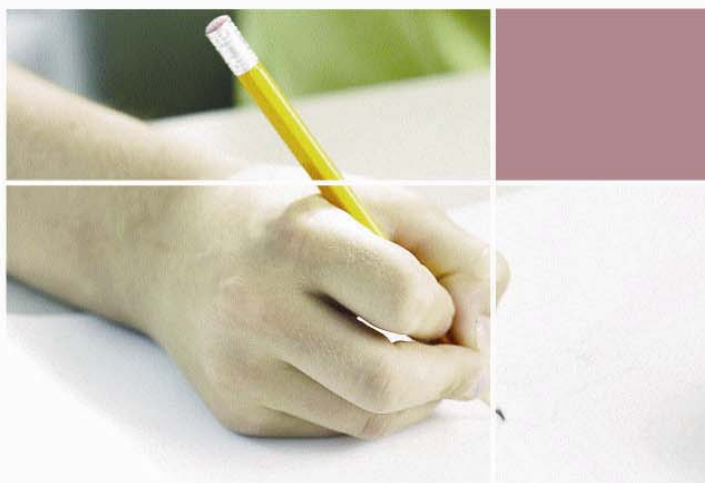


EQAO's Technical Report

for the 2006–2007 Assessments

*Assessments of Reading, Writing and Mathematics,
Primary Division (Grades 1–3) and Junior Division (Grades 4–6),
Grade 9 Assessment of Mathematics, and
Ontario Secondary School Literacy Test*



CHAPTER 1: PURPOSE AND OVERVIEW OF THIS REPORT	4
The EQAO Assessment Program: Primary (Grades 1–3), Junior (Grades 4–6), Grade 9 and Ontario Secondary School Literacy Test (OSSLT)	4
CHAPTER 2: ASSESSMENT DESIGN AND DEVELOPMENT	6
Assessment Frameworks	6
Assessment Blueprints	6
Item Development.....	6
Item Developers	6
Training for Item Developers.....	7
EQAO Content Specialist and Education Officer Review	7
Item Tryouts.....	8
Assessment Development and Sensitivity Review Committees.....	8
EQAO Assessment Development Committees.....	8
EQAO Sensitivity Committee.....	9
Field Testing	9
Test Construction: Selecting Items for the Operational Form	10
Questionnaire Construction	10
CHAPTER 3: TEST ADMINISTRATION AND PARTICIPATION	12
Assessment Administration	12
The Administration Guides.....	12
Administering the Assessment to Students.....	12
Support for English-Language Learners and Students with Special Education Needs	12
Definition of “Accommodation”	13
Special Version Assessments for Accommodated Students	13
Special Provisions for English-Language Learners	13
Exemptions	13
Deferrals (OSSLT only).....	14
Quality Assurance	14
CHAPTER 4: SCORING	15
The Range Finding Process	15
Pre-Range Finding Stage.....	15
Overview of the Pre-Range Finding Process	15
Range Finding Stage	16
Overview of the Range Finding Process.....	17
Preparing Training Materials for Scoring: Anchors, Training Papers, Qualifying Tests, Validity Papers.....	17
Field-Test Scoring.....	17
Scoring Open-Response Field-Test Responses	18
Training Field-Test Scoring Leaders and Scorers	18
Developing Additional Scoring Training Materials Before Scoring Operational Items ..	18
Scoring Open-Response Operational Items	19
Scoring Rooms for Scoring Open-Response Operational Items.....	19
Training for Scoring Open-Response Operational Items.....	20
Training of Scoring Leaders and Scoring Supervisors for Scoring Open-Response Operational Items.....	20
Training of Scorers for Scoring Open-Response Operational Items	20

Procedures at the Scoring Site	22
Evidence of Students at Risk	22
Reporting Procedure	22
Reporting of Inappropriate Content, Teacher Interference and Other Issues	22
Exceptions Room (Issues).....	22
Offensive Content	22
Different Handwriting.....	23
Damaged or Misprinted Booklets	23
Ongoing Daily Training.....	23
Daily Morning Review of Anchors.....	23
Daily Afternoon Calibration Activity	23
Daily Scoring Centre Reports for Monitoring the Quality of Open-Response Item	
Scoring	23
Daily and Cumulative Validity Reports.....	24
Daily and Cumulative Mean Score and Score-Point-Distribution Reports	25
Daily and Cumulative Scoring Centre Reliability Reports.....	25
Daily and Cumulative Productivity Reports	25
Aggregated Daily and Cumulative Individual Scorer Data Reports.....	26
Required Actions: Consequences of the Review and Analysis of Daily Scoring Centre	
Data Reports.....	26
Audits	26
Marker Reliability and Validity	27
Marker Reliability	27
Assessment of Reading, Writing and Mathematics: Primary Division	27
Assessment of Reading, Writing and Mathematics: Junior Division	31
Grade 9 Assessment of Mathematics: Academic and Applied.....	33
Ontario Secondary School Literacy Test	35
Marking Validity	37
Assessment of Reading, Writing and Mathematics: Primary Division	38
Assessment of Reading, Writing and Mathematics: Junior Division	41
Grade 9 Assessment of Mathematics	44
Ontario Secondary School Literacy Test	46
CHAPTER 5: EQUATING.....	49
IRT Models	49
Equating Design.....	50
Equating Samples.....	50
Calibration.....	50
Assessments of Reading, Writing and Mathematics: Primary and Junior Divisions	51
Description of the IRT Model.....	51
Equating Sample: Exclusion Rules.....	51
Equating Steps	52
Eliminating Items and Collapsing of Score Categories	52
Equating Results	53
Grade 9 Assessment of Mathematics.....	56
Description of the IRT Model.....	56
Equating Sample	56
Equating Steps	56
Eliminating Items and Collapsing of Score Categories	57

Equating Results	57
Ontario Secondary School Literacy Test (OSSLT)	58
Description of the IRT Model	58
Equating Sample	58
Equating Steps	59
Eliminating Items and Collapsing of Score Categories	60
Equating Results	60
CHAPTER 6: REPORTING RESULTS	61
Reporting the Results of the Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions.....	61
Reporting the Results of the Grade 9 Assessment of Mathematics	62
Reporting the Results of the OSSLT	62
Interpretation Guides	63
CHAPTER 7: STATISTICAL AND PSYCHOMETRIC SUMMARIES	64
Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions.....	64
Classical Test Theory (CTT) Analyses.....	64
Item Response Theory (IRT) Analyses.....	65
Item Parameter Estimates For Classical Test Theory (CTT) and Item Response Theory (IRT).....	79
Grade 9 Assessment of Mathematics	80
Classical Test Theory (CTT) Analyses.....	80
Item Response Theory (IRT) Analyses.....	81
Item Parameter Estimates For Classical Test Theory (CTT) and Item Response Theory (IRT).....	86
Ontario Secondary School Literacy Test (OSSLT)	88
Item Response Theory (IRT) Analyses.....	88
Item Parameter Estimates For Classical Test Theory (CTT) and Item Response Theory (IRT).....	91
APPENDICES	93

CHAPTER 1: PURPOSE AND OVERVIEW OF THIS REPORT

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

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

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

The assessments provide individual student, school and school board 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 also publishes annual provincial reports in English and in French for education stakeholders and the general public on the EQAO Web site. The assessment results provide valuable information that support improvement planning by schools, school boards and the Ontario Ministry of Education.

The annual Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions measure how well elementary school students have met the reading, writing and mathematics curriculum expectations assessed by EQAO and outlined in *The Ontario Curriculum, Grades 1–8: Language* (revised 2006) and *The Ontario Curriculum, Grades 1–8: Mathematics* (revised 2005). The reading component requires students to demonstrate 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 requires students to demonstrate their skill at organizing main ideas and supporting details using correct spelling, grammar and punctuation for written communication required by the curriculum. The mathematics component requires students to demonstrate their knowledge and skill across the five mathematical strands: number sense and numeration; measurement; geometry and spatial sense; patterning and algebra; and data management and probability.

The Grade 9 Assessment of Mathematics measures how well students have met the expectations outlined in *The Ontario Curriculum, Grades 9 and 10: Mathematics* (revised 2005). The Grade 9 assessment is conducted twice annually to accommodate secondary schools with semestered programs. EQAO develops separate versions of the Grade 9 Assessment of Mathematics for students in academic and applied courses. Grade 9 academic students are assessed based on their knowledge and skills in number sense and algebra; linear relations; analytic geometry; and measurement and geometry.

Grade 9 applied students are assessed based on their knowledge and skills in number sense and algebra; linear relations; and measurement and geometry. Two parallel forms of the Grade 9 Assessment of Mathematics are developed for each course (academic and applied), with one form administered toward the end of the first semester and the second form administered toward the end of the second semester.

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

EQAO education officers involve educators across the province in most aspects of EQAO assessments, including design; development of items and item-specific scoring rubrics; review of items for curriculum content and sensitivity; administration of the assessments in schools; scoring student responses to open-response items; and reporting assessment results. Educators are selected to participate in EQAO activities based on the following criteria:

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

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

CHAPTER 2: ASSESSMENT DESIGN AND DEVELOPMENT

Assessment Frameworks

EQAO posts the current framework for each large-scale assessment on its Web site to provide educators, students, parents and the general public with a detailed description of the assessment, including an explanation of how the assessment relates to the *Ontario Curriculum* expectations. The English and French frameworks for the four EQAO 2006–2007 assessments can be found on the EQAO Web site 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 coverage of *Ontario Curriculum* expectations and the difficulty of the assessments are comparable each year. It should be noted that not all expectations can be measured in a large-scale assessment. Measurable curriculum expectations are clustered by topic, and question items are then mapped to these clusters of expectations. Not all of the measurable expectations in a cluster are measured in any one assessment; however, over time all measurable expectations in a cluster are assessed. Therefore, it is important that teachers consider all of the curriculum expectations in their teaching.

The blueprints can be found in EQAO's assessment frameworks. A more detailed version of the blueprints is available in the *Development Specifications Guide for EQAO Assessments* provided to item developers.

Item Development

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

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

Item Developers

EQAO recruits and trains educators who are experts in the areas of English- and French-language (reading and writing) and mathematics to participate in its item-writing committees. The item-writing committee for each assessment comprises 15 to 20 educators, who serve for terms of one to three years, with a maximum of five years. Committee members meet twice a year to write and review items, discuss results of item tryouts and review items that will be considered for use in subsequent operational assessments.

Item developers construct multiple-choice items in reading, writing and mathematics; open-response items in reading and mathematics; and writing prompts for short and long writing tasks, all referenced to the *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, writing prompts and their item-specific scoring rubrics.

Item writers for EQAO assessments are selected based on their

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

Training for Item Developers

The matrix material for 2006–2007 was developed by EQAO in partnership with Educational Testing Service (ETS) Canada. EQAO and ETS staff led a three-day workshop and development session for item writers. They spent half a day introducing item writers to the criteria for effective item writing. EQAO provided an overview to the assessments, including a description of the frameworks, then ETS provided details on the elements of effective item writing. The remaining time involved a guided item-writing session structured by EQAO, in consultation with ETS.

In general, each item writer is assigned items to write based on the EQAO blueprint for the assessment. At the end of the item-writing session, the draft material is examined by EQAO and ETS staff. ETS further refines the items prior to their presentation to the Assessment Development and Sensitivity review committees. Further revisions also take place following the committee reviews, so that the items can be submitted to EQAO for matrix-item selection.

EQAO Content Specialist and Education Officer Review

When the first draft of the items and item-specific scoring rubrics are developed by the item writers, they are reviewed by the content specialists and education officers. The EQAO staff members ensure that each item is referenced correctly to curriculum expectations and difficulty levels. For the multiple-choice items, they consider the clarity and completeness of the stem, the correctness of the correct answer and the plausibility of the three incorrect options. For the open-response items, they consider

the correspondence between the reading and mathematics items or the writing prompts and their scoring rubrics and determine if the items will elicit the range of responses expected. They also review the results of the item tryouts conducted by the item writers in their classes and schools.

Item Tryouts

Item writers use cognitive labs in their own classes and schools to try out the items they develop. These item tryouts allow item writers to see quickly if their items are working as intended. The student responses are used to inform the editing and refining of stems of multiple-choice items, open-response items, multiple-choice options and item-specific scoring rubrics for open-response items. The results of these item tryouts are provided to EQAO content specialists and education officers to help them when reviewing, revising and editing the items.

Assessment Development and Sensitivity Review Committees

EQAO recruits and trains Ontario educators with expertise in the areas of English- and French-language (reading and writing), mathematics and equity issues to participate in its Assessment Development and Sensitivity committees. The goal of these committees is to ensure that items on the Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions, the Grade 9 Assessment of Mathematics and the OSSLT assess literacy and mathematics standards based on *Ontario Curriculum* expectations and that these items are appropriate, fair and accessible to the broadest range of students in Ontario.

Assessment Development Committee and Sensitivity Committee members are selected based on their

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

EQAO Assessment Development Committees

The Assessment Development Committee for each subject in each assessment comprises 10 to 12 Ontario educators, who serve for terms of one to three years, with a maximum of five years. Members meet up to four times a year to ensure that items appropriately assess standards of literacy and mathematics based on *Ontario Curriculum* expectations, and to make recommendations to EQAO for the content development for particular assessments.

The members of the Assessment Development Committee work collaboratively under the guidance of an EQAO content specialist to ensure that the materials (e.g., reading selections, reading and mathematics items and/or 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.

The members also provide expert advice from a specialized content and assessment perspective on the quality and fairness of materials being proposed for EQAO assessments. In addition, the members make suggestions for the inclusion and/or revision of the items or for their exclusion. As such, the members of the Assessment Development Committee must possess, in addition to the aforementioned qualities, expertise in and current experience with the curriculum and students in at least one of the subjects in the grade being assessed:

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

EQAO Sensitivity Committee

The Sensitivity Committee, which considers all four EQAO assessments, comprises 20 to 24 Ontario educators, who serve for terms of one to three years, with a maximum of five years. Members meet in focused subgroups up to four times a year to make recommendations that will assist EQAO in ensuring the fairness of its assessments.

The members of the Sensitivity Committee work collaboratively, under the guidance of an EQAO content specialist, to review assessment materials (e.g., items, reading selections) in various stages of development, to ensure that no particular group of students is unfairly advantaged or disadvantaged on any item. The members also provide expert advice from a specialized equity perspective to ensure that assessment materials are fair for a wide range of students, and they make suggestions for the inclusion and/or revision of the items or for their exclusion.

EQAO requires that members of the Sensitivity Committee possess

- expertise in and current experience with equity issues in education (such as multicultural issues, English-language learners and students with special education needs) and
- expertise in and current experience with the students and the curriculum in one or all of the grades being assessed:
 - primary assessment: language and mathematics in the primary division (Grades 1 to 3)
 - junior assessment: language and mathematics in the junior division (Grades 4 to 6)
 - Grade 9 assessment: mathematics in the intermediate division (Grades 7 to 10)
 - OSSLT: literacy across the curriculum in the intermediate division (Grades 7 to 10)

Field Testing

EQAO field tests all newly developed items, except for the long writing tasks, on the primary and junior assessments and the OSSLT. If items are found to be acceptable based on the field-test results, they are used in future assessments. The field-test items are embedded in each year's operational assessments, to guarantee 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 prevents the lack of motivation that occurs among students when new items are field tested outside the operational form. Scores on the field-test items are not used in determining student, school, school board or provincial results.

A matrix-sample design is used to embed field-test items in each operational form. The field-test items are divided into subsets that are inserted into the operational assessment. The number of field-test items

in each subset ensures that no more than 20% of the assessment items given to a student are field-test items. The field-test items are placed in the booklets according to psychometrically sound principles to ensure that valid and reliable data are obtained for each field-test item.

Field testing of assessment materials ensures that assessment items selected for future operational assessments are psychometrically sound and fair for all students. Field testing also provides data to equate each year's assessment with the previous year's assessment, so that assessment results can be validly compared across time.

Test Construction: Selecting Items for the Operational Form

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

- *Data:* The data for individual items, groups of items and test characteristic curves (based on the selection of 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 subject content, gender representations and aspects of provincial demographics (e.g., urban/rural, north/south).
- *Curriculum coverage:* As mentioned earlier, it is important to note that while items are mapped to clusters of curriculum expectations, and not all expectations within a cluster are measured in any one assessment, over time, all measurable expectations in a cluster are included on an assessment.

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

Questionnaire Construction

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

The Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions include student, teacher and principal questionnaires. Students are asked about their age; language spoken at home; attitudes toward reading, writing and mathematics; home help with reading, writing and mathematics; use of computers at school for reading, writing and mathematics learning; and use of calculators at school for mathematics learning. Teachers are asked about the size and makeup of their class (e.g., single, split or multi-grade); the usefulness of the resources in their classroom; the accessibility of their classroom resources and the frequency with which they are used; collaboration with colleagues to plan and discuss curriculum and assessment; their use of assessment-related materials from the EQAO Web site; and their gender, experience, qualifications and participation in professional development. Principals are asked about the school's characteristics (e.g., grades taught, enrolment, student absenteeism, number of staff and communication with parents and guardians about EQAO); and their gender; education; participation in professional development; knowledge about instruction and assessment, and use of data for school improvement planning; communication with parents and staff about EQAO testing; and use of the *EQAO Guide to School and Board Improvement Planning* (2005). (Copies of this guide can be obtained from the EQAO Web site.)

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

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

CHAPTER 3: TEST ADMINISTRATION AND PARTICIPATION

Assessment Administration

The Administration Guides

The administration guide for each 2006–2007 EQAO assessment can be found on the EQAO Web site at www.eqao.com.

Administering the Assessment to Students

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-language or French-language administration guide for training teachers to administer the assessment. The guide outlines in detail what is expected of educators involved in the administration, including

- the procedures to follow (e.g., preparation of materials for distribution to students, do's and don'ts during the assessment);
- the accommodations and special provisions allowed for English-language learners and students with special education needs;
- 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 work independently to answer multiple-choice questions and write their responses to open-response items. Students must work in a quiet environment and be supervised at all times.

Support for English-Language Learners and Students with Special Education Needs

The 2006–2007 administration 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 exemptions and deferrals. Students with special education needs are allowed accommodations and English-language learners are provided with special provisions to ensure that they can participate in the assessment and demonstrate the full extent of their skills. In cases where the accommodations or special provisions do not address a student's needs, exemption from participation in the assessment is allowed. Each year, EQAO reviews and updates these provisions and accommodations to ensure that they reflect new developments in supports for students.

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

Definition of “Accommodation”

Accommodations are defined in the 2005–2006 administration guides (which are based on Policy/Program Memorandum 127) as follows:

“Accommodations” are supports and services that enable students with special 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. They do not alter the content of the assessment or affect the validity or reliability.

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

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

Special Version Assessments for Accommodated Students

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

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

Special Provisions for English-Language Learners

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

Exemptions

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

- reading, if the student has to be read to by a teacher or another adult.
- writing, if the student has to be scribed for (i.e., cannot write independently).
- mathematics, if mathematics terms have to be defined.

If a student is unable to participate in the Grade 9 Assessment of Mathematics even with accommodations, the student may be exempted.

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

- the student has not yet acquired a level of English proficiency sufficient to participate in the assessment and is registered in a Level 1, 2, 3 or 4 ESL or ELD course or a Level 5 ESL (Bridge to English) course.
- the student has been identified as exceptional by an Identification, Placement and Review Committee and is not able to participate in the assessment, even with permitted accommodations.
- the student has not yet acquired the reading and writing skills appropriate for Grade 9.
- the student is new to the board, requires accommodations and such appropriate accommodations cannot be provided.

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

Quality Assurance

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

- *Quality assurance monitors:* EQAO contracts quality assurance monitors to visit and observe the administration of the assessments (in a random sample of schools), to determine the extent to which EQAO guidelines are being followed.
- *Database analyses:* EQAO conducts two types of statistical analyses of student response data. The first analysis identifies student response patterns that suggest the possibility of collusion between two or more students. The second examines overall patterns of school results (for open-response items) and data on the proportion of students performing at or above the provincial standard (Level 3) over time. In the case of the OSSLT, the data are analyzed for unusual changes in the passing rate of a school.
- *Examination of test materials:* EQAO looks for evidence of possible irregularities in the administration of assessments in schools, following each assessment. This is done through an examination of test materials from a random sample of schools, prior to scoring.

CHAPTER 4: SCORING

EQAO follows rigorous scoring procedures to ensure that its assessment results are valid and reliable. All responses to open-response field-test and operational reading and mathematics items, as well as writing prompts are scored by trained scorers. The multiple-choice items, however, are machine scored.

Item-specific scoring rubrics and anchors are the key tools used for scoring open-response reading and mathematics items and writing prompts. In order to maintain consistency across items and years, item-specific rubrics for open-response items are based on generic rubrics. Each student response to an open-response item is scored according to its best match, with one of the score points in the rubric for the item. EQAO scoring rubrics describe work at different levels of student performance. Anchors illustrate the descriptors for each score point in the rubrics. The anchors are chosen and validated by expert educators from across the province during the range finding process. Scorers are trained to constantly refer to the anchors to ensure consistent scoring. Scorers match student work to the score point descriptors in the rubrics, and anchors that illustrate student work for the score points. 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 scoring assessments) are included in each 2006-2007 EQAO framework document.

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 score point in each scoring rubric, to arrive at consensus scores for student responses. This range is used to train scorers. The process is completed in two stages: the pre-range finding stage and the range finding stage.

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. The long writing prompts on the primary and junior assessments, and on the OSSLT are field trialed with a limited number of students (and not field tested). As such, range finding for these prompts uses student responses to operational items and occurs just prior to operational scoring.

Pre-Range Finding Stage

Pre-range finding selects responses that represent the full range of score points for each item or prompt. These responses are used by the range-finding committee.

Overview of the Pre-Range Finding Process

1. EQAO education officers and content specialists are responsible for pre-range finding, as well as reviewing items, prompts and generic rubrics, prior to the pre-range finding process.

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

Range Finding Stage

Range finding

- brings in subject experts from the Ontario education system to prepare high-quality scoring tools and training materials for scorers, in order to ensure the accurate, reliable and consistent scoring of open-response items on EQAO assessments;
- selects representative samples of student responses to define and illustrate the range of student performances within the scoring rubrics codes and
- provides expert consensus coding of student responses used to train scorers of open-response items.

Range-finding committees, comprised of 8 – 10 Ontario educators, meet up to three times a year to make recommendations about student responses which will be used as anchors during scoring. They also discuss training for scorers, the qualifying test for scorers following training, and the selection of papers for calibration activities during scoring.

The qualifications for range-finding committee members include

- expertise and experience in the application of rubrics based on the Achievement Charts in *The Ontario Curriculum* (to identify varying levels of student performance in language and mathematics);
- the ability to explain clearly and concisely the reasons why a student response is at one of the levels in a rubric and
- expertise in and current experience with the curriculum and the grades being assessed:
 - primary assessment: language and mathematics in the primary division (Grades 1 to 3)
 - junior assessment: language and mathematics in the junior division (Grades 4 to 6)
 - Grade 9 assessment: mathematics in the intermediate division (Grades 7 to 10)
 - OSSLT: literacy across the curriculum in the intermediate division (Grades 7 to 10).

Members of the range-finding committees:

- use their scoring expertise to assign the appropriate generic rubric and/or item-specific rubric codes to a set of student responses for each group of assessment items;
- share the codes they have assigned with other members of the committee;

- work collaboratively with other members of the committee, under the guidance of an EQAO content specialist, 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 rationales explaining the codes assigned.

Overview of the Range Finding Process

1. Range finding committee members (including literacy experts and current classroom teachers) are recruited and selected for each assessment item.
2. Range finding committee meetings are run by EQAO content specialists and education officers.
 - To ensure productivity and efficiency during meetings, the committee is divided into groups (usually of three or four members).
 - Each group discusses a set of items, prompts and associated item-specific scoring rubrics and identifies, through scoring, the appropriate responses to be used to train scorers of each task.
 - The discussions focus on
 - content and requirements of each item or task;
 - consensus on scores for student responses for the group and
 - scoring rules, as required, to ensure consistent scoring of each item or task.

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

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

The training materials include

- anchors that solidly represent the score points in scoring rubrics;
- training papers that represent both solid score-point responses and unusual and/or less-solid responses (e.g., shorter than normal, atypical approaches, mix of very low and very high attributes, and difficult-to-decipher handwriting);
- rationales for each anchor and training paper used;
- solid score-point responses for one or more of the qualifying tests;
- papers used for ongoing training during the daily calibration activity (operational scoring only) and
- papers used for monitoring validity (operational scoring only).

Education officers arrange for sufficient copies of materials to train scorers of field-test items.

Field-Test Scoring

Field-test scoring occurs following operational scoring. Field-test items are scored according to the high standards applied to the scoring of operational items, since these items will be used in future assessments. The most reliable and productive scoring leaders and scorers of operational items are selected to score items similar to those scored on the operational assessment, to ensure consistency of year-to-year scoring and to reduce the time required to train scorers.

Scoring Open-Response Field-Test Responses

A sample of 1500 demographically representative English-language and French-language student responses for each field-test item or prompt is scored. Student confidentiality is maintained at all times, and all training materials are kept secure.

In-depth training for the first item and prompt are provided to scorers by their scoring leader. Scorers then write a qualifying test on that item or prompt, before the scoring begins. Scorers are trained on each item, and complete scoring of an item before proceeding to the next item. There are no qualifying tests for subsequent items and prompts. All OSSLT field-test item responses are double scored.

Item analysis statistics reports are prepared following field-test scoring. These reports, together with reports related to field-test item and prompt performance, are used to inform test construction.

Training Field-Test Scoring Leaders and Scorers

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

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

- an introduction to the purpose of field-test scoring;
- an explanation of the need to report suspected abuse to the Children's Aid Society;
- a grounding in field-test scoring procedures (using the first item or prompt and its scoring rubric, anchors and training papers);
- a qualifying test on the first item or prompt;
- an introduction to subsequent items and prompts and their scoring rubric, anchors and training papers prior to the scoring of each item.

Standards for passing the qualifying test and for scoring operational items are the same. All scoring leaders and scorers are required to qualify on the first item in order to participate in field-test scoring.

Developing Additional Scoring Training Materials Before Scoring Operational Items

Sometimes a full range of training materials for short writing prompts and open-response reading and mathematics items is not found during the range finding process for field-test scoring. When this occurs, EQAO pulls field-test booklets with responses that received identical scores during field-test scoring, so that additional scoring materials can be developed. Education officers then bundle the booklets so that range finders can select additional scorer-training materials for operational scoring such as anchors, training papers, qualifying tests for first items or prompts and validity papers.

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

Scoring Open-Response Operational Items

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

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

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

Scoring Rooms for Scoring Open-Response Operational Items

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

Scorers sit alone at tables and score individually. Each bundle of booklets is scored twice by two separate scorers. Scorers can discuss anomalous responses with their scoring leader or supervisor.

Open-response reading, writing and mathematics items for primary and junior assessments and operational mathematics items for the Grade 9 assessment are single scored. To measure overall scorer reliability, 10% of the total daily scores are randomly routed for a second round of scoring.

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

Exact agreement means that the marks assigned to an open-response item by a pair of markers are exactly the same. Adjacent agreement means that there is a difference of one score point between the marks assigned to an open-response item by a pair of markers. Non-adjacent means that there is a difference larger than one score point between the marks assigned to an open-response item by a pair of scorers.

Training for Scoring Open-Response Operational Items

The purpose of training is to develop a clear and common understanding of the scoring materials so that each scoring leader, scoring supervisor and scorer interprets and applies the scoring materials in the same way, resulting in consistent, reliable and valid student scores (regardless of which scorer scores the items).

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 to align their thinking and judgment to the procedures and scoring materials for the items being scored. The responsibilities of scoring leaders include

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

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

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

The training for scoring leaders and scoring supervisors is conducted in the days immediately preceding scoring. EQAO education officers train scoring leaders and oversee the training of scoring supervisors. Scoring leaders then train scoring supervisors. Supervisor training is substantially similar to the training and qualifying that scorers receive. The only difference is that supervisors receive additional training regarding scoring materials, room management problems and issues that may arise during scoring.

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

Training of Scorers for Scoring Open-Response Operational Items

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

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

EQAO scoring procedures and processes have been developed to ensure consistency in scoring. The purpose of training is to develop a clear and common understanding of the scoring materials so that each scoring leader, scoring supervisor and scorer interprets and applies the scoring materials in the same way, resulting in reliable, consistent, valid, and accurate scoring of student responses (regardless of which scorer scores the responses). All scorers in a room are trained to score using the same scoring materials. During training, scorers are told they may have to adjust their thinking about scoring student performance in order to accept EQAO's standards and practices for scoring assessments.

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

- a brief overview of the assessment being scored (including the purpose and goal of the assessment and the unique features of the assessment and/or assessment population).
- general instructions about the security, confidentiality and suitability of the scoring materials.
- instruction on entering scores into the Personal Digital Assistant (PDA) used to collect scoring data:
 - prior to entering the scores, scorers will scan the unique student booklet barcodes into the PDA (which has a built in barcode scanner) in order to link student names to their corresponding scores.
 - scorers will enter their scores for student responses into the PDA, then synchronize the PDA in a cradle connected to a laptop which uploads the data to the server.
- a thorough review and discussion of the scoring materials for each item to be scored (item, generic rubric, item-specific rubric, anchors and their rationales). These scoring materials are approved by EQAO and cannot be altered by scorers.
 - 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 score point on the rubric.
 - the anchors consist of responses that are typical (rather than unusual or uncommon), solid (rather than controversial or borderline) and true (meaning they cannot be changed by anyone).
- scoring of a series of training papers, consisting of selected expertly scored student responses.
 - training paper sets typically contain responses that are solid, as well as responses that are unusual and/or less solid (e.g., are shorter than normal, employ atypical approaches, contain both very low and very high attributes or are written in difficult-to-decipher handwriting).
 - scorers will first score the responses, then compare their scores with the correct scores provided along with rationales.
 - scorers will then discuss the rationale and characteristics of each correct response with their scoring leader and supervisor.
 - during this process, the scorer's job is to internalize the rubric and adjust his or her individual scoring to conform to it. Once all training papers have been scored and discussed, scorers begin the qualifying process.

Scorers are also trained to

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

Following training and prior to scoring, scorers must pass a qualifying test consisting of 14 to 20 student responses to all the items to be scored in their rooms. 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 they 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

Evidence of Students at Risk

On occasion, a student's response to an open-response question will contain evidence that he or she may be at risk (e.g., content that states or implies threats of violence to self or others or possible abuse or neglect).

Reporting Procedure

Copies of student responses that raise concerns are sent to the student's local Children's Aid Society. It is the legal responsibility and duty of scorers, in consultation with their site managers, to inform the Children's Aid Society of such cases.

Reporting of Inappropriate Content, Teacher Interference and Other Issues

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

Exceptions Room (Issues)

Offensive content, variations in handwriting (in a single student response booklet) and damaged or misprinted booklets are sent to the Exceptions Room.

Offensive Content

Obscene, racist or sexist content in student response booklets are further reviewed by scorers, who use their experience with young people to determine whether the school should be contacted. If the offensive content warrants the notification of the school, the booklets are sent to EQAO who contacts the school.

Different Handwriting

Different types of handwriting in a booklet may indicate some form of irregularity (e.g., the student was cheating) or the student had a scribe. Booklets containing variations in handwriting (two or more different types of handwriting) or corrections are reviewed by EQAO staff, who determine whether the school should be notified.

Damaged or Misprinted Booklets

Sometimes booklets are torn, misstapled, 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 who determine whether the results of these booklets should be pro-rated based on the results of undamaged booklets.

Ongoing Daily Training

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

Daily Morning Review of Anchors

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

Daily Afternoon Calibration Activity

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

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

Scoring leaders and supervisors receive daily data reports showing daily and cumulative validity, reliability and productivity data for individual scorers and for groups of scorers in their room. The five data reports and the data they contain are outlined below.

Daily and Cumulative Validity Reports

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

Exact agreement means that the marks assigned to an open-response item by a pair of markers are exactly the same. Adjacent agreement means that there is a difference of one score point between the marks assigned to an open-response item by a pair of markers. Non-adjacent means that there is a difference larger than one score point between the marks assigned to an open-response item by a pair of scorers.

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

The reports also include a cumulative-trend review and are summarized by item or item set, rubric, room, group and scorer. Scorers are listed from low to high validity. Scorers not meeting the exact-agreement requirement of 75% on three-point rubrics, 70% on four-point rubrics or 60% on six-point rubrics are highlighted in the report.

