-0.56	0.19		-0.13	-0.50	0.25	
16293	3(1)	14	0.80	0.45	1.16		0.60	0.25	0.95	
14675	NR	NR	-0.33	-0.66	-0.01		-0.12	-0.44	0.21	
22409	NR	NR	0.11	-0.28	0.49		0.02	-0.37	0.41	
20116	NR	NR	-0.35	-0.75	0.05		-0.12	-0.52	0.29	
23663	NR	NR	-0.05	-0.41	0.30		0.27	-0.09	0.63	
13334	NR	NR	0.48	0.13	0.82		0.25	-0.09	0.60	
20202	NR	NR	0.24	-0.13	0.61		0.11	-0.26	0.48	
12795	NR	NR	-0.33	-0.75	0.08		-0.26	-0.67	0.16	
22405	NR	NR	0.26	-0.06	0.58		0.10	-0.21	0.42	
22441	3(2)	2	-0.25	-0.66	0.16		0.02	-0.40	0.44	
26478	3(2)	7	1.64	1.29	1.99	C+	1.51	1.16	1.86	C+
11657	3(2)	12	-0.45	-0.78	-0.11		-0.62	-0.96	-0.29	
18016	3(2)	13	1.27	0.91	1.63	B+	1.37	1.01	1.73	B+
14716	3(2)	15	0.04	-0.27	0.35		0.03	-0.28	0.34	
15927	3(2)	16	0.12	-0.21	0.46		-0.10	-0.44	0.23	
11517	3(2)	17	0.25	-0.09	0.59		0.45	0.11	0.78	
16340	3(2)	18	-0.12	-0.49	0.26		-0.33	-0.71	0.04	
16353	NR	NR	-0.05	-0.38	0.29		0.05	-0.28	0.39	
22400	NR	NR	-0.06	-0.43	0.32		0.17	-0.21	0.54	
18314	NR	NR	0.16	-0.17	0.50		-0.02	-0.35	0.31	
14714	NR	NR	-0.72	-1.10	-0.33		-0.69	-1.07	-0.32	
20104	NR	NR	-0.20	-0.55	0.15		-0.08	-0.43	0.27	
22508	NR	NR	0.02	-0.41	0.45		-0.06	-0.49	0.37	

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
22491	3(1)	8	0.00	0.13		0.00	0.07	
23315	3(1)	9	0.01	0.00		0.01	0.05	
11577	NR	NR	-0.09	0.00		-0.10	0.00	
20198	NR	NR	-0.06	0.00		-0.05	0.06	
22490	3(2)	10	-0.09	0.00		-0.06	0.01	
23316	3(2)	11	0.03	0.00		0.03	0.00	
20150	NR	NR	0.01	0.01		0.00	0.16	
20167	NR	NR	-0.10	0.00		-0.12	0.00	

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
22247	3(1)	1	-0.27	-0.65	0.11		-0.17	-0.54	0.20	
22295	3(1)	2	0.16	-0.19	0.50		0.00	-0.34	0.34	
12568	3(1)	3	0.33	-0.01	0.67		0.27	-0.07	0.61	
16784	3(1)	4	0.37	-0.20	0.94		0.27	-0.29	0.82	
19247	3(1)	5	0.33	0.01	0.65		-0.19	-0.51	0.13	
16839	3(1)	7	0.50	0.17	0.84		0.54	0.21	0.87	
10728	3(1)	13	-0.21	-0.58	0.16		-0.24	-0.61	0.12	
22289	3(1)	16	0.59	0.27	0.92		0.65	0.33	0.98	
22352	NR	NR	-0.47	-0.79	-0.15		-0.52	-0.84	-0.20	
22286	NR	NR	-0.69	-1.05	-0.34		-0.51	-0.87	-0.15	
16666	NR	NR	-0.86	-1.23	-0.50		-0.73	-1.10	-0.36	
13144	NR	NR	-0.18	-0.51	0.15		-0.25	-0.58	0.07	
22358	NR	NR	0.26	-0.20	0.71		0.43	-0.02	0.88	
19226	NR	NR	0.07	-0.28	0.41		0.19	-0.16	0.53	
19259	3(2)	6	0.70	0.36	1.04		0.79	0.45	1.13	
17235	3(2)	12	-0.06	-0.46	0.34		0.35	-0.05	0.75	
16846	3(2)	14	-0.26	-0.60	0.08		0.04	-0.29	0.37	
22349	3(2)	15	-0.07	-0.40	0.27		0.22	-0.11	0.56	
22250	3(2)	17	0.06	-0.33	0.45		0.25	-0.14	0.64	
22235	3(2)	18	0.42	0.01	0.84		0.40	-0.02	0.81	
20966	NR	NR	0.14	-0.26	0.55		0.22	-0.18	0.63	
22311	NR	NR	-0.48	-0.84	-0.12		-0.40	-0.76	-0.04	
15099	NR	NR	-0.17	-0.49	0.14		0.09	-0.22	0.40	
15107	NR	NR	0.48	0.09	0.87		0.39	0.01	0.78	
11993	NR	NR	-0.27	-0.62	0.08		-0.32	-0.66	0.03	
17424	NR	NR	-0.11	-0.45	0.23		-0.16	-0.50	0.18	
22316	NR	NR	0.64	0.26	1.02		0.30	-0.07	0.67	
19254	NR	NR	-0.19	-0.54	0.15		-0.06	-0.40	0.29	

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
22301	3(1)	8	-0.04	0.54		-0.05	0.08	
22321	3(1)	9	0.04	0.00		0.08	0.00	
22644	NR	NR	0.13	0.00		0.09	0.00	
22254	NR	NR	0.04	0.22		0.03	0.80	
15096	3(2)	10	-0.02	0.26		-0.03	0.84	
19252	3(2)	11	-0.02	0.78		0.00	0.97	
26475	NR	NR	-0.05	0.01		-0.05	0.29	
19269	NR	NR	-0.05	0.00		-0.04	0.05	

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
22217	3(1)	1	0.19	-0.14	0.52		0.29	-0.04	0.61	
23597	3(1)	2	0.36	-0.17	0.89		0.32	-0.20	0.85	
20522	3(1)	3	-0.19	-0.56	0.18		-0.13	-0.50	0.24	
22223	3(1)	4	-0.06	-0.42	0.29		-0.09	-0.44	0.26	
22219	3(1)	5	0.31	-0.02	0.65		0.21	-0.12	0.55	
22379	3(1)	6	0.14	-0.20	0.47		0.43	0.09	0.76	
22470	3(1)	7	-0.09	-0.59	0.41		0.39	-0.11	0.89	
15035	3(1)	12	0.28	-0.09	0.65		0.23	-0.14	0.61	
22369	3(1)	13	-0.37	-0.70	-0.05		-0.31	-0.64	0.01	
22365	NR	NR	-0.44	-0.82	-0.07		-0.67	-1.04	-0.30	
22218	NR	NR	-0.04	-0.38	0.30		0.12	-0.22	0.46	
15016	NR	NR	-0.01	-0.37	0.36		0.33	-0.03	0.70	
12663	NR	NR	-0.28	-0.65	0.10		-0.17	-0.54	0.20	
12720	NR	NR	0.91	0.53	1.30		0.81	0.43	1.19	
22225	3(2)	14	-0.20	-0.54	0.14		-0.18	-0.52	0.16	
22260	3(2)	15	-0.15	-0.51	0.21		-0.26	-0.62	0.10	
20492	3(2)	16	-0.95	-1.28	-0.62		-0.97	-1.30	-0.64	
15020	3(2)	17	0.00	-0.35	0.36		0.08	-0.28	0.43	
22274	3(2)	18	-0.33	-0.69	0.02		-0.17	-0.53	0.19	
22325	NR	NR	-0.32	-0.74	0.10		-0.49	-0.89	-0.08	
22214	NR	NR	-0.72	-1.07	-0.37		-1.09	-1.44	-0.74	B-
22330	NR	NR	1.22	0.88	1.55	B+	1.28	0.95	1.62	B+
20484	NR	NR	-0.45	-0.78	-0.13		-0.21	-0.53	0.11	
22216	NR	NR	0.32	-0.03	0.66		0.39	0.04	0.73	
20500	NR	NR	-0.06	-0.40	0.28		-0.07	-0.41	0.27	
23599	NR	NR	0.14	-0.25	0.53		0.18	-0.20	0.56	
17141	NR	NR	-0.58	-1.02	-0.14		-0.23	-0.67	0.20	
22270	NR	NR	0.85	0.40	1.30		0.72	0.29	1.16	

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
22386	3(1)	8	0.09	0.00		0.05	0.00	
22276	NR	NR	-0.25	0.00	C-	-0.22	0.00	B-
22343	NR	NR	0.12	0.00		0.12	0.00	
23606	NR	NR	0.04	0.24		0.07	0.06	
20635	3(2)	9	0.03	0.23		0.06	0.05	
20469	3(2)	10	-0.03	0.11		-0.02	0.09	
22384	3(2)	11	0.13	0.00		0.12	0.00	
20529	NR	NR	-0.07	0.02		-0.11	0.00	

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
17780	3(1)	1	0.35	-0.13	0.82		-0.06	-0.53	0.40	
26538	3(1)	2	-0.23	-0.68	0.22		-0.33	-0.78	0.13	
16440	3(1)	3	-0.16	-0.55	0.22		0.04	-0.35	0.44	
22119	3(1)	4	-0.18	-0.57	0.21		0.00	-0.39	0.40	
14619	3(1)	7	0.09	-0.24	0.42		-0.05	-0.38	0.28	
19721	3(1)	12	0.46	0.14	0.79		0.42	0.10	0.74	
19672	3(1)	14	-0.56	-0.91	-0.21		-0.27	-0.62	0.08	
23415	3(1)	15	-0.10	-0.45	0.24		0.10	-0.25	0.45	
12613	3(1)	16	-0.17	-0.56	0.22		0.04	-0.35	0.43	
22178	NR	NR	0.23	-0.23	0.69		0.21	-0.25	0.68	
26537	NR	NR	-0.43	-0.82	-0.05		-0.36	-0.75	0.02	
14637	NR	NR	0.28	-0.12	0.67		0.42	0.03	0.81	
16422	NR	NR	-0.14	-0.50	0.23		-0.01	-0.38	0.35	
19782	NR	NR	0.39	-0.06	0.83		0.20	-0.24	0.64	
22124	3(2)	5	0.06	-0.32	0.44		0.24	-0.14	0.61	
22191	3(2)	6	0.12	-0.45	0.69		0.02	-0.54	0.58	
14639	3(2)	13	-0.14	-0.47	0.20		-0.05	-0.38	0.28	
22156	3(2)	17	0.34	-0.08	0.77		-0.05	-0.46	0.37	
22158	3(2)	18	0.16	-0.20	0.51		0.30	-0.06	0.66	
23458	NR	NR	0.42	0.07	0.78		0.34	-0.01	0.69	
19791	NR	NR	-0.17	-0.84	0.50		0.01	-0.67	0.70	
16322	NR	NR	-0.15	-0.53	0.23		-0.18	-0.56	0.20	
16380	NR	NR	0.63	0.09	1.17		0.66	0.12	1.21	
12646	NR	NR	0.15	-0.21	0.52		0.17	-0.19	0.54	
22064	NR	NR	0.19	-0.16	0.54		0.08	-0.27	0.43	
20833	NR	NR	-0.39	-0.79	0.01		-0.40	-0.80	-0.01	
16331	NR	NR	0.14	-0.20	0.48		-0.03	-0.37	0.32	
19816	NR	NR	-0.26	-0.65	0.14		-0.20	-0.59	0.20	

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
19784	3(1)	8	-0.01	0.14		-0.04	0.11	
14654	3(1)	11	0.00	0.95		0.03	0.41	
22185	NR	NR	-0.03	0.59		-0.06	0.08	
19837	NR	NR	-0.01	0.01		-0.01	0.45	
16385	3(2)	9	0.00	0.09		-0.01	0.14	
19705	3(2)	10	0.01	0.01		0.02	0.41	
14598	NR	NR	-0.03	0.87		-0.01	0.35	
22193	NR	NR	0.05	0.04		0.05	0.02	

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
12817	3(1)	1	-0.08	-0.51	0.35		0.25	-0.19	0.69	
20156	3(1)	3	0.07	-0.44	0.59		-0.17	-0.68	0.34	
20161	3(1)	4	-0.60	-1.12	-0.08		-0.34	-0.85	0.17	
22486	3(1)	5	0.43	-0.06	0.92		0.63	0.14	1.11	
22478	3(1)	6	-0.05	-0.54	0.44		-0.15	-0.63	0.34	
16293	3(1)	14	0.05	-0.40	0.51		-0.22	-0.68	0.24	
14675	NR	NR	0.15	-0.27	0.58		-0.05	-0.48	0.38	
22409	NR	NR	-0.13	-0.64	0.38		-0.41	-0.92	0.09	
20116	NR	NR	0.28	-0.24	0.79		0.03	-0.49	0.55	
23663	NR	NR	0.02	-0.45	0.49		0.39	-0.08	0.85	
13334	NR	NR	0.35	-0.10	0.80		0.57	0.12	1.02	
20202	NR	NR	-0.40	-0.88	0.07		-0.18	-0.67	0.31	
12795	NR	NR	0.44	-0.09	0.97		0.28	-0.25	0.82	
22405	NR	NR	-0.17	-0.59	0.25		-0.23	-0.65	0.19	
22441	3(2)	2	-0.12	-0.64	0.41		-0.28	-0.81	0.25	
26478	3(2)	7	-0.50	-0.95	-0.05		-0.52	-0.97	-0.08	
11657	3(2)	12	0.01	-0.41	0.44		0.19	-0.24	0.61	
18016	3(2)	13	-0.09	-0.54	0.36		-0.07	-0.52	0.39	
14716	3(2)	15	0.06	-0.35	0.46		0.34	-0.07	0.74	
15927	3(2)	16	0.21	-0.23	0.65		-0.03	-0.47	0.41	
11517	3(2)	17	0.48	0.03	0.92		0.59	0.14	1.03	
16340	3(2)	18	-0.41	-0.90	0.08		-0.67	-1.16	-0.18	
16353	NR	NR	-0.04	-0.48	0.41		-0.20	-0.64	0.24	
22400	NR	NR	0.18	-0.29	0.65		0.35	-0.12	0.83	
18314	NR	NR	0.25	-0.19	0.69		0.16	-0.28	0.59	
14714	NR	NR	0.40	-0.08	0.89		0.60	0.11	1.09	
20104	NR	NR	0.07	-0.40	0.54		0.18	-0.29	0.64	
22508	NR	NR	0.32	-0.22	0.86		0.28	-0.27	0.83	

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
22491	3(1)	8	-0.01	0.25		-0.03	0.07	
23315	3(1)	9	-0.03	0.17		-0.06	0.10	
11577	NR	NR	0.04	0.43		0.02	0.53	
20198	NR	NR	-0.03	0.16		-0.02	0.00	
22490	3(2)	10	0.00	0.93		0.00	0.12	
23316	3(2)	11	0.00	0.56		-0.03	0.69	
20150	NR	NR	-0.02	0.08		0.01	0.41	
20167	NR	NR	-0.03	0.32		-0.05	0.17	

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
21500	1	1.01	KU	N	3	65.14	0.42	-0.27	0.87
15517	2	1.03	AP	N	1	55.10	0.34	0.34	0.64
15545	3	1.05	TH	N	2	71.75	0.31	-0.68	0.59
18952	5	2.01	KU	N	2	67.32	0.20	-0.68	0.32
21556	6	2.03	AP	N	2	42.62	0.24	1.19	0.53
21505	9	2.01	AP	R	4	53.40	0.26	0.45	0.48
23365	10	2.02	AP	R	3	60.04	0.36	-0.12	0.68
19649	11	2.03	KU	R	4	82.54	0.25	-1.72	0.51
21581	13	3.01	TH	R	2	50.83	0.31	0.51	0.66
21507	14	3.05	AP	R	3	52.16	0.41	0.34	0.98
14818	15	4.01	KU	R	1	67.78	0.43	-0.42	0.96
14830	17	4.05	TH	R	1	31.15	0.32	1.45	0.99
21586	20	2.08	AP	N	4 [†]	49.97 (2.00)	0.53	-0.21	0.39
19625	22	3.04	AP	R	4 [†]	53.79 (2.15)	0.58	-0.81	0.55
21588	24	2.02	TH	M	4 [†]	56.94 (2.78)	0.53	-0.76	0.37
15560	26	1.02	KU	M	4	77.87	0.15	-2.32	0.27
15550	27	2.03	TH	M	3	34.39	0.30	1.32	0.89
15561	28	2.05	AP	M	3	34.68	0.31	1.24	0.78
22556	31	3.02	KU	M	1	78.22	0.29	-1.24	0.56
15597	32	1.06	TH	N	2	46.27	0.27	0.87	0.56
21555	NR	2.02	KU	N	4	65.22	0.37	-0.31	0.70
21568	NR	1.05	TH	N	4 [†]	86.56 (3.46)	0.45	-2.05	0.44
21576	NR	1.04	AP	R	3	66.84	0.22	-0.63	0.36
21560	NR	3.03	KU	R	4	32.74	0.31	1.34	1.01
19453	NR	4.03	TH	R	4	44.60	0.31	0.81	0.79
21564	NR	4.06	AP	R	4	32.73	0.24	1.63	0.74
21570	NR	1.01	AP	R	4 [†]	68.17 (2.73)	0.51	-1.55	0.37
14845	NR	4.02	TH	R	4 [†]	54.35 (2.17)	0.57	-0.86	0.53
10136	NR	2.05	TH	M	2	60.65	0.40	-0.04	0.86
22553	NR	3.01	AP	M	3	53.92	0.31	0.34	0.62
21534	NR	3.02	AP	M	4 [†]	52.38 (2.10)	0.53	-0.36	0.43

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
21586	20	% of Students	8.29	2.63	30.33	25.09	15.29	18.37
		Parameters			-2.22	0.04	0.95	0.38
19625	22	% of Students	0.69	0.37	35.15	26.09	22.92	14.78
		Parameters			-4.70	-0.06	0.33	1.21
21588	24	% of Students	4.19	1.59	34.42	13.52	18.85	27.44
		Parameters			-3.62	1.14	-0.45	-0.10
21568	NR	% of Students	1.23	0.31	7.38	8.79	7.91	74.39
		Parameters			-3.52	-1.24	-0.40	-3.02
21570	NR	% of Students	2.00	0.10	24.32	16.26	13.40	43.91
		Parameters			-4.89	0.11	0.19	-1.60
14845	NR	% of Students	1.00	0.12	35.43	24.39	23.08	15.99
		Parameters			-4.84	0.04	0.23	1.12
21534	NR	% of Students	6.34	0.76	23.29	39.49	13.20	16.92
		Parameters			-2.44	-1.04	1.69	0.33

Note. The total number of students is 13 868; 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
15517	2	1.03	AP	N	1	53.36	0.31	0.34	0.64
15545	3	1.05	TH	N	2	70.26	0.32	-0.68	0.59
21519	4	1.06	TH	N	1	49.04	0.36	0.57	0.82
18952	5	2.01	KU	N	2	66.32	0.19	-0.68	0.32
26863	7	2.06	KU	N	4	44.45	0.28	0.97	0.62
21540	8	1.01	AP	R	3	81.28	0.27	-1.53	0.52
23365	10	2.02	AP	R	3	64.12	0.35	-0.12	0.68
21506	12	3.01	KU	R	3	43.80	0.33	0.87	0.77
14818	15	4.01	KU	R	1	70.63	0.43	-0.42	0.96
12836	16	4.03	AP	R	3	70.80	0.30	-0.67	0.56
14830	17	4.05	TH	R	1	30.45	0.31	1.45	0.99
21545	18	4.06	TH	R	4	61.04	0.30	-0.05	0.55
21585	19	1.04	TH	N	4 [†]	65.75 (2063)	0.60	-0.78	0.50
19624	21	2.03	AP	R	4 [†]	60.64 (2.43)	0.47	-1.99	0.41
21551	23	4.01	TH	R	4 [†]	70.73 (2.83)	0.55	-1.16	0.48
21572	25	3.02	AP	M	4 [†]	56.15 (2.25)	0.52	-0.82	0.36
15560	26	1.02	KU	M	4	79.33	0.16	-2.32	0.27
15561	28	2.05	AP	M	3	38.94	0.33	1.24	0.78
19638	29	2.05	TH	M	3	51.41	0.23	0.67	0.46
22557	30	3.01	AP	M	2	63.31	0.22	-0.27	0.37
15543	NR	1.01	KU	N	2	50.11	0.40	0.47	0.95
21575	NR	2.04	AP	N	3	72.19	0.42	-0.55	0.92
21530	NR	2.02	AP	N	4 [†]	42.08 (1.68)	0.47	0.21	0.35
19424	NR	2.02	KU	R	4	67.37	0.27	-0.54	0.45
14809	NR	2.03	AP	R	3	37.86	0.24	1.38	0.68
21580	NR	3.03	AP	R	2	30.40	0.18	1.65	1.09
21543	NR	3.04	TH	R	1	43.32	0.38	0.77	1.02
21531	NR	3.05	AP	R	4 [†]	42.83 (1.71)	0.51	0.11	0.42
21584	NR	2.05	TH	M	2	48.91	0.39	0.52	0.99
22550	NR	3.01	KU	M	2	72.39	0.40	-0.59	0.84
19627	NR	2.03	TH	M	4 [†]	67.50 (2.70)	0.56	-1.02	0.55

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
21585	19	% of Students	5.66	1.58	19.15	16.68	17.23	39.70
		Parameters			-2.13	-0.36	-0.10	-0.55
19624	21	% of Students	0.10	0.01	18.74	35.55	29.65	15.94
		Parameters			-8.36	-1.39	0.32	1.47
21551	23	% of Students	2.63	0.81	16.27	14.47	25.62	40.21
		Parameters			-3.05	-0.50	-0.87	-0.21
21572	25	% of Students	4.67	0.94	35.21	21.90	3.58	33.71
		Parameters			-3.74	0.50	3.11	-3.17
21530	NR	% of Students	9.65	1.57	44.49	18.91	15.51	9.88
		Parameters			-2.83	1.39	0.72	1.56
21531	NR	% of Students	9.08	0.90	43.56	22.90	12.27	11.28
		Parameters			-2.64	0.83	1.28	0.98
19627	NR	% of Students	1.81	0.34	15.24	15.91	43.89	22.81
		Parameters			-3.33	-0.72	-1.20	1.18

Note. The total number of students is 14 510; 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
21610	1	1.03	AP	N	4	52.22	0.26	0.48	0.53
15688	4	2.07	KU	N	4	72.14	0.35	-0.68	0.68
21594	5	2.09	TH	N	4	56.98	0.32	0.13	0.58
21632	6	1.01	TH	R	2	46.40	0.40	0.55	1.07
21634	7	2.03	AP	R	2	83.36	0.30	-1.53	0.64
10263	9	2.05	TH	R	1	80.71	0.31	-1.23	0.63
23636	11	3.04	AP	R	4	69.87	0.38	-0.70	0.75
15680	12	2.02	AP	N	4 [†]	73.70 (2.95)	0.54	-1.40	0.38
19587	13	2.04	AP	R	4 [†]	72.71 (2.91)	0.56	-1.63	0.44
21661	15	2.04	AP	G	4 [†]	55.26 (2.21)	0.62	-0.48	0.60
26861	18	3.01	AP	M	4 [†]	66.13 (2.65)	0.60	-1.01	0.47
23637	19	1.01	KU	G	3	77.93	0.33	-1.09	0.74
21653	21	2.01	TH	G	1	51.73	0.34	0.40	0.75
19461	22	2.03	AP	G	1	56.26	0.41	0.11	0.85
21600	23	3.01	KU	G	4	73.79	0.31	-0.91	0.58
19563	24	3.04	AP	G	3	75.08	0.46	-0.75	1.10
15623	26	1.02	TH	M	2	66.11	0.38	-0.40	0.69
10287	27	1.04	TH	M	3	68.87	0.22	-0.78	0.39
21656	28	2.02	KU	M	4	81.54	0.38	-1.17	0.85
14954	30	3.01	KU	M	4	64.10	0.36	-0.30	0.70
21609	NR	1.02	KU	N	2	75.10	0.34	-0.89	0.69
21630	NR	2.03	AP	N	2	73.75	0.32	-0.90	0.60
14892	NR	2.02	KU	R	2	89.32	0.23	-2.40	0.53
14898	NR	3.03	KU	R	3	72.57	0.38	-0.72	0.75
21642	NR	3.01	TH	R	4 [†]	69.05 (2.76)	0.55	-1.48	0.45
21598	NR	2.02	AP	G	3	67.99	0.37	-0.47	0.72
26308	NR	3.02	TH	G	1	81.66	0.45	-1.03	1.14
15702	NR	3.05	TH	G	4 [†]	68.31 (2.73)	0.51	-1.45	0.44
22537	NR	2.03	AP	M	3	75.48	0.28	-1.14	0.50
22545	NR	3.01	AP	M	2	61.64	0.35	-0.14	0.64
21644	NR	2.06	TH	M	4 [†]	73.50 (2.94)	0.52	-1.72	0.41