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

The daily validity report on scoring leaders and supervisors doing OSSLT validation or resolution scoring, summarizes daily level-of-agreement data for validity booklets scored during adjudication and resolution scoring. This report is sorted by rubric, room, scoring leader and supervisor. Exact, adjacent and non-adjacent agreement on validity booklets with preset scores are reported. The report sorts scoring leaders and supervisors from low to high validity. Scoring leaders and supervisors not meeting the exact-agreement requirement of 85% on three-point rubrics and 80% on four- and six-point rubrics are highlighted in the report.

Daily and Cumulative Mean Score and Score-Point-Distribution Reports

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

This report identifies and summarizes (by item or item set, room, group, rubric and scorer) the daily and cumulative mean score, the distribution of assigned score points and the high or low scoring among individual scorers.

Daily and Cumulative Scoring Centre Reliability Reports

During scoring, EQAO tracks reliability and consistency by means of blind (unidentified) reinsertion of booklets. Open-response primary, junior and Grade 9 assessment items in reading, writing and mathematics are single scored and 10% of all student responses are routed for a second scoring (to track reliability). All open-response OSSLT items are routed for a second scoring, which is used to monitor interrater reliability. Scoring leaders and supervisors use this secondary source of data, in conjunction with validity data, to determine whether any individual scorer or group of scorers requires retraining.

The reports identify and summarize daily and cumulative levels of interrater agreement on two independent scorings, including exact, adjacent, and high and low non-adjacent agreement. The reports are summarized by item or item set, room, group, rubric and scorer and sort scorers from low to high reliability. Scorers not meeting the exact-agreement requirement of 75% on three-point rubrics, 70% on four-point rubrics or 60% on six-point rubrics are highlighted in the report.

Daily and Cumulative Productivity Reports

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 scorers in each room taking into consideration the subset of items they are scoring. Productivity targets for scoring supervisors are 50% to 60% of scorer targets.

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

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

Aggregated Daily and Cumulative Individual Scorer Data Reports

These reports combine validity data with secondary data for each scorer. The reports include not only daily and cumulative validity data, but daily and cumulative reliability, mean score and productivity data. The reports sort scorers from low to high validity. Scorers not meeting the exact-agreement requirement of 75% on three-point rubrics, 70% on four-point rubrics or 60% on six-point rubrics are highlighted in this report. As such, this report assists the scoring leaders in identifying the scorers that require retraining.

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

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

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

Scoring leaders share the data and data-related issues with the appropriate scoring supervisors so that interventions can be planned to ensure the quality of individual and group scoring.

The following occurs when scorers are not meeting the validity metrics:

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

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

Audits

EQAO audits individual student marksheets for inconsistencies that may indicate incomplete scoring.

Booklets containing blank responses to reading or writing items in primary or junior language or OSSLT assessments are pulled and checked. Similarly, OSSLT booklets, containing blank responses from students with a computer accommodation, are pulled and checked for computer printout pages stapled into the booklet.

Additionally, booklets containing blank responses to primary, junior, Grade 9 or OSSLT multiple-choice or open-response items are pulled and checked.

Marker Reliability and Validity

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

Marker Reliability

Test reliability is affected by different sources of measurement error. In the case of open-response items, the inconsistency in marking is the source of error. As part of its attempt to minimize this type of error, EQAO has implemented a reinsertion procedure for the primary, junior and Grade 9 assessments to help ensure marking consistency (see above). The reinsertion procedure involves the random rescoring of pre-selected booklets. The booklets identified for reinsertion are circulated during the scoring session so that each open-response item is scored by pairs of randomly assigned scorers.

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

Exact agreement means that the marks assigned to an open-response item by a pair of markers are exactly the same. Adjacent agreement means that there is a difference of one score point between the marks assigned to an open-response item by a pair of markers. Non-adjacent agreement means that there is a difference larger than one score point between the marks assigned to an open-response item by a pair of scorers.

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

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

Assessment of Reading, Writing and Mathematics: Primary Division

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

For reading, the percentage of exact agreement for English-language individual items ranged from 70.3% to 81.1% and from 61.1% to 86.0% for French-language items (Tables 4.1 and 4.5). The percentage of exact-plus-adjacent agreement was larger and less variable, ranging from 94.9% to

98.3% for English-language and 95.9% to 99.7% for French-language items. The corresponding aggregate values were 76.2% and 96.5% for the English-language assessment and 72.7% and 97.5% for the French-language assessment. With the exception of one item on the French-language assessment, these indices met EQAO's targets of 70% exact agreement and 95% for exact-plus-adjacent agreement.

For writing (Tables 4.2 and 4.6), the percentages of exact agreement for French-language were considerably smaller than the percentages for English-language, as reflected by the aggregate percentages of exact agreement (48.2% for French and 71.2% for English). However, as with reading, the percentages of exact-plus-adjacent agreement were larger and less variable. They ranged from 93.1% to 98.7% for English-language and from 82.0% to 96.0% for French-language. The percentage of exact-agreement for all items on the English-language assessment met the EQAO targets (60% for the six-point rubrics for topic development and 70% for the four-point rubrics for conventions). The corresponding percentage for the French-language assessment was below the EQAO targets for all items.

For the mathematics component of the primary assessment, the English-language assessment instrument is translated into French and administered to students in the French Immersion program in some school boards. Since the student booklets for these French Immersion students are scored by a separate group of markers, reliability results are reported separately. The percentages of exact- and exact-plus-adjacent agreement for the mathematics items were large (Tables 4.3, 4.4 and 4.7). The aggregate percentages of agreement were, respectively, 82.8% and 97.5% for English-language items, 77.0% and 96.4% for French Immersion and 78.5% and 95.7% for French-language items. The indices for all items except one French item met the EQAO target of 70% exact-agreement and the indices for all except three French-language items met the 95% target for exact-plus-adjacent agreement. For these three items, the indices were one or two points lower than the target.

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non- Adjacent
Booklet 1					
1	6 886	81.1	96.4	15.3	3.6
5	8 766	77.4	97.3	19.9	2.7
6	8 765	80.4	98.3	17.9	1.7
11	7 563	78.8	98.3	19.6	1.7
12	7 563	76.5	97.6	21.1	2.4
Booklet 2					
12	6 886	79.7	96.3	16.6	3.7
21	11 154	77.5	95.6	18.1	4.4
22	11 153	72.1	96.5	24.4	3.5
28	10 777	73.4	95.1	21.7	4.9
29	10 777	70.3	94.9	24.6	5.1
Aggregate	90 290	76.2	96.5	20.3	3.5

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
23 Topic Development	7 800	66.0	93.4	27.4	6.7
23 Conventions	7 801	72.8	98.1	25.3	1.9
Aggregate Long Writing	15 601	69.4	95.7	26.3	4.3
13 Topic Development	8 296	71.5	95.3	23.9	4.7
13 Conventions	8 298	76.7	98.2	21.6	1.8
14 Topic Development	8 703	65.4	93.1	27.7	6.9
14 Conventions	8 702	74.5	98.7	24.2	1.3
Aggregate Short Writing	33 999	72.0	96.3	24.4	3.7
Aggregate Across All Items	49 600	71.2	96.1	25.0	3.9

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
8	7 389	78.0	95.7	17.6	4.3
9	7 389	86.3	96.1	9.8	3.9
10	7 390	85.8	99.1	13.3	0.9
11	7 390	86.2	97.7	11.5	2.3
30	9 448	76.2	98.0	21.7	2.1
31	9 450	84.4	96.8	12.4	3.2
32	9 450	82.8	97.8	15.1	2.2
33	9 448	83.9	98.8	14.9	1.2
Aggregate	67 354	82.8	97.5	14.7	2.5

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
8	1 204	73.0	94.6	21.6	5.4
9	1 204	78.8	95.5	16.7	4.5
10	1 204	79.2	97.1	17.9	2.9
11	1 204	80.9	94.7	13.8	5.3
30	1 329	69.5	97.4	27.8	2.6
31	1 329	80.0	96.2	16.3	3.8
32	1 329	78.6	98.3	19.7	1.7
33	1 329	76.5	96.9	20.5	3.1
Aggregate	10 132	77.0	96.4	19.4	3.6

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
Booklet 1					
5	684	76.3	98.3	21.9	1.8
6	684	71.9	97.7	25.7	2.3
11	684	86.0	99.7	13.7	0.3
12	684	76.9	99.4	22.5	0.6
Booklet 2					
11	1 383	70.5	95.4	25.0	4.6
12	1 382	61.1	96.5	35.5	3.5
21	291	77.0	99.0	22.0	1.0
22	291	80.8	99.0	18.2	1.0
28	291	76.0	98.6	22.7	1.4
29	291	73.9	95.9	22.0	4.1
Aggregate	6 665	72.7	97.5	24.9	2.5

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
23 Topic Development	1 408	49.9	89.3	39.4	10.7
23 Conventions	1 408	54.4	96.0	41.6	4.0
Aggregate Long Writing	2 816	52.2	92.7	40.5	7.4
13 Topic Development	1 412	43.6	82.0	38.4	18.0
13 Conventions	1 412	47.5	91.8	44.3	8.2
14 Topic Development	1 511	45.2	85.8	40.6	14.2
14 Conventions	1 512	48.9	93.4	44.4	6.6
Aggregate Short Writing	5 847	46.3	88.3	42.0	11.7
Aggregate Across All Items	8 663	48.2	89.7	41.5	10.3

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
8	1 306	87.7	97.0	9.3	3.0
9	1 306	70.4	94.1	23.7	5.9
11	1 307	77.4	93.4	16.1	6.6
12	1 307	73.8	94.7	20.9	5.3
30	1 429	86.7	97.3	10.6	2.7
31	1 429	61.4	97.1	35.7	2.9
32	1 429	91.7	97.8	6.0	2.2
33	1 429	78.7	93.5	14.8	6.5
Aggregate	10 942	78.5	95.7	17.1	4.4

Assessment of Reading, Writing and Mathematics: Junior Division

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

The aggregate percentages of exact agreement for English-language reading, writing and mathematics were 73.4%, 65.4% and 81.9%, respectively. The percentages for French-language were 64.0%, 48.6% and 72.3%. The percentages of exact-plus-adjacent agreement for each subject in both languages were large and less variable; the aggregate percentages ranged from 89.7% to 97.3%.

The EQAO target of 70% exact agreement was met for all but three items on the English-language reading component and for half the items on the French-language reading component. The EQAO target of 95% exact-plus-adjacent agreement was met for all items in both assessments.

The percentages of exact agreement met the EQAO target for two out of three prompts on the English writing component but were below the EQAO target for all prompts on the French writing component. The percentage of exact-plus-adjacent agreement met the EQAO target for half the items on both assessments.

The EQAO targets of 70% exact agreement and 95% exact-plus-adjacent agreement were met for all mathematics items on the English assessment. These targets were met for five out of eight mathematics items on the French assessment.

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non Adjacent
Booklet 1					
5	5 585	72.6	97.0	24.4	3.0
6	5 585	69.4	96.1	26.8	3.9
11	9 315	80.0	96.6	16.6	3.4
12	9 316	78.7	97.6	18.9	2.4
Booklet 2					
11	12 248	66.2	97.1	31.0	2.9
12	12 252	63.8	94.9	31.1	5.1
21	10 175	79.4	98.9	19.6	1.1
22	10 175	72.9	98.1	25.2	1.9
28	8 394	76.3	98.0	21.7	2.0
29	8 393	78.2	98.9	20.7	1.1
Aggregate	91 438	73.4	97.3	23.9	2.7

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
23 Topic Development	10 025	64.5	91.4	26.9	8.6
23 Conventions	10 025	73.7	98.4	24.7	1.7
Aggregation Long Writing	20 050	69.1	94.9	25.8	5.1
13 Topic Development	8 215	62.0	91.5	29.5	8.5
13 Conventions	8 215	70.2	97.7	27.6	2.3
14 Topic Development	11 778	57.0	90.6	33.6	9.4
14 Conventions	11 778	66.6	97.5	30.9	2.5
Aggregation Short Writing	39 986	63.6	94.3	30.7	5.7
Aggregate Across All Items	60 036	65.4	94.5	29.1	5.5

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
8	10 965	80.2	97.6	17.3	2.5
9	10 972	79.6	95.9	16.2	4.2
10	10 963	84.6	97.7	13.1	2.3
11	10 965	82.0	97.6	15.7	2.4
30	9 373	82.2	97.4	15.2	2.6
31	9 373	84.1	97.9	13.8	2.1
32	9 373	82.6	95.9	13.3	4.1
33	9 373	80.1	97.3	17.2	2.7
Aggregate	81 357	81.9	97.2	15.3	2.8

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
Booklet 1					
5	2 212	78.2	98.6	20.3	1.5
6	2 212	62.3	97.6	35.3	2.4
11	2 212	57.7	95.5	37.8	4.5
12	2 212	59.2	96.6	37.3	3.4
Booklet 2					
11	1 424	60.7	96.3	35.6	3.7
12	1 422	55.9	95.6	39.7	4.4
21	239	85.4	99.6	14.2	0.4
22	239	77.8	98.7	20.9	1.3
28	239	74.1	97.9	23.9	2.1
29	239	75.7	98.3	22.6	1.7
Aggregate	12 650	64.0	96.9	32.9	3.1

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
23 Topic Development	1 500	46.3	87.2	40.9	12.8
23 Conventions	1 500	56.0	96.1	40.1	3.9
Aggregate Long Writing	3 000	51.1	91.7	40.5	8.3
13 Topic Development	1 463	39.6	81.5	41.9	18.5
13 Conventions	1 460	54.0	95.0	41.0	5.0
14 Topic Development	1 349	45.3	85.7	40.4	14.3
14 Conventions	1 349	55.7	96.9	41.1	3.1
Aggregate Short Writing	5 621	48.6	89.7	41.1	10.3
Aggregate Across All Items	8 621	49.5	90.4	40.9	9.6

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

Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non Adjacent
8	1 612	67.2	91.8	24.6	8.3
9	1 612	75.1	97.4	22.3	2.6
11	1 612	68.4	93.1	24.7	6.9
12	1 612	80.6	97.6	17.1	2.4
30	1 309	70.3	96.2	25.9	3.8
31	1 309	78.9	96.3	17.4	3.7
32	1 309	71.8	94.0	22.2	6.0
33	1 309	66.0	96.4	30.4	3.6
Aggregate	11 684	72.3	95.3	23.0	4.7

Grade 9 Assessment of Mathematics: Academic and Applied

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

As shown in these tables, the marking was highly consistent. For example, the aggregate percentages of exact agreement among pairs of markers ranged from 79.0% to 89.3% across the languages, programs and administrations. The aggregate percentages of exact-plus-adjacent agreement ranged from 96.2% to 98.8%. The percentages of exact agreement and exact-plus-adjacent agreement exceeded the EQAO targets of 70% and 95% for all items on the assessments of both languages, with the exception of exact agreement for one French item in the applied program. The percentage of exact agreement was above 80% for 43 of the 56 items across all languages, programs and administrations.

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

Administration	Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
	8	2 359	88.7	96.7	8.0	3.3
	9	2 359	81.4	97.5	16.0	2.5
	11	2 359	86.2	97.4	11.2	2.6
Winter	26	2 358	86.1	96.0	9.9	4.0
	27	2 359	86.2	98.3	12.1	1.7
	28	2 359	85.4	96.1	10.7	3.9
	29	2 359	88.3	98.0	9.7	2.0
Aggregate		16 512	86.0	97.1	11.1	2.9
	8	2 863	85.2	95.6	10.5	4.4
	9	2 863	87.1	98.0	10.9	2.0
	11	2 863	93.0	98.2	5.2	1.8
Spring	26	3 377	90.1	96.8	6.7	3.2
	27	3 377	88.6	97.6	9.0	2.4
	28	3 377	88.8	97.7	8.9	2.3
	29	3 377	91.9	99.2	7.4	0.8
Aggregate		22 097	89.3	97.6	8.3	2.4
Aggregate Across Administrations		38 609	87.9	97.4	9.5	2.6

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

Administration	Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
	8	2 251	78.1	97.2	19.1	2.8
	9	2 252	76.6	96.5	19.9	3.6
	11	2 252	83.8	95.6	11.8	4.4
Winter	26	2 278	81.3	97.1	15.8	2.9
	27	2 278	86.6	96.6	10.0	3.4
	28	2 278	77.5	96.9	19.4	3.1
	29	2 279	87.9	99.3	11.4	0.8
Aggregate		15 868	81.7	97.0	15.3	3.0
	8	2 655	82.6	95.9	13.3	4.1
	9	2 655	83.1	96.3	13.3	3.7
	11	2 655	87.5	97.5	10.1	2.5
Spring	26	4 026	86.4	97.6	11.1	2.4
	27	4 026	76.4	97.5	21.1	2.5
	28	4 026	74.9	94.5	19.6	5.5
	29	4 026	80.6	94.9	14.3	5.1
Aggregate		24 069	81.2	96.3	15.1	3.7
Aggregate Across Administrations		39 937	81.4	96.6	15.2	

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

Administration	Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
	8	814	74.5	95.7	21.3	4.3
	9	809	80.2	97.4	17.2	2.6
	11	813	66.4	96.1	29.6	3.9
Winter	26	680	86.6	95.9	9.3	4.1
	27	680	84.7	96.8	12.1	3.2
	28	680	81.8	95.0	13.2	5.0
	29	680	81.8	98.5	16.8	1.5
Aggregate		5 156	79.0	96.5	17.5	3.5
	8	872	89.6	98.4	8.8	1.6
	9	872	89.0	98.6	9.6	1.4
	11	873	88.2	99.5	11.3	0.5
Spring	26	575	88.4	97.6	9.2	2.4
	27	573	83.3	98.4	15.2	1.6
	28	575	92.2	99.3	7.1	0.7
	29	574	91.5	99.7	8.2	0.4
Aggregate		4 914	88.9	98.8	9.9	1.2
Aggregate Across Administrations		10 070	83.8	97.6	13.8	2.4

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

Administration	Item No.	No. of Pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
	8	1 506	87.3	97.5	10.2	2.5
	9	1 507	75.2	97.2	22.0	2.8
	11	1 506	82.5	97.3	14.9	2.7
Winter	26	1 251	83.9	96.6	12.7	3.4
	27	1 251	79.1	98.5	19.4	1.5
	28	1 251	77.2	96.6	19.3	3.4
	29	1 251	76.3	96.5	20.1	3.5
Aggregate		9 523	80.3	97.2	16.9	2.8
	8	1 703	83.9	96.7	12.7	3.4
	9	1 703	76.7	97.1	20.4	2.9
	11	1 703	84.3	94.6	10.3	5.4
Spring	26	1 306	83.2	97.9	14.6	2.1
	27	1 308	78.8	97.0	18.2	3.0
	28	1 306	81.3	94.6	13.3	5.4
	29	1 306	81.6	95.7	14.2	4.3
Aggregate		10 335	81.4	96.2	14.8	3.8
Aggregate Across Administrations		19 858	80.9	96.7	15.8	3.3

Ontario Secondary School Literacy Test (OSSLT)

The marking reliability estimates for the OSSLT are presented in Tables 4.18 to 4.21. All of the open-response items included in both language forms of the OSSLT are marked twice. The percentages of exact agreement for reading ranged from 53.1% to 65.7% for English and from 58.3% to 72.1% for

French. As with the other assessments, the percentages of exact-plus-adjacent agreement were larger and less variable. The aggregate percentages were 91.7% for English and 96.3% for French. The percentages of exact agreement for all reading items in both English and French were below the EQAO targets of 75% for three-point rubrics. The percentage of exact-plus-adjacent agreement met the EQAO targets of 95% for one out of four items in English and three out of four items in French.

For writing, each open-response item is marked for content and use of conventions. For long writing, a six-point rubric is used for content and a four-point rubric is used for conventions. For the short writing, the content and conventions rubrics were combined into a five-point scale. For long writing, the percentages of exact agreement ranged from 36.1% to 44.4% for content and from 53.6% to 61.4% for conventions. For short writing, the percentages of exact agreement ranged from 55.0% to 78.7%. The percentages of exact-plus-adjacent agreement ranged from 80.7% to 98.1%.

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

In the following tables, the letters in the item column designate the sections of the assessment booklets in which the items were positioned.

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

Item No.	No. of pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
A: Item 6	196 288	53.1	89.2	36.1	10.8
F: Item 7	196 288	63.0	87.2	24.1	12.9
J: Item 6	192 075	65.7	94.0	28.3	6.0
J: Item 7	192 075	59.8	96.4	36.6	3.6
Aggregate	776 726	60.4	91.7	31.3	8.3

Table 4.19 Interrater Reliability Estimates for Reading: OSSLT (French)

Item No.	No. of pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
A: Item 6	6 899	71.8	96.6	24.7	3.5
F: Item 7	6 899	72.1	94.3	22.2	5.7
J: Item 6	6 798	62.3	97.0	34.8	3.0
J: Item 7	6 798	58.3	97.2	38.9	2.8
Aggregate	27 394	66.2	96.3	30.1	3.7

Table 4.20 Interrater Reliability Estimates for Writing: OSSLT (English)

Item No.	No. of pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non-Adjacent
E-Content	202 142	39.6	85.2	45.6	14.8
E-Conventions	202 142	55.1	96.1	40.9	3.9
G-Content	191 036	44.4	90.0	45.6	10.0
G-Conventions	191 036	53.6	95.0	41.5	5.0
Aggregate Long Writing	786 356	48.2	91.5	43.4	8.5
D-Content & Conventions	197 131	78.7	95.8	17.1	4.2
K-Content & Conventions	192 476	77.5	96.3	18.8	3.7
Aggregate Short Writing	389 607	78.1	96.1	17.9	3.9
Aggregate Across All Items	1 175 963	58.1	93.0	35.0	7.0

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

Item No.	No. of pairs	% Exact	% Exact Plus Adjacent	% Adjacent	% Non Adjacent
E-Content	7 144	39.2	83.1	43.9	16.9
E-Conventions	7 144	61.4	98.1	36.7	1.9
G-Content	7 387	36.1	80.7	44.7	19.3
G-Conventions	7 387	55.0	96.3	41.3	3.7
Aggregate Long Writing	29 062	47.9	89.6	41.7	10.5
D-Content & Conventions	7 266	55.0	94.0	39.1	6.0
K-Content & Conventions	7 411	63.0	96.8	33.8	3.2
Aggregate Short Writing	14 677	59.0	95.4	36.4	4.6
Aggregate Across All Items	43 739	51.6	91.5	39.9	8.5

Marking Validity

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

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

Assessment of Reading, Writing and Mathematics: Primary Division

The validity estimates for reading, writing and mathematics for the primary assessment are provided in Tables 4.22 to 4.28. For reading, the percentages of exact agreement ranged from 56.4% to 75.6% for English and from 69.2% to 86.8% for French. The English-language markers tended to mark one score point lower than the expert panel more often than one score point higher (aggregates were 19.6% and 12.8%, respectively, as shown in Table 4.22). Among French-language markers, the assigned marks above or below the expert panel were evenly divided (9.8% for adjacent-low and 11.2% for adjacent-high). The percentages of exact-plus-adjacent agreement were larger and less variable, ranging from 92.7% to 99.2% for English and 96.3% to 99.5% for French. The percentages of exact agreement met the EQAO target of 70% for four-point rubrics for three out of 10 English items and for nine out of 10 French items. The percentage of exact-plus-adjacent agreement met the EQAO target of 95% for seven out of 10 English items and all French items.

The percentages of exact agreement for writing were smaller than that for reading (see Tables 4.23 and 4.27). They ranged from 50.2% to 67.1% for English and from 69.3% to 74.2% for French. English-language scorers had a greater tendency to mark lower than the expert panel did for long writing (26.1% lower and 12.0% higher) but were more evenly divided for the short items (18.3% lower and 18.5% higher). French-language markers had a greater tendency to mark lower than the expert panel for long writing (16.5% lower and 10.2% higher) and for short writing (13.7% lower and 9.5% higher). The percentages of exact-plus-adjacent agreement were large for both long and short writing. The aggregates were 93.6% and 95.7%, respectively for English, and 98.0% and 94.6% respectively, for French. The percentage of exact agreement for all items in French met the EQAO targets of 60% for topic development and 70% for conventions. The percentages for all English items were below the EQAO targets. The percentage of exact-plus-adjacent agreement met the EQAO target of 95% for three out of six English items and for four out of six French items.

The percentages of exact agreement were larger for mathematics than for reading and writing (see Tables 4.24, 4.25 and 4.28). The aggregate percentages of exact agreement were 82.4% for English, 78.3% for French Immersion and 86.3% for French. The aggregate percentages of exact-plus-adjacent agreement were 98.1%, 96.8% and 98.3%. The aggregate percentage adjacent-low was larger than the aggregate percentage adjacent-high for the scoring of the French Immersion responses, but similar for the scoring of the English and French responses. The percentage of exact agreement met the EQAO target of 70% for four-point rubrics on all items. This index was over 80% for a number of items; except for one item in the French Immersion scoring. The percentages of exact-plus-adjacent agreement met the EQAO target of 95%.

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
Booklet 1							
5	660	75.6	99.2	23.6	12.6	11.1	0.8
6	665	74.1	98.5	24.4	14.4	9.9	1.5
11	1 347	59.2	98.4	39.1	25.0	14.1	1.6
12	1 344	56.4	95.8	39.4	25.5	14.0	4.2
Booklet 2							
11	725	63.2	94.8	31.6	23.5	8.1	5.2
12	734	62.3	94.0	31.7	20.2	11.6	6.0
21	1 201	70.0	96.8	26.8	13.6	13.2	3.2
22	1 153	59.8	95.6	35.7	19.1	16.7	4.4
28	1 132	64.3	92.7	28.4	15.3	13.1	7.3
29	1 099	59.5	93.2	33.7	21.7	12.0	6.8
Aggregate	10 060	63.4	95.8	32.4	19.6	12.8	4.2

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
23 Topic Development	4 431	50.2	90.4	40.2	24.8	15.5	9.6
23 Conventions	4 562	60.6	96.8	36.2	27.5	8.7	3.2
Aggregate Long Writing	8 993	55.5	93.6	38.2	26.1	12.0	6.4
13 Topic Development	1 694	53.8	94.0	40.2	19.2	21.0	6.0
13 Conventions	1 645	63.5	97.9	34.4	18.6	15.8	2.1
14 Topic Development	1 964	51.4	91.9	40.5	22.1	18.5	8.1
14 Conventions.	2 029	67.1	99.2	32.1	13.5	18.6	0.8
Aggregate Short Writing	7 332	59.0	95.7	36.7	18.3	18.5	4.3
Aggregate Across All Items	16 325	57.1	94.6	37.5	22.6	14.9	5.4

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
8	1 759	76.2	97.0	20.8	9.6	11.2	3.0
9	1 723	88.7	98.2	9.5	6.0	3.5	1.8
10	1 688	85.7	98.2	12.6	6.5	6.0	1.8
11	1 690	83.6	98.3	14.7	6.1	8.6	1.7
30	1 885	70.7	97.2	26.5	17.5	9.0	2.8
31	1 834	86.2	98.6	12.4	5.8	6.6	1.4
32	1 807	85.3	98.5	13.1	7.5	5.6	1.6
33	1 863	83.5	98.9	15.4	9.8	5.5	1.1
Aggregate	14 249	82.4	98.1	15.7	8.7	7.0	1.9

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
8	213	70.4	94.8	24.4	20.2	4.2	5.2
9	215	88.4	97.2	8.8	6.1	2.8	2.8
10	213	80.3	98.1	17.8	11.7	6.1	1.9
11	216	77.3	95.8	18.5	13.4	5.1	4.2
30	247	70.9	94.3	23.5	16.2	7.3	5.7
31	244	85.7	98.4	12.7	3.7	9.0	1.6
32	243	81.1	98.8	17.7	12.4	5.4	1.2
33	244	72.5	96.7	24.2	12.7	11.5	3.3
Aggregate	1 835	78.3	96.8	18.5	12.0	6.5	3.2

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
Booklet 1							
5	312	76.9	99.4	22.4	10.9	11.5	0.6
6	346	79.2	98.8	19.7	10.4	9.3	1.2
11	364	86.8	99.5	12.6	9.3	3.3	0.6
12	328	79.3	98.8	19.5	12.2	7.3	1.2
Booklet 2							
11	211	86.3	97.6	11.4	3.8	7.6	2.4
12	211	69.2	97.6	28.4	9.0	19.4	2.4
21	236	74.6	97.5	22.9	6.8	16.1	2.5
22	205	75.1	99.5	24.4	8.8	15.6	0.5
28	218	70.6	98.6	28.0	13.8	14.2	1.4
29	242	70.3	96.3	26.0	11.2	14.9	3.7
Aggregate	2 673	77.5	98.5	21.0	9.8	11.2	1.5

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
23 Topic Development	122	72.1	98.4	26.2	18.0	8.2	1.6
23 Conventions	132	70.5	97.7	27.3	15.2	12.1	2.3
Aggregate Long Writing	254	71.3	98.0	26.8	16.5	10.2	2.0
13 Topic Development	176	69.3	94.3	25.0	15.3	9.7	5.7
13 Conventions	196	72.5	96.9	24.5	10.2	14.3	3.1
14 Topic Development	131	69.5	89.3	19.9	16.0	3.8	10.7
14 Conventions	163	74.2	96.3	22.1	14.1	8.0	3.7
Aggregate Short Writing	666	71.5	94.6	23.1	13.7	9.5	5.4
Aggregate Across All Items	920	71.4	95.5	24.1	14.5	9.7	4.5

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non Adjacent
8	196	96.9	99.0	2.0	2.0	0.0	1.0
9	173	76.9	98.8	22.0	8.7	13.3	1.2
11	182	84.6	96.7	12.1	7.1	5.0	3.3
12	173	82.1	98.3	16.2	14.5	1.7	1.7
30	224	91.1	99.6	8.5	1.8	6.7	0.5
31	179	77.7	98.3	20.7	10.1	10.6	1.7
32	243	93.4	97.9	4.5	3.3	1.2	2.1
33	172	82.0	97.1	15.1	6.4	8.7	2.9
Aggregate	1 542	86.3	98.3	12.0	6.4	5.6	1.8

Assessment of Reading, Writing and Mathematics: Junior Division

The pattern of percentage agreement for the junior assessment was similar to that of the primary assessment, but the percentages of exact agreement tended to be smaller for the junior assessment (see Tables 4.29 to 4.34). As with the primary assessment, the percentages of exact agreement for the junior assessment were larger for mathematics, followed by reading and then writing.

For reading, the percentages of exact agreement ranged from 46.7% to 77.5% for English and from 58.0% to 77.7% for French (see Tables 4.29 and 4.32). Percentages of exact-plus-adjacent agreement were larger and less variable, ranging from 90.1% to 99.1% for English and 95.2% to 100% for French. The percentages of exact agreement met the EQAO target of 70% for four-point rubrics on one out of ten English items and on two out of ten French items. The percentages of exact-plus-adjacent agreement met the EQAO target of 95% on seven out of ten English items and all French items.

The percentages of exact agreement for writing were smaller than those for reading (see Tables 4.30 and 4.33). The percentages of exact agreement ranged from 48.8% to 63.7% for English and from 65.2% to 76.0% for French. Percentages of exact-plus-adjacent agreement were larger and less variable, ranging from 90.6% to 98.0% for English and 94.6% to 100% for French. The percentages of exact agreement on all items in French met the EQAO targets of 60% for topic development and 70% for conventions. These percentages on all English items were below the EQAO targets. The percentages of exact-plus-adjacent agreement met the EQAO target of 95% on three out of six English items and on all French items.

For mathematics, the aggregate percentages of exact agreement were 80.7% for English and 77.2% for French (see Tables 4.31 and 4.34). The aggregate percentages of exact-plus-adjacent agreement were 98.2% and 98.1%. The percentages of exact agreement met the EQAO target of 70% for four-point rubrics on all items in English and on all but one item in French. The percentages of exact-plus-adjacent agreement met the EQAO target of 95% on all items.

There was little difference between the adjacent-low percentage and the adjacent-high percentage for English-language mathematics and for French-language reading and writing. The adjacent-low percentage was considerably larger than the adjacent-high percentage for English-language reading and French-language mathematics. The adjacent-high percentage was considerably larger than the adjacent-low percentage for English-language writing.