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
15680	12	% of Students	4.01	0.85	17.59	13.10	6.82	57.64
		Parameters			-3.02	-0.19	0.76	-3.16
19587	13	% of Students	1.03	0.46	16.93	20.44	11.54	49.61
		Parameters			-4.37	-0.95	0.50	-1.72
21661	15	% of Students	4.74	0.50	25.85	31.51	17.42	19.98
		Parameters			-2.55	-0.63	0.69	0.55
26861	18	% of Students	4.51	0.67	22.93	17.46	11.01	43.41
		Parameters			-2.87	-0.23	0.45	-1.39
21642	NR	% of Students	1.09	0.41	18.29	21.57	19.81	38.84
		Parameters			-4.40	-0.87	-0.08	-0.57
15702	NR	% of Students	0.87	0.24	13.30	22.36	37.71	25.53
		Parameters			-4.45	-1.37	-0.87	0.88
21644	NR	% of Students	1.11	0.26	13.51	18.74	22.51	43.87
		Parameters			-4.42	-1.19	-0.53	-0.73

Note. The total number of students is 40 424; 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
15668	2	1.04	KU	N	3	60.39	0.35	0.01	0.71
21666	3	2.05	AP	N	4	36.61	0.30	1.11	1.22
15688	4	2.07	KU	N	4	71.71	0.37	-0.68	0.68
21633	8	2.04	KU	R	3	66.75	0.41	-0.31	0.84
10263	9	2.05	TH	R	1	78.53	0.33	-1.23	0.63
21635	10	3.03	KU	R	1	74.90	0.37	-0.76	0.77
23636	11	3.04	AP	R	4	75.76	0.37	-0.70	0.75
26868	14	3.02	TH	R	4 [†]	71.83 (2.87)	0.57	-1.90	0.53
14943	16	3.02	TH	G	4 [†]	63.62 (2.54)	0.70	-0.71	0.74
19608	17	2.03	TH	M	4 [†]	72.91 (2.92)	0.50	-1.55	0.48
23637	19	1.01	KU	G	3	80.57	0.37	-1.09	0.74
21652	20	1.03	AP	G	2	65.83	0.44	-0.24	0.98
19461	22	2.03	AP	G	1	58.21	0.40	0.11	0.85
19563	24	3.04	AP	G	3	77.69	0.47	-0.75	1.10
21637	25	3.05	TH	G	4	69.59	0.42	-0.43	0.90
15623	26	1.02	TH	M	2	67.99	0.36	-0.40	0.69
21602	29	2.04	AP	M	2	66.76	0.43	-0.28	0.94
14954	30	3.01	KU	M	4	66.59	0.37	-0.30	0.70
19005	31	3.01	AP	M	1	83.04	0.36	-1.23	0.82
21665	NR	1.01	AP	N	1	56.77	0.43	0.16	1.01
12913	NR	2.02	TH	N	2	66.61	0.28	-0.41	0.51
19567	NR	2.03	AP	N	4 [†]	81.11 (3.24)	0.53	-1.77	0.45
15653	NR	1.01	TH	R	3	78.17	0.24	-1.48	0.42
15250	NR	2.03	AP	R	4	57.67	0.45	0.11	1.03
26865	NR	2.03	AP	R	4 [†]	76.35 (3.05)	0.46	-2.84	0.38
12884	NR	2.04	TH	G	2	56.22	0.44	0.18	1.01
21673	NR	3.03	KU	G	3	91.72	0.27	-2.17	0.71
21624	NR	2.04	AP	G	4 [†]	82.62 (3.30)	0.52	-2.03	0.46
15265	NR	2.01	KU	M	4	79.05	0.41	-0.93	0.89
22546	NR	2.03	TH	M	2	78.39	0.38	-0.98	0.77
19591	NR	3.02	AP	M	4 [†]	63.25 (2.53)	0.65	-1.35	0.66

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
26868	14	% of Students	0.18	0.01	11.61	24.96	27.15	36.08
		Parameters			-5.78	-1.58	-0.30	0.06
14943	16	% of Students	3.31	0.63	26.10	15.31	20.85	33.81
		Parameters			-2.58	-0.12	-0.28	0.15
19608	17	% of Students	0.74	0.11	6.87	20.80	42.76	28.72
		Parameters			-3.82	-2.12	-1.11	0.85
19567	NR	% of Students	1.17	0.45	9.37	14.20	12.57	62.24
		Parameters			-3.54	-1.37	-0.22	-1.97
26865	NR	% of Students	0.08	0.01	10.49	14.99	32.80	41.63
		Parameters			-8.53	-1.26	-1.46	-0.13
21624	NR	% of Students	0.73	0.05	10.43	9.70	15.76	63.34
		Parameters			-4.63	-0.78	-1.04	-1.69
19591	NR	% of Students	0.26	0.02	30.54	17.81	18.67	32.71
		Parameters			-5.33	-0.05	-0.05	0.05

Note. The total number of students is 45 304; 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
20372	1	01	CC	N	3	88.29	0.10	-2.45	0.47
21884	3	08	CC	N	1	75.61	0.33	-0.83	0.69
14374	4	08	MA	N	1	58.05	0.38	0.43	0.67
14383	6	12	HP	N	2	49.27	0.19	1.00	0.37
21972	7	03	CC	R	4	34.15	0.47	0.79	1.94
14394	8	04	HP	R	4	62.44	0.36	0.11	0.73
21748	9	06	CC	R	1	66.34	0.45	-0.32	1.00
20442	11	14	MA	R	2	75.12	0.40	-0.61	1.02
22012	12	02	MA	N	4 [†]	68.17 (2.73)	0.49	-1.08	0.33
20429	15	07	MA	R	4 [†]	66.46 (2.66)	0.49	-0.88	0.41
22019	16	11	HP	R	4 [†]	51.59 (2.06)	0.59	-0.18	0.58
20366	17	04	MA	M	4 [†]	69.76 (2.79)	0.49	-1.57	0.35
14429	19	06	CC	M	2	77.56	0.20	-1.49	0.40
15320	21	05	HP	M	2	38.54	0.20	1.05	0.58
22005	22	14	MA	M	2	34.63	0.17	2.33	0.42
15303	24	17b	CC	M	1	78.54	0.27	-1.53	0.51
15292	NR	02	MA	N	4	42.93	0.27	1.15	0.57
10000	NR	05	CC	N	3	54.63	0.12	0.83	0.22
21990	NR	10	MA	N	4	50.73	0.36	0.52	0.76
20391	NR	10	HP	N	4 [†]	50.73 (2.03)	0.48	-0.34	0.37
20414	NR	03	MA	R	2	31.71	0.42	1.14	1.29
21750	NR	09	MA	R	3	55.12	0.58	0.18	1.86
9990	NR	12	CC	R	1	74.63	0.37	-0.58	0.64
21877	NR	18	HP	R	4	28.29	0.11	2.08	0.87
20426	NR	03	MA	R	4 [†]	50.00 (2.00)	0.58	-0.42	0.53
20395	NR	03	HP	M	1	63.90	0.33	-0.21	0.56
22002	NR	09	MA	M	3	72.68	0.31	-0.75	0.59
15369	NR	14	HP	M	2	60.49	0.33	0.01	0.61
20365	NR	17a	MA	M	2	75.12	0.38	-1.00	0.76
14454	NR	17a	HP	M	4	53.66	0.34	0.30	0.75
18496	NR	15	HP	M	4 [†]	66.22 (2.65)	0.52	-1.28	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 7.1.58 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
22012	12	% of Students	4.39	1.46	15.12	19.02	20.49	39.51
		Parameters			-2.50	-0.86	-0.18	-0.80
20429	15	% of Students	1.95	0.00	8.78	29.27	41.46	18.54
		Parameters			-3.10	-2.06	-0.24	1.87
22019	16	% of Students	3.41	0.98	25.37	39.02	21.95	9.27
		Parameters			-2.69	-0.73	0.91	1.79
20366	17	% of Students	0.98	0.98	17.56	22.44	15.61	42.44
		Parameters			-4.59	-0.91	0.53	-1.33
20391	NR	% of Students	3.41	2.44	27.80	39.02	12.20	15.12
		Parameters			-3.12	-0.73	2.11	0.38
20426	NR	% of Students	0.98	2.93	43.41	23.41	7.32	21.95
		Parameters			-3.41	0.50	1.62	-0.40
18496	NR	% of Students	1.95	0.49	17.56	31.22	10.24	38.54
		Parameters			-3.88	-1.34	1.66	-1.56

Note. The total number of students is 205; 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
20372	1	01	CC	N	3	88.05	0.25	-2.45	0.47
21692	2	02	MA	N	3	45.05	0.30	0.83	0.74
14374	4	08	MA	N	1	51.08	0.32	0.43	0.67
14381	5	12	MA	N	1	49.03	0.41	0.49	0.84
14394	8	04	HP	R	4	56.54	0.38	0.11	0.73
21748	9	06	CC	R	1	66.78	0.48	-0.32	1.00
13001	10	09	MA	R	3	61.09	0.38	-0.06	0.78
20442	11	14	MA	R	2	73.15	0.45	-0.61	1.02
22016	13	10	HP	N	4 [†]	53.67 (2.15)	0.51	-0.56	0.45
20369	14	01	MA	R	4 [†]	73.35 (2.93)	0.55	-1.36	0.48
20429	15	07	MA	R	4 [†]	62.12 (2.48)	0.45	-0.88	0.41
21684	18	13	HP	M	4 [†]	76.17 (3.05)	0.45	-2.09	0.31
21959	20	06	MA	M	3	58.36	0.26	0.11	0.45
15320	21	05	HP	M	2	45.16	0.27	1.05	0.58
22011	23	14	MA	M	1	57.34	0.43	0.12	0.99
15303	24	17b	CC	M	1	81.34	0.27	-1.53	0.51
15360	NR	06	CC	N	1	75.54	0.31	-1.06	0.55
9701	NR	05	CC	N	2	44.14	0.32	0.82	0.82
15365	NR	11	HP	N	4	70.53	0.29	-0.75	0.50
15307	NR	02	MA	N	4 [†]	66.87 (2.67)	0.48	-1.48	0.36
21886	NR	03	CC	R	1	61.43	0.34	-0.09	0.62
20414	NR	03	MA	R	2	34.81	0.31	1.14	1.29
9990	NR	12	CC	R	1	68.94	0.33	-0.58	0.64
21851	NR	15	HP	R	4	31.06	0.39	1.23	1.22
20448	NR	13	HP	R	4 [†]	48.49 (1.94)	0.46	-0.15	0.40
20435	NR	04	HP	M	3	67.46	0.48	-0.33	1.10
20364	NR	08	CC	M	3	71.22	0.32	-0.76	0.54
15279	NR	14	HP	M	2	41.18	0.33	0.98	0.79
20365	NR	17a	MA	M	2	79.07	0.36	-1.00	0.76
14454	NR	17a	HP	M	4 [†]	54.15 (2.79)	0.38	0.30	0.75

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 7.1.60 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
22016	13	% of Students	3.64	0.80	19.91	50.51	6.83	18.32
		Parameters			-2.87	-1.57	2.82	-0.61
20369	14	% of Students	1.37	1.14	9.56	20.59	26.73	40.61
		Parameters			-2.92	-1.72	-0.58	-0.23
20429	15	% of Students	2.28	0.57	11.83	33.45	37.77	14.11
		Parameters			-3.10	-2.06	-0.24	1.87
21684	18	% of Students	1.02	0.46	12.97	14.68	21.16	49.72
		Parameters			-5.12	-0.89	-0.93	-1.42
15307	NR	% of Students	1.48	0.46	16.15	26.62	23.09	32.20
		Parameters			-4.50	-1.37	0.15	-0.19
20448	NR	% of Students	3.07	0.23	35.72	29.81	26.05	5.12
		Parameters			-4.28	0.04	0.48	3.18
21787	NR	% of Students	2.05	1.25	17.97	19.00	15.47	44.25
		Parameters			-3.87	-0.65	0.21	-1.51

Note. The total number of students is 879; 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
21685	1	03	MA	N	3	76.74	0.39	-1.04	0.79
20253	2	07	CC	N	1	51.91	0.37	0.40	0.75
15424	3	10	MA	N	1	73.48	0.41	-0.45	0.96
15451	4	01	MA	R	2	86.74	0.21	-2.44	0.43
9681	6	07	CC	R	4	63.15	0.34	-0.18	0.66
15429	7	08	MA	R	2	49.33	0.42	0.39	1.27
20362	9	11	HP	R	3	61.46	0.30	-0.22	0.65
20269	11	20	MA	N	4 [†]	78.68 (3.15)	0.56	-1.72	0.47
18490	12	15	HP	R	4 [†]	60.31 (2.41)	0.55	-0.82	0.43
22714	14	07	MA	G	4 [†]	65.56 (2.62)	0.56	-1.04	0.48
15441	16	20a	HP	M	4 [†]	70.84 (2.83)	0.52	-1.49	0.42
23611	17	03	CC	G	4	72.02	0.40	-0.54	0.82
20340	19	07	HP	G	2	60.79	0.29	-0.06	0.55
20301	20	09	HP	G	4	55.73	0.39	0.07	0.95
15391	21	03	HP	M	2	53.60	0.31	0.37	0.57
22603	22	12	CC	M	4	47.98	0.26	0.83	0.62
14544	23	17	CC	M	3	61.80	0.34	-0.18	0.60
15241	25	20b	CC	M	4	74.38	0.34	-0.87	0.63
9679	NR	08	CC	N	1	76.63	0.35	-0.95	0.70
14479	NR	16	MA	N	1	76.52	0.37	-1.19	0.57
20275	NR	20	HP	N	4	60.56	0.41	-0.20	0.70
20330	NR	03	MA	N	4 [†]	63.09 (2.52)	0.51	-0.99	0.42
20336	NR	03	HP	R	1	39.44	0.33	0.97	0.97
20338	NR	12	MA	R	1	81.35	0.32	-1.32	0.66
20307	NR	04	MA	R	4 [†]	76.49 (3.06)	0.56	-1.44	0.51
26857	NR	04	MA	G	2	65.06	0.45	-0.24	0.99
21946	NR	10	CC	G	4	48.76	0.45	0.44	1.02
15436	NR	09	CC	M	4	86.18	0.26	-2.05	0.52
15423	NR	17	HP	M	2	56.97	0.30	0.19	0.57
15456	NR	20a	MA	M	1	67.53	0.33	-0.47	0.59
22030	NR	05	HP	M	4 [†]	69.02 (2.76)	0.43	-1.60	0.32

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 7.1.62 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
20269	11	% of Students	0.90	0.67	12.81	14.16	12.25	59.21
		Parameters			-3.91	-0.96	-0.18	-1.83
18490	12	% of Students	3.60	1.80	23.71	26.97	12.13	31.80
		Parameters			-3.42	-0.41	1.05	-0.48
22714	14	% of Students	2.36	1.01	16.07	28.76	18.54	33.26
		Parameters			-3.01	-1.30	0.46	-0.29
15441	16	% of Students	1.01	0.79	7.75	33.71	18.76	37.98
		Parameters			-3.20	-2.73	0.63	-0.66
20330	NR	% of Students	1.57	1.57	12.25	38.20	21.91	24.49
		Parameters			-2.92	-2.12	0.75	0.32
20307	NR	% of Students	1.69	0.90	5.73	25.06	16.40	50.22
		Parameters			-2.26	-2.56	0.15	-1.08
22030	NR	% of Students	1.24	0.90	8.09	33.93	23.26	32.58
		Parameters			-3.43	-3.20	0.57	-0.32

Note. The total number of students is 890; 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
21685	1	03	MA	N	3	80.08	0.37	-1.04	0.79
20253	2	07	CC	N	1	52.31	0.35	0.40	0.75
15424	3	10	MA	N	1	68.82	0.43	-0.45	0.96
15447	5	04	HP	R	4	54.43	0.29	0.36	0.53
15429	7	08	MA	R	2	49.48	0.50	0.39	1.27
21770	8	14	MA	R	2	91.98	0.28	-2.13	0.77
20362	9	11	HP	R	3	64.64	0.35	-0.22	0.65
20331	10	05	MA	N	4 [†]	56.30 (2.25)	0.50	-0.97	0.40
18490	12	15	HP	R	4 [†]	58.04 (2.32)	0.56	-0.82	0.43
20346	13	04	MA	G	4 [†]	70.79 (2.83)	0.55	-1.38	0.50
21968	15	06	HP	M	4 [†]	60.74 (2.43)	0.59	-0.88	0.51
23611	17	03	CC	G	4	70.17	0.40	-0.54	0.82
14510	18	08	MA	G	4	67.75	0.36	-0.43	0.68
20301	20	09	HP	G	4	59.42	0.42	0.07	0.95
22603	22	12	CC	M	4	45.50	0.31	0.83	0.62
14544	23	17	CC	M	3	63.53	0.32	-0.18	0.60
14549	24	18	HP	M	3	54.97	0.22	0.42	0.40
21920	26	20a	MA	M	1	78.33	0.37	-0.97	0.79
14472	NR	09	CC	N	4	47.69	0.44	0.49	1.11
14479	NR	16	MA	N	1	78.46	0.27	-1.19	0.57
20275	NR	20	HP	N	4	64.44	0.36	-0.20	0.70
20289	NR	13	MA	N	4 [†]	69.94 (2.80)	0.50	-1.41	0.45
9942	NR	02	MA	R	1	78.70	0.39	-0.98	0.83
20277	NR	06	CC	R	2	69.94	0.23	-0.81	0.40
20287	NR	03	MA	R	4 [†]	77.82 (3.11)	0.58	-1.67	0.54
20260	NR	05	HP	G	2	66.87	0.47	-0.32	1.02
21898	NR	09	CC	G	4	53.45	0.39	0.31	0.83
9814	NR	04	HP	M	3	66.13	0.29	-0.40	0.53
21933	NR	06	CC	M	2	76.95	0.34	-0.97	0.69
9838	NR	20a	CC	M	4	73.00	0.34	-0.76	0.64
15399	NR	20a	HP	M	4 [†]	72.40 (2.90)	0.51	-1.28	0.35

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 7.1.64 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
20331	10	% of Students	1.11	1.08	23.93	42.60	9.03	22.24
		Parameters			-4.34	-1.20	2.39	-0.71
18490	12	% of Students	2.83	1.04	27.91	27.40	13.82	27.00
		Parameters			-3.42	-0.41	1.05	-0.48
20346	13	% of Students	0.98	0.57	9.61	30.40	21.03	37.41
		Parameters			-3.37	-2.07	0.24	-0.32
21968	15	% of Students	1.04	1.45	26.12	21.71	25.28	24.40
		Parameters			-3.70	-0.27	-0.15	0.60
20289	NR	% of Students	1.01	0.30	7.95	27.40	36.33	27.00
		Parameters			-3.55	-2.32	-0.54	0.77
20287	NR	% of Students	0.67	0.34	12.20	16.38	15.30	55.11
		Parameters			-4.02	-1.18	-0.29	-1.20
15399	NR	% of Students	2.22	3.34	14.66	14.19	15.81	49.78
		Parameters			-2.57	-0.53	-0.36	-1.67

Note. The total number of students is 2967; 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
21500	11101	1	0.08	-0.26	0.43		0.34	0.00	0.69	
15517	11101	2	0.28	-0.03	0.60		0.53	0.21	0.84	
15545	11101	3	0.49	0.14	0.83		0.29	-0.06	0.64	
18952	11101	5	-0.90	-1.22	-0.58		-0.95	-1.28	-0.63	
21556	11101	6	0.21	-0.10	0.53		0.19	-0.12	0.50	
21505	11101	9	0.36	0.05	0.67		0.03	-0.28	0.34	
23365	11101	10	0.57	0.23	0.90		0.55	0.23	0.88	
19649	11101	11	-0.13	-0.52	0.26		0.05	-0.36	0.45	
21581	11101	13	0.17	-0.15	0.49		0.17	-0.14	0.49	
21507	11101	14	-0.52	-0.85	-0.19		-0.40	-0.73	-0.08	
14818	11101	15	0.06	-0.30	0.42		-0.02	-0.38	0.33	
14830	11101	17	0.36	0.02	0.71		-0.02	-0.37	0.32	
15560	11101	26	-0.79	-1.15	-0.42		-0.53	-0.89	-0.17	
15550	11101	27	0.06	-0.28	0.39		0.38	0.05	0.71	
15561	11101	28	-0.61	-0.94	-0.27		-0.56	-0.89	-0.23	
22556	11101	31	0.50	0.13	0.87		0.36	-0.02	0.74	
15597	11101	NR	0.55	0.24	0.86		0.53	0.22	0.84	
21555	11101	NR	-0.05	-0.38	0.29		-0.10	-0.43	0.24	
21576	11101	NR	0.47	0.15	0.80		0.54	0.22	0.87	
21560	11101	NR	-0.23	-0.56	0.11		0.10	-0.24	0.43	
19453	11101	NR	0.06	-0.26	0.39		0.20	-0.12	0.51	
21564	11101	NR	0.50	0.17	0.83		0.56	0.23	0.89	
10136	11101	NR	-0.34	-0.68	0.00		-0.44	-0.77	-0.10	
22553	11101	NR	1.03	0.71	1.35	B+	0.80	0.49	1.12	

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
21586	11101	20	-0.14	0.00		-0.14	0.00	
19625	11101	22	0.14	0.00		0.11	0.00	
21588	11101	24	-0.21	0.00	B-	-0.20	0.00	B-
21568	11101	NR	0.01	0.42		0.02	0.64	
21570	11101	NR	-0.06	0.00		-0.03	0.00	
14845	11101	NR	0.18	0.00	B+	0.18	0.00	B+
21534	11101	NR	-0.02	0.25		0.04	0.35	

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
15517	11201	2	0.62	0.31	0.94		0.40	0.09	0.71	
15545	11201	3	0.35	0.01	0.70		0.36	0.02	0.71	
21519	11201	4	0.72	0.40	1.04		0.70	0.38	1.02	
18952	11201	5	-0.91	-1.23	-0.59		-0.75	-1.07	-0.44	
26863	11201	7	-0.70	-1.02	-0.39		-0.40	-0.71	-0.08	
21540	11201	8	-0.30	-0.69	0.08		-0.22	-0.61	0.16	
23365	11201	10	0.78	0.45	1.11		0.64	0.31	0.97	
21506	11201	12	0.37	0.05	0.69		0.38	0.06	0.70	
14818	11201	15	0.16	-0.20	0.52		0.33	-0.03	0.69	
12836	11201	16	0.19	-0.16	0.53		0.33	-0.01	0.68	
14830	11201	17	-0.02	-0.36	0.32		-0.09	-0.44	0.25	
21545	11201	18	0.33	0.01	0.66		0.28	-0.04	0.60	
15560	11201	26	-0.68	-1.04	-0.31		-0.58	-0.94	-0.22	
15561	11201	28	-0.49	-0.81	-0.16		-0.51	-0.84	-0.18	
19638	11201	29	0.40	0.09	0.71		0.26	-0.05	0.56	
22557	11201	30	0.30	-0.02	0.61		0.24	-0.07	0.56	
15543	11201	NR	0.11	-0.21	0.44		0.15	-0.18	0.48	
21575	11201	NR	-0.38	-0.75	-0.01		-0.60	-0.97	-0.23	
19424	11201	NR	0.00	-0.33	0.33		0.07	-0.26	0.39	
14809	11201	NR	0.92	0.60	1.24		0.79	0.47	1.10	
21580	11201	NR	-0.32	-0.66	0.01		-0.44	-0.77	-0.11	
21543	11201	NR	-0.04	-0.37	0.29		-0.18	-0.51	0.15	
21584	11201	NR	0.40	0.07	0.73		0.15	-0.18	0.48	
22550	11201	NR	0.30	-0.07	0.66		0.38	0.02	0.74	