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
Booklet 1							
5	805	54.4	92.8	38.4	31.7	6.7	7.2
6	791	46.7	90.1	43.5	39.4	4.1	9.9
11	1 499	65.6	96.8	31.2	14.3	16.9	3.2
12	1 498	68.8	96.8	28.0	14.5	13.5	3.2
Booklet 2							
11	1 133	55.2	95.0	39.8	25.2	14.6	5.0
12	1 123	56.5	91.6	35.1	26.0	9.1	8.4
21	1 182	77.5	99.1	21.6	11.5	10.1	0.9
22	1 195	65.1	98.2	33.1	12.0	21.1	1.8
28	985	58.3	95.4	37.2	25.6	11.6	4.6
29	1 229	63.9	97.8	33.9	23.4	10.5	2.2
Aggregate	11 440	62.4	95.7	33.4	20.9	12.4	4.3

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
23 Topic Development	2 968	48.8	90.6	41.8	23.5	18.3	9.4
23 Conventions	3 109	54.2	96.0	41.8	12.4	29.5	4.0
Aggregate Long Writing	6 077	51.6	93.4	41.8	17.8	24.0	6.6
13 Topic Development	2 803	49.0	90.7	41.7	22.4	19.3	9.4
13 Conventions	2 742	63.0	98.0	35.0	12.9	22.1	2.0
14 Topic Development	1 201	54.9	92.2	37.3	17.6	19.7	7.8
14 Conventions	1 202	63.7	97.8	34.1	18.8	15.3	2.2
Aggregate Short Writing	7 948	57.0	94.5	37.6	17.8	19.7	5.5
Aggregate Across All Items	14 025	54.6	94.0	39.4	17.8	21.6	6.0

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
8	2 465	80.6	98.3	17.7	10.1	7.6	1.7
9	2 491	73.5	96.4	22.8	13.5	9.3	3.7
10	2 397	88.2	98.6	10.4	5.8	4.6	1.4
11	2 325	80.7	98.8	18.2	13.0	5.2	1.2
30	2 114	81.8	98.1	16.3	7.9	8.4	1.9
31	2 070	86.7	99.2	12.5	7.2	5.4	0.8
32	2 095	75.3	97.6	22.3	12.0	10.3	2.4
33	1 881	78.8	98.5	19.7	9.1	10.5	1.5
Aggregate	17 838	80.7	98.2	17.5	9.9	7.6	1.9

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
Booklet 1							
5	440	64.1	96.8	32.7	13.2	19.6	3.2
6	320	66.3	96.9	30.6	16.3	14.4	3.1
11	446	67.7	97.8	30.0	18.4	11.7	2.2
12	462	58.0	95.7	37.7	21.7	16.0	4.3
Booklet 2							
11	205	66.3	96.6	30.2	15.1	15.1	3.4
12	229	67.7	96.9	29.3	18.3	10.9	3.1
21	188	77.7	100.0	22.3	8.5	13.8	0.0
22	191	67.5	95.3	27.8	10.5	17.3	4.7
28	165	66.7	95.2	28.5	19.4	9.1	4.9
29	195	71.8	96.4	24.6	18.0	6.7	3.6
Aggregate	2 841	66.2	96.8	30.6	16.5	14.1	3.2

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
23 Topic Development	92	65.2	94.6	29.4	20.7	8.7	5.4
23 Conventions	100	76.0	100.0	24.0	13.0	11.0	0.0
Aggregate Long Writing	192	70.8	97.4	26.6	16.7	9.9	2.6
13 Topic Development	130	66.2	94.6	28.5	24.6	3.9	5.4
13 Conventions	150	74.0	97.3	23.3	7.3	16.0	2.7
14 Topic Development	74	74.3	94.6	20.3	14.9	5.4	5.4
14 Conventions.	88	71.6	96.6	25.0	1.1	23.9	3.4
Aggregate Short Writing	442	71.3	95.9	24.7	12.4	12.2	4.1
Aggregate Across All Items	634	71.1	96.4	25.2	13.7	11.5	3.6

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

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
8	199	74.9	99.5	24.6	12.6	12.1	0.5
9	200	81.5	99.0	17.5	8.5	9.0	1.0
11	165	72.7	97.0	24.2	15.8	8.5	3.0
12	179	86.6	100.0	13.4	9.5	3.9	0.0
30	183	74.9	100.0	25.1	12.0	13.1	0.0
31	245	80.0	96.7	16.7	7.4	9.4	3.3
32	184	79.4	97.8	18.5	13.0	5.4	2.2
33	204	67.2	95.6	28.4	20.1	8.3	4.4
Aggregate	1 559	77.2	98.1	21.0	12.2	8.8	1.9

Grade 9 Assessment of Mathematics

The validity study results for Grade 9 mathematics are reported in Tables 4.35 to 4.38. The percentages of exact agreement between the expert panel and the markers were large for both programs, both languages and both administrations. The aggregates percentage of exact-agreement ranged from 6.6% to 91.2%. The percentages of exact-plus-adjacent agreement were larger, with the aggregate values ranging from 98.3% to 100%. The percentages of exact agreement and exact-plus-adjacent agreement met the EQAO targets on all items. There were few validity papers for each item for the French-language scoring and no validity papers circulated for the two French-language items.

The percentages of adjacent-low agreement were larger than the percentages of adjacent-high agreement for all English-language marking groups and for applied French-language assessments. For the winter session of the academic-level French assessment, the percentage adjacent-low agreement was smaller than the percentage the adjacent-high agreement. For the spring session of the academic French assessment, the two aggregate percentages were equal.

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

Administration	Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
	8	122	80.3	99.2	18.9	16.4	2.5	0.8
	9	114	93.0	100.0	7.0	2.6	4.4	0.0
	11	154	83.8	97.4	13.6	9.1	4.6	2.6
Winter	26	92	65.2	100.0	34.8	27.2	7.6	0.0
	27	116	84.5	98.3	13.8	8.6	5.2	1.7
	28	94	89.4	100.0	10.6	6.4	4.3	0.0
	29	115	78.3	100.0	21.7	15.7	6.1	0.0
Aggregate		807	82.4	99.1	16.7	11.9	4.8	0.9
	8	165	81.2	98.8	17.6	11.5	6.1	1.2
	9	124	83.9	100.0	16.1	12.1	4.0	0.0
	11	151	88.1	99.3	11.3	9.9	1.3	0.7
Spring	26	219	75.8	99.1	23.3	17.4	5.9	0.9
	27	177	87.6	97.2	9.6	3.4	6.2	2.8
	28	194	77.8	99.5	21.7	8.8	12.9	0.5
	29	180	82.8	99.4	16.7	11.7	5.0	0.6
Aggregate		1 210	82.0	99.0	17.0	10.8	6.2	1.0
Aggregate Across Administrations		2 017	82.2	99.1	16.9	11.3	5.7	0.9

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

Administration	Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
	8	543	78.6	98.0	19.3	12.3	7.0	2.0
	9	325	74.5	99.4	24.9	12.9	12.0	0.6
	11	301	79.1	95.4	16.3	8.3	8.0	4.7
Winter	26	248	86.7	98.0	11.3	6.1	5.2	2.0
	27	191	78.0	99.5	21.5	14.7	6.8	0.5
	28	292	72.6	98.6	26.0	9.9	16.1	1.4
	29	229	89.5	100.0	10.5	2.6	7.9	0.0
Aggregate		2 129	79.3	98.3	19.0	10.0	9.0	1.7
	8	365	80.8	98.4	17.5	14.3	3.3	1.6
	9	342	74.3	98.5	24.3	13.2	11.1	1.5
	11	383	82.3	97.9	15.7	7.1	8.6	2.1
Spring	26	286	79.7	99.3	19.6	8.7	10.8	0.7
	27	249	68.7	99.6	30.9	8.4	22.5	0.4
	28	264	73.1	99.6	26.5	16.3	10.2	0.4
	29	261	73.2	98.9	25.7	21.1	4.6	1.2
Aggregate		2 150	76.6	98.8	22.2	12.5	9.7	1.2
Aggregate Across Administrations		4 279	77.9	98.5	20.6	11.2	9.4	1.5

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

Administration	Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
	8	12	75.0	100.0	25.0	8.3	16.7	0.0
	9	10	80.0	100.0	20.0	0.0	20.0	0.0
	11	13	84.6	100.0	15.4	15.4	0.0	0.0
Winter	26	5	80.0	100.0	20.0	20.0	0.0	0.0
	27	8	75.0	100.0	25.0	0.0	25.0	0.0
	28	7	57.1	100.0	42.9	28.6	14.3	0.0
	29	9	77.8	100.0	22.2	22.2	0.0	0.0
Aggregate		64	76.6	100.0	23.4	12.5	10.9	0.0
	8	3	100.0	100.0	0.0	0.0	0.0	0.0
	9	8	87.5	100.0	12.5	12.5	0.0	0.0
	11	6	83.3	100.0	16.7	0.0	16.7	0.0
Spring	26	4	100.0	100.0	0.0	0.0	0.0	0.0
	27	3	100.0	100.0	0.0	0.0	0.0	0.0
	28	2	100.0	100.0	0.0	0.0	0.0	0.0
	29	8	87.5	100.0	12.5	12.5	0.0	0.0
Aggregate		34	91.2	100.0	8.8	5.9	2.9	0.0
Aggregate Across Administrations		98	81.6	100.0	18.4	10.2	8.2	0.0

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

Administration	Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
	8	37	86.5	97.3	10.8	5.4	5.4	2.7
	9	18	83.3	100.0	16.7	0.0	16.7	0.0
	11	15	86.7	100.0	13.3	0.0	13.3	0.0
Winter	26	46	76.1	97.8	21.7	8.7	13.0	2.2
	27	12	100.0	100.0	0.0	0.0	0.0	0.0
	28	12	100.0	100.0	0.0	0.0	0.0	0.0
	29	19	73.7	100.0	26.3	10.5	15.8	0.0
Aggregation		159	83.7	98.7	15.1	5.0	10.1	1.3
	8	0	n/a	n/a	n/a	n/a	n/a	n/a
	9	36	86.1	100.0	13.9	0.0	13.9	0.0
	11	37	91.9	97.3	5.4	2.7	2.7	2.7
Spring	26	0	n/a	n/a	n/a	n/a	n/a	n/a
	27	24	95.8	100.0	4.2	4.2	0.0	0.0
	28	23	87.0	100.0	13.0	4.4	8.7	0.0
	29	17	70.6	100.0	29.4	29.4	0.0	0.0
Aggregate		137	87.6	99.3	11.7	5.8	5.8	0.7
Aggregate Across Administrations		296	85.5	99.0	13.5	5.4	8.1	1.0

Ontario Secondary School Literacy Test (OSSLT)

The marking validity estimates for the OSSLT are reported in Tables 4.39 to 4.42. For reading, the aggregate percentage of exact agreement was 70.0% for French and 66.9% for English. The aggregate percentage of exact-plus-adjacent agreement was 95.1% for French and 93.8% for English. The percentages of exact agreement met the EQAO target of 75% for three-point rubrics on two out of four French items, but did not meet the EQAO target on any of the English items. Among French-language markers, the aggregate adjacent-high percentage was considerably larger than the aggregate adjacent-low percentage. Among English-language markers, the aggregate adjacent-low percentage was slightly larger than the aggregate adjacent-high percentage.

For long writing, the percentages of exact agreement ranged from 31.5% to 46.6% for content and from 50.8% to 59.6% for conventions. For short writing, the percentages of exact agreement ranged from 71.0% to 88.5%. The percentages of exact-plus-adjacent agreement ranged from 83.8% to 99.6%.

The percentages of exact agreement did not meet the EQAO targets on any long writing items in the English-language or French-language assessments, but did meet the EQAO target for the short writing prompts in both English- and French-language assessments.

Table 4.39 Validity Estimates for Reading: OSSLT (English)

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
A: Item 6	6 341	54.1	89.7	35.6	17.1	18.4	10.3
F: Item 7	6 341	70.4	90.1	19.7	10.7	9.0	9.9
J: Item 6	8 032	73.4	95.8	22.5	10.1	12.3	4.2
J: Item 7	8 032	67.6	97.8	30.3	19.7	10.6	2.2
Aggregate	28 746	66.9	93.8	26.9	14.5	12.5	6.2

Table 4.40 Validity Estimates for Reading: OSSLT (French)

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
A: Item 6	270	83.0	98.9	15.9	8.5	7.4	1.1
F: Item 7	270	77.0	91.9	14.8	5.6	9.3	8.2
J: Item 6	209	61.7	91.9	30.1	12.4	17.7	8.1
J: Item 7	209	52.6	97.6	45.0	11.5	33.5	2.4
Aggregate	958	70.0	95.1	25.1	9.2	15.9	4.9

Table 4.41 Validity Estimates for Writing: OSSLT (English)

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
E-Content	10 794	38.8	84.2	45.4	22.0	23.4	15.8
E-Conventions	10 794	58.0	97.0	39.0	22.3	16.7	3.0
G-Content	10 399	46.6	92.0	45.4	21.2	24.3	8.0
G-Conventions	10 399	59.6	96.8	37.3	22.2	15.0	3.2
Aggregate Long Writing	42 386	50.7	92.5	41.8	21.9	19.9	7.5
D-Content & Conventions	7 553	86.8	98.4	11.6	8.6	3.0	1.6
K-Content & Conventions	11 857	88.5	98.9	10.4	6.7	3.8	1.1
Aggregate Short Writing	19 410	87.8	98.7	10.9	7.4	3.5	1.3
Aggregate Across All Items	61 796	62.4	94.4	32.1	17.4	14.7	5.6

Table 4.42 Validity Estimates for Writing: OSSLT (French)

Item No.	No. of Marks	% Exact	% Exact Plus Adjacent	% Adjacent	% Adjacent-Low	% Adjacent-High	% Non-Adjacent
E-Content	389	32.4	84.6	52.2	26.5	25.7	15.4
E-Conventions	389	55.5	96.9	41.4	14.9	26.5	3.1
G-Content	419	31.5	83.8	52.3	20.5	31.7	16.2
G-Conventions	419	50.8	95.2	44.4	11.9	32.5	4.8
Aggregate Long Writing	1 616	42.5	90.1	47.6	18.4	29.2	9.9
D-Content & Conventions	287	72.5	98.6	26.1	19.2	7.0	1.4
K-Content & Conventions	231	71.0	99.6	28.6	19.9	8.7	0.4
Aggregate Short Writing	518	71.8	99.0	27.2	19.5	7.7	1.0
Aggregate Across All Items	2 134	49.6	93.3	42.6	18.7	24.0	7.7

CHAPTER 5: EQUATING

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

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

The processes used in equating and scaling are similar but their purposes are different. Equating is used to adjust for differences in difficulty among tests that are similar in content and statistical specifications. Scaling is used to link two tests that are different in content and statistical specifications (Kolen & Brennan, 2004). Since there were no significant changes to the test specifications from 2006 to 2007, only equating procedures were used in 2007.

The following sections describe the Item Response Theory (IRT) models, equating design, equating samples and calibration procedures used during the 2006–2007 school year for the various EQAO assessments.

IRT Models

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

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

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

The Generalized Partial Credit Model (GPCM) is given by

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

where $P_{ix}(\theta)$ is the probability of a student (with ability θ) receiving a particular score x on item i ;
 α_i is the slope parameter for item i and
 δ_{ij} is the item i step difficulty parameter associated with a category score j .

Equating Design

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

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

Equating Samples

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

Calibration

Calibration is the process of estimating the item parameters that determine the relationship between ability and the probability of answering a multiple-choice item correctly or receiving a particular score on a polytomously scored open-response item. For each assessment, the calibration of the items for the

English-language and the French-language populations are conducted separately. The calibrations are conducted using the program PARSCALE 4.1 (Muraki & Bock, 2003). However, before conducting the analyses, the p -values of common items are plotted to identify outliers, which are then eliminated. Also, common items showing large differences in their p -values between the two years are eliminated from the equating process.

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

Description of the IRT Model

A modified 3PL model (equation 1) is used for the primary and junior assessments. The pseudo-guessing parameter is fixed to $0.20 \left(\frac{1}{k+1} \right)$, where k is the number of options) for the multiple-choice items to reflect the possibility of students with a very low ability to correctly answer an item. The GPCM (equation 2) is used for the constructed-response items.

Equating Sample: Exclusion Rules

The following students were excluded from the equating samples for 2005–2006 and 2006–2007:

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

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

1. Students from the 2005–2006 population who responded to both the operational 2005–2006 test items and the field test items that were brought forward to form the 2006–2007 operational tests and who were not excluded by the rules stated above (calibration sample).
2. Students from the 2005–2006 population who wrote the operational test and who were not excluded by the rules stated above (equating sample).
3. Students from the 2006–2007 population who wrote the operational test and who were not excluded by the rules stated below (calibration/equating sample).

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

Table 5.1 Number of Students in the Calibration and Equating Samples for the 2005–2006 and the 2006–2007 Primary and Junior Assessments (English and French)

Component	2006–2007 Calibration/ Equating Sample	2005–2006 Calibration Sample	2005–2006 Equating Sample
Primary Mathematics (English)	109 545	104 159	113 568
Junior Mathematics (English)	129 730	118 328	133 023
Primary Mathematics (French)	5 310	6 192	6 192
Junior Mathematics (French)	5 588	6 112	6 116
Primary Reading (English)	111 099	36 019	115 535
Junior Reading (English)	125 794	39 419	131 815
Primary Reading (French)	5 458	3 709	6 139
Junior Reading (French)	6 100	3 104	6 059
Primary Writing (English)	113 730	45 815	118 886
Junior Writing (English)	125 455	52 992	133 157
Primary Writing (French)	5 587	3 743	6 194
Junior Writing (French)	6 168	4 851	6 091

Equating Steps

In equating the 2006–2007 tests to the 2005–2006 tests, the EQAO forward-fixed parameter equating procedure was used. The following steps were implemented:

1. The 2006–2007 operational items were calibrated independently to obtain item parameters and student ability scores for the 2006–2007 equating sample.
2. The 2005–2006 operational items were calibrated together with the field test items that were brought forward to the 2006–2007 operational tests. In this calibration, the item parameters of the field test items were fixed at the values obtained from the 2006–2007 calibration runs (step 1).
3. The 2005–2006 equating sample was scored using the operational item parameters obtained in step 2.
4. The percentage of students in each achievement level was determined for the 2005–2006 equating sample from the levels assigned in 2005–2006. The theta value cut points that replicated this distribution were identified for each boundary (0/1, 1/2, 2/3, and 3/4).
5. These theta values were then used as the cut points for 2006–2007.
6. The operational item parameters of 2006–2007 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 2006–2007 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 assessments, two items (a primary French mathematics item and a junior English reading item) were completely dropped from the assessment. Careful inspection of the items revealed concerns and item-total correlations were found to be unacceptable. In addition, six multiple-choice items and nine open-response items across the 12 tests were dropped from the equating link. These items were either modified between field and operational test administrations or were not field

tested in the 2005–2006 assessment year. The seven open-response items included four long writing prompts that were not field tested during the administration of the operational tests. Two short writing prompts (for primary French) and a short writing prompt (for junior French) were dropped due to a curriculum change. The number of items not used in the equating process and the number of dropped items are presented for each test in Table 5.2.

Table 5.2 Items Dropped from the Primary and Junior Assessments and from the Equating Process (2006–2007)

Assessment Component	Items Not Used in Equating*		Items Dropped from the Assessment
	Multiple-Choice	Open-Response	
Primary Mathematics (English)	1		
Junior Mathematics (English)	1		
Primary Mathematics (French)	2		1
Junior Mathematics (French)	1		
Primary Reading (English)			
Junior Reading (English)			1
Primary Reading (French)			
Junior Reading (French)			
Primary Writing (English)		1	
Junior Writing (English)		1	
Primary Writing (French)	1	3	
Junior Writing (French)		2	

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

In contrast to the OSSLT (see below), there were no open-response items for which score categories were collapsed.

Equating Results

The results of the equating process are provided in Tables 5.3 to 5.8 below. The theta cut scores and the number of students at each achievement level in 2006 and 2007 are reported for both English-language and French-language students. For example, the equated cut scores for primary English-language reading were 1.28 for Levels 3 and 4, –0.37 for Levels 2 and 3, –1.46 for Levels 1 and 2 and –2.41 for NE1 and Level 1.

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	
		2006	2007		2006	2007
Number of Students		115 535	111 099		6 139	5 458
Level 4		7.1%	7.0%		7.7%	8.9%
Level 3	1.28	61.4%	60.7%	1.22	52.0%	48.3%
Level 2	-0.37	25.2%	25.5%	-0.16	33.7%	37.0%
Level 1	-1.46	5.35%	5.7%	-1.47	6.1%	5.4%
NE1	-2.41	1.03%	1.1%	-2.57	0.5%	0.4%
% of Students at or Above the Provincial Standard		68.5%	67.7%		59.7%	57.2%

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	
		2006	2007		2006	2007
Number of Students		131 815	125 794		6 059	6 100
Level 4		8.6%	8.3%		9.8%	9.6%
Level 3	1.22	61.0%	62.2%	1.20	63.4%	62.3%
Level 2	-0.46	24.9%	24.0%	-0.55	25.0%	26.6%
Level 1	-1.52	5.1%	5.0%	-1.98	1.7%	1.4%
NE1	-2.66	0.4%	0.5%	-2.78	0.2%	0.2%
% of Students at or Above the Provincial Standard		69.6%	70.5%		73.2%	71.9%

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	
		2006	2007		2006	2007
Number of Students		118 886	113 730		6 194	5 587
Level 4		6.3%	6.5%		16.2%	15.8%
Level 3	1.36	62.9%	63.4%	0.92	60.3%	60.5%
Level 2	-0.45	30.4%	29.7%	-0.64	22.8%	23.2%
Level 1	-2.53	0.3%	0.0%	-2.37	0.6%	0.4%
NE1	-2.88	0.1%	0.1%	-3.26	0.0%	0.0%
% of Students at or Above the Provincial Standard		69.2%	69.9%		76.5%	76.3%

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

	Junior Writing (English)			Junior Writing (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2006	2007		2006	2007
Number of Students		133 157	125 455		6 091	6 168
Level 4		7.5%	7.5%		9.2%	9.2%
Level 3	1.31	59.4%	58.9%	1.23	69.7%	69.6%
Level 2	-0.37	31.9%	32.6%	-0.72	20.9%	21.1%
Level 1	-2.10	1.1%	0.0%	-2.71	0.2%	0.1%
NE1	-2.92	0.1%	0.0%	-3.16	0.0%	0.0%
% of Students at or Above the Provincial Standard		66.9%	66.4%		78.7%	78.8%

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

	Primary Mathematics (English)			Primary Mathematics (French)		
	Theta Cut Scores	Students at Each Level for Each Equating Sample		Theta Cut Scores	Students at Each Level for Each Equating Sample	
		2006	2007		2006	2007
Number of Students		113 568	109 545		6 192	5 310
Level 4		12.1%	12.4%		7.2%	7.2%
Level 3	1.08	61.0%	61.2%	1.33	55.0%	56.6%
Level 2	-0.58	24.0%	24.0%	-0.29	33.9%	32.8%
Level 1	-1.90	2.5%	2.2%	-1.71	3.6%	3.2%
NE1	-2.60	0.4%	0.3%	-2.52	0.3%	0.2%
% of Students at or Above the Provincial Standard		73.2%	73.5%		62.2%	63.8%

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	
		2006	2007		2006	2007
Number of Students		133 023	129 730		6 116	5 588
Level 4		11.7%	12.5%		30.2%	29.6%
Level 3	1.12	54.3%	51.9%	0.52	50.7%	50.8%
Level 2	-0.37	28.0%	30.7%	-0.83	18.7%	19.1%
Level 1	-1.55	5.9%	4.8%	-2.30	0.3%	0.3%
NE1	-3.03	0.1%	0.1%	-2.54	0.1%	0.3%
% of Students at or Above the Provincial Standard		66.0%	64.4%		80.9%	80.3%

Grade 9 Assessment of Mathematics

Description of the IRT Model

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

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

Equating Sample

Equating samples for 2005–2006 and 2006–2007 were identified using a common set of selection rules. The equating samples for the academic and the applied programs were selected separately. However, the selection and exclusion rules for both programs and across the two administration years were the same. Students were excluded if they

- did not attend a publicly funded school;
- were home-schooled;
- produced no work by booklet and by question type;
- received accommodations except for ESL/ELD students who received accommodation for setting;
- received special provisions or
- 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 samples for the 2005–2006 and the 2006–2007 tests.

Table 5.9 Number of Grade 9 Students in the Equating Samples

Grade 9	2005–2006		2006–2007	
	English	French	English	French
Academic	97 570	3 799	96 221	3 894
Applied	36 785	1 268	34 441	1 202

Equating Steps

The calibration and equating of the 2005–2006 and 2006–2007 tests for the English-language and the French-language student populations in both academic and applied programs were conducted using the following steps:

- Calibration and equating were conducted simultaneously for the 2006–2007 winter and spring samples.
- A concurrent calibration was conducted for the 2005–2006 winter and spring samples (including the field test items that were brought forward to the 2006–2007 operational test). In this calibration,

the item parameters of the field test items that were brought forward to the 2006–2007 operational tests were fixed at the values obtained from the 2006–2007 calibration runs (step 1). The parameters of the 2005–2006 operational items that were repeated on the 2006–2007 operational tests were also fixed at the values obtained from the 2006–2007 calibration runs (step 1). The 2005–2006 equating samples (operational items only) were scored using the scaled 2005–2006 operational item parameters.

3. The percentage of students in each achievement level was determined for the 2005–2006 equating sample from the levels assigned in 2005–2006. The theta value cut points that replicated this distribution were identified for each boundary (0/1, 1/2, 2/3, and 3/4).
4. These theta values were then used as the cut points for the 2006–2007 assessments.
5. The item parameters for the 2006–2007 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 2006–2007 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

No items in the English-language Grade 9 mathematics assessments were eliminated from the equating process. Two multiple-choice items were removed from the equating link for the French-language Applied Program because of modifications between the 2006 field test and 2007 operational administrations. None of the polytomously scored item categories were collapsed.

Equating Results

The equating results for the Grade 9 applied mathematics assessment are summarized in Table 5.10 and the results for the Grade 9 academic mathematics assessment are summarized in Table 5.11. The theta cut scores and number of students in 2006 and 2007 at each achievement level is reported for both the English-language and French-language students. For example, the equated cut scores for English applied were 1.45 for Levels 3 and 4, 0.15 for Levels 2 and 3, –0.90 for Levels 1 and 2 and –1.65 for Below Level 1 and Level 1.

Table 5.10 Equating Results for the Grade 9 Applied Mathematics Program

	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	
		2006	2007		2006	2007
Number of Students		36 785	34 441		1 268	1 202
Level 4		5.5%	5.8%		5.6%	5.2%
Level 3	1.45	37.2%	37.1%	1.54	32.3%	31.2%
Level 2	0.15	39.7%	41.2%	0.32	43.1%	44.0%
Level 1	–0.90	13.6%	12.9%	–0.82	14.1%	15.6%
Below Level 1	–1.65	3.9%	3.1%	–1.52	5.0%	4.0%
% of Students at or Above the Provincial Standard		42.7%	42.9%		37.9%	36.4%

Table 5.11 Equating Results for the Grade 9 Academic Mathematics Program

	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	
		2006	2007		2006	2007
Number of Students		97 570	96 221		3 799	3 894
Level 4		5.9%	6.4%		3.1%	4.3%
Level 3	1.50	67.7%	66.9%	1.66	69.0%	67.9%
Level 2	-0.62	17.7%	18.7%	-0.58	18.2%	19.0%
Level 1	-1.25	8.3%	7.8%	-1.23	9.0%	8.6%
Below Level 1	-2.36	0.5%	0.2%	-2.24	0.6%	0.3%
% of Students at or Above the Provincial Standard		73.5%	73.3%		72.2%	72.1%

Ontario Secondary School Literacy Test (OSSLT)

Description of the IRT Model

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

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

where, $P_{ix}(\theta)$ is the probability of a student (with ability θ) receiving a particular score x on item i ,
 δ_{ij} is the item i step difficulty parameter associated with a category score j .

Equating Sample

First-time eligible students from both publicly funded and private schools were selected for the 2006 and 2007 equating samples. The following categories of students were excluded from the equating sample:

- students with no work or incomplete work in a major section of the test;
- students receiving accommodations, except for students receiving an accommodation for extra time;
- previously eligible students;
- students who were exempted, deferred or taking the Ontario Secondary School Literacy Course (OSSLC) and
- students who were home-schooled.

Table 5.12 presents the number of first-time eligible students in the OSSLT equating samples for the March 2006 and March 2007 tests.

Table 5.12 Number of First-Time Eligible OSSLT Students in the Equating Samples

OSSLT	2006		2007	
	English	French	English	French
First-Time Eligible	149 589	5 260	136 498	5 159

Equating Steps

The 2007 OSSLT was shortened to minimize the administration time. A shortened version of the 2006 OSSLT was created to match the 2007 test blueprint. The equating process was conducted using the 2007 operational test and the 2006 shortened test.

The following steps were used to calibrate and equate the 2006 and 2007 OSSLT:

1. The parameters of the operational items administered in 2007 were calibrated using the 2007 equating sample.
2. The operational items that formed the shortened test for 2006 and the field test items brought forward to the 2007 test were recalibrated using the 2006 equating sample (and fixing the common-item parameters to the values obtained in step 1.)
3. The operational item parameters of the 2006 shortened test, obtained in step 2, were used to score the 2006 equating sample data.
4. The percentage of successful students was determined for the 2005–2006 equating sample from the student results reported in 2005–2006. The theta value cut point that replicated this percentage was identified for the distribution of scores in the 2005–2006 equating sample.
5. This theta value was applied to student scores for the 2006–2007 assessments to determine which students would be successful if this were to be used as the cut point. This result is presented in Table 5.13.
6. This was not the final decision for the cut point. For the OSSLT, EQAO has adopted a scale that is generated from a conversion table in which all students with the same raw score are assigned the same scale score. However, there is no one-to-one correspondence between student raw scores and theta values because EQAO uses a modified Rasch model with a fixed *c*-parameter of 0.20 for OSSLT multiple-choice items. An additional step is required to select a single raw score as the cut point for the successful-unsuccessful classification.
7. To assist in the determination of the raw score cut point, a cross-tabulation of raw scores (with successful and unsuccessful decisions based on the theta cuts) was generated. Through an examination of this raw score cross-tabulation, a raw score of 58 was selected for the English-language assessment cut point. A raw score of 56 was selected for the French-language assessment cut-point. Both tests had a maximum possible raw score of 81.
8. The cut-score points identified in step 7 were applied to the raw score distribution of the 2007 full student population data and the percentage of students who were successful on the 2007 test, was identified.

Eliminating Items and Collapsing of Score Categories

None of the items in the English-language OSSLT were eliminated from the equating process. However, one short writing item was eliminated from the equating process, (but not from the scoring process) in the French-Language OSSLT due to differences between the scoring of the 2006 field test responses and the 2007 operational responses.

The score 1.0 in the scoring rubric for the long writing items was collapsed with the score 1.5 for both long writing items in the English-language test. The score 1.0 was collapsed with the score 1.5 for one long writing item in the French-language test.

Equating Results

The equating results based on the equating samples for the OSSLT are summarized in Table 5.13. The theta cut score (see step 5 above) and the percentage of successful and unsuccessful students in 2006 and 2007 are reported for English-language and French-language students. For example, the equated cut score for English was -0.89 . The percentage of successful students in the equating samples was 86.0% in 2006 and 85.1% in 2007.

Table 5.13 Equating Results for the OSSLT

	English-Language			French-Language		
	Theta Cut Point	Equating Sample 2006	Equating Sample 2007	Theta Cut Point	Equating Sample 2006	Equating Sample 2007
Number of Students		149 589	136 498		5 260	5 159
% Successful	-0.89	86.0%	85.1%	-0.86	82.6%	86.3%
% Unsuccessful		14.0%	14.9%		17.4%	13.7%

References

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

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

CHAPTER 6: REPORTING RESULTS

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

EQAO publishes an annual provincial report for education stakeholders and the general public. The following reports for the 2007 assessments are available on the EQAO Web site (www.eqao.com):

- *EQAO's Provincial Report on the Results of the 2006–2007 Assessments of Reading, Writing and Mathematics: Primary Division (Grades 1–3) and Junior Division (Grades 4–6), and the Grade 9 Assessment of Mathematics (September 2007)*
- *EQAO's Provincial Report on the Results of the 2006–2007 Ontario Secondary School Literacy Test.*

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

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

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

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

Further detailed results are posted on a secure Web site available only to school board and school personnel through user identification numbers and passwords. These reports include results for small *n*-counts that are suppressed in the public reports. In addition, schools and school boards receive data files with individual student achievement results for all students in the school and the board. School boards also receive data files with detailed results for each school, the board and the province.

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

The four achievement levels defined in *The Ontario Curriculum* are used by EQAO to report student achievement in reading, writing and mathematics. Level 3 has been established as the provincial standard. Levels 1 and 2 indicate achievement below the provincial standard and Level 4 indicates achievement above the provincial standard. In addition to the four performance levels, the following three reporting categories are also included:

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

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

The Individual Student Reports for students in Grades 3 and 6 show the overall achievement level for each subject (reading, writing and mathematics) at one of the five possible positions within the achievement level. School, school board and provincial results are included on the Individual Student Report to provide a context for interpreting student results.