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
21585	11201	19	-0.07	0.00		0.00	-0.06	
19624	11201	21	-0.06	0.00		0.01	-0.04	
21551	11201	23	0.05	0.00		0.00	0.08	
21572	11201	25	0.05	0.09		0.05	0.01	
21530	11201	NR	-0.10	0.00		0.02	-0.04	
21531	11201	NR	0.10	0.00		0.00	0.08	
19627	11201	NR	-0.12	0.00		0.00	-0.12	

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
21610	12101	1	0.01	-0.30	0.32		-0.07	-0.38	0.24	
15688	12101	4	-0.29	-0.64	0.07		-0.39	-0.74	-0.04	
21594	12101	5	1.38	1.06	1.70	B+	1.21	0.89	1.53	B+
21632	12101	6	0.60	0.27	0.94		0.46	0.12	0.79	
21634	12101	7	-0.04	-0.45	0.38		0.03	-0.38	0.43	
10263	12101	9	-0.12	-0.52	0.28		-0.24	-0.63	0.16	
23636	12101	11	1.18	0.83	1.54	B+	1.17	0.81	1.53	B+
23637	12101	19	-0.66	-1.03	-0.29		-0.77	-1.16	-0.38	
21653	12101	21	0.10	-0.22	0.42		0.24	-0.09	0.56	
19461	12101	22	1.13	0.80	1.46	B+	0.97	0.64	1.31	
21600	12101	23	-0.70	-1.05	-0.34		-0.62	-0.98	-0.26	
19563	12101	24	0.33	-0.06	0.71		0.28	-0.12	0.67	
15623	12101	26	0.20	-0.13	0.54		0.08	-0.26	0.42	
10287	12101	27	-0.17	-0.49	0.16		-0.33	-0.66	0.00	
21656	12101	28	-0.38	-0.78	0.03		-0.80	-1.22	-0.37	
14954	12101	30	-0.06	-0.39	0.27		0.40	0.07	0.73	
21609	12101	NR	0.57	0.20	0.94		0.43	0.06	0.80	
21630	12101	NR	-0.46	-0.82	-0.10		-0.59	-0.94	-0.23	
14892	12101	NR	-0.20	-0.67	0.27		-0.64	-1.12	-0.16	
14898	12101	NR	-0.19	-0.55	0.16		-0.29	-0.65	0.07	
21598	12101	NR	0.03	-0.31	0.37		-0.07	-0.42	0.28	
26308	12101	NR	-0.12	-0.55	0.31		-0.11	-0.55	0.32	
22537	12101	NR	-0.35	-0.71	0.01		-0.40	-0.76	-0.04	
22545	12101	NR	-0.96	-1.29	-0.63		-0.52	-0.85	-0.20	

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
15680	12101	12	-0.11	0.00		0.00	-0.07	
19587	12101	13	0.03	0.59		0.03	0.03	
21661	12101	15	-0.14	0.00		0.00	-0.13	
26861	12101	18	-0.06	0.05		0.15	-0.04	
21642	12101	NR	0.17	0.00		0.00	0.15	
15702	12101	NR	0.17	0.00		0.00	0.16	
21644	12101	NR	-0.12	0.00		0.00	-0.11	

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
15668	12201	2	0.20	-0.12	0.53		0.01	-0.31	0.34	
21666	12201	3	-0.07	-0.40	0.26		-0.08	-0.41	0.26	
15688	12201	4	-0.63	-0.99	-0.28		-0.38	-0.74	-0.02	
21633	12201	8	-0.48	-0.83	-0.14		-0.59	-0.93	-0.24	
10263	12201	9	0.02	-0.36	0.40		0.12	-0.26	0.49	
21635	12201	10	-0.27	-0.64	0.10		-0.02	-0.40	0.35	
23636	12201	11	1.23	0.84	1.61	B+	1.34	0.96	1.72	B+
23637	12201	19	-0.83	-1.24	-0.43		-0.50	-0.90	-0.10	
21652	12201	20	-0.05	-0.40	0.30		-0.08	-0.43	0.27	
19461	12201	22	0.99	0.66	1.33		0.94	0.60	1.27	
19563	12201	24	0.68	0.27	1.09		0.29	-0.12	0.70	
21637	12201	25	0.59	0.23	0.95		0.01	-0.35	0.38	
15623	12201	26	-0.01	-0.36	0.33		0.20	-0.14	0.54	
21602	12201	29	0.52	0.17	0.88		0.35	0.00	0.70	
14954	12201	30	0.41	0.07	0.75		0.39	0.05	0.74	
19005	12201	31	-0.08	-0.50	0.34		-0.72	-1.15	-0.29	
21665	12201	NR	-0.13	-0.47	0.20		-0.15	-0.49	0.19	
12913	12201	NR	0.52	0.19	0.85		0.52	0.19	0.85	
15653	12201	NR	0.69	0.33	1.06		0.96	0.58	1.34	
15250	12201	NR	0.25	-0.09	0.59		0.25	-0.09	0.59	
12884	12201	NR	-0.33	-0.67	0.01		-0.59	-0.93	-0.25	
21673	12201	NR	0.31	-0.24	0.86		0.22	-0.33	0.77	
15265	12201	NR	-0.21	-0.61	0.18		-0.01	-0.42	0.39	
22546	12201	NR	0.00	-0.40	0.39		-0.17	-0.56	0.22	

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
26868	12201	14	0.11	0.00		0.00	0.14	
14943	12201	16	-0.14	0.00		0.00	-0.10	
19608	12201	17	-0.01	0.00		0.00	-0.07	
19567	12201	NR	-0.01	0.03		0.01	-0.01	
26865	12201	NR	0.06	0.00		0.00	0.04	
21624	12201	NR	-0.08	0.00		0.00	-0.11	
19591	12201	NR	-0.11	0.00		0.00	-0.11	

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
20372	21101	1	-1.25	-3.43	0.94	B-
21884	21101	3	-0.88	-2.57	0.81	
14374	21101	4	0.52	-0.93	1.98	
14383	21101	6	-0.63	-1.98	0.71	
21972	21101	7	-0.37	-2.09	1.34	
14394	21101	8	-0.68	-2.16	0.79	
21748	21101	9	0.48	-1.11	2.08	
20442	21101	11	-1.08	-2.80	0.64	B-
14429	21101	19	-1.57	-3.22	0.07	B-
15320	21101	21	-1.26	-2.69	0.17	B-
22005	21101	22	1.96	0.54	3.37	C+
15303	21101	24	-0.80	-2.52	0.93	
15292	21101	NR	0.23	-1.15	1.61	
10000	21101	NR	0.31	-1.03	1.64	
21990	21101	NR	-1.34	-2.79	0.10	B-
20414	21101	NR	-0.59	-2.22	1.04	
21750	21101	NR	-0.51	-2.20	1.18	
9990	21101	NR	-1.84	-3.56	-0.13	C-
21877	21101	NR	0.72	-0.76	2.20	
20395	21101	NR	0.38	-1.08	1.85	
22002	21101	NR	0.01	-1.54	1.56	
15369	21101	NR	-0.25	-1.68	1.18	
20365	21101	NR	-0.98	-2.70	0.73	
14454	21101	NR	0.43	-0.98	1.84	

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
22012	21101	12	0.06	0.67	
20429	21101	15	0.03	0.69	
22019	21101	16	0.24	0.03	B+
20366	21101	17	-0.16	0.15	
20391	21101	NR	-0.07	0.60	
20426	21101	NR	0.35	0.01	C+
18496	21101	NR	0.04	0.55	

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
20372	21201	1	-1.51	-2.56	-0.47	C-
21692	21201	2	0.73	0.05	1.40	
14374	21201	4	-0.60	-1.28	0.08	
14381	21201	5	-0.22	-0.92	0.49	
14394	21201	8	0.65	-0.04	1.34	
21748	21201	9	0.05	-0.73	0.83	
13001	21201	10	-0.48	-1.19	0.23	
20442	21201	11	-1.25	-2.07	-0.42	B-
21959	21201	20	0.04	-0.63	0.71	
15320	21201	21	0.33	-0.33	1.00	
22011	21201	23	1.04	0.32	1.76	B+
15303	21201	24	0.54	-0.30	1.39	
15360	21201	NR	-0.55	-1.33	0.23	
9701	21201	NR	-0.40	-1.08	0.28	
15365	21201	NR	1.49	0.76	2.22	B+
21886	21201	NR	-0.63	-1.33	0.06	
20414	21201	NR	-0.40	-1.11	0.32	
9990	21201	NR	0.45	-0.28	1.18	
21851	21201	NR	0.09	-0.67	0.85	
20435	21201	NR	0.20	-0.58	0.98	
20364	21201	NR	0.20	-0.54	0.94	
15279	21201	NR	0.88	0.19	1.57	
20365	21201	NR	0.10	-0.74	0.94	

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
22016	21201	13	-0.07	0.27	
20369	21201	14	-0.09	0.00	
20429	21201	15	-0.14	0.02	
21684	21201	18	-0.11	0.01	
15307	21201	NR	0.19	0.00	B+
20448	21201	NR	0.26	0.00	C+
21787	21201	NR	-0.19	0.00	B-

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
21685	22101	1	-0.05	-0.87	0.78	
20253	22101	2	-0.15	-0.83	0.54	
15424	22101	3	-0.70	-1.51	0.10	
15451	22101	4	0.19	-0.76	1.15	
9681	22101	6	-0.69	-1.39	0.01	
15429	22101	7	1.13	0.42	1.84	B+
20362	22101	9	1.44	0.75	2.14	B+
23611	22101	17	0.52	-0.27	1.30	
20340	22101	19	-0.48	-1.16	0.20	
20301	22101	20	0.14	-0.56	0.84	
15391	22101	21	0.76	0.09	1.44	
22603	22101	22	-0.03	-0.69	0.64	
14544	22101	23	0.09	-0.61	0.79	
15241	22101	25	0.75	-0.03	1.54	
9679	22101	27	-0.84	-1.65	-0.02	
14479	22101	NR	-1.18	-2.00	-0.37	B-
20275	22101	NR	0.32	-0.40	1.03	
20336	22101	NR	0.36	-0.34	1.06	
20338	22101	NR	0.49	-0.39	1.37	
26857	22101	NR	-1.27	-2.03	-0.50	B-
21946	22101	NR	0.31	-0.41	1.03	
15436	22101	NR	-0.09	-1.04	0.86	
15423	22101	NR	0.14	-0.53	0.82	
15456	22101	NR	-0.35	-1.07	0.36	

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
20269	22101	11	-0.23	0.00	B-
18490	22101	12	0.18	0.00	B+
22714	22101	14	-0.09	0.38	
15441	22101	16	0.06	0.51	
20330	22101	NR	0.07	0.27	
20307	22101	NR	0.07	0.18	
22030	22101	NR	0.00	0.07	

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
21685	22201	1	-0.16	-0.63	0.31	
20253	22201	2	-0.32	-0.69	0.06	
15424	22201	3	-0.35	-0.77	0.07	
15447	22201	5	1.47	1.11	1.84	B+
15429	22201	7	1.18	0.77	1.59	B+
21770	22201	8	0.29	-0.38	0.96	
20362	22201	9	0.80	0.41	1.19	
23611	22201	17	-0.42	-0.84	-0.01	
14510	22201	18	-0.52	-0.92	-0.13	
20301	22201	20	-0.18	-0.57	0.21	
22603	22201	22	-0.49	-0.86	-0.13	
14544	22201	23	0.53	0.15	0.91	
14549	22201	24	-0.27	-0.63	0.09	
21920	22201	26	0.28	-0.18	0.74	
14472	22201	NR	-0.36	-0.75	0.03	
14479	22201	NR	0.42	-0.02	0.86	
20275	22201	NR	0.43	0.04	0.82	
9942	22201	NR	-0.08	-0.54	0.39	
20277	22201	NR	-0.19	-0.58	0.20	
20260	22201	NR	-0.63	-1.06	-0.21	
21898	22201	NR	-0.70	-1.09	-0.32	
9814	22201	NR	0.26	-0.12	0.64	
21933	22201	NR	-0.12	-0.56	0.31	
9838	22201	NR	0.62	0.20	1.05	

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
20331	22201	10	0.01	0.00	
18490	22201	12	0.17	0.00	
20346	22201	13	-0.23	0.00	B-
21968	22201	15	-0.09	0.00	
20289	22201	NR	0.16	0.00	
20287	22201	NR	0.02	0.42	
15399	22201	NR	-0.02	0.03	

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
21500	11101	1	-0.19	-0.67	0.30		-0.11	-0.60	0.37	
15517	11101	2	-1.45	-1.91	-0.98	B-	-1.15	-1.61	-0.70	B-
15545	11101	3	0.13	-0.35	0.62		0.37	-0.12	0.86	
18952	11101	5	-0.64	-1.11	-0.18		-0.73	-1.19	-0.27	
21556	11101	6	-0.61	-1.05	-0.17		-0.76	-1.20	-0.31	
21505	11101	9	0.06	-0.37	0.49		0.10	-0.34	0.54	
23365	11101	10	1.30	0.85	1.76	B+	1.04	0.58	1.49	B+
19649	11101	11	-0.83	-1.40	-0.26		-0.58	-1.16	0.00	
21581	11101	13	-0.16	-0.61	0.29		-0.04	-0.49	0.41	
21507	11101	14	0.20	-0.27	0.68		0.15	-0.33	0.62	
14818	11101	15	0.08	-0.43	0.58		-0.38	-0.88	0.13	
14830	11101	17	0.42	-0.08	0.92		0.05	-0.45	0.55	
15560	11101	26	0.05	-0.45	0.55		0.28	-0.23	0.78	
15550	11101	27	0.62	0.14	1.10		0.59	0.11	1.07	
15561	11101	28	-0.55	-1.03	-0.08		-0.21	-0.68	0.27	
22556	11101	31	0.47	-0.03	0.97		0.86	0.35	1.37	
15597	11101	NR	-0.02	-0.46	0.42		-0.03	-0.47	0.41	
21555	11101	NR	0.22	-0.25	0.70		0.08	-0.40	0.55	
21576	11101	NR	0.28	-0.18	0.74		0.20	-0.25	0.65	
21560	11101	NR	-0.54	-1.02	-0.06		-0.45	-0.93	0.03	
19453	11101	NR	-0.16	-0.61	0.30		-0.12	-0.58	0.33	
21564	11101	NR	0.40	-0.08	0.87		0.47	-0.01	0.94	
10136	11101	NR	-0.31	-0.79	0.17		-0.39	-0.87	0.10	
22553	11101	NR	0.46	0.01	0.91		0.35	-0.10	0.79	

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
21586	11101	20	-0.11	0.00		-0.13	0.00	
19625	11101	22	0.07	0.01		0.06	0.00	
21588	11101	24	-0.03	0.26		-0.05	0.01	
21568	11101	NR	0.07	0.01		0.06	0.00	
21570	11101	NR	0.07	0.13		0.11	0.02	
14845	11101	NR	0.16	0.00		0.15	0.00	
21534	11101	NR	-0.01	0.95		-0.06	0.28	

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
15517	11201	2	-1.74	-2.21	-1.27	C-	-1.41	-1.87	-0.94	B-
15545	11201	3	0.43	-0.06	0.92		0.21	-0.27	0.69	
21519	11201	4	-0.07	-0.53	0.40		0.11	-0.36	0.57	
18952	11201	5	-1.08	-1.56	-0.61	B-	-0.87	-1.35	-0.39	
26863	11201	7	-1.88	-2.34	-1.42	C-	-2.12	-2.58	-1.65	C-
21540	11201	8	1.10	0.57	1.63	B+	1.09	0.56	1.62	B+
23365	11201	10	1.37	0.90	1.83	B+	1.54	1.07	2.00	C+
21506	11201	12	-0.58	-1.04	-0.12		-0.80	-1.26	-0.33	
14818	11201	15	-0.34	-0.86	0.18		-0.68	-1.21	-0.15	
12836	11201	16	0.59	0.12	1.06		0.48	0.00	0.95	
14830	11201	17	0.36	-0.14	0.87		0.55	0.05	1.05	
21545	11201	18	0.75	0.28	1.21		0.40	-0.06	0.86	
15560	11201	26	0.12	-0.39	0.64		0.09	-0.42	0.59	
15561	11201	28	-0.70	-1.18	-0.22		-0.51	-0.99	-0.04	
19638	11201	29	0.04	-0.40	0.48		0.29	-0.16	0.73	
22557	11201	30	0.50	0.06	0.95		0.78	0.34	1.23	
15543	11201	NR	-1.24	-1.73	-0.76	B-	-1.00	-1.48	-0.52	B-
21575	11201	NR	-0.26	-0.79	0.28		-0.63	-1.16	-0.10	
19424	11201	NR	1.62	1.16	2.08	C+	1.28	0.82	1.73	B+
14809	11201	NR	0.38	-0.07	0.84		0.09	-0.38	0.55	
21580	11201	NR	-0.20	-0.68	0.27		-0.32	-0.79	0.16	
21543	11201	NR	-0.14	-0.62	0.35		-0.08	-0.56	0.39	
21584	11201	NR	0.15	-0.33	0.63		0.18	-0.30	0.65	
22550	11201	NR	0.54	0.02	1.05		0.15	-0.36	0.66	

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
21585	11201	19	-0.05	0.26		-0.03	0.68	
19624	11201	21	0.09	0.03		0.10	0.02	
21551	11201	23	0.28	0.00	C+	0.24	0.00	B+
21572	11201	25	-0.05	0.02		-0.08	0.00	
21530	11201	NR	-0.24	0.00	B-	-0.22	0.00	B-
21531	11201	NR	0.13	0.00		0.18	0.00	B+
19627	11201	NR	0.10	0.00		0.07	0.02	

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
21610	12101	1	-1.10	-1.55	-0.65	B-	-1.06	-1.52	-0.61	B-
15688	12101	4	-1.04	-1.57	-0.51	B-	-1.13	-1.65	-0.60	B-
21594	12101	5	0.61	0.16	1.05		0.24	-0.21	0.69	
21632	12101	6	-0.27	-0.75	0.20		0.03	-0.45	0.50	
21634	12101	7	0.27	-0.32	0.86		-0.24	-0.83	0.34	
10263	12101	9	-0.21	-0.78	0.37		-0.33	-0.89	0.23	
23636	12101	11	0.70	0.20	1.20		1.03	0.54	1.53	B+
23637	12101	19	-0.43	-0.97	0.10		-0.11	-0.65	0.44	
21653	12101	21	-0.13	-0.60	0.33		-0.60	-1.05	-0.14	
19461	12101	22	-0.08	-0.55	0.39		0.11	-0.36	0.58	
21600	12101	23	-0.21	-0.71	0.29		0.19	-0.31	0.69	
19563	12101	24	0.37	-0.18	0.92		0.29	-0.27	0.84	
15623	12101	26	0.29	-0.19	0.78		0.23	-0.26	0.72	
10287	12101	27	0.43	-0.04	0.90		0.65	0.18	1.12	
21656	12101	28	-0.37	-0.97	0.23		-0.21	-0.81	0.39	
14954	12101	30	0.25	-0.23	0.73		-0.21	-0.69	0.27	
21609	12101	NR	-0.46	-0.99	0.07		-0.60	-1.13	-0.07	
21630	12101	NR	-0.85	-1.39	-0.31		-0.39	-0.92	0.13	
14892	12101	NR	-0.06	-0.74	0.62		0.26	-0.42	0.95	
14898	12101	NR	-0.22	-0.74	0.30		-0.08	-0.60	0.44	
21598	12101	NR	0.64	0.15	1.12		0.53	0.04	1.01	
26308	12101	NR	-0.72	-1.36	-0.09		-1.13	-1.78	-0.48	B-
22537	12101	NR	0.13	-0.38	0.65		0.10	-0.41	0.61	
22545	12101	NR	-0.07	-0.54	0.40		-0.17	-0.64	0.30	

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
15680	12101	12	0.05	0.42		0.01	0.33	
19587	12101	13	-0.04	0.00		0.00	0.23	
21661	12101	15	-0.06	0.14		-0.06	0.03	
26861	12101	18	-0.04	0.37		-0.04	0.42	
21642	12101	NR	0.12	0.00		0.13	0.00	
15702	12101	NR	0.12	0.00		0.12	0.00	
21644	12101	NR	0.01	0.91		0.01	0.87	

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
15668	12201	2	-0.88	-1.35	-0.40		-1.00	-1.47	-0.53	B-
21666	12201	3	-2.01	-2.48	-1.54	C-	-2.08	-2.55	-1.61	C-
15688	12201	4	-1.22	-1.75	-0.69	B-	-1.45	-1.98	-0.92	B-
21633	12201	8	0.71	0.24	1.18		0.58	0.11	1.05	
10263	12201	9	0.09	-0.45	0.63		0.28	-0.26	0.82	
21635	12201	10	-0.97	-1.50	-0.43		-0.86	-1.40	-0.31	
23636	12201	11	1.29	0.78	1.81	B+	1.45	0.93	1.97	B+
23637	12201	19	-0.46	-1.04	0.11		-0.33	-0.90	0.24	
21652	12201	20	-0.87	-1.39	-0.36		-0.99	-1.50	-0.47	
19461	12201	22	0.22	-0.25	0.68		0.13	-0.34	0.61	
19563	12201	24	0.69	0.13	1.25		1.09	0.53	1.65	B+
21637	12201	25	0.73	0.22	1.23		0.68	0.18	1.18	
15623	12201	26	0.06	-0.43	0.54		0.20	-0.29	0.68	
21602	12201	29	-0.71	-1.22	-0.19		-0.59	-1.11	-0.08	
14954	12201	30	-0.42	-0.91	0.06		-0.31	-0.79	0.16	
19005	12201	31	-0.85	-1.47	-0.23		-0.94	-1.57	-0.31	
21665	12201	NR	-2.39	-2.89	-1.89	C-	-2.23	-2.73	-1.72	C-
12913	12201	NR	-0.19	-0.66	0.28		-0.01	-0.48	0.46	
15653	12201	NR	1.11	0.61	1.61	B+	0.86	0.36	1.35	
15250	12201	NR	0.10	-0.38	0.59		-0.03	-0.52	0.47	
12884	12201	NR	0.29	-0.19	0.77		0.67	0.19	1.14	
21673	12201	NR	1.21	0.50	1.93	B+	0.90	0.21	1.60	
15265	12201	NR	-0.73	-1.31	-0.14		-0.59	-1.18	-0.01	
22546	12201	NR	0.82	0.29	1.36		0.81	0.27	1.35	

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
26868	12201	14	0.18	0.00	B+	0.16	-0.16	
14943	12201	16	0.02	0.41		0.01	-0.01	
19608	12201	17	0.07	0.01		0.07	-0.07	
19567	12201	NR	-0.18	0.00	B-	-0.23	0.23	B-
26865	12201	NR	0.05	0.29		0.10	-0.10	
21624	12201	NR	0.12	0.00		0.18	-0.18	B+
19591	12201	NR	0.11	0.00		0.10	-0.10	

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*	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
21190_555	I	1	R2	1	0.49	0.26	0.78
21194_555	I	2	R1	2	0.83	0.39	-1.40
21191_555	I	3	R1	3	0.76	0.33	-0.86
21195_555	I	4	R2	4	0.62	0.16	0.10
23677_555	I	5	R3	3	0.80	0.29	-1.13
21197_555	I	6	R3	3 [†]	0.78 (2.33)	0.43	-1.75
22051	II	1	W1	2	0.61	0.26	0.09
22050	II	2	W2	3	0.84	0.21	-1.48
22094	II	3	W3	2	0.73	0.36	-0.67
22086	II	4	W3	2	0.85	0.36	-1.69
26495_T	III	1	W4	3 [†]	0.84 (2.52)	0.36	-1.92
26495_V	III	1	W3	2 [†]	0.96 (1.92)	0.31	-2.69
26731_T	IV	1	W4	6 [†]	0.72 (4.34)	0.52	-0.96
26731_V	IV	1	W3	4 [†]	0.89 (3.55)	0.51	-2.09
18691_303	V	1	R3	2	0.82	0.29	-1.21
15722_303	V	2	R2	1	0.68	0.17	-0.25
18590_303	V	3	R1	3	0.69	0.40	-0.46
15721_303	V	4	R1	1	0.80	0.37	-1.23
16604_303	V	5	R2	2	0.72	0.30	-0.52
15723_303	V	6	R2	4	0.73	0.41	-0.67
17429_303	V	7	R2	3 [†]	0.67 (2.00)	0.37	-0.79
23367_T	NR	NR	W4	6 [†]	0.75 (4.53)	0.57	-1.46
23367_V	NR	NR	W3	4 [†]	0.89 (3.56)	0.52	-2.46
22084	NR	NR	W1	3	0.84	0.35	-1.59
22055	NR	NR	W2	3	0.91	0.35	-2.43
19390	NR	NR	W3	2	0.65	0.35	-0.30
22092	NR	NR	W3	2	0.66	0.38	-0.29
22924_565	NR	NR	R1	3	0.86	0.23	-1.72
23633_565	NR	NR	R2	2	0.93	0.22	-2.50
21291_565	NR	NR	R2	1	0.75	0.32	-0.79
23352_565	NR	NR	R3	3	0.74	0.36	-0.76
22918_565	NR	NR	R1	3	0.77	0.40	-0.99

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.