The five possible positions within the achievement levels are created from the distribution of student theta scores. Through the equating process described in Chapter 5, four theta values are identified that define the cut points between adjacent levels (NE1/Level 1, Level 1/Level 2, Level 2/Level 3 and Level 3/Level 4). The width of each sub-category in a given level is determined by the range of theta values represented in the level divided by five. These results are designated in student data files as 1.1, 1.3, 1.5, 1.7, 1.9, 2.1, 2.3, and so on to 4.9.

The provincial, board and school reports provide the following:

- overall jurisdictional results for each subject (reading, writing and mathematics);
- longitudinal data showing jurisdictional results over time;
- overall jurisdictional results for each subject (reading, writing and mathematics) by gender and characteristics (such as ESL/ELD learners and students with special needs);
- areas of strength within the curriculum and areas for improvement;
- contextual data and
- student questionnaire results.

Results for the teacher and principal questionnaires are reported at the board and provincial levels.

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

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 assessments. The same four achievement levels with five sub-categories are used to report student achievement. The set of reports described above for the primary and junior assessments is also produced for the Grade 9 assessments. However, there are some differences between the reporting for the primary and junior assessments and the Grade 9 assessment. For instance, the option to exempt students from the Grade 9 mathematics assessment was removed in 2007. Moreover, the reporting category, “Not Enough Evidence For Level 1,” is called “Below Level 1” in the Grade 9 assessment. In addition to the disaggregations identified for the primary and junior assessments, results for the Grade 9 mathematics assessment are reported for semester 1, semester 2 and the full year. Furthermore, the Individual Student Report for Grade 9 does not include school and board results and there is no principal questionnaire for the Grade 9 assessments.

Reporting the Results of the OSSLT

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

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

The OSSLT Individual Student Report provides the following:

- specification of successful or unsuccessful results and
- feedback to help improve the level of achievement among unsuccessful students.

Student scale scores are also reported to students who receive unsuccessful results. Students are also informed that a successful result requires a scale score of 300.

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

- overall successful/unsuccessful jurisdictional results;
- overall successful/unsuccessful jurisdictional results by gender and characteristics such as ESL/ELD learners and students with special needs;
- results by exceptionality category and for students receiving accommodations;
- results by type of English or French course (academic, applied, locally developed or ESL/ELD);
- longitudinal data showing jurisdictional results over time;
- areas of strength within the curriculum and areas for improvement;
- contextual data and
- results for the student questionnaire.

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

Interpretation Guides

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

- *Using Data to Promote Student Success: A Brief Guide to Assist School Administrators in Interpreting Their Data*
 - *Assessments of Reading, Writing and Mathematics, Primary Division (Grades 1–3) and Junior Division (Grades 4–6) (August 2007)*
 - *Grade 9 Assessment of Mathematics (August 2007)*
 - *Ontario Secondary School Literacy Test (June 2007)*
- *Guide to Interpreting EQAO Skill-Focused Reports*
 - *Assessments of Reading, Writing and Mathematics, Primary and Junior Divisions (Spring 2007)*
 - *Grade 9 Assessment of Mathematics (Spring 2007)*
 - *Ontario Secondary School Literacy Test (OSSLT) (March 2007)*

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

CHAPTER 7: STATISTICAL AND PSYCHOMETRIC SUMMARIES

A variety of statistical and psychometric analyses were conducted in 2007 for the Assessments of Reading, Writing and Mathematics: Primary and Junior Divisions, the Grade 9 Assessment of Mathematics and the Ontario Secondary School Literacy Test (OSSLT). These were analyses using Classical Test Theory (CTT) and Item Response Theory (IRT). All IRT item parameters were obtained from the calibration process used for the equating samples (described in Chapter 5).

The results from these analyses are summarized in this chapter. Detailed data for individual items are provided in Appendix 7.1.

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

Classical Test Theory (CTT) Analyses

Table 7.1 presents descriptive statistics, Cronbach's alpha estimates of test score reliability and the standard error of measurement (SEM) for the English-language and French-language primary and junior assessments. The test means (converted to percentages) ranged from 58.2% (for primary, French-language reading) to 67.4% (for junior, English-language writing). The corresponding distribution of scores are provided in Appendix 7.1. Inspection of these score distributions reveals that the scores are negatively skewed (slightly). This indicates that few students had low scores (demonstrated by the stretching out of the left tail on the line curves).

Reliability and the corresponding standard error of measurement refer to the precision of test scores with higher reliability coefficients and lower standard errors of measurement, which indicate higher levels of precision. For the primary and junior assessments, Cronbach's alpha estimates range from 0.86 and 0.88 for reading, 0.77 to 0.80 for writing and 0.82 to 0.89 for mathematics. The corresponding standard errors range from 4.9% to 5.1% of the possible maximum score for reading, 6.1% to 6.8% of the possible maximum score for writing and 6.1% and 6.4% of the possible maximum score for mathematics. The reliabilities for writing are a little lower than the ones for reading and mathematics; and the corresponding standard errors are a little larger. This is attributable, in part, to the smaller number of writing items and the subjectivity in scoring writing performance. Taking these two factors into account¹, the obtained reliabilities and standard errors of measurement are acceptable and indicate that the test scores from these assessments provide a satisfactory level of precision.

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

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

Grade and Subject	No. of Items	Item Type		Possible Max. Score	No. of Students	Min.	Max.	Mean	STD	Cronbach's Coefficient Alpha	SEM
		MC	OR								
Primary Reading (English)	36	26	10	66	120 957	0	64	39.24	9.24	0.88	3.23
Primary Reading (French)	36	26	10	66	6 203	3	65	38.38	8.85	0.86	3.34
Junior Reading (English)	35	25	10	65	139 317	0	65	40.54	8.19	0.86	3.09
Junior Reading (French)	36	26	10	66	6 397	5	65	40.19	8.52	0.86	3.24
Primary Writing (English)	14	8	6	38	121 211	0	38	24.90	5.17	0.80	2.31
Primary Writing (French)	14	8	6	38	6 201	1	37	22.24	5.35	0.77	2.59
Junior Writing (English)	14	8	6	38	139 322	0	38	25.61	5.50	0.80	2.46
Junior Writing (French)	14	8	6	38	6 389	0	38	23.06	5.64	0.79	2.57
Primary Mathematics (English)	36	28	8	60	125 115	0	60	36.99	9.83	0.86	3.66
Primary Mathematics (French)	35	27	8	59	5 793	7	59	37.23	8.85	0.82	3.76
Junior Mathematics (English)	36	28	8	60	139 165	0	60	34.98	11.32	0.89	3.80
Junior Mathematics (French)	36	28	8	60	5 858	2	60	35.48	10.62	0.87	3.77

MC = Multiple-Choice

OR = Open-Response

STD = Standard Deviation

SEM = Standard Error of Measurement

Item Response Theory (IRT) Analyses

In the IRT analysis, item parameters were estimated from the student responses used in the equating calibration sample (see Chapter 5). The estimated item parameters were then used to score all student responses. The descriptive statistics for IRT scores reported in Table 7.2 cover the total population. The mean student ability scores are less than zero and the standard deviations are less than one (due to the inclusion of all students). The item parameters for individual items are presented in Appendix 7.1.

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

Grade and Subject	N	Minimum	Maximum	Mean	STD
Primary Reading (English)	120 957	−3.79	3.30	−0.07	0.98
Primary Reading (French)	6 203	−3.83	3.47	−0.04	0.96
Junior Reading (English)	139 317	−3.90	3.48	−0.08	0.98
Junior Reading (French)	6 397	−3.87	3.21	−0.04	0.96
Primary Writing (English)	121 211	−3.77	2.80	−0.06	0.95
Primary Writing (French)	6 201	−3.72	2.73	−0.03	0.94
Junior Writing (English)	139 322	−3.71	2.57	−0.07	0.96
Junior Writing (French)	6 389	−3.85	2.82	−0.04	0.95
Primary Mathematics (English)	125 115	−3.75	2.94	−0.04	0.96
Primary Mathematics (French)	5 793	−3.13	2.65	−0.04	0.93
Junior Mathematics (English)	139 165	−3.59	2.66	−0.06	0.97
Junior Mathematics (French)	5 858	−3.55	2.76	−0.03	0.96

N = Number of students

STD = Standard Deviation

The Test Characteristic Curves (TCC) and Test Information Functions (TIF) are provided in Figures 7.1 to 7.24. The test characteristic curves slope upwards from the low left to the upper right. The curves can be used to translate a student ability score in IRT, to a student ability score in CTT. For example, a primary student with a theta score of −0.94 in reading is expected to have a total observed score of 30.00. These curves can also be used to determine the amount of information and the precision of the scores at the cut-points used for reporting, which are provided in each TIF. The information curves indicate where the test contains the most information. For example, the maximum information provided by the primary reading test is at $\theta = -0.63$. The precision of the scores is greatest at this point.

Figure 7.1 Test Characteristic Curve for Reading: Primary Division (English)

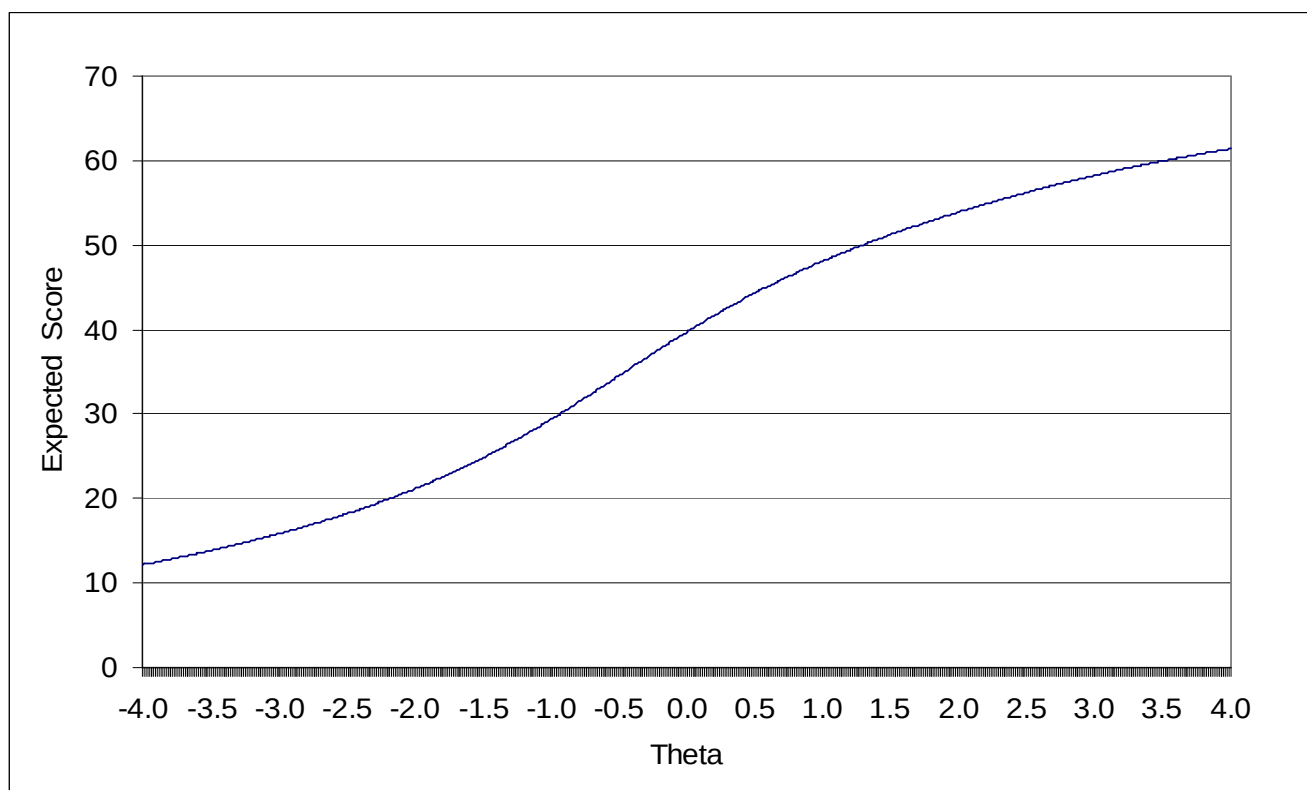


Figure 7.2 Test Characteristic Curve for Reading: Primary Division (French)

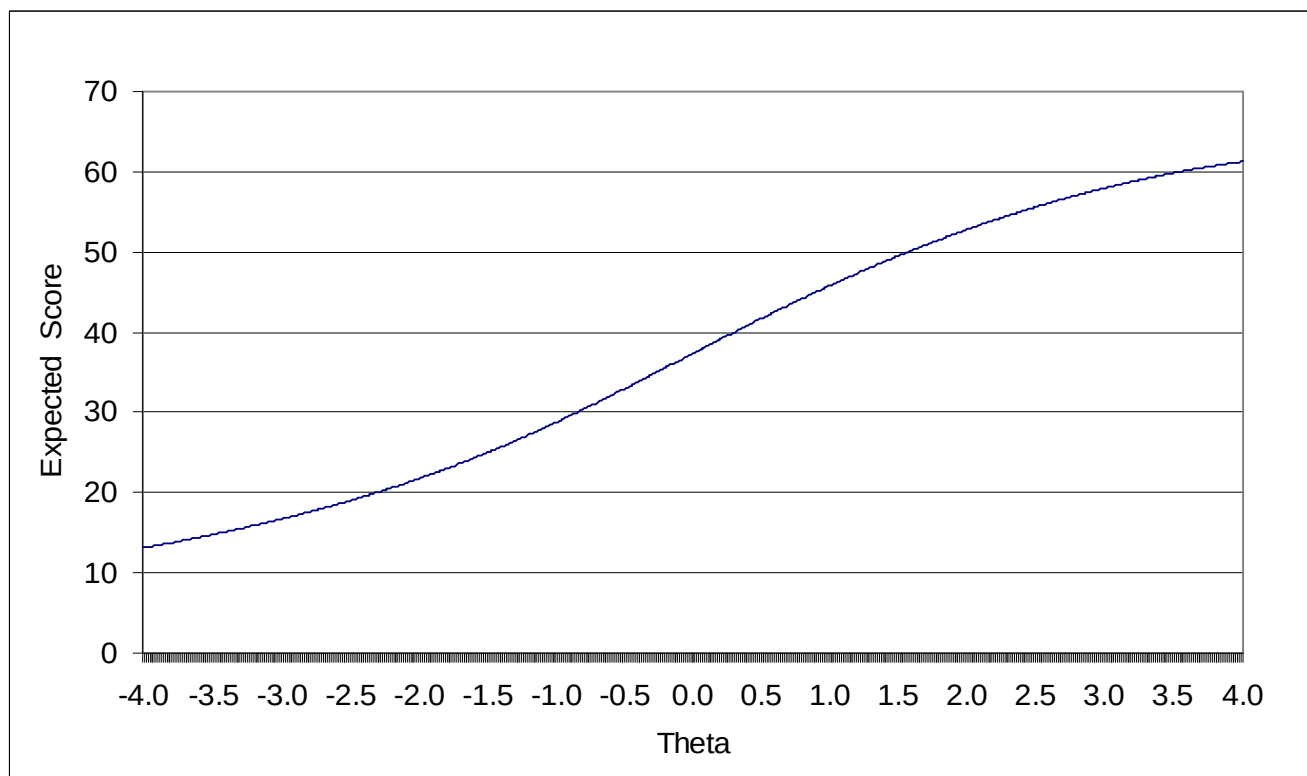


Figure 7.3 Test Characteristic Curve for Reading: Junior Division (English)

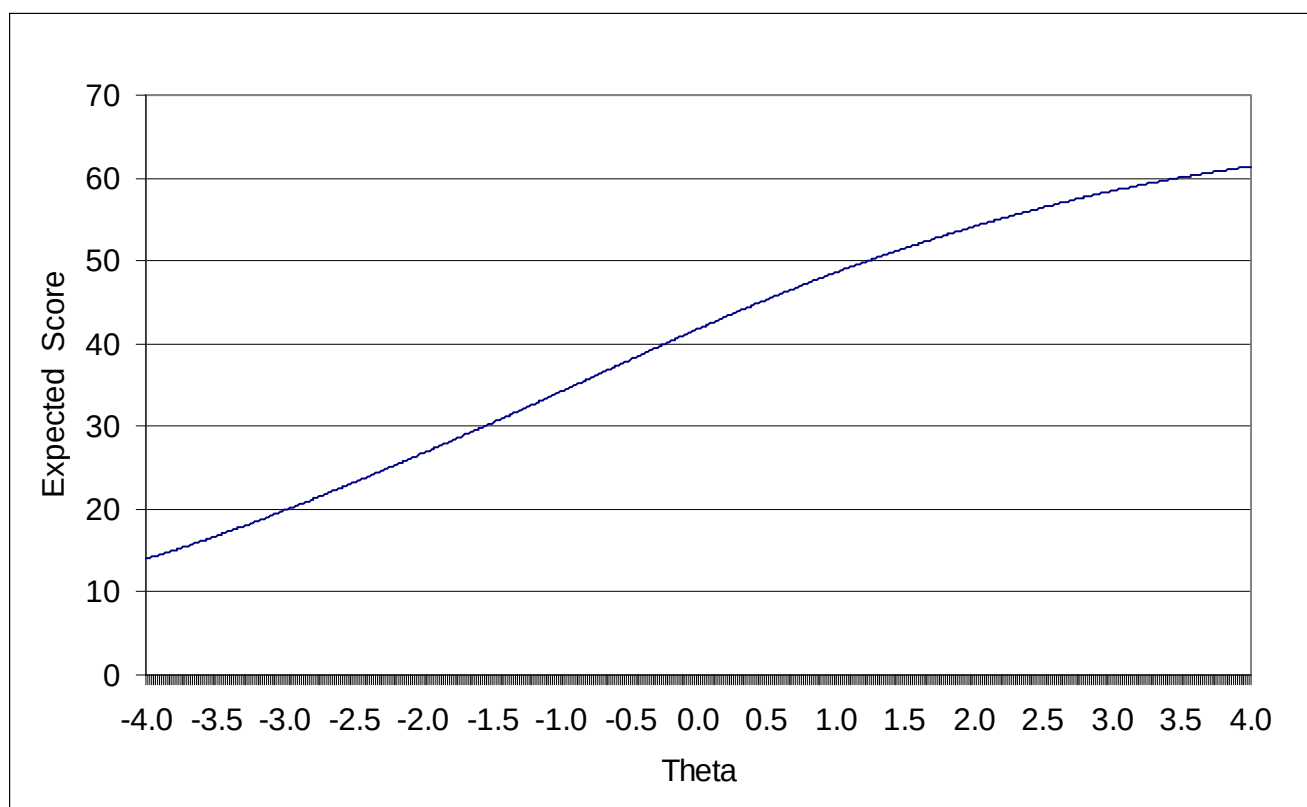


Figure 7.4 Test Characteristic Curve for Reading: Junior Division (French)

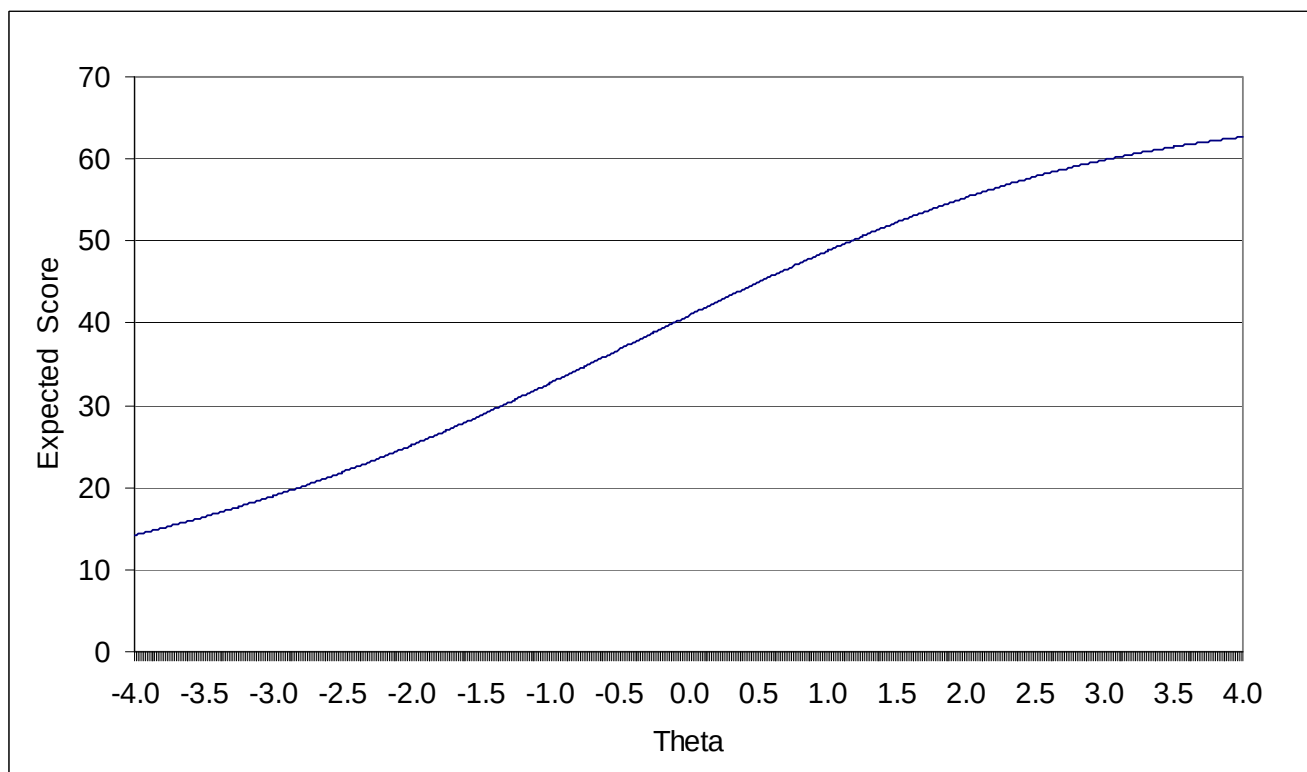


Figure 7.5 Test Characteristic Curve for Writing: Primary Division (English)

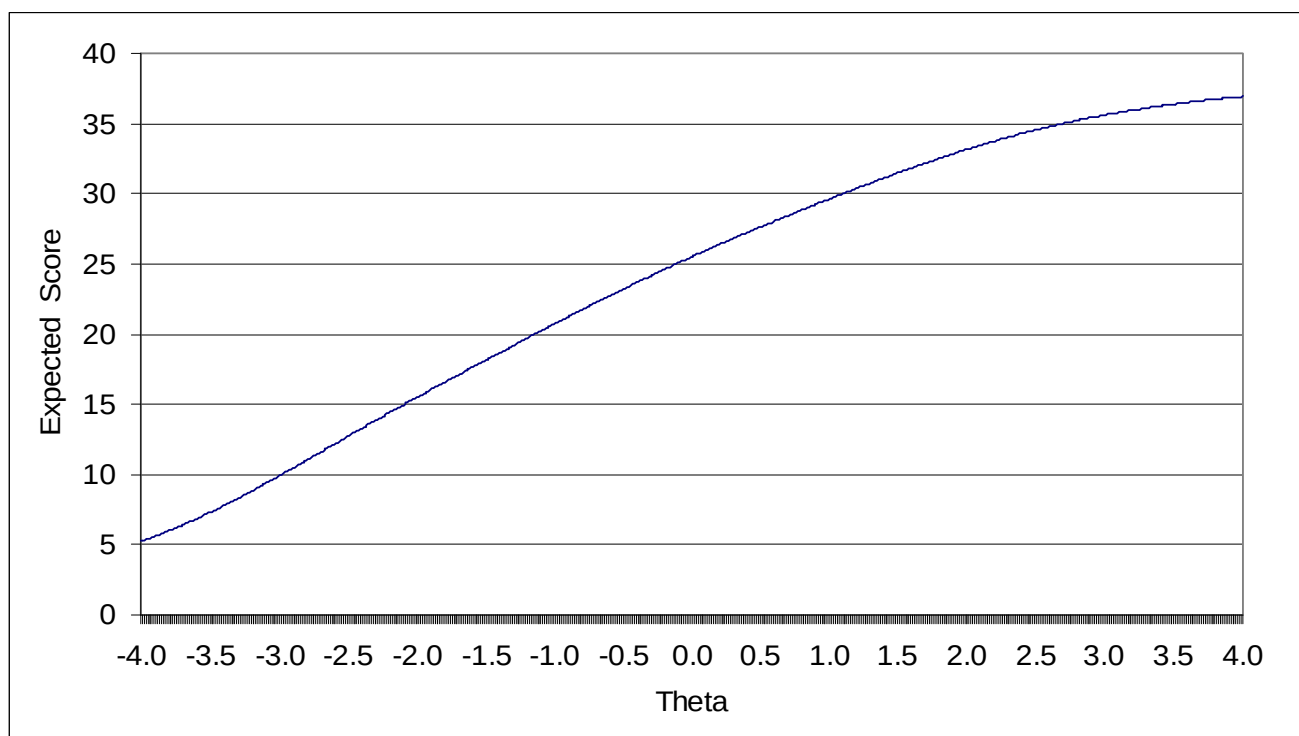


Figure 7.6 Test Characteristic Curve for Writing: Primary Division (French)

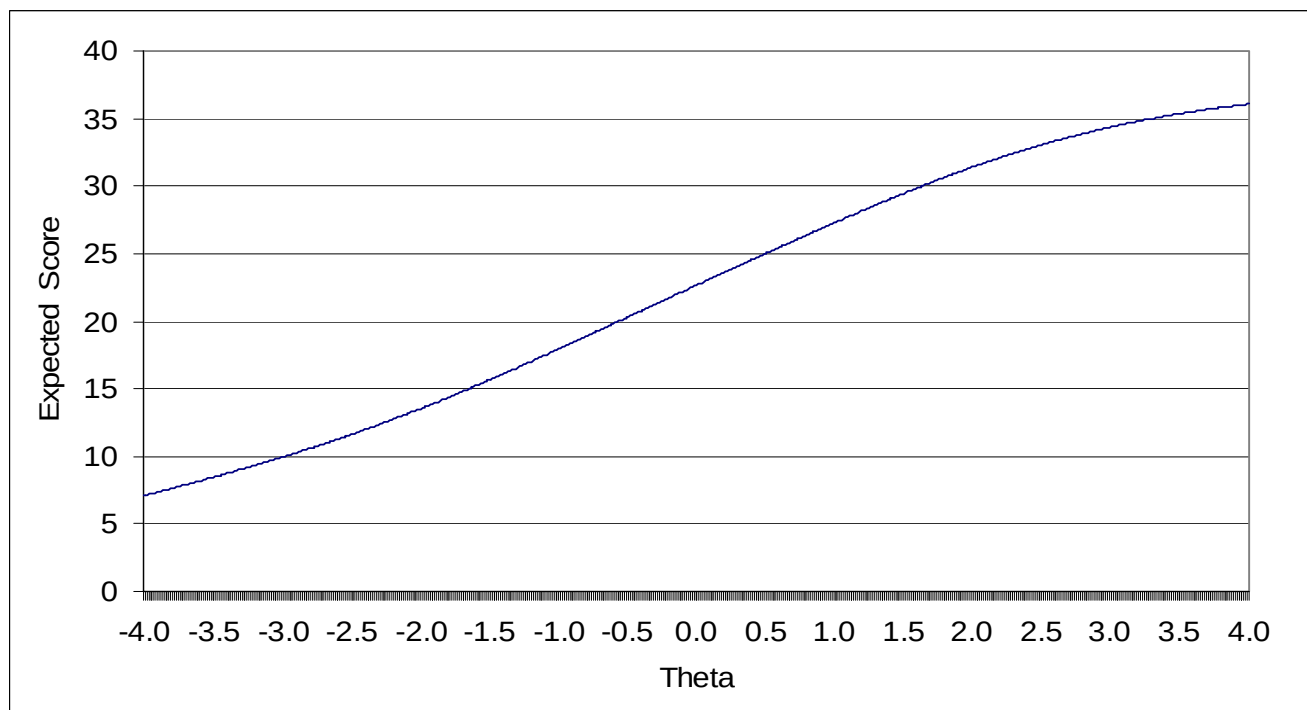


Figure 7.7 Test Characteristic Curve for Writing: Junior Division (English)

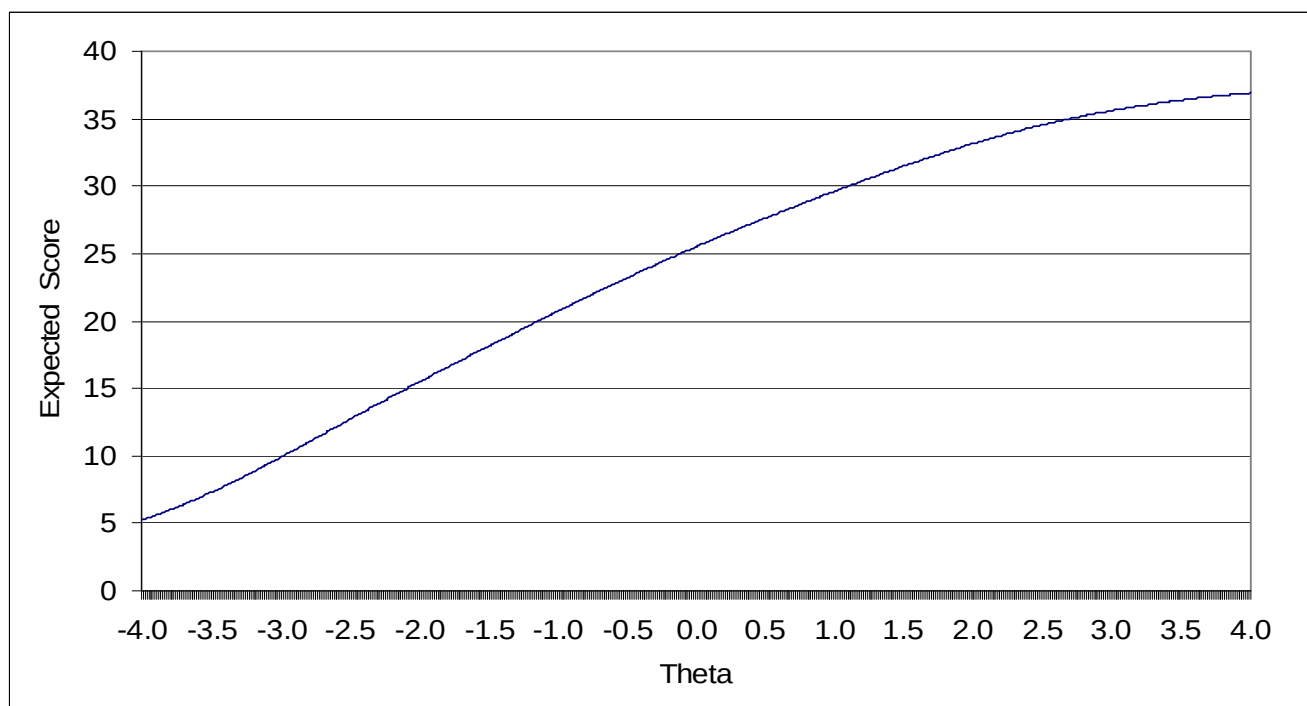


Figure 7.8 Test Characteristic Curve for Writing: Junior Division (French)