*Answer 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*	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
21299 565	NR	NR	R3	3	0.67	0.30	-0.33
21298 565	NR	NR	R2	2	0.68	0.36	-0.36
21295 565	NR	NR	R2	4	0.91	0.31	-2.16
23087 565	NR	NR	R2	4	0.77	0.42	-1.00
21315 567	NR	NR	R2	1	0.76	0.27	-0.80
23123 567	NR	NR	R2	4	0.83	0.38	-1.42
21317 567	NR	NR	R2	2	0.82	0.23	-1.28
23671 567	NR	NR	R3	3	0.36	0.29	1.55
23673 567	NR	NR	R2	1	0.83	0.34	-1.36
21321 567	NR	NR	R2	3 [†]	0.78 (2.35)	0.47	-1.61
21322 567	NR	NR	R3	3 [†]	0.73 (2.19)	0.44	-1.17
23689 T	NR	NR	W4	3 [†]	0.81 (2.44)	0.38	-1.59
23689 V	NR	NR	W3	2 [†]	0.93 (1.86)	0.37	-2.28
18405 320	NR	NR	R2	1	0.73	0.22	-0.58
26817 320	NR	NR	R1	4	0.81	0.30	-1.15
26816 320	NR	NR	R2	3	0.81	0.29	-1.25
16091 320	NR	NR	R2	3	0.84	0.31	-1.51
16092 320	NR	NR	R2	4	0.66	0.37	-0.29
17748 320	NR	NR	R3	2	0.64	0.25	-0.06

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.

*Answer 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 7.1.7 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
21197 _555	I	6	% of Students	N/A	N/A	0.32	0.45	0.01	7.29	49.80	42.12
			Parameters						-3.58	-2.16	0.51
26495 _T	III	1	% of Students	N/A	N/A	0.32	0.68	0.02	5.02	35.37	58.60
			Parameters						-3.23	-2.26	-0.28
26495 _V	III	1	% of Students	0.29	0.12	0.32	0.68	0.02	5.26	93.32	
			Parameters						-3.03	-2.34	
17429 _303	V	7	% of Students	N/A	N/A	0.64	3.66	0.05	22.55	42.33	30.77
			Parameters						-2.51	-0.79	0.94
21321 _567	NR	NR	% of Students	N/A	N/A	0.71	0.69	0.01	7.43	45.70	45.46
			Parameters						-3.14	-2.01	0.31
21322 _567	NR	NR	% of Students	N/A	N/A	0.92	1.25	0.04	11.16	52.05	34.58
			Parameters						-2.74	-1.61	0.85
23689 _T	NR	NR	% of Students	N/A	N/A	0.40	1.44	0.04	4.04	42.63	51.45
			Parameters						-2.69	-2.15	0.06
23689 _V	NR	NR	% of Students	0.51	0.12	0.40	1.44	0.04	8.93	88.57	
			Parameters						-2.70	-1.84	

Note. The total number of students is 127 817; 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
26731_T	IV	1	% of Students	0.05	0.52	0.02	0.36	0.24	0.70	1.37
			Parameters				-2.95*		-2.03	-1.84
26731_V	IV	1	% of Students	0.00	0.52	0.02	0.42	0.18	1.39	4.33
			Parameters				-2.95*		-2.59	-2.34
23367_T	NR	NR	% of Students	0.04	0.28	0.01	0.09	0.20	1.03	1.93
			Parameters				-3.41*		-2.74	-2.34
23367_V	NR	NR	% of Students	0.00	0.28	0.01	0.31	0.25	1.67	5.16
			Parameters				-3.78*		-3.06	-2.63
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
26731_T	IV	1	% of Students	4.69	10.75	24.85	25.99	16.97	10.19	3.30
			Parameters	-1.70	-1.48	-1.15	-0.55	0.05	0.59	1.50
26731_V	IV	1	% of Students	14.95	33.51	44.68				
			Parameters	-2.11	-1.71	-0.85				
23367_T	NR	NR	% of Students	4.96	9.66	16.75	21.96	19.48	15.94	7.66
			Parameters	-2.08	-1.76	-1.41	-1.00	-0.54	-0.06	0.80
23367_V	NR	NR	% of Students	14.16	32.64	45.52				
			Parameters	-2.30	-1.89	-1.06				

Note. The total number of students is 127 817.

*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*	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
21117_581	I	1	R2	3	0.81	0.22	-1.29
21110_581	I	2	R1	2	0.77	0.33	-1.01
21112_581	I	3	R2	4	0.85	0.31	-1.74
23712_581	I	4	R1	4	0.81	0.45	-1.35
21113_581	I	5	R3	1	0.86	0.33	-1.78
21119_581	I	6	R3	3 [†]	0.79 (2.36)	0.30	-2.08
21079	II	1	W1	4	0.87	0.23	-1.74
21060	II	2	W3	3	0.70	0.32	-0.57
21059	II	3	W2	2	0.89	0.30	-2.12
21167	II	4	W3	4	0.66	0.33	-0.39
21121_T	III	1	W4	3 [†]	0.79 (2.37)	0.43	-1.58
21121_V	III	1	W3	2 [†]	0.84 (1.67)	0.44	-1.97
26724_T	IV	1	W4	6 [†]	0.70 (4.21)	0.54	-1.16
26724_V	IV	1	W3	4 [†]	0.76 (3.02)	0.55	-1.71
21132_582	V	1	R2	4	0.34	0.29	1.74
22782_582	V	2	R1	4	0.82	0.39	-1.40
21124_582	V	3	R1	3	0.77	0.31	-0.97
21130_582	V	4	R2	2	0.85	0.21	-1.65
21129_582	V	5	R2	2	0.62	0.37	-0.08
21125_582	V	6	R3	3	0.89	0.36	-2.02
21828_582	V	7	R2	3 [†]	0.73 (2.20)	0.33	-1.05
41028_T	NR	NR	W4	6 [†]	0.80 (4.78)	0.55	-1.85
41028_V	NR	NR	W3	4 [†]	0.77 (3.10)	0.56	-2.11
21080	NR	NR	W1	3	0.83	0.26	-1.51
21151	NR	NR	W3	4	0.75	0.23	-0.89
21149	NR	NR	W2	2	0.46	0.34	0.86
21082	NR	NR	W3	3	0.87	0.26	-1.76
21040_577	NR	NR	R1	4	0.68	0.33	-0.43
23733_577	NR	NR	R3	1	0.68	0.32	-0.48
21044_577	NR	NR	R2	4	0.62	0.24	0.03
21041_577	NR	NR	R1	1	0.81	0.36	-1.43
21050_577	NR	NR	R2	3	0.43	0.45	0.94
21042_577	NR	NR	R2	1	0.84	0.25	-1.52
21052_577	NR	NR	R3	4	0.56	0.11	0.50
23734_577	NR	NR	R2	1	0.43	0.22	1.18
21048_577	NR	NR	R2	4	0.93	0.40	-2.76

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.

*Answer 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*	CTT Item Statistics		IRT Item Parameters
					Difficulty	Item-Total Correlation	Location
23152_580	NR	NR	R2	3	0.94	0.19	-2.79
21103_580	NR	NR	R2	1	0.76	0.10	-0.82
23153_580	NR	NR	R2	3	0.62	0.33	-0.04
23338_580	NR	NR	R3	3	0.92	0.27	-2.43
21105_580	NR	NR	R2	4	0.70	0.33	-0.58
21106_580	NR	NR	R2	3 [†]	0.68 (2.03)	0.42	-1.95
23729_580	NR	NR	R3	3 [†]	0.75 (2.25)	0.40	-1.57
21154_T	NR	NR	W4	3 [†]	0.81 (2.44)	0.45	-1.78
21154_V	NR	NR	W3	2 [†]	0.75 (1.50)	0.46	-1.65
21094_579_12	NR	NR	R2	4	0.82	0.15	-1.35
21096_579_12	NR	NR	R3	3	0.79	0.30	-1.21
21088_579_12	NR	NR	R2	4	0.84	0.32	-1.67
21091_579_12	NR	NR	R2	3	0.90	0.20	-2.23
21089_579_12	NR	NR	R1	1	0.68	0.32	-0.44
23330_579_12	NR	NR	R2	4	0.88	0.36	-1.99

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.

*Answer 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
21119_581	I	6	% of Students	N/A	N/A	0.19	0.23	0.02	4.98	52.50	42.09
			Parameters						-3.99	-2.65	0.39
21121_T	III	1	% of Students	N/A	N/A	0.72	0.59	0.00	7.55	44.21	46.93
			Parameters						-2.94	-1.95	0.14
21121_V	III	1	% of Students	0.78	0.45	0.72	0.59	0.00	27.91	69.55	
			Parameters						-3.09	-0.84	
21828_582	V	7	% of Students	N/A	N/A	1.17	3.41	0.21	17.85	30.22	47.14
			Parameters						-2.39	-0.93	0.20
21106_580	NR	NR	% of Students	N/A	N/A	0.08	0.19	0.00	25.51	45.21	29.01
			Parameters						-5.00	-1.01	0.98
23729_580	NR	NR	% of Students	N/A	N/A	0.66	0.78	0.00	16.14	37.93	44.49
			Parameters						-3.56	-1.39	0.25
21154_T	NR	NR	% of Students	N/A	N/A	0.28	0.95	0.00	5.92	40.67	52.18
			Parameters						-3.16	-2.11	-0.05
21154_V	NR	NR	% of Students	1.48	0.83	0.28	0.95	0.00	43.07	53.39	
			Parameters						-3.23	-0.08	

Note. The total number of students is 5284. 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
26724_T	IV	1	% of Students	0.02	0.23	0.00	0.25	0.15	0.78	1.08
			Parameters				-3.66*		-2.41	-2.09
26724_V	IV	1	% of Students	0.00	0.23	0.00	0.51	0.66	8.01	15.84
			Parameters				-3.83*		-3.04	-2.13
41028_T	NR	NR	% of Students	0.00	0.09	0.02	0.08	0.09	0.42	0.89
			Parameters				-3.78*		-2.98	-2.70
41028_V	NR	NR	% of Students	0.00	0.09	0.02	0.45	1.34	8.12	18.11
			Parameters				-4.81*		-3.21	-2.26
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
26724_T	IV	1	% of Students	6.70	11.32	36.37	21.69	11.71	6.30	3.41
			Parameters	-1.95	-1.61	-1.29	-0.35	0.16	0.58	1.13
26724_V	IV	1	% of Students	44.04	19.70	11.01				
			Parameters	-1.53	-0.25	0.55				
41028_T	NR	NR	% of Students	3.27	6.79	14.70	20.04	18.89	21.59	13.12
			Parameters	-2.50	-2.14	-1.80	-1.36	-0.95	-0.54	0.28
41028_V	NR	NR	% of Students	28.33	27.02	16.50				
			Parameters	-1.56	-0.85	0.08				

Note. The total number of students is 5284.