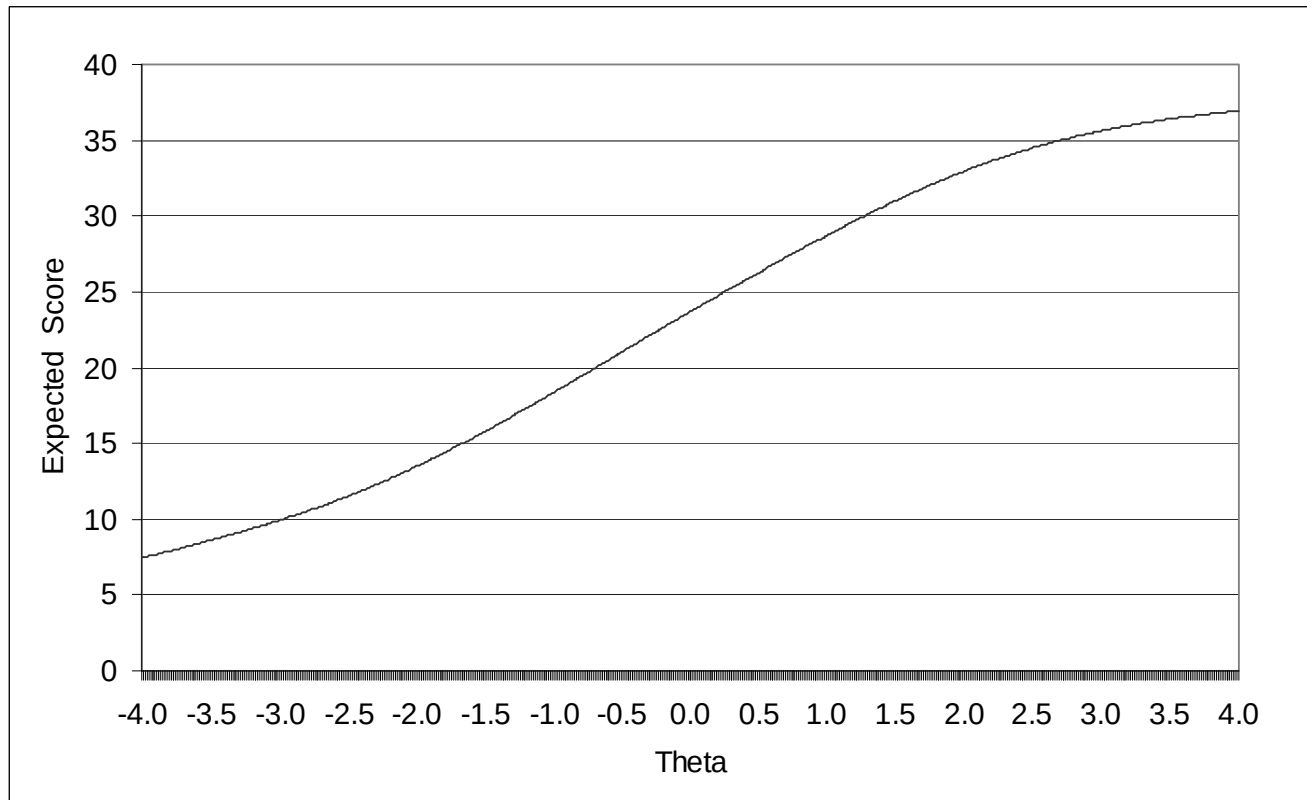


Figure 7.9 Test Characteristic Curve for Mathematics: Primary Division (English)

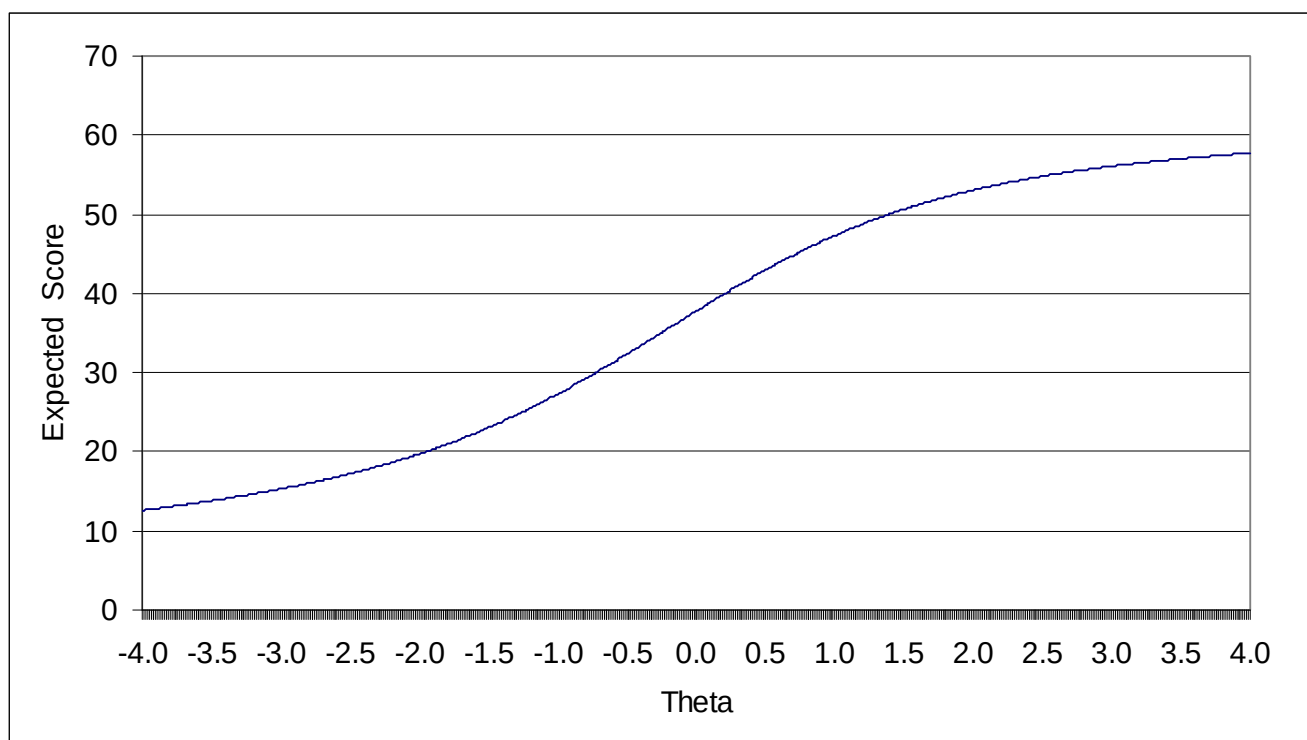


Figure 7.10 Test Characteristic Curve for Mathematics: Primary Division (French)

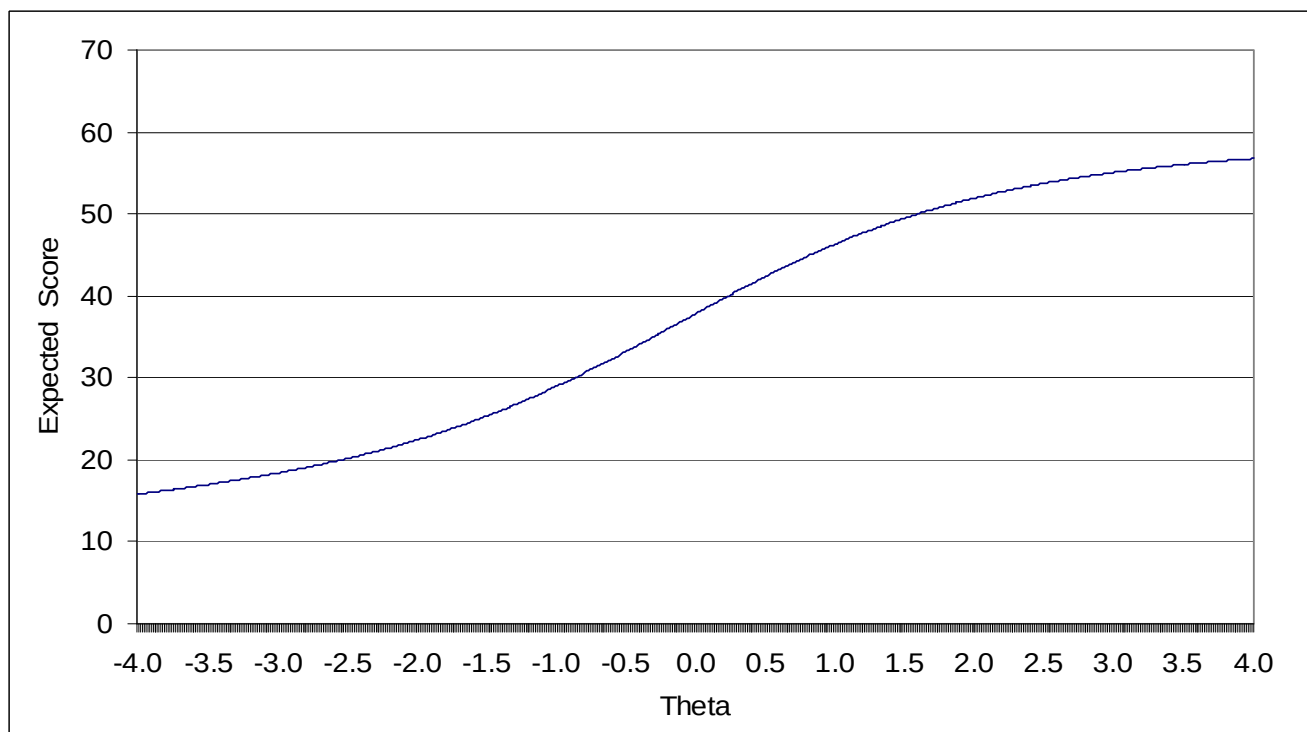


Figure 7.11 Test Characteristic Curve for Mathematics: Junior Division (English)

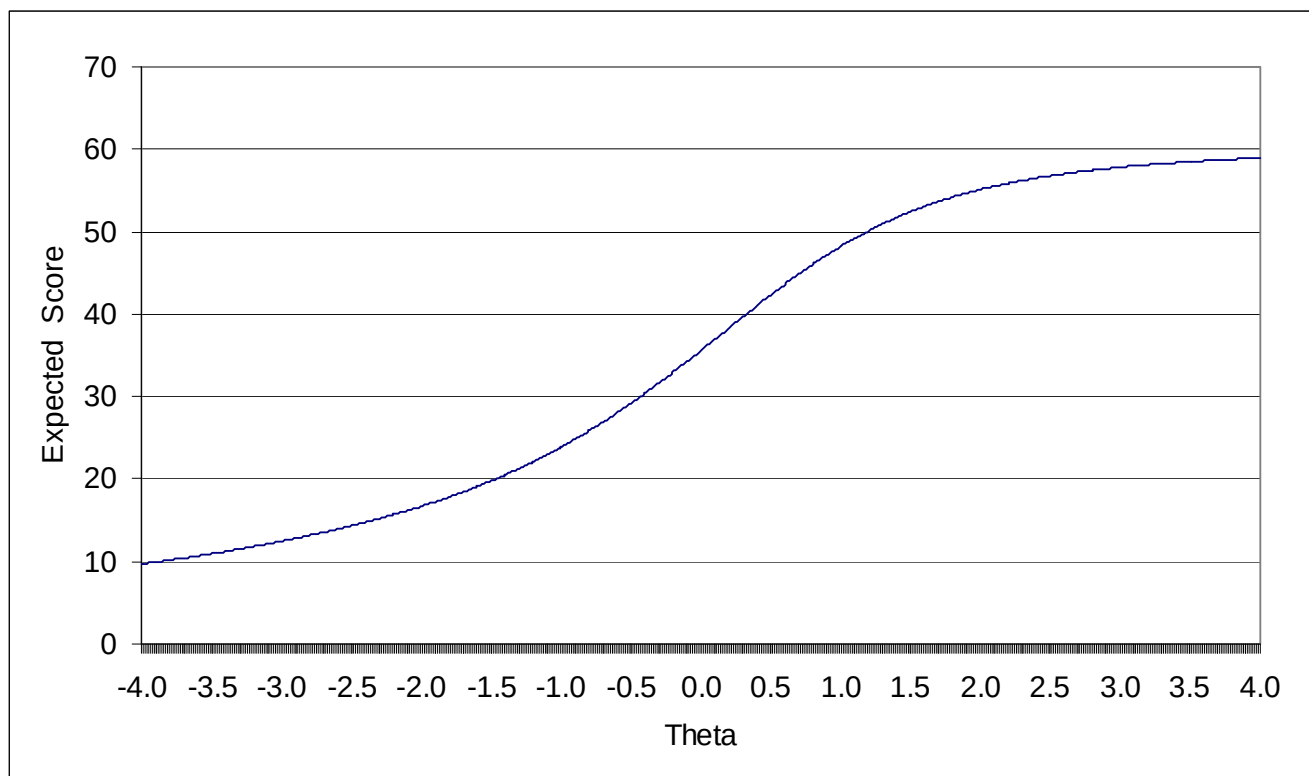


Figure 7.12 Test Characteristic Curve for Mathematics: Junior Division (French)

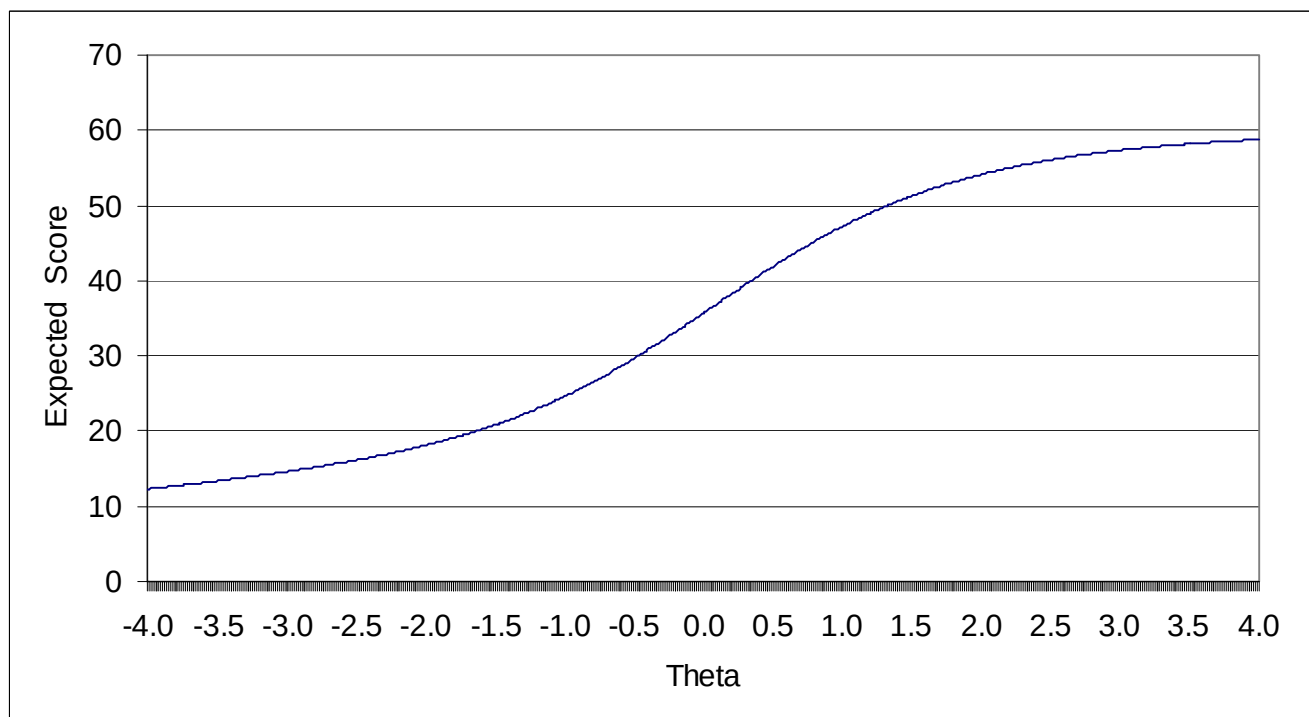


Figure 7.13 Test Information Function for Reading: Primary Division (English)

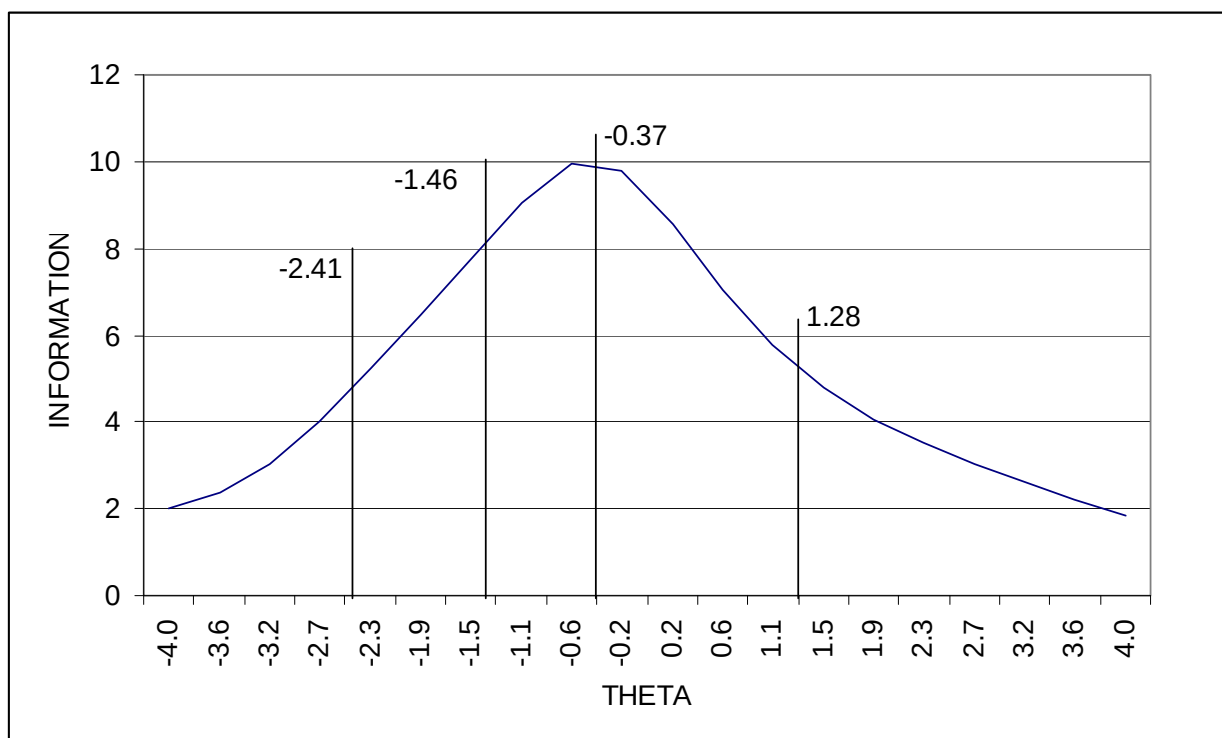


Figure 7.14 Test Information Function for Reading: Primary Division (French)

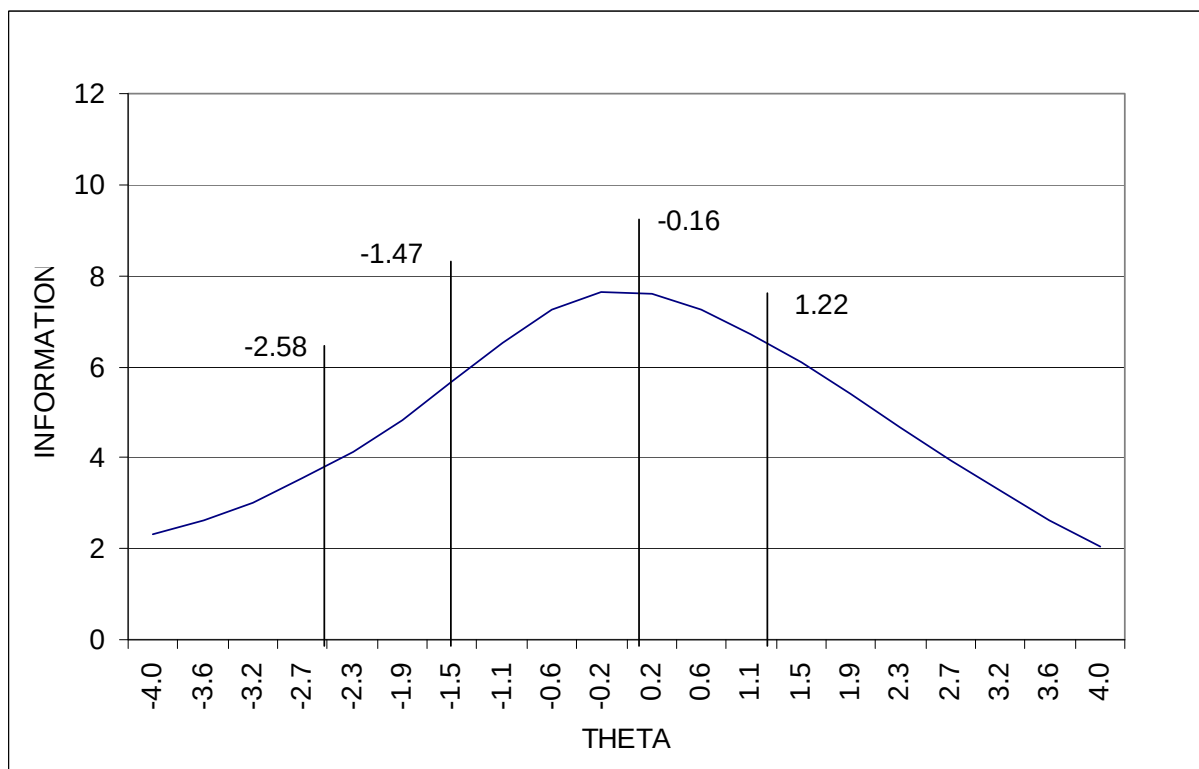


Figure 7.15 Test Information Function for Reading: Junior Division (English)

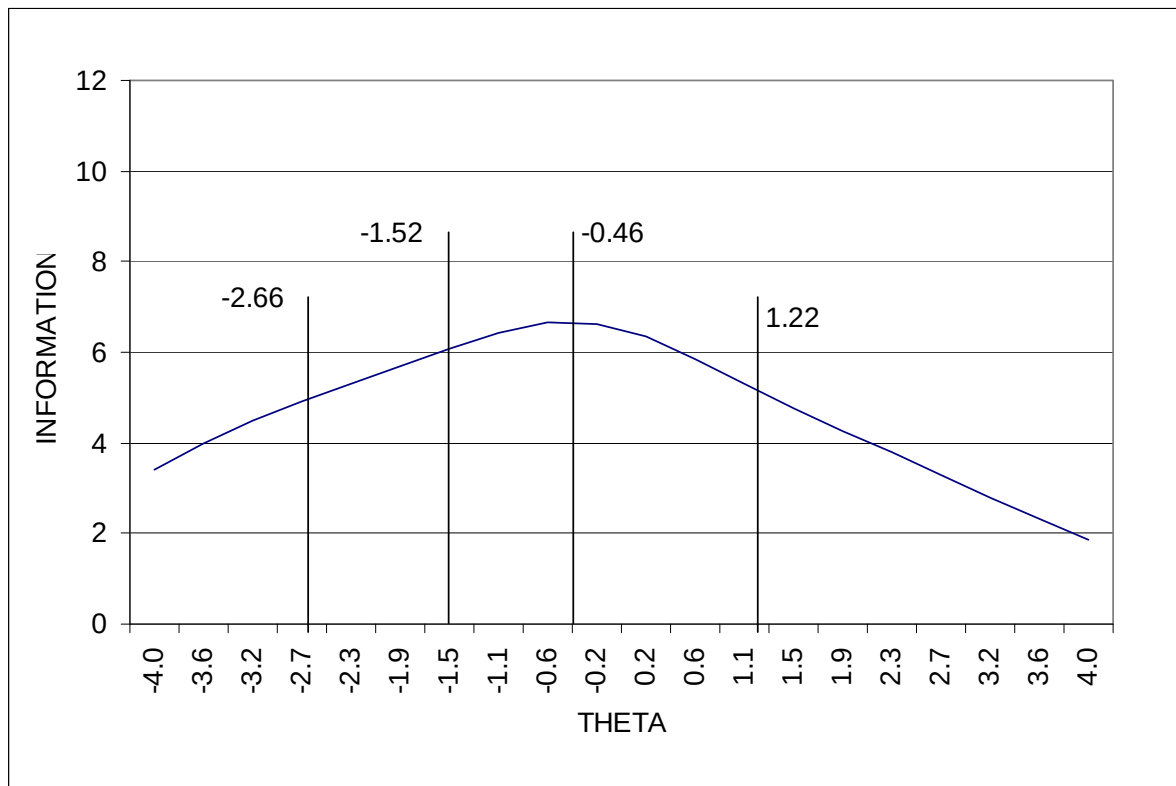


Figure 7.16 Test Information Function for Reading: Junior Division (French)

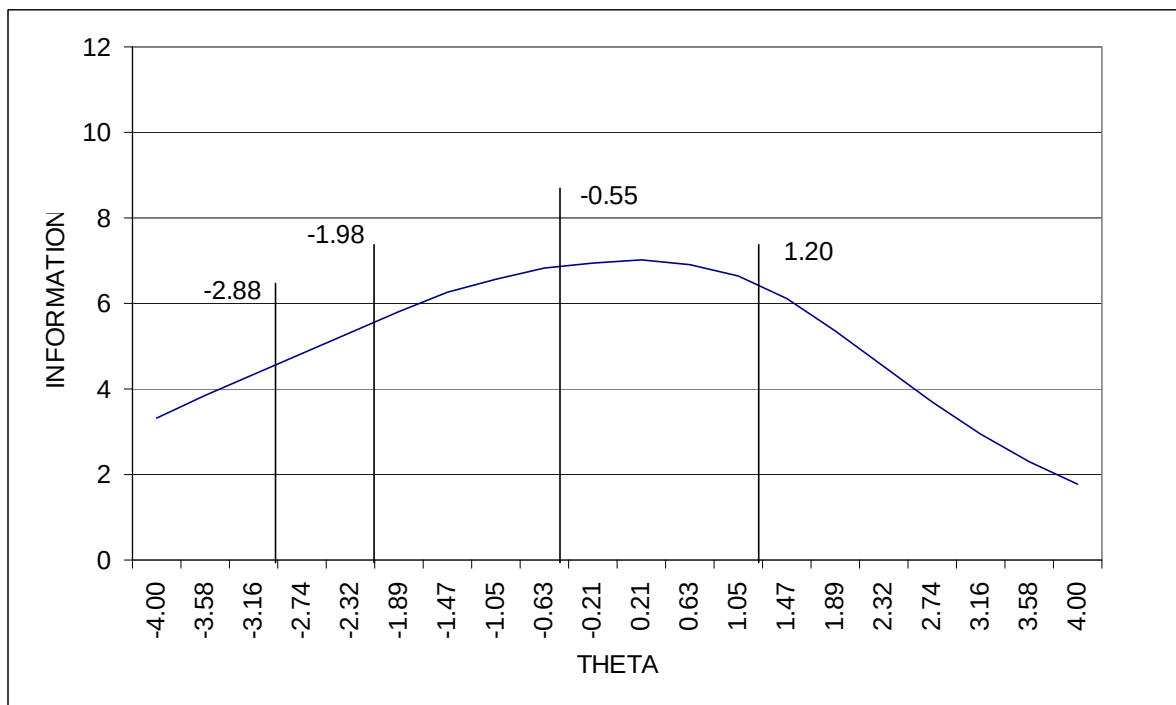


Figure 7.17 Test Information Function for Writing: Primary Division (English)

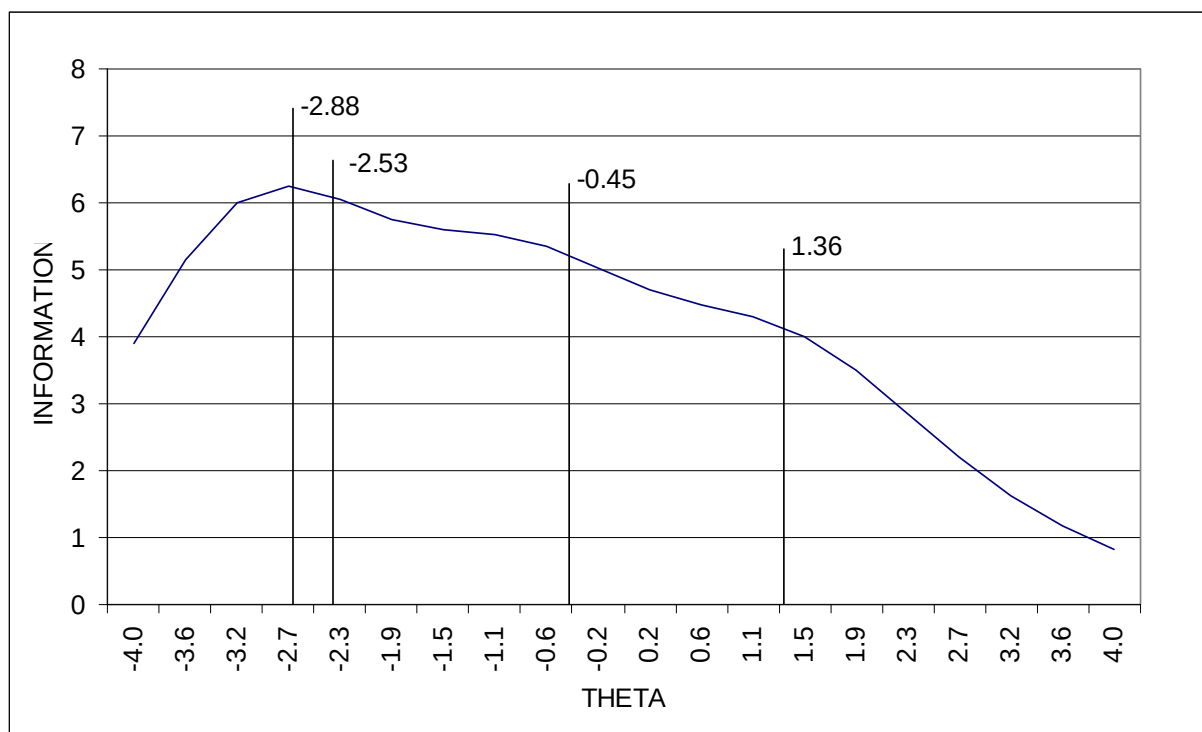


Figure 7.18 Test Information Function for Writing: Primary Division (French)

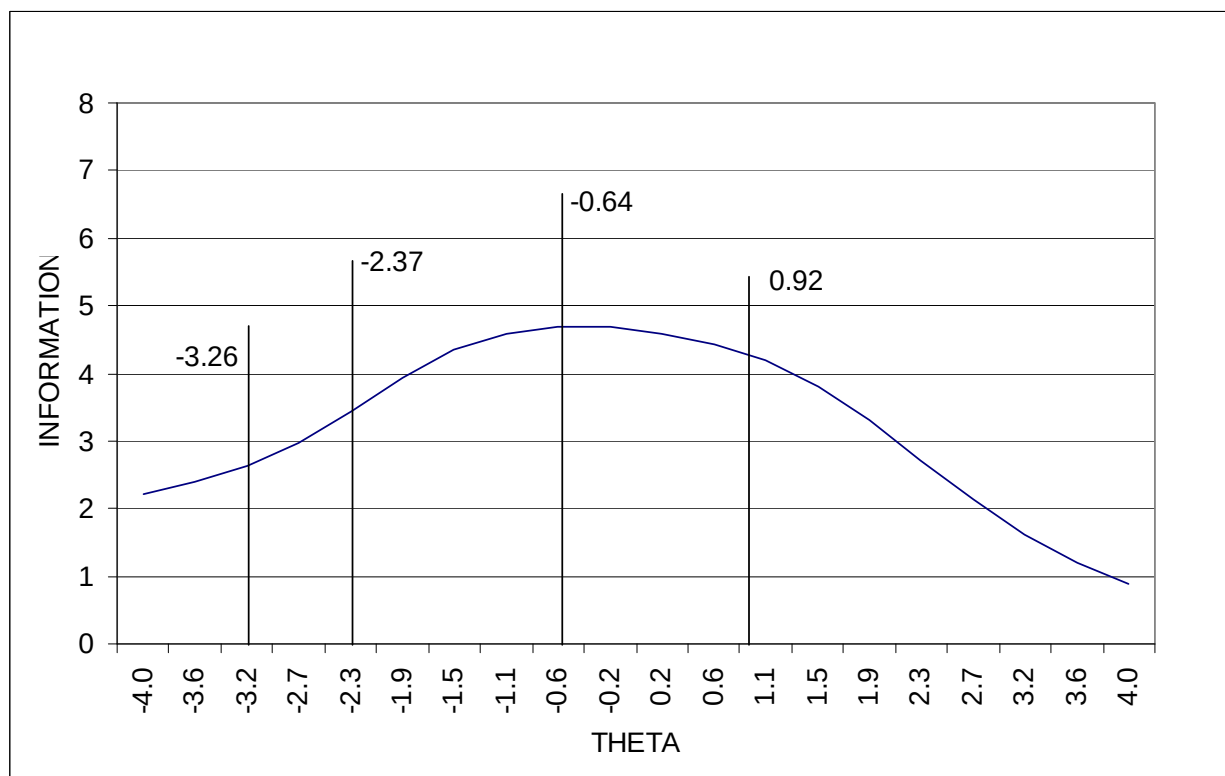


Figure 7.19 Test Information Function for Writing: Junior Division (English)