*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 1000 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
21190_555	I	1	1.05	0.73	1.37	B+	0.99	0.67	1.31	
21194_555	I	2	0.38	-0.07	0.83		-0.06	-0.49	0.37	
21191_555	I	3	0.17	-0.21	0.54		0.18	-0.20	0.56	
21195_555	I	4	-0.23	-0.54	0.09		-0.13	-0.44	0.19	
23677_555	I	5	1.49	1.06	1.92	B+	1.68	1.29	2.08	C+
22051	II	1	0.18	-0.14	0.50		-0.08	-0.40	0.24	
22050	II	2	0.74	0.33	1.15		0.42	0.00	0.84	
22094	II	3	0.02	-0.35	0.38		-0.10	-0.47	0.27	
22086	II	4	-0.17	-0.63	0.29		-0.67	-1.13	-0.20	
18691_303	V	1	1.02	0.61	1.43	B+	0.80	0.39	1.21	
15722_303	V	2	0.35	0.03	0.68		0.66	0.33	0.99	
18590_303	V	3	-0.87	-1.22	-0.52		-0.61	-0.98	-0.25	
15721_303	V	4	0.14	-0.29	0.56		0.18	-0.23	0.59	
16604_303	V	5	0.97	0.61	1.32		1.01	0.65	1.37	B+
15723_303	V	6	0.13	-0.26	0.51		0.28	-0.09	0.66	
22084	NR	NR	-0.31	-0.75	0.14		-0.74	-1.19	-0.29	
22055	NR	NR	-0.90	-1.47	-0.32		-1.27	-1.87	-0.67	B-
19390	NR	NR	0.76	0.42	1.10		0.58	0.23	0.94	
22092	NR	NR	-0.64	-0.99	-0.30		-0.55	-0.90	-0.20	
22924_565	NR	NR	-0.29	-0.75	0.17		-0.06	-0.52	0.40	
23633_565	NR	NR	0.33	-0.24	0.90		0.65	0.05	1.25	
21291_565	NR	NR	1.73	1.34	2.12	C+	1.82	1.45	2.20	C+
23352_565	NR	NR	0.62	0.25	0.98		0.67	0.29	1.04	
22918_565	NR	NR	0.94	0.55	1.33		1.36	0.95	1.77	B+
21299_565	NR	NR	-0.42	-0.76	-0.07		-0.55	-0.90	-0.21	
21298_565	NR	NR	0.70	0.35	1.06		1.01	0.65	1.36	B+
21295_565	NR	NR	-0.75	-1.34	-0.16		-0.07	-0.64	0.49	
23087_565	NR	NR	-0.06	-0.46	0.34		0.50	0.10	0.90	
21315_567	NR	NR	0.72	0.35	1.09		0.49	0.14	0.85	
23123_567	NR	NR	1.69	1.22	2.15	C+	1.81	1.36	2.26	C+
21317_567	NR	NR	0.81	0.40	1.22		0.62	0.22	1.02	
23671_567	NR	NR	0.94	0.60	1.27		0.52	0.19	0.86	
23673_567	NR	NR	0.41	-0.04	0.87		0.21	-0.21	0.63	
18405_320	NR	NR	0.43	0.08	0.77		0.48	0.13	0.83	
26817_320	NR	NR	2.07	1.64	2.50	C+	2.20	1.79	2.61	C+
26816_320	NR	NR	1.15	0.74	1.57	B+	1.43	1.02	1.84	B+
16091_320	NR	NR	-0.45	-0.89	-0.02		-0.03	-0.48	0.41	
16092_320	NR	NR	0.25	-0.09	0.60		0.00	-0.35	0.35	
17748_320	NR	NR	0.33	0.01	0.65		0.10	-0.23	0.43	

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
21197_555	I	6	-0.10	0.00		-0.16	0.00	
26495_T	III	1	-0.08	0.00		-0.12	0.00	
26495_V	III	1	-0.02	0.23		-0.03	0.04	
26731_T	IV	1	-0.10	0.00		-0.07	0.06	
26731_V	IV	1	-0.20	0.00	B-	-0.15	0.00	
17429_303	V	7	-0.07	0.00		-0.11	0.01	
23367_T	NR	NR	-0.21	0.00	B-	-0.21	0.00	B-
23367_V	NR	NR	-0.25	0.00	B-	-0.20	0.00	B-
21321_567	NR	NR	-0.13	0.00		-0.20	0.00	B-
21322_567	NR	NR	-0.11	0.00		-0.11	0.00	
23689_T	NR	NR	0.02	0.04		0.00	0.16	
23689_V	NR	NR	-0.06	0.00		-0.07	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
21117_581	I	1	0.74	0.35	1.13	
21110_581	I	2	0.74	0.36	1.13	
21112_581	I	3	1.10	0.64	1.56	B+
23712_581	I	4	0.59	0.15	1.03	
21113_581	I	5	1.48	1.01	1.95	B+
21079	II	1	0.77	0.32	1.22	
21060	II	2	-0.06	-0.41	0.29	
21059	II	3	-0.07	-0.59	0.44	
21167	II	4	-0.05	-0.39	0.29	
21132_582	V	1	0.38	0.04	0.72	
22782_582	V	2	-0.34	-0.77	0.09	
21124_582	V	3	1.91	1.51	2.30	C+
21130_582	V	4	0.29	-0.14	0.72	
21129_582	V	5	0.73	0.39	1.07	
21125_582	V	6	1.05	0.53	1.57	B+
21080	NR	NR	0.05	-0.37	0.47	
21151	NR	NR	-0.57	-0.93	-0.21	
21149	NR	NR	0.23	-0.10	0.55	
21082	NR	NR	-0.20	-0.66	0.26	
21040_577	NR	NR	0.93	0.58	1.28	

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
23733_577	NR	NR	0.89	0.55	1.24	
21044_577	NR	NR	0.53	0.21	0.85	
21041_577	NR	NR	1.00	0.57	1.42	
21050_577	NR	NR	1.36	1.00	1.72	B+
21042_577	NR	NR	0.54	0.10	0.98	
21052_577	NR	NR	0.21	-0.10	0.51	
23734_577	NR	NR	-0.01	-0.33	0.31	
21048_577	NR	NR	-0.09	-0.75	0.57	
23152_580	NR	NR	0.22	-0.42	0.86	
21103_580	NR	NR	0.28	-0.07	0.63	
23153_580	NR	NR	0.13	-0.20	0.46	
23338_580	NR	NR	0.36	-0.20	0.93	
21105_580	NR	NR	0.23	-0.12	0.58	
21094_579_12	NR	NR	0.43	0.04	0.82	
21096_579_12	NR	NR	0.70	0.31	1.10	
21088_579_12	NR	NR	0.93	0.49	1.37	
21091_579_12	NR	NR	-0.14	-0.66	0.37	
21089_579_12	NR	NR	-0.21	-0.56	0.13	
23330_579_12	NR	NR	0.02	-0.47	0.52	

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
21119_581	I	6	-0.05	0.00	
21121_T	III	1	-0.10	0.00	
21121_V	III	1	-0.16	0.00	
26724_T	IV	1	0.00	0.88	
26724_V	IV	1	-0.06	0.03	
21828_582	V	7	-0.12	0.00	
41028_T	NR	NR	-0.17	0.00	
41028_V	NR	NR	-0.23	0.00	B-
21106_580	NR	NR	-0.18	0.00	B-
23729_580	NR	NR	-0.06	0.09	
21154_T	NR	NR	-0.13	0.00	
21154_V	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
21190_555	I	1	0.71	0.26	1.16		0.74	0.29	1.19	
21194_555	I	2	1.23	0.65	1.80	B+	0.88	0.35	1.41	
21191_555	I	3	0.59	0.09	1.09		0.74	0.25	1.23	
21195_555	I	4	0.31	-0.13	0.75		0.24	-0.20	0.68	
23677_555	I	5	1.28	0.75	1.81	B+	1.07	0.57	1.57	B+
22051	II	1	-0.60	-1.05	-0.15		-0.66	-1.13	-0.20	
22050	II	2	-0.23	-0.80	0.34		-0.59	-1.19	0.02	
22094	II	3	0.31	-0.18	0.80		0.52	0.03	1.00	
22086	II	4	1.48	0.88	2.08	B+	0.95	0.38	1.52	
18691_303	V	1	1.01	0.46	1.56	B+	1.03	0.48	1.58	B+
15722_303	V	2	0.00	-0.45	0.45		0.03	-0.43	0.49	
18590_303	V	3	0.18	-0.30	0.66		0.27	-0.22	0.75	
15721_303	V	4	-0.18	-0.75	0.39		-0.43	-0.98	0.12	
16604_303	V	5	-0.09	-0.59	0.40		-0.15	-0.65	0.35	
15723_303	V	6	0.26	-0.25	0.77		0.73	0.24	1.22	
22084	NR	NR	0.11	-0.46	0.68		0.43	-0.15	1.01	
22055	NR	NR	-0.05	-0.83	0.72		-0.31	-1.06	0.44	
19390	NR	NR	-1.11	-1.60	-0.61	B-	-1.19	-1.68	-0.70	B-
22092	NR	NR	0.28	-0.20	0.76		0.24	-0.23	0.72	
22924_565	NR	NR	0.06	-0.56	0.67		-0.08	-0.69	0.52	
23633_565	NR	NR	0.53	-0.25	1.32		-0.51	-1.26	0.23	
21291_565	NR	NR	0.15	-0.36	0.66		-0.12	-0.60	0.36	
23352_565	NR	NR	-0.27	-0.80	0.26		-0.03	-0.54	0.49	
22918_565	NR	NR	-0.22	-0.76	0.33		-0.61	-1.17	-0.05	
21299_565	NR	NR	-0.33	-0.80	0.14		0.17	-0.30	0.63	
21298_565	NR	NR	1.13	0.64	1.61	B+	0.34	-0.14	0.81	
21295_565	NR	NR	1.40	0.67	2.13	B+	2.29	1.64	2.95	C+
23087_565	NR	NR	-0.70	-1.25	-0.15		-0.14	-0.67	0.38	
21315_567	NR	NR	0.80	0.30	1.30		0.83	0.35	1.31	
23123_567	NR	NR	1.43	0.84	2.02	B+	1.29	0.74	1.84	B+
21317_567	NR	NR	0.92	0.36	1.48		0.15	-0.38	0.68	
23671_567	NR	NR	0.31	-0.18	0.79		0.41	-0.06	0.89	
23673_567	NR	NR	2.60	2.05	3.14	C+	2.47	1.94	2.99	C+
18405_320	NR	NR	-0.02	-0.50	0.46		-0.10	-0.58	0.39	
26817_320	NR	NR	0.61	0.08	1.15		0.88	0.35	1.40	
26816_320	NR	NR	-0.30	-0.84	0.24		-0.68	-1.23	-0.13	
16091_320	NR	NR	0.01	-0.58	0.60		0.40	-0.18	0.98	
16092_320	NR	NR	-0.28	-0.75	0.19		-0.52	-1.01	-0.04	
17748_320	NR	NR	-0.6	-1.07	-0.14		-0.48	-0.94	-0.02	

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
21197_555	I	6	-0.04	0.01		-0.05	0.08	
26495_T	III	1	-0.17	0.00	B-	-0.20	0.00	B-
26495_V	III	1	0.09	0.00		0.13	0.02	
26731_T	IV	1	-0.08	0.05		-0.10	0.14	
26731_V	IV	1	0.05	0.01		0.14	0.00	
17429_303	V	7	-0.10	0.04		-0.26	0.00	C-
23367_T	NR	NR	-0.22	0.00	B-	-0.23	0.00	B-
23367_V	NR	NR	0.05	0.00		0.24	0.00	B+
21321_567	NR	NR	-0.04	0.11		0.03	0.89	
21322_567	NR	NR	-0.12	0.00		-0.16	0.00	
23689_T	NR	NR	-0.24	0.00	B-	-0.21	0.00	B-
23689_V	NR	NR	0.01	0.02		0.04	0.00	

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