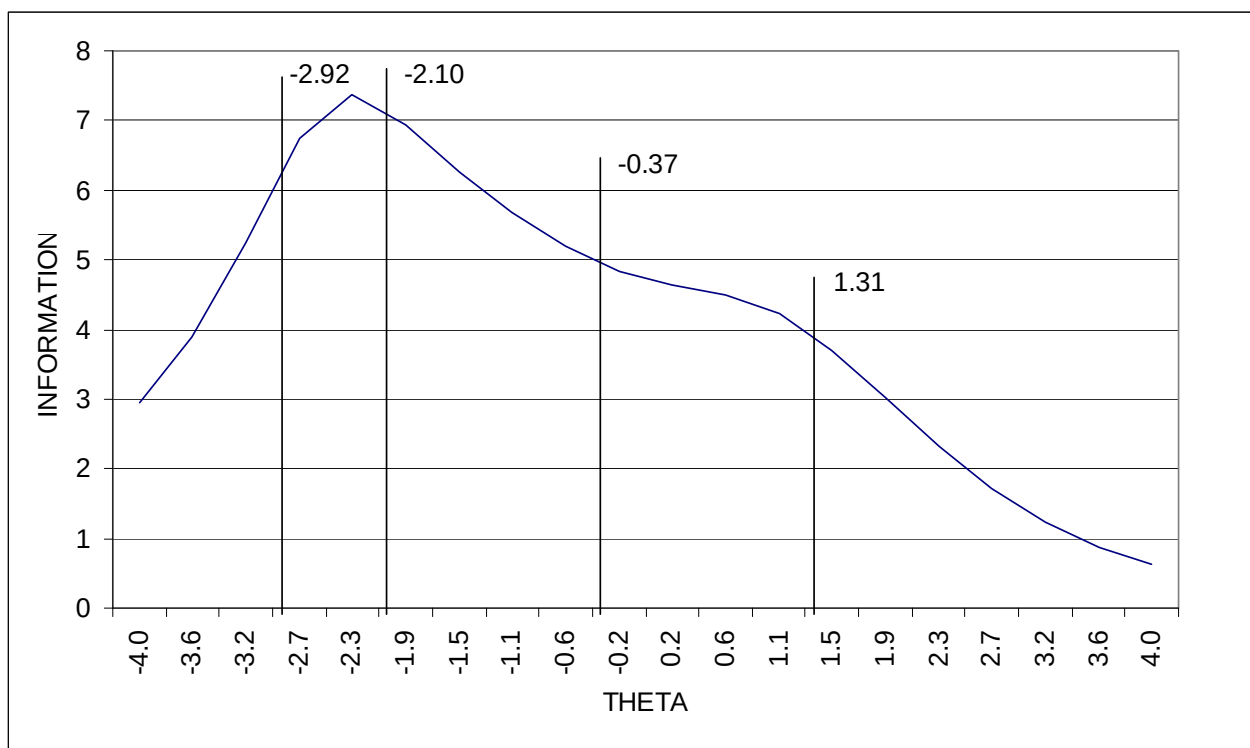


Figure 7.20 Test Information Function for Writing: Junior Division (French)

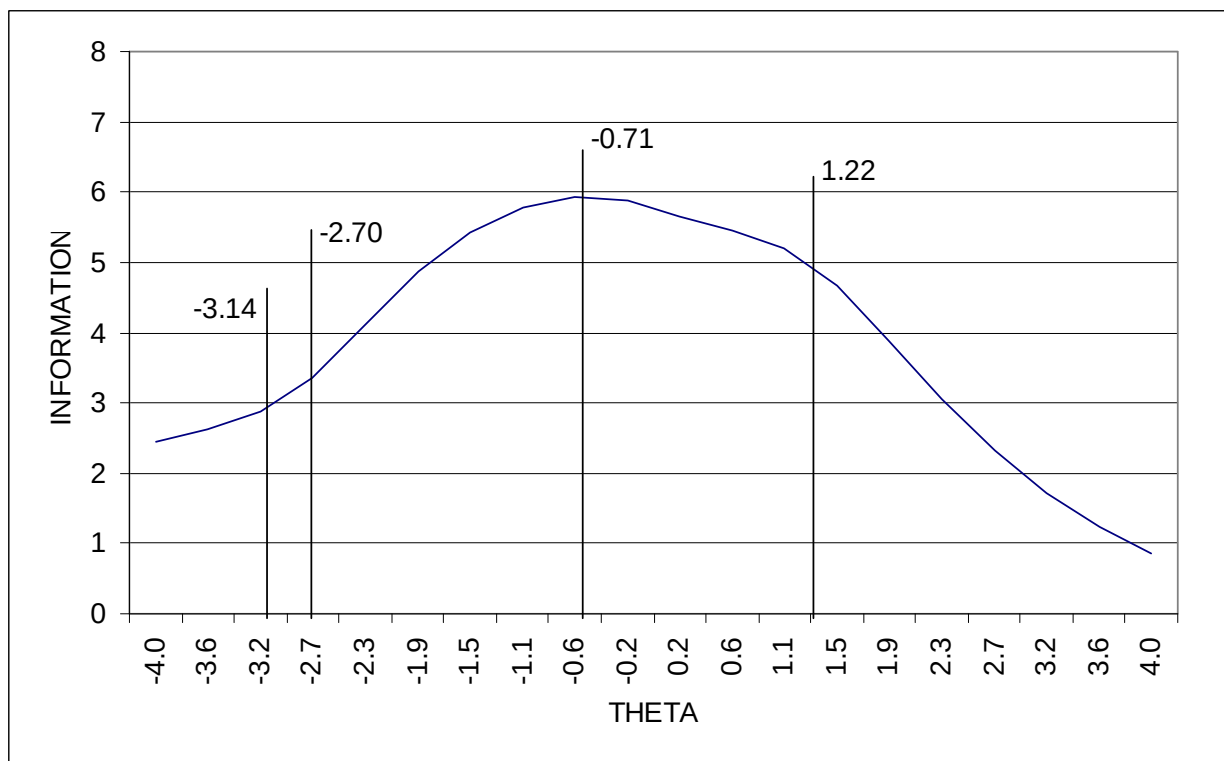


Figure 7.21 Test Information Function for Mathematics: Primary Division (English)

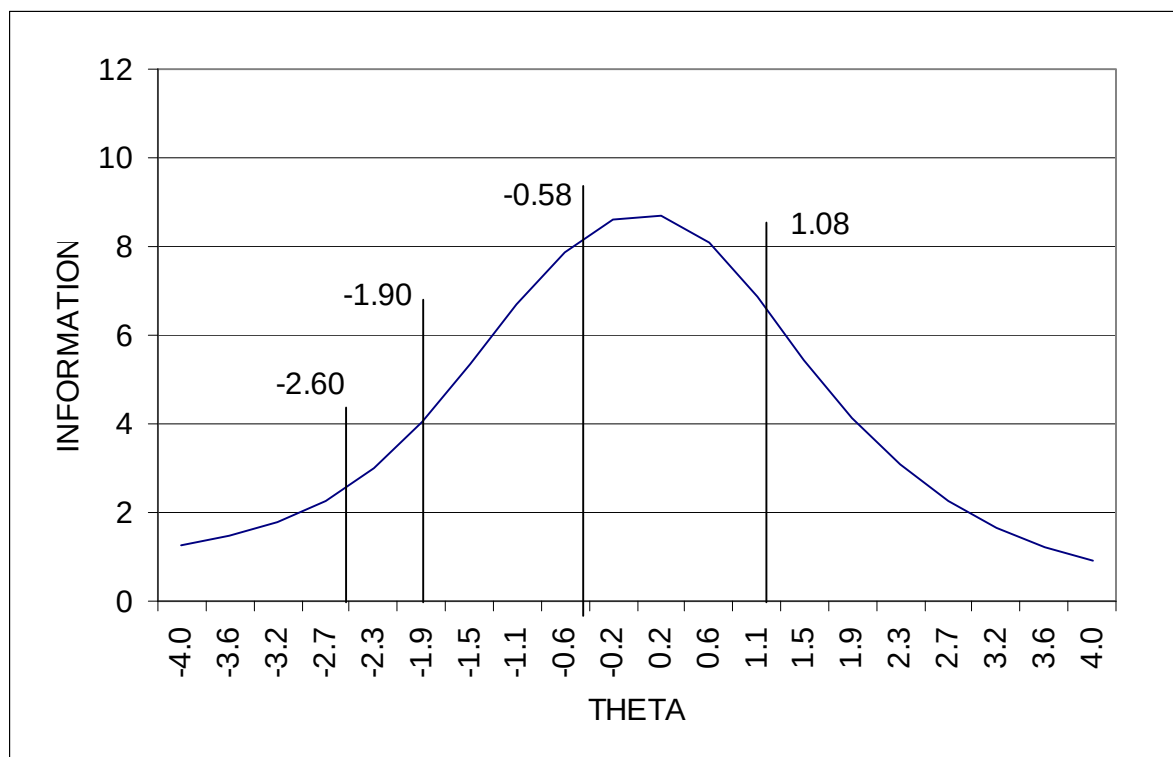


Figure 7.22 Test Information Function for Mathematics: Primary Division (French)

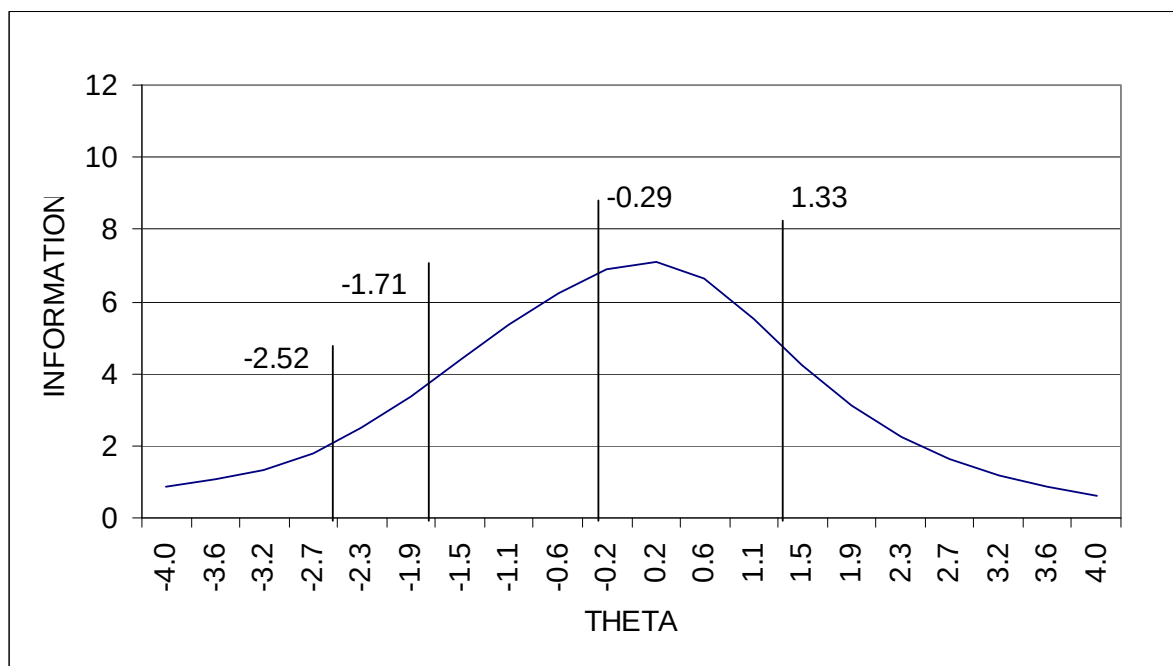


Figure 7.23 Test Information Function for Mathematics: Junior Division (English)

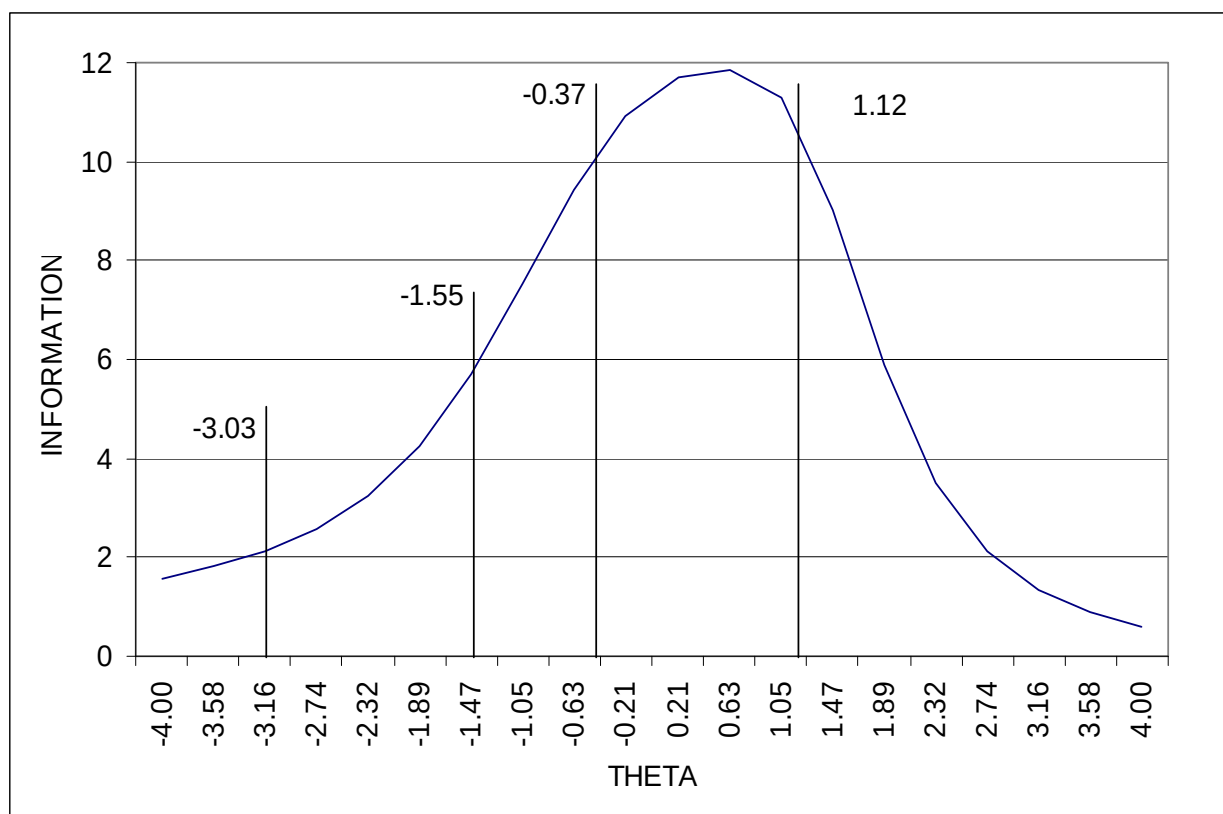
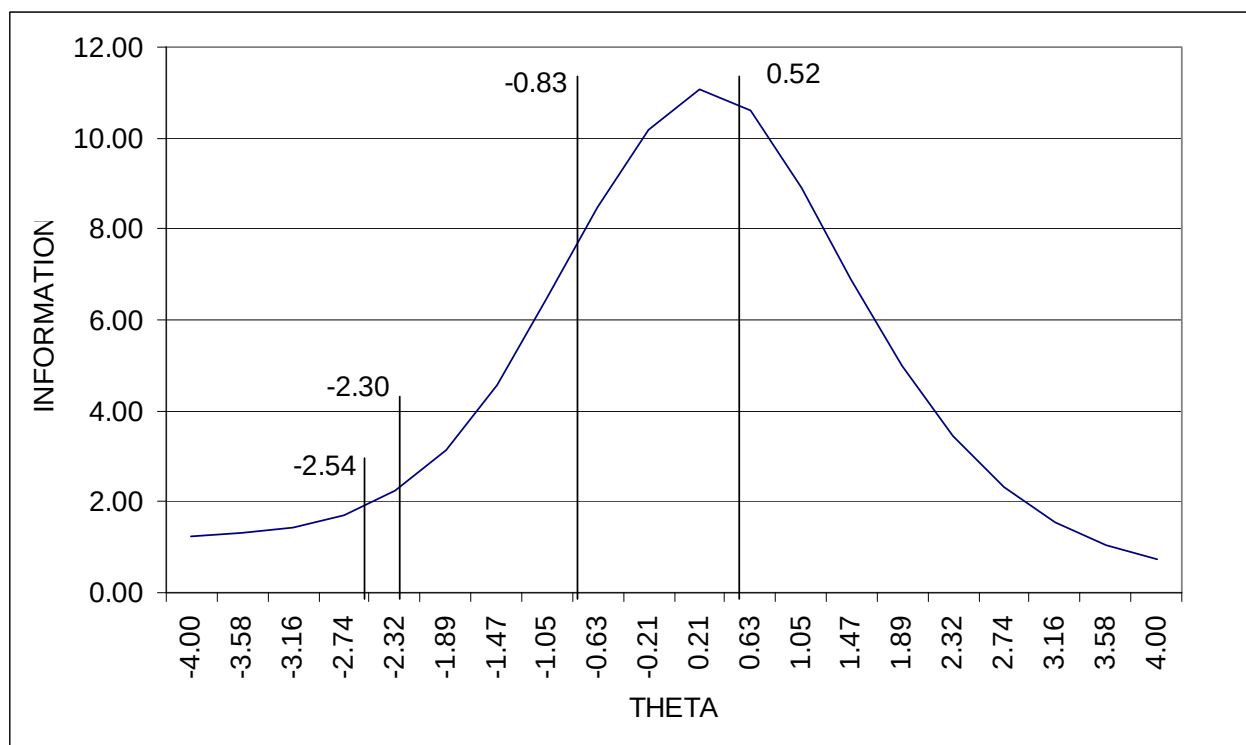


Figure 7.24 Test Information Function for Mathematics: Junior Division (French)



Item Parameter Estimates 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 primary and junior assessments. These statistics were computed using the equating sample (see Chapter 5). As expected, there is an inverse relationship between the item difficulty and the location, 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 and the IRT slope parameter. Statistics for individual items are presented in Appendix 7.1.

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

Grade and Subject	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*	
			Item Difficulty	Item-Total Correlation	Location	Slope
Primary Reading (English)	36	Min	0.40	0.18	−2.33	0.34
		Max	0.96	0.55	0.55	1.41
		Mean	0.66	0.40	−0.63	0.72
		STD	0.15	0.08	0.75	0.30
Primary Reading (French)	36	Min	0.43	0.11	−2.55	0.16
		Max	0.93	0.52	1.24	0.91
		Mean	0.63	0.35	−0.49	0.57
		STD	0.15	0.11	1.00	0.18
Junior Reading (English)	35	Min	0.42	0.16	−4.52	0.24
		Max	0.98	0.53	0.52	0.89
		Mean	0.69	0.35	−1.04	0.55
		STD	0.14	0.10	1.08	0.16
Junior Reading (French)	36	Min	0.42	0.14	−3.09	0.22
		Max	0.97	0.55	0.93	0.84
		Mean	0.64	0.34	−0.71	0.55
		STD	0.15	0.12	1.07	0.15
Primary Writing (English)	14	Min	0.45	0.23	−1.76	0.36
		Max	0.88	0.62	−0.45	0.97
		Mean	0.72	0.42	−1.05	0.65
		STD	0.15	0.14	0.43	0.18
Primary Writing (French)	14	Min	0.42	0.19	−2.33	0.31
		Max	0.89	0.60	0.34	0.89
		Mean	0.70	0.38	−1.25	0.54
		STD	0.17	0.15	0.86	0.16
Junior Writing (English)	14	Min	0.47	0.24	−2.07	0.39
		Max	1.00	0.63	0.67	0.94
		Mean	0.72	0.42	−1.09	0.61
		STD	0.16	0.15	0.69	0.15
Junior Writing (French)	14	Min	0.15	0.16	−2.04	0.26
		Max	0.89	0.65	2.60	1.04
		Mean	0.62	0.40	−0.67	0.61
		STD	0.20	0.18	1.15	0.23

STD = Standard Deviation

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

Grade and Subject	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*	
			Item Difficulty	Item-Total Correlation	Location	Slope
Primary Mathematics (English)	36	Min	0.19	0.13	-2.52	0.29
		Max	0.91	0.58	2.10	1.13
		Mean	0.62	0.36	-0.44	0.62
		STD	0.15	0.10	1.04	0.20
Primary Mathematics (French)	35	Min	0.43	0.05	-4.06	0.09
		Max	0.93	0.47	0.75	1.20
		Mean	0.66	0.33	-0.97	0.58
		STD	0.14	0.09	1.21	0.27
Junior Mathematics (English)	36	Min	0.26	0.13	-1.50	0.20
		Max	0.84	0.63	1.44	1.68
		Mean	0.59	0.41	-0.25	0.71
		STD	0.14	0.12	0.81	0.30
Junior Mathematics (French)	36	Min	0.32	0.21	-1.95	0.29
		Max	0.84	0.59	1.75	1.40
		Mean	0.59	0.39	-0.24	0.70
		STD	0.13	0.10	0.94	0.25

Note: * Guessing parameter was set as a constant of 0.2 for multiple-choice items

STD = Standard Deviation

Grade 9 Assessment of Mathematics

Classical Test Theory (CTT) Analyses

Table 7.4 presents descriptive statistics, Cronbach's alpha estimates of test score reliability and the standard error of measurement (SEM) for the English-language and French-language applied and academic versions of the Grade 9 mathematics assessment. The assessment's mean percentages ranged from 51.2% (English, applied, spring) to 61.6% (French, applied, spring) for applied mathematics. The mean percentages ranged from 53.5% (English, academic, winter) to 62.5% (French, academic, spring) for academic mathematics. The corresponding score distributions are provided in Appendix 7.1. Inspection reveals that the score distributions for applied mathematics are essentially symmetrical, albeit with one exception: French, spring is negatively skewed (slightly). In contrast, all four distributions for academic mathematics are negatively skewed (slightly).

Cronbach's alpha estimates were 0.81 or 0.82 for applied mathematics and 0.84 or 0.86 for academic mathematics. The corresponding standard errors were likewise similar, ranging from 7.2% to 7.7% of the possible maximum score for applied mathematics and 7.0% to 7.3% of the possible maximum score for academic mathematics. The obtained reliabilities and standard errors of measurement were acceptable. This indicates that the test scores from these assessments provide a satisfactory level of precision.

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

Program	No. of Items	Item Type		Possible Max. Score	No. of Students	Min.	Max.	Mean	STD	Reliability	SEM
		MC	OR								
Applied, Winter Administration (English)	31	24	7	52	20 159	0	52	27.04	8.54	0.81	3.75
Applied, Spring Administration (English)	31	24	7	52	24 641	0	52	26.67	8.89	0.82	3.81
Academic, Winter Administration (English)	31	24	7	52	43 485	1	52	27.83	9.08	0.84	3.66
Academic, Spring Administration (English)	31	24	7	52	57 948	0	52	29.59	9.72	0.86	3.69
Applied, Winter Administration (French)	31	24	7	52	804	0	51	26.63	8.85	0.81	3.89
Applied, Spring Administration (French)	31	24	7	52	724	5	50	30.81	9.14	0.82	3.83
Academic, Winter Administration (French)	31	24	7	52	1 923	6	52	31.50	9.24	0.84	3.74
Academic, Spring Administration (French)	31	24	7	52	2 131	4	52	32.50	9.58	0.84	3.82

MC = Multiple-Choice

OR = Open-Response

STD = Standard Deviation

SEM = Standard Error of Measurement

Item Response Theory (IRT) Analyses

The descriptive statistics were computed for both IRT scores and item parameter estimates after the 2006 and 2007 tests were equated. The estimated item parameters were then used to score all student responses. The descriptive statistics reported in Table 7.5 cover the total population of students. The mean student ability scores are less than zero and the standard deviations are less than one, due to the inclusion of all students. The item parameters for individual items are presented in Appendix 7.1.

Table 7.5 Descriptive Statistics of IRT Scores: Grade 9 Mathematics

Program	N	Minimum	Maximum	Mean	STD
Applied, Winter Administration (English)	20 159	-3.38	3.11	-0.17	1.00
Applied, Spring Administration (English)	24 645	-3.31	3.03	-0.16	1.02
Academic, Winter Administration (English)	43 485	-3.27	2.78	-0.10	0.94
Academic, Spring Administration (English)	57 949	-3.34	2.69	0.01	0.98
Applied, Winter Administration (French)	804	-3.17	2.85	-0.16	0.93
Applied, Spring Administration (French)	724	-3.08	2.42	0.01	0.97
Academic, Winter Administration (French)	1 923	-2.98	2.61	-0.06	0.93
Academic, Spring Administration (French)	2 131	-3.05	2.50	0.01	0.94

N = Number of students

STD = Standard Deviation

The TCCs and TIFs are displayed in Figures 7.25 to 7.32. The TCCs follow the expected S-shaped distribution. The TCCs for the winter and spring tests (for all of the assessments) either overlap or are very close to each other. This indicates that the winter and spring applied and academic tests possessed the same difficulty level. The TIFs indicate that the academic tests provided more information than the applied tests provided. Further, the amount of information at the 2/3 cut point was either at the maximum point (i.e., French applied) or nearly at the maximum point of the curves. The amount of information at the remaining 3 cut points was acceptable, with perhaps the exception of the 0/1 cut point for English academic.

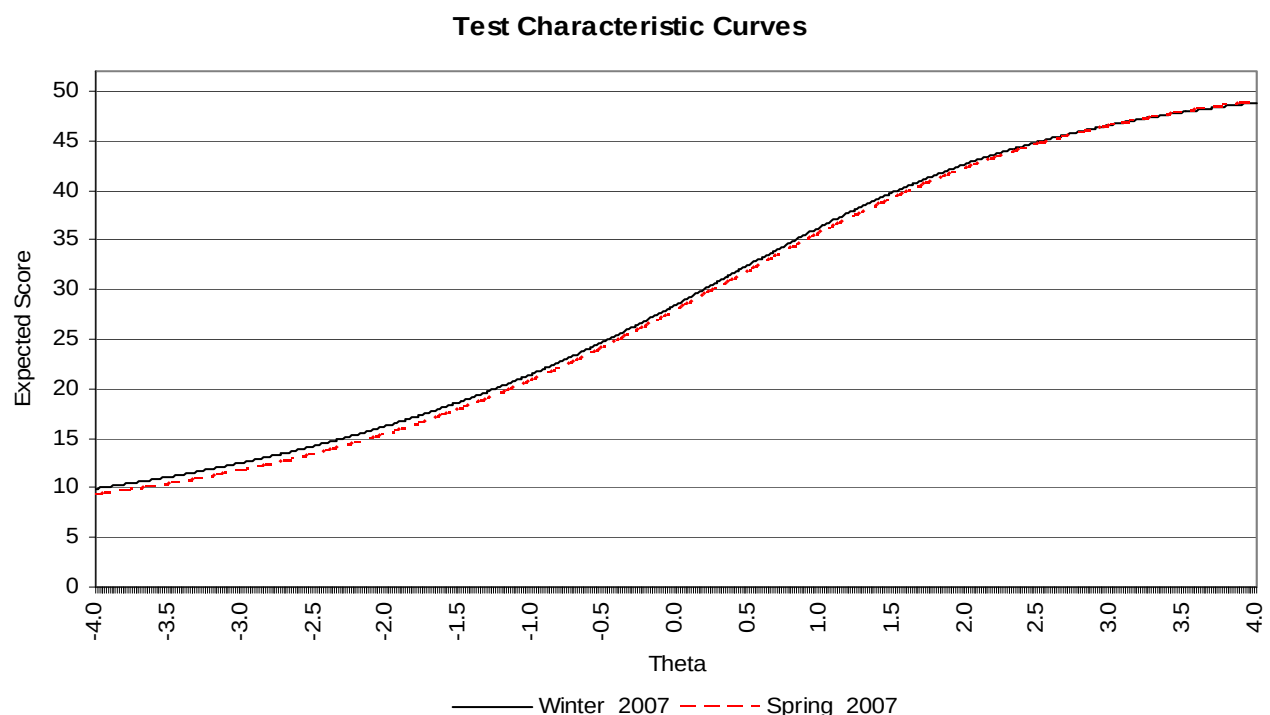
Figure 7.25 Test Characteristic Curves for Grade 9 Mathematics: Applied Program (English)

Figure 7.26 Test Characteristic Curves for Grade 9 Mathematics: Academic Program (English)

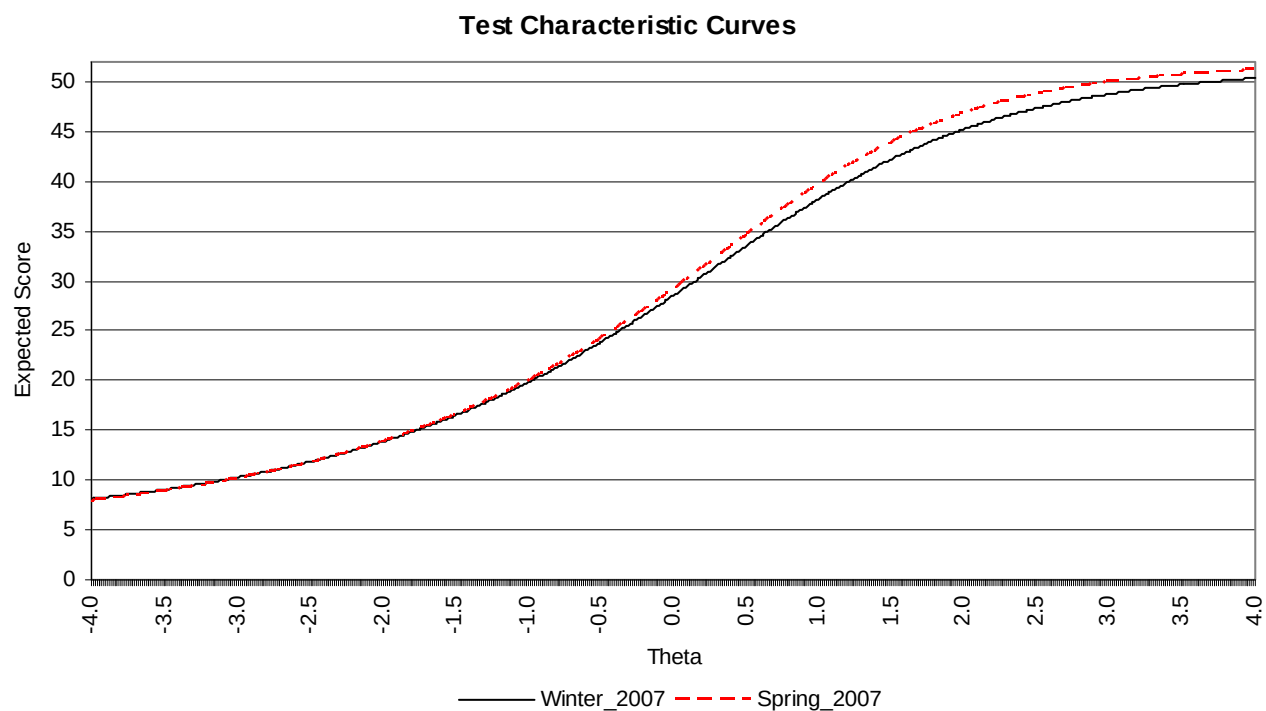


Figure 7.27 Test Characteristic Curves for Grade 9 Mathematics: Applied Program (French)

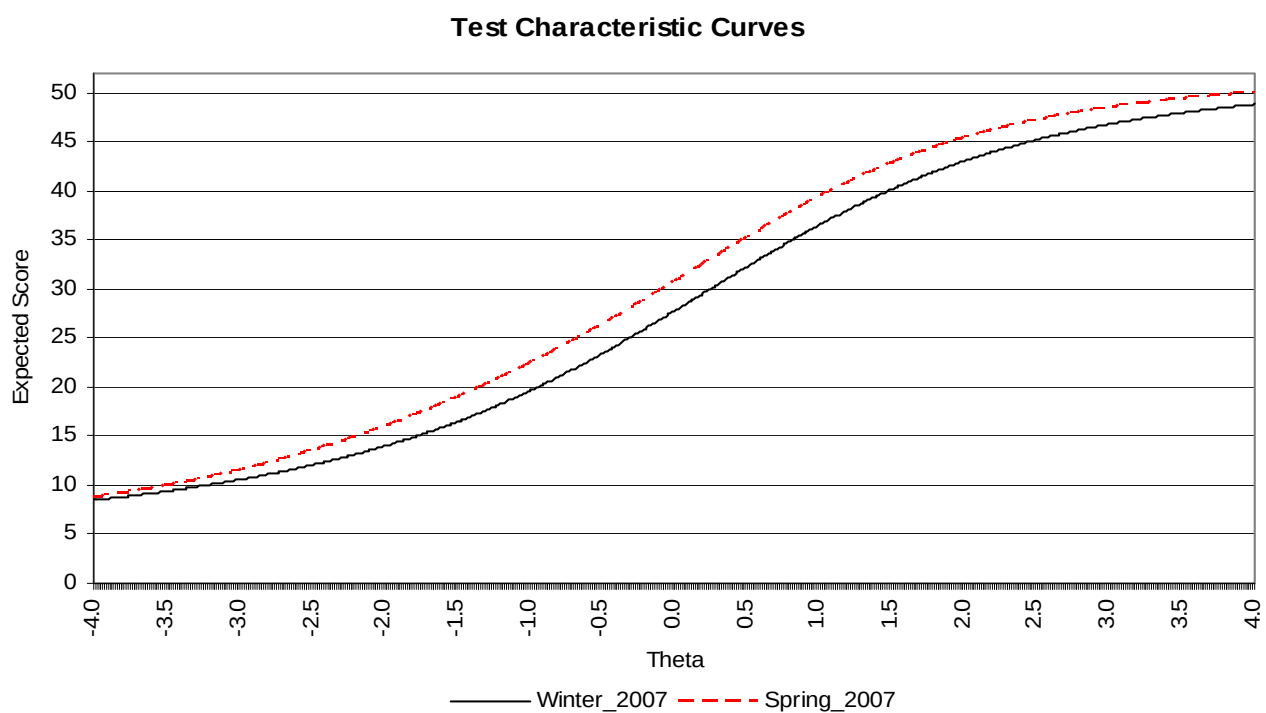


Figure 7.28 Test Characteristic Curves for Grade 9 Mathematics: Academic Program (French)

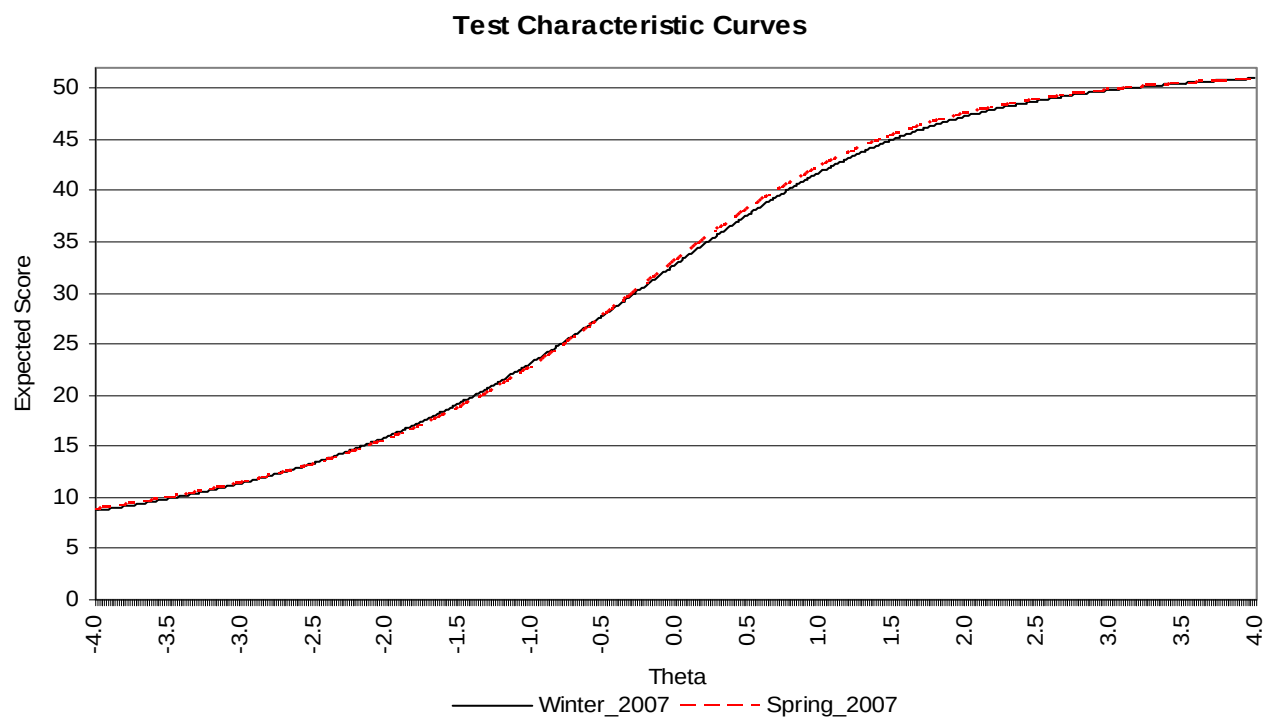


Figure 7.29 Test Information Function for Grade 9 Applied Mathematics: Applied Program (English)

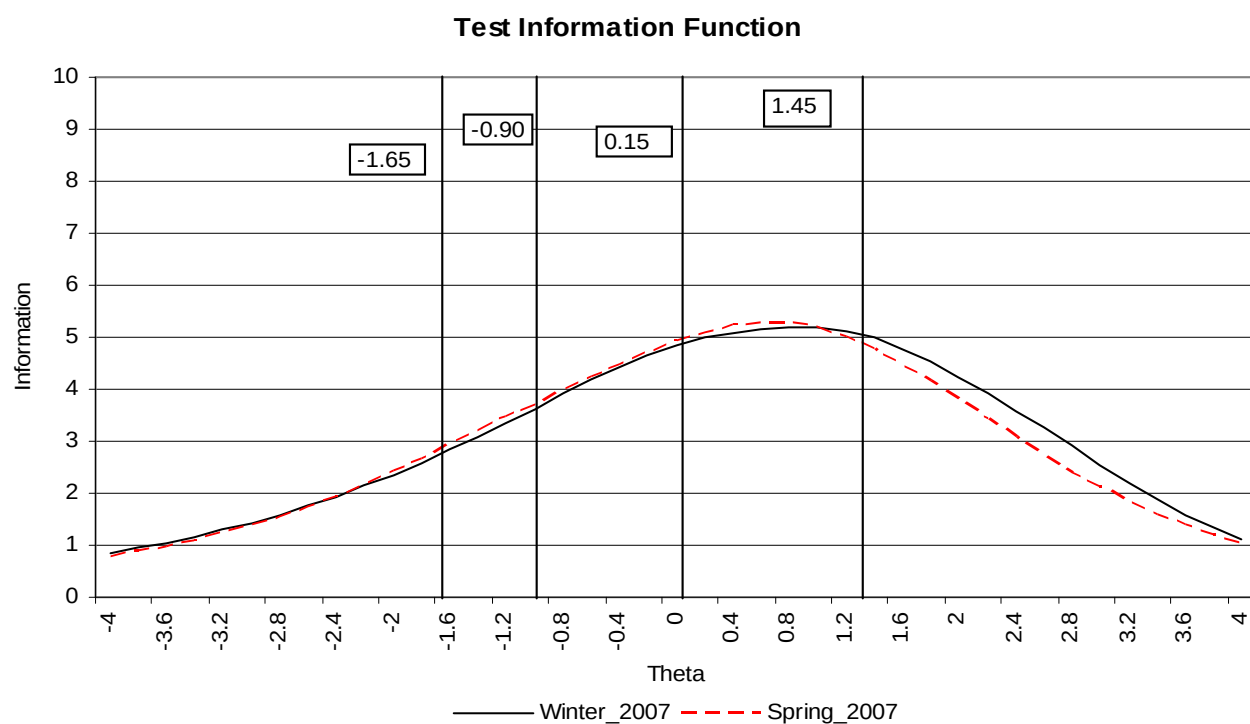


Figure 7.30 Test Information Function for Grade 9 Mathematics: Academic Program (English)

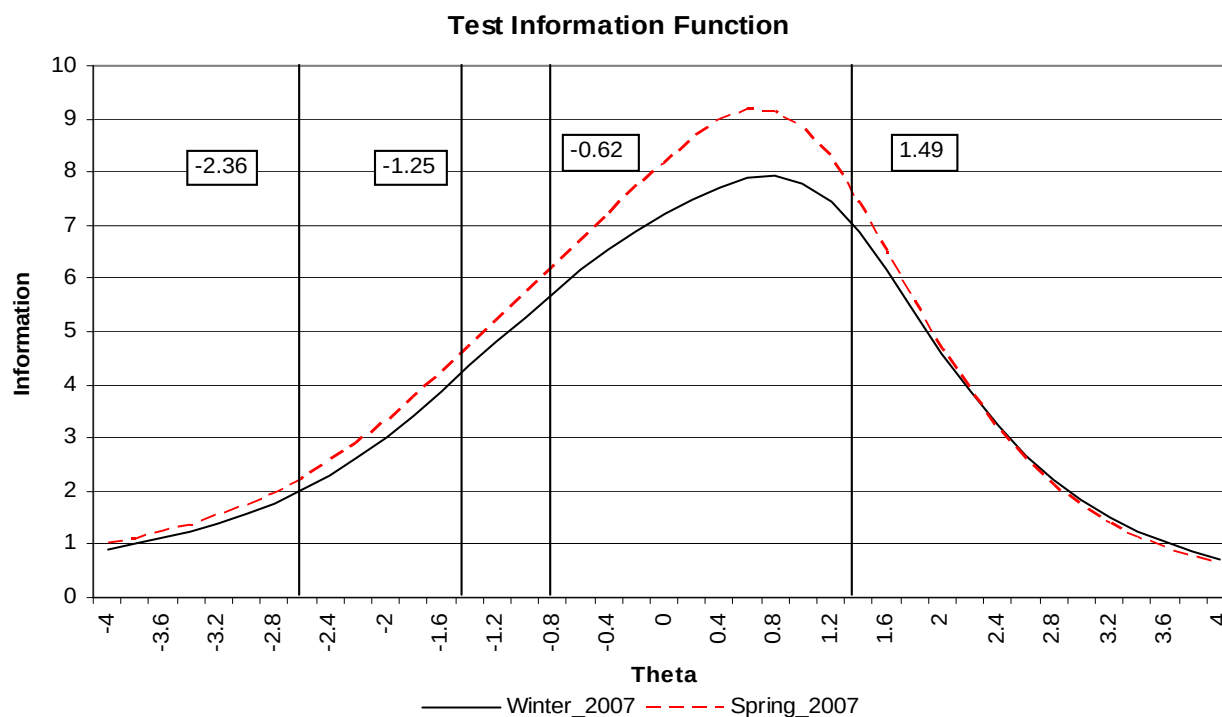


Figure 7.31 Test Information Function for Grade 9 Mathematics: Applied Program (French)

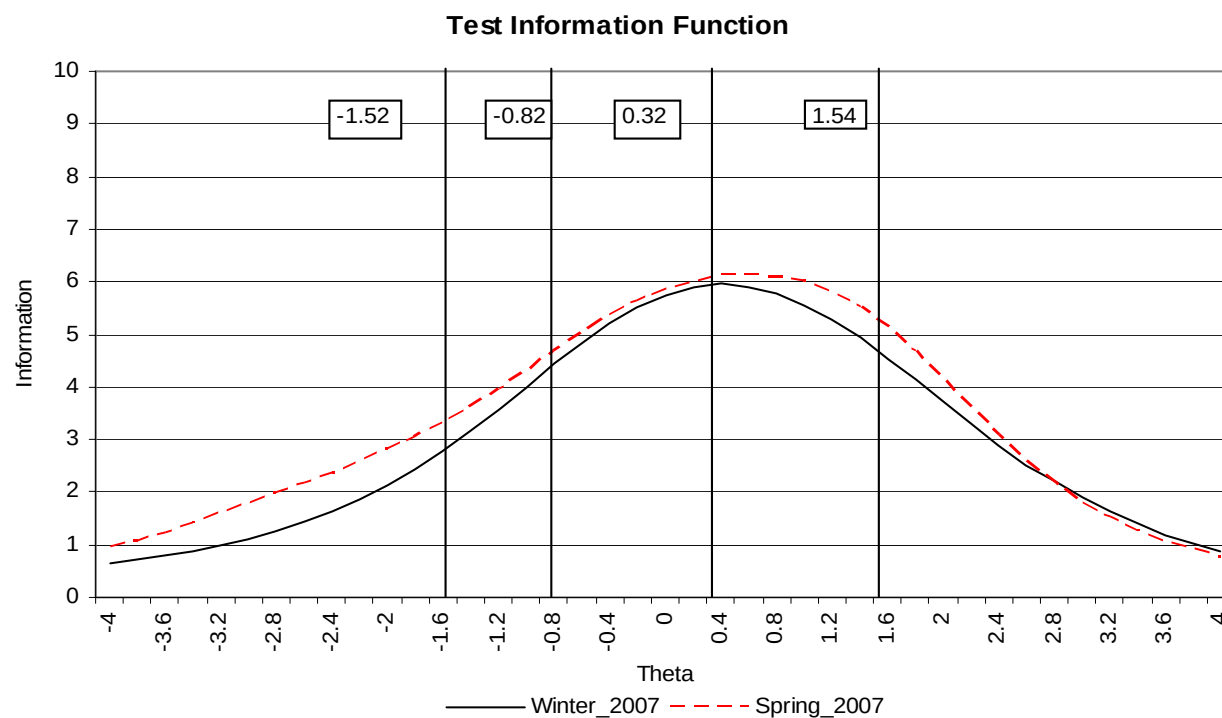
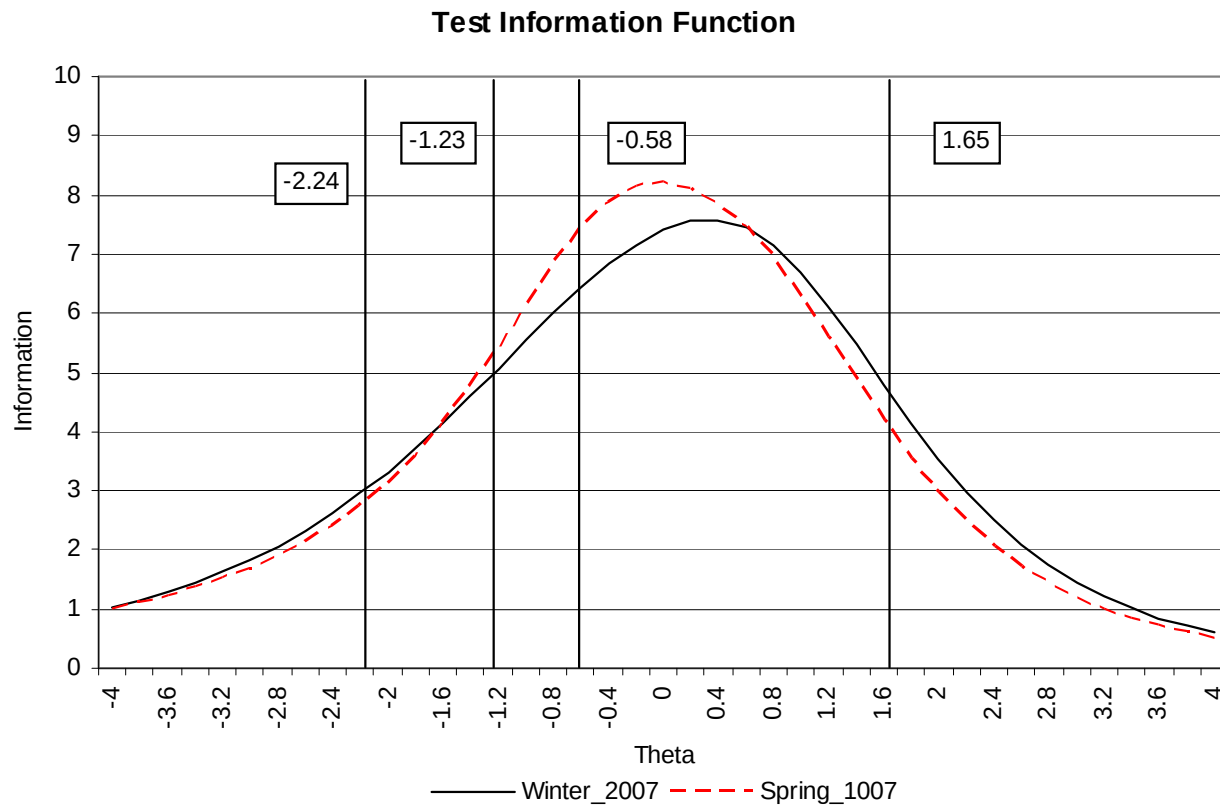


Figure 7.32 Test Information Function for Grade 9 Mathematics: Academic Program (French)



Item Parameter Estimates For Classical Test Theory (CTT) and Item Response Theory (IRT)

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

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

Program	No. of Items	Descriptive Statistics	CTT Item Statistics		IRT Item Parameters*	
			Item Difficulty	Item-Total Correlation	Location	Slope
Applied, Winter Administration (English)	31	Min	0.11	0.15	-2.496	0.258
		Max	0.87	0.60	2.582	1.149
		Mean	0.54	0.31	-0.080	0.510
		STD	0.18	0.11	1.370	0.250
Applied, Spring Administration (English)	31	Min	0.28	0.13	-2.496	0.221
		Max	0.87	0.56	2.141	0.836
		Mean	0.56	0.32	-0.100	0.490
		STD	0.17	0.11	1.240	0.180
Academic, Winter Administration (English)	31	Min	0.31	0.15	-1.638	0.302
		Max	0.86	0.55	2.180	1.199
		Mean	0.55	0.35	-0.050	0.630
		STD	0.16	0.10	0.970	0.250
Academic, Spring Administration (English)	31	Min	0.39	0.23	-1.782	0.339
		Max	0.86	0.62	1.242	1.174
		Mean	0.60	0.37	-0.190	0.650
		STD	0.15	0.11	0.930	0.220
Applied, Winter Administration (French)	31	Min	0.21	0.16	-2.496	0.258
		Max	0.75	0.58	2.582	1.149
		Mean	0.53	0.31	-0.080	0.510
		STD	0.12	0.11	1.370	0.250
Applied, Spring Administration (French)	31	Min	0.28	0.10	-2.496	0.221
		Max	0.83	0.55	2.141	0.836
		Mean	0.58	0.27	-0.100	0.490
		STD	0.14	0.13	1.240	0.180
Applied, Winter Administration (French)	31	Min	0.36	0.21	-2.087	0.275
		Max	0.86	0.55	1.355	0.973
		Mean	0.60	0.35	-0.300	0.600
		STD	0.13	0.10	0.880	0.210
Applied, Spring Administration (French)	31	Min	0.44	0.18	-2.087	0.280
		Max	0.78	0.62	0.823	0.973
		Mean	0.62	0.36	-0.330	0.570
		STD	0.10	0.11	0.770	0.210

Note: * Guessing parameter was set as a constant of 0.2 for multiple-choice items
STD = Standard Deviation

Ontario Secondary School Literacy Test (OSSLT)

Table 7.7 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 OSSLT (for first-time eligible and previously eligible students). Inspection of these distributions reveals that the distributions are somewhat negatively skewed. The mean percentages for first-time eligible students, (80.9% for English-language and 78.1% for French-language) were greater than the corresponding means for previously eligible students (69.7% and 69.0% respectively). The corresponding score distributions are provided in Appendix 7.1.

Cronbach's alpha estimates ranged from 0.85 to 0.87. The corresponding standard errors were likewise similar, ranging from 4.0% to 4.9% of the possible maximum score. The obtained reliabilities and standard errors of measurement were acceptable and indicate that test scores from these assessments were at a satisfactory level of precision.

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

Language		No. of Items	Item Type				Possible Max. Score	N	Min.	Max.	Mean	STD	R	SEM
			MC	OR	SW	LW								
English	First-Time Eligible Students	47	39	4	2	2	81	146 150	7.0	81.0	65.51	8.84	0.87	3.24
	Previously Eligible Students	47	39	4	2	2	81	29 966	2.0	80.5	56.46	10.78	0.87	3.94
French	First-Time Eligible Students	47	39	4	2	2	81	5 455	15.5	79.5	63.26	8.53	0.85	3.36
	Previously Eligible Students	47	39	4	2	2	81	1 013	0.0	78.5	55.85	10.35	0.86	3.85

MC = Multiple-Choice

SW = Short writing

N = Number of students

STD = Standard Deviation

OR = Open-Response

LW = Long writing

R = Reliability

SEM = Standard Error of Measurement

Item Response Theory (IRT) Analysis

The descriptive statistics were computed for both IRT scores and item parameter estimates after the 2006 and 2007 tests were equated (based on the equating samples described in Chapter 5). The estimated item parameters were then used to score all student responses. The descriptive statistics reported in Table 7.8 cover the total student population. Due to the inclusion of all students, the student ability mean scores are less than zero and the standard deviations are less than one. The item parameters for individual items are presented in Appendix 7.1.

Table 7.8 Descriptive Statistics for IRT Scores: OSSLT

Language		Number of Students	Minimum	Maximum	Mean	Standard Deviation
English	First-Time Eligible Students	146 150	−3.83	2.56	−0.08	0.87
	Previously Eligible Students	29 966	−3.85	2.39	−0.92	0.84
French	First-Time Eligible Students	5 455	−3.43	2.36	−0.05	0.82
	Previously Eligible Students	1 013	−3.94	2.07	−0.73	0.87

The TCCs are displayed in Figures 7.33 and 7.34. The TCCs follow the expected S-shaped distribution. The TIF plots for the English-language and French-language test administrations are shown in Figures 7.35 and 7.36. The TIFs indicate that the amount of information at the cut point was satisfactory, even though it was not at the maximum point.

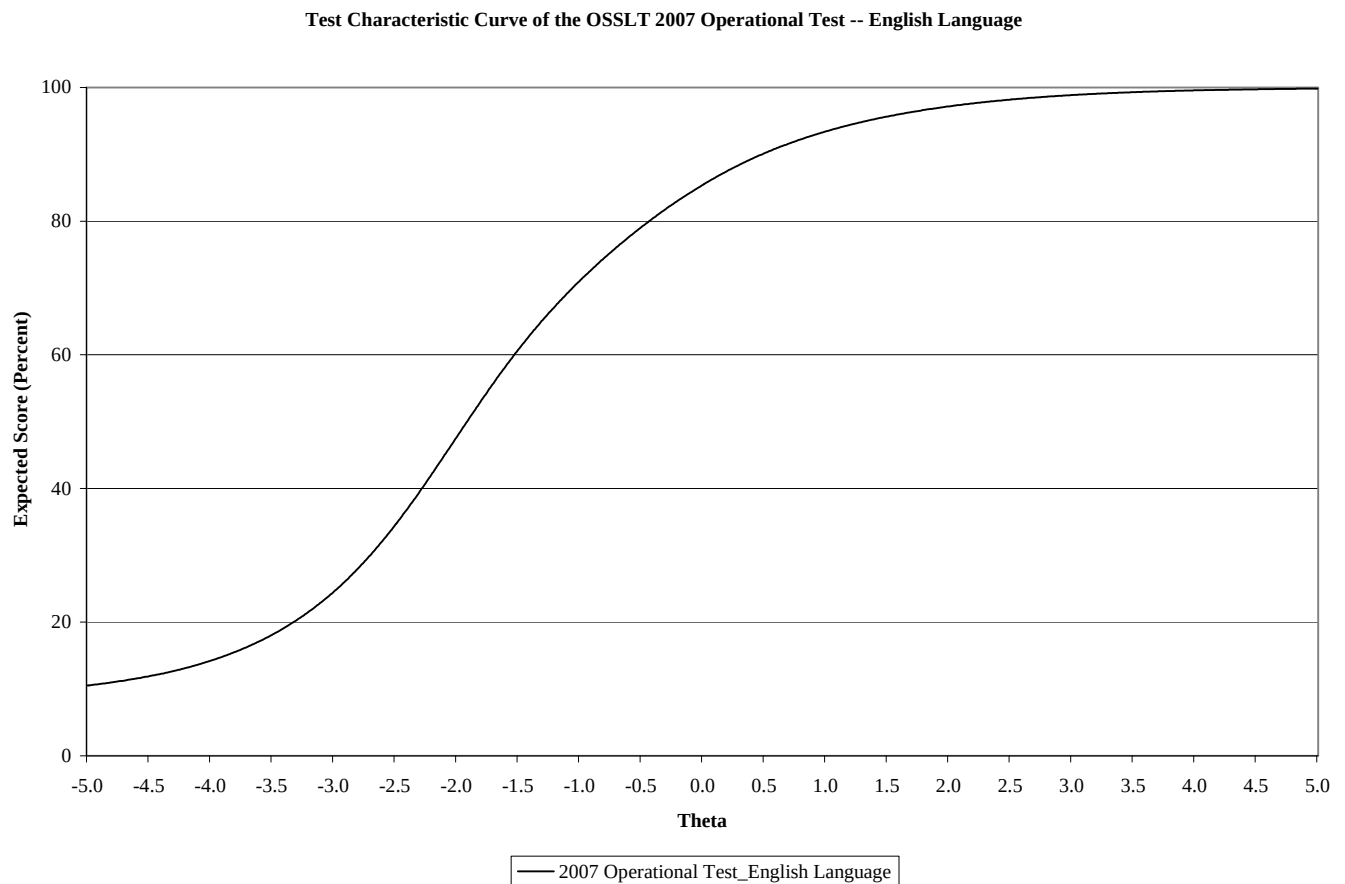
Figure 7.33 Test Characteristic Curve: OSSLT (English)

Figure 7.34 Test Characteristic Curve: OSSLT (French)

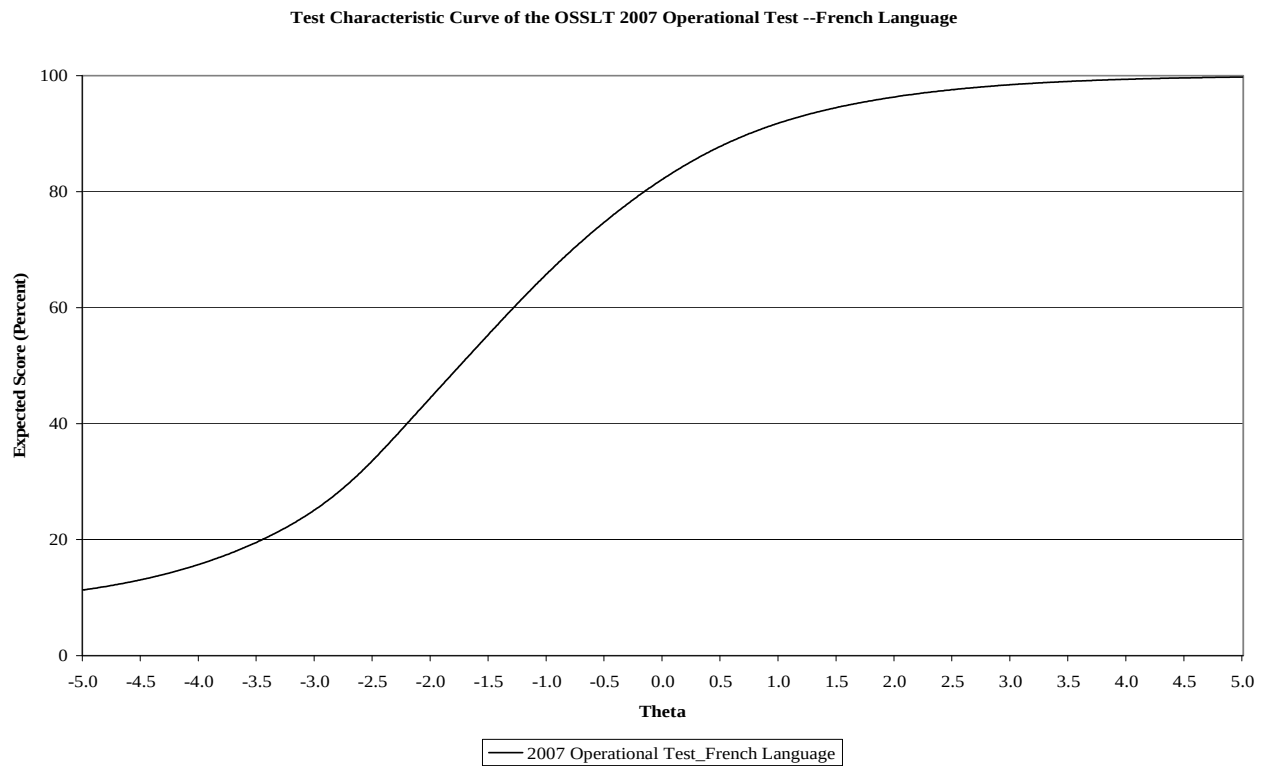


Figure 7.35 Test Information Function: OSSLT (English)

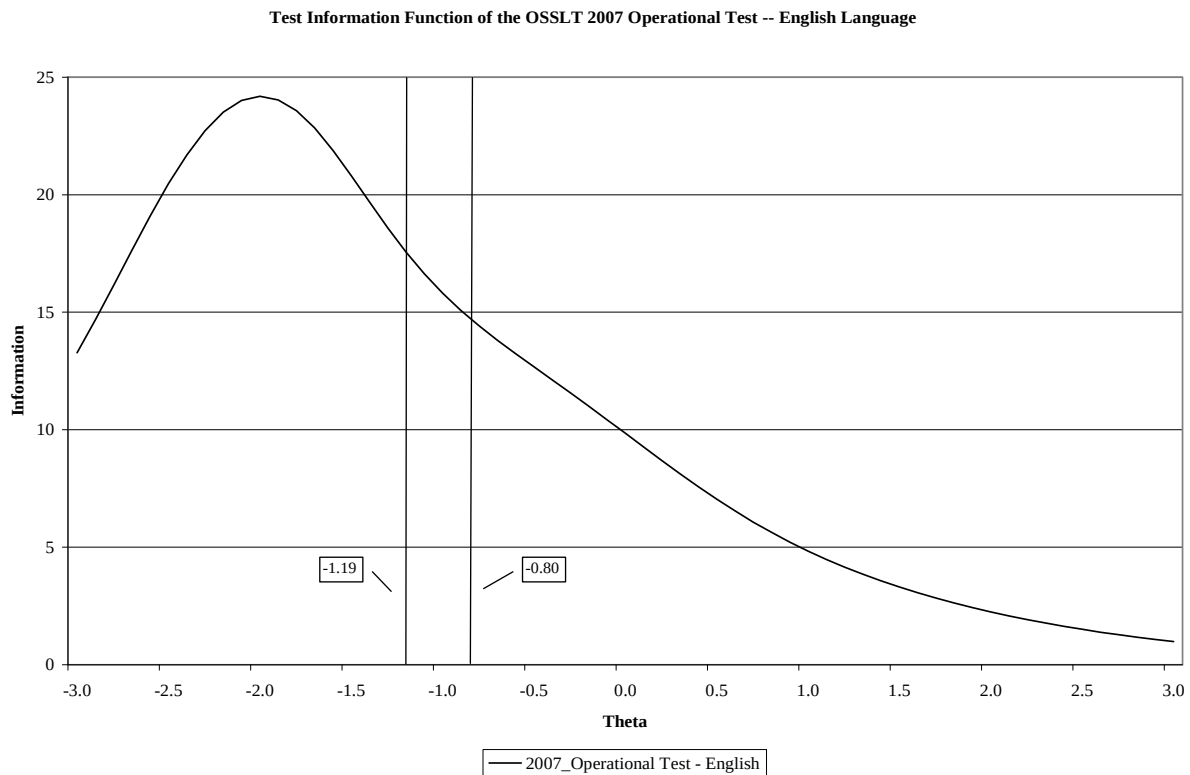
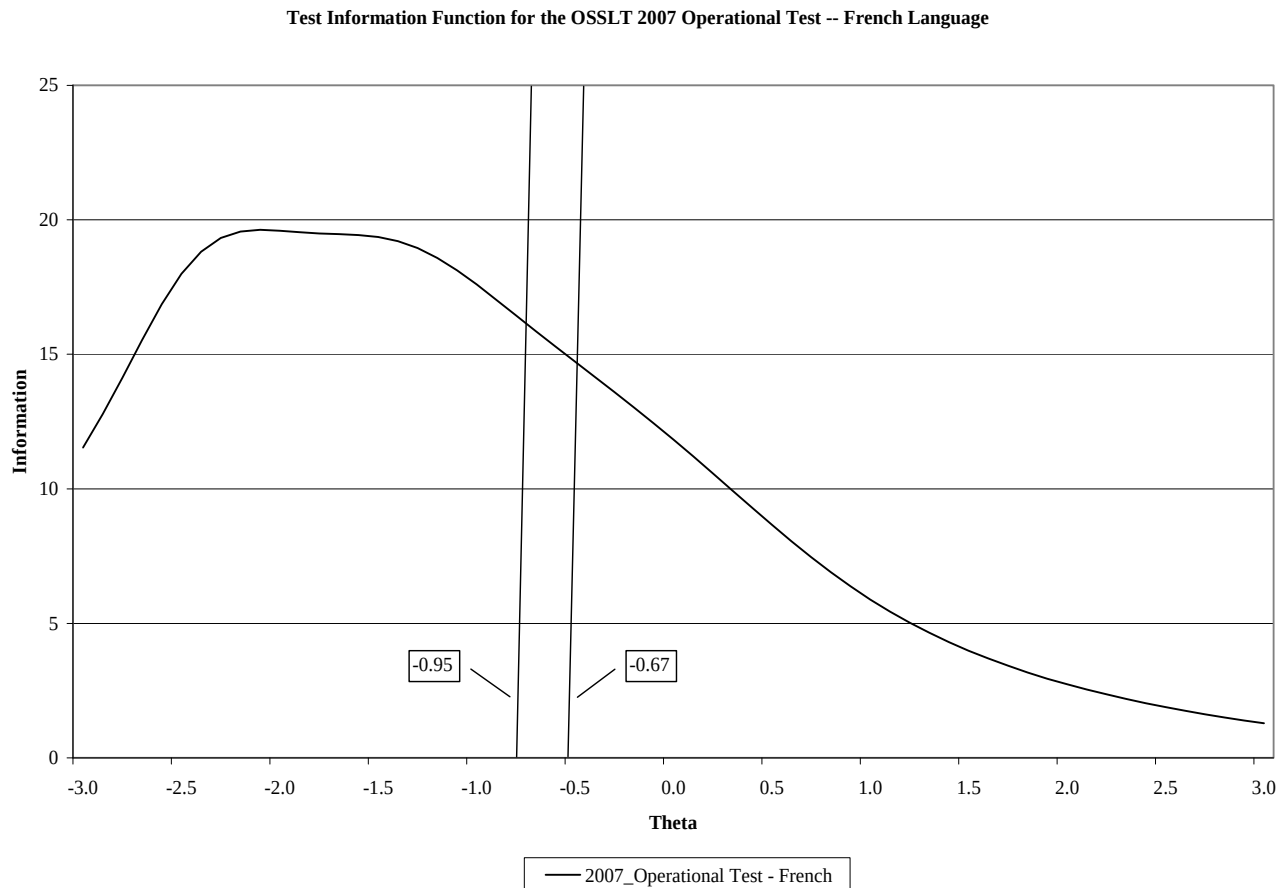


Figure 7.36 Test Information Function: OSSLT (French)



Item Parameter Estimates For Classical Test Theory (CTT) and Item Response Theory (IRT)

Table 7.9 contains a summary of both the CTT and IRT item statistics for the multiple-choice items included in the primary and junior assessments. As with the primary and junior assessments, care must be taken to avoid matching the minimum CTT p -value with the minimum IRT location value, and the item-total correlation with the slope. As expected, there is an inverse relationship between the item difficulty and the location, due to the difference between the definitions of difficulty in the two approaches. Unlike the primary, junior and Grade 9 assessments, the value of the a -parameter was set to a constant value. Hence, it is not possible to determine the nature of the relationship between the CTT item-total correlation and the IRT slope parameter. However, the low values for the minimum point biserial correlations suggest that some of the items did not reach the expected level (0.20) but the item difficulty values were within an acceptable range. Statistics for individual items are presented in Appendix 7.1.

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	Location
English	47	Min.	36.96	0.08	-3.53
		Max.	98.47	0.60	1.68
		Mean	80.03	0.32	-1.57
		STD	14.05	0.11	1.11
French	47	Min.	26.43	-0.09	-4.54
		Max.	98.30	0.57	2.27
		Mean	77.49	0.29	-1.47
		STD	17.50	0.12	1.45

Note: * The slope was set at 0.588 for all items and the guessing parameter was set to 0.20 for all multiple-choice items.
 STD = Standard Deviation

Appendix 7.1: Score Distributions and Item Statistics for the 2007 EQAO Assessments

This appendix presents the following:

- graphs showing the distribution of raw scores for each assessment;
- classical statistics for the operational items on the 2007 assessments and
- IRT parameters for the operational items on the 2007 assessments.

The score distribution graphs are presented in Figures 7.1.1 to 7.1.12 for the primary and junior assessments, in Figures 7.1.13 to 7.1.20 for the Grade 9 mathematics assessment and in Figures 7.1.21 to 7.1.24 for the OSSLT.

Classical item statistics and IRT item parameters are combined into tables for each assessment: Tables 7.1.1 to 7.1.12 for the primary and junior assessments, Tables 7.1.13 to 7.1.20 for the Grade 9 mathematics assessment and Tables 7.1.25 and 7.1.26 for the OSSLT.

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

Score Distributions

Figure 7.1.1 Score Distribution for Primary Reading (English)

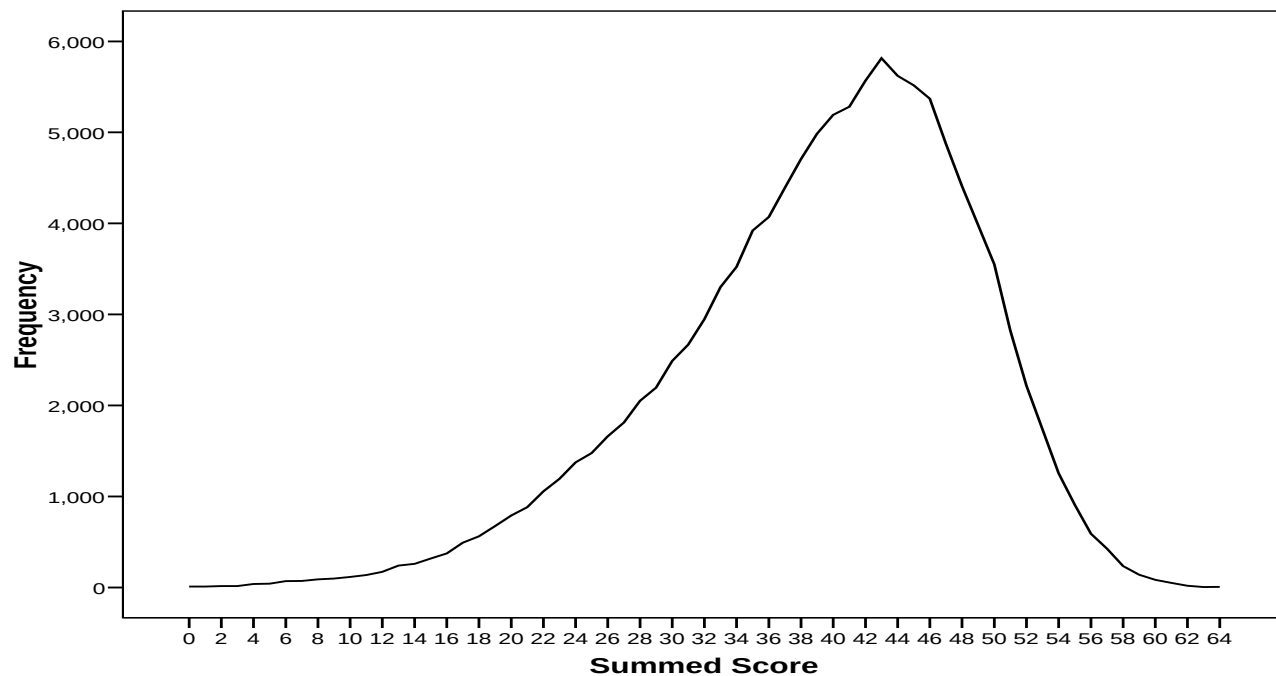


Figure 7.1.2 Score Distribution for Primary Reading (French)

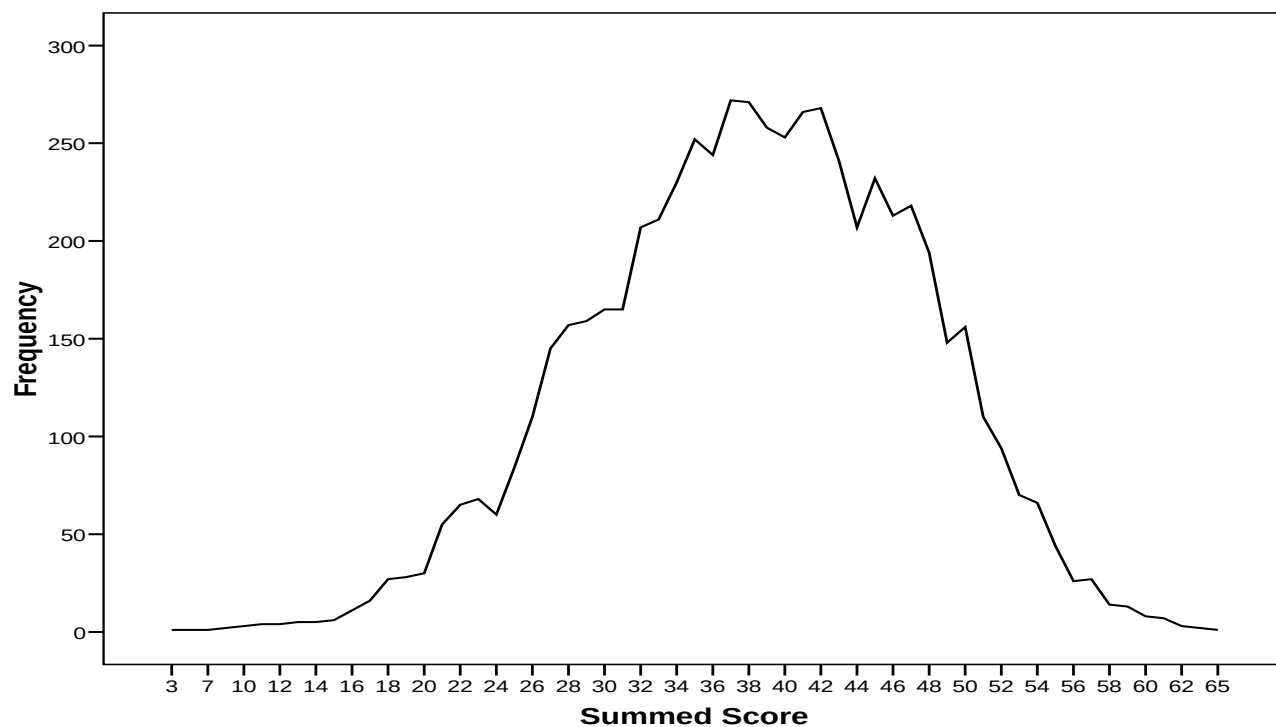


Figure 7.1.3 Score Distribution for Junior Reading (English)

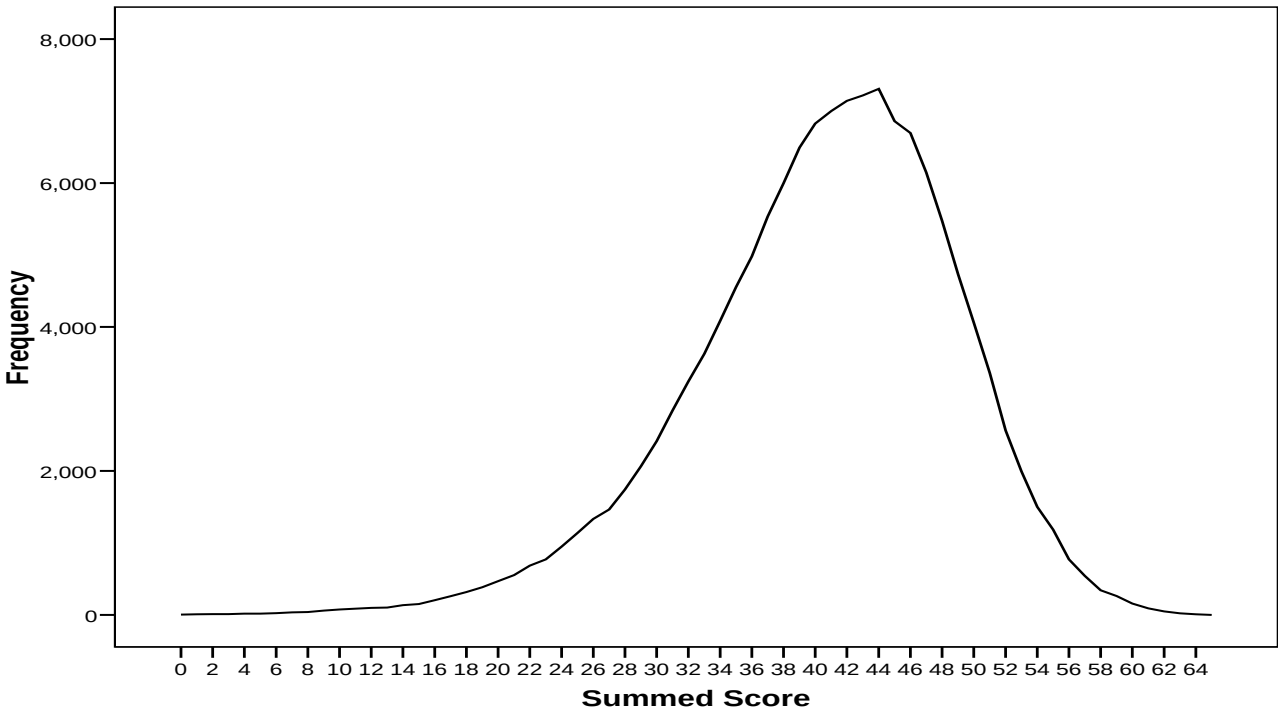


Figure 7.1.4 Score Distribution for Primary Reading (French)

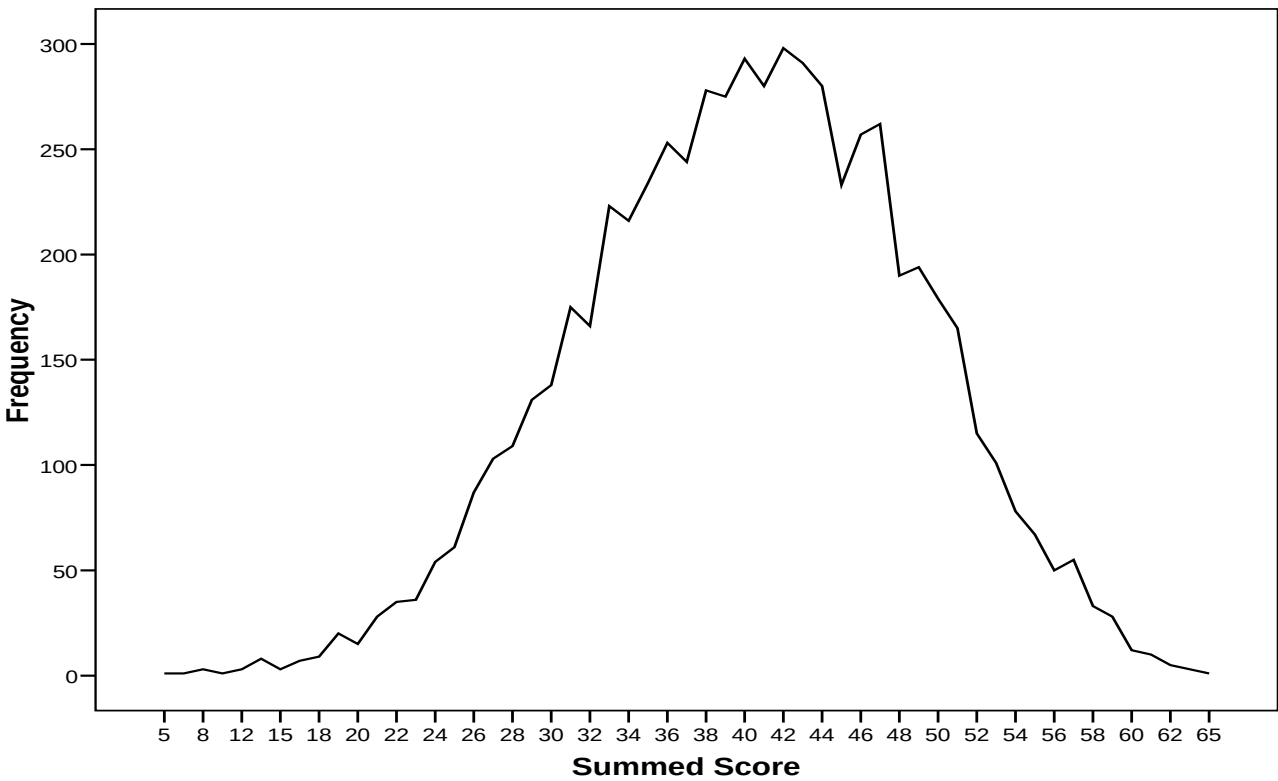


Figure 7.1.5 Score Distribution for Primary Writing (English)

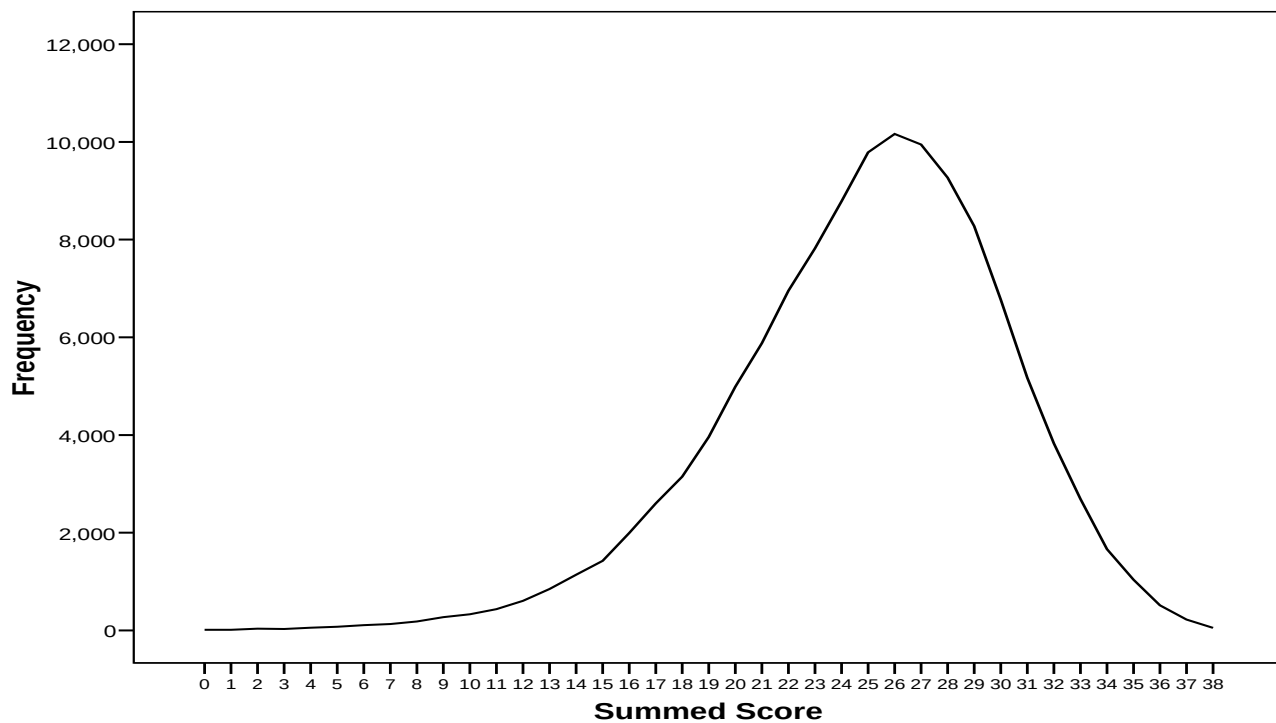


Figure 7.1.6 Score Distribution for Primary Writing (French)

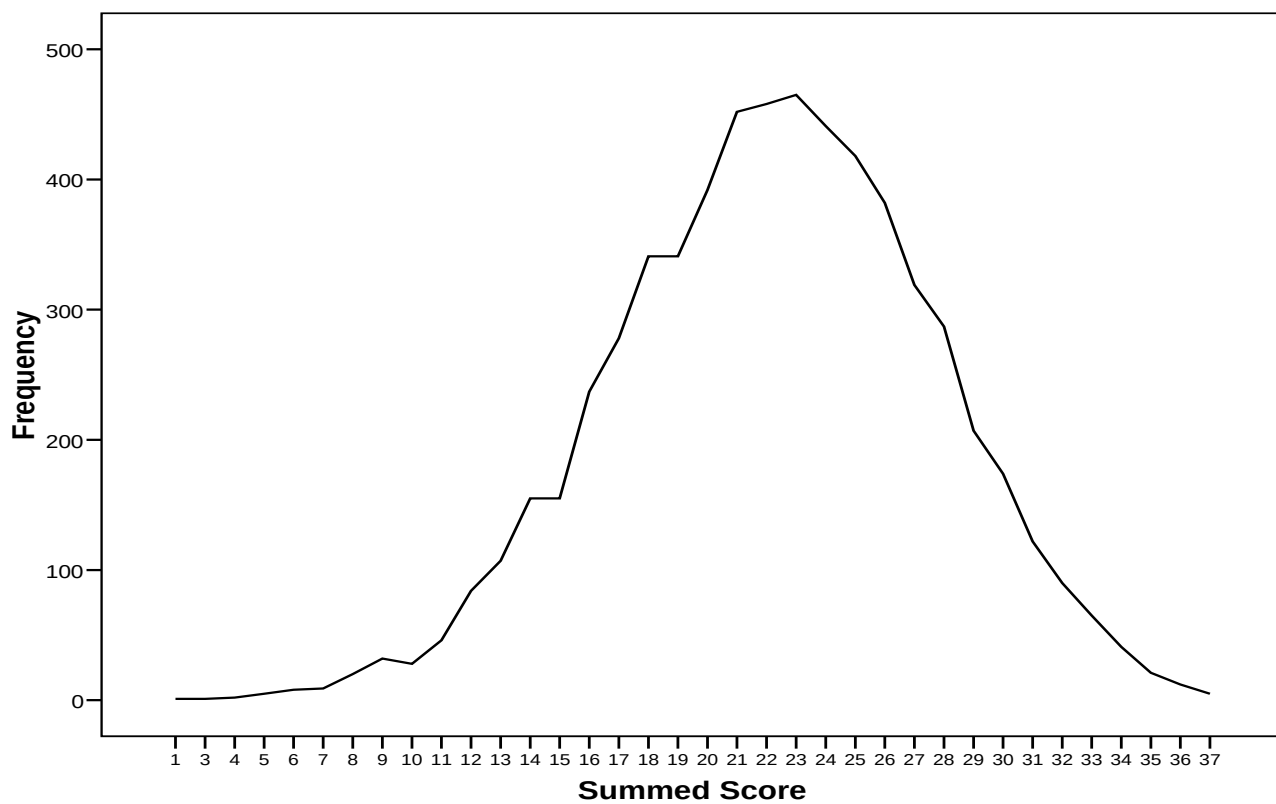


Figure 7.1.7 Score Distribution for Junior Writing (English)

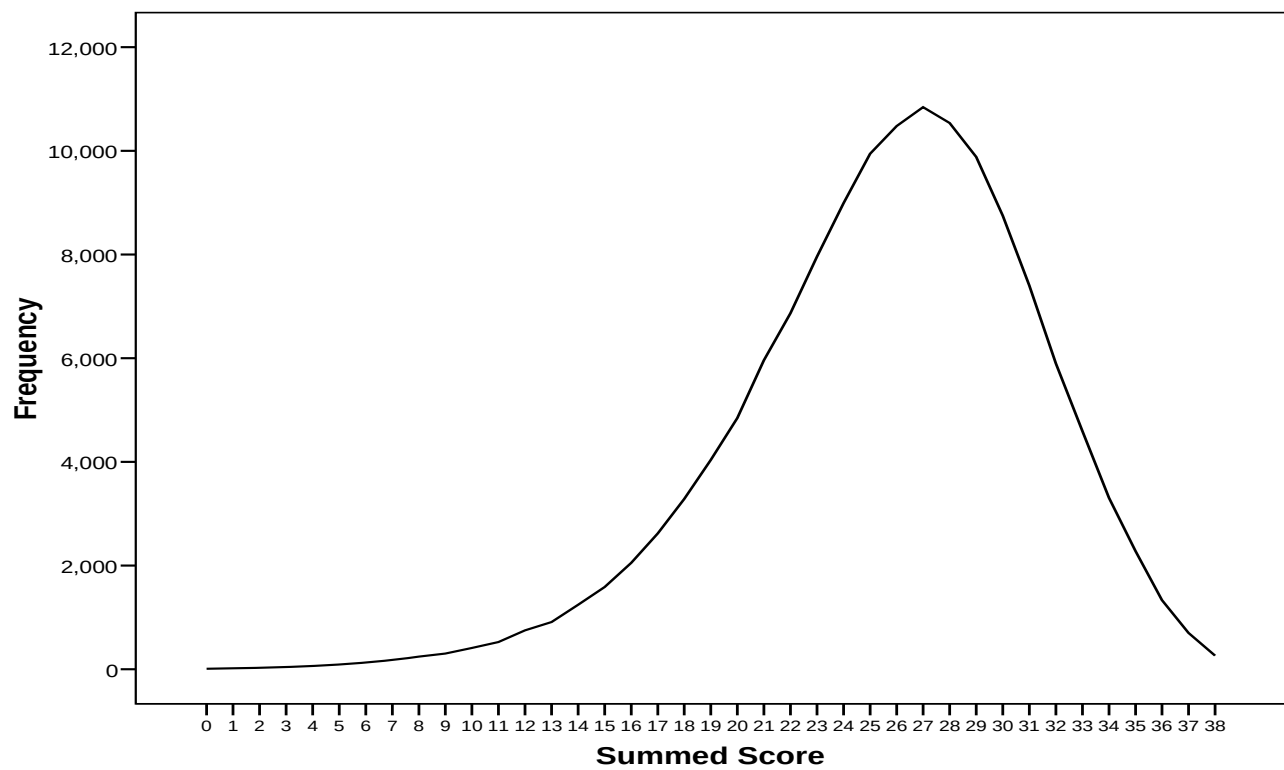


Figure 7.1.8 Score Distribution for Junior Writing (French)

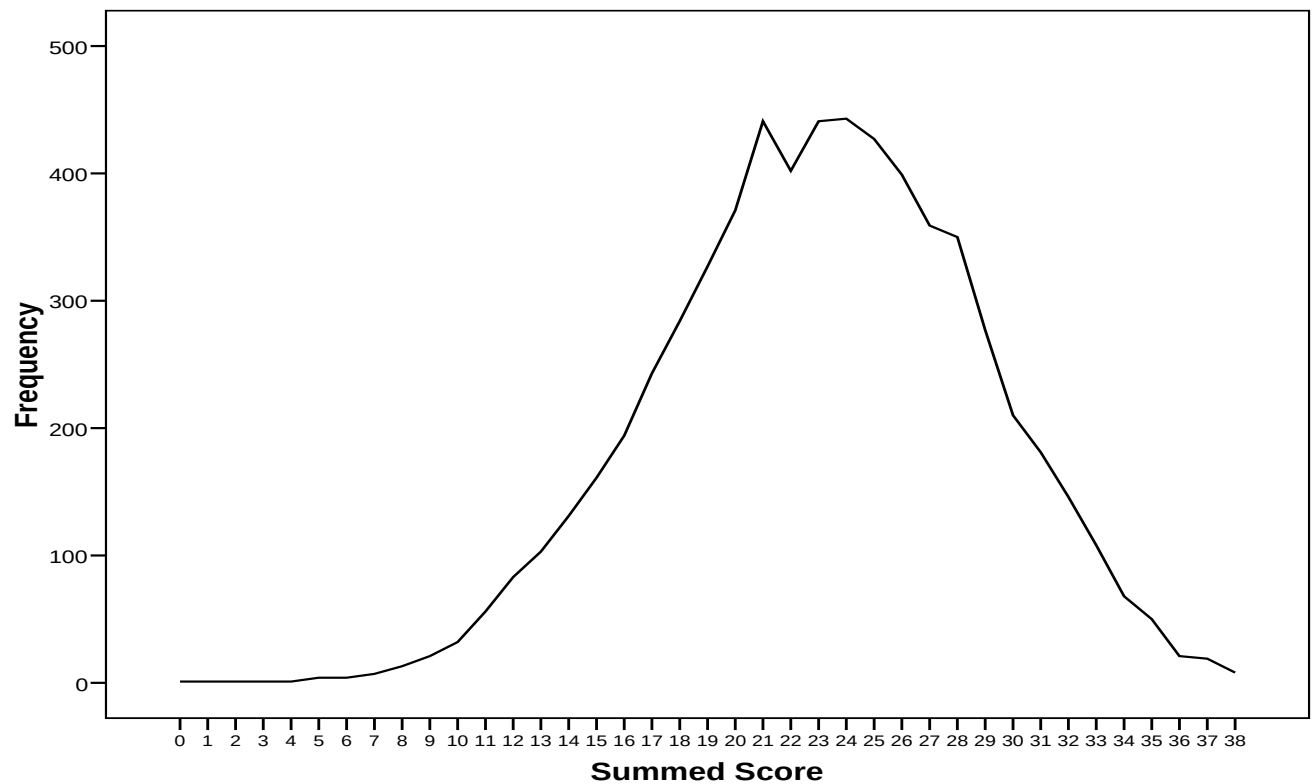


Figure 7.1.9 Score Distribution for Primary Math (English)

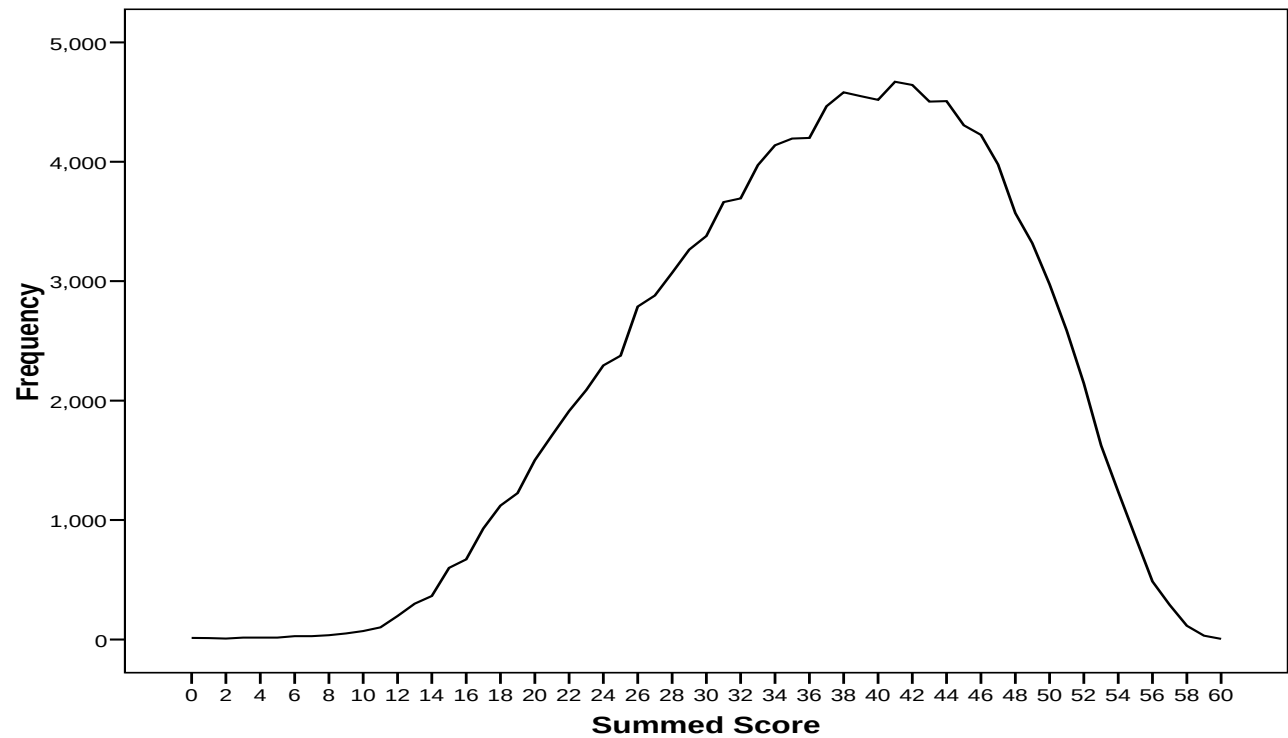


Figure 7.1.10 Score Distribution for Primary Math (French)

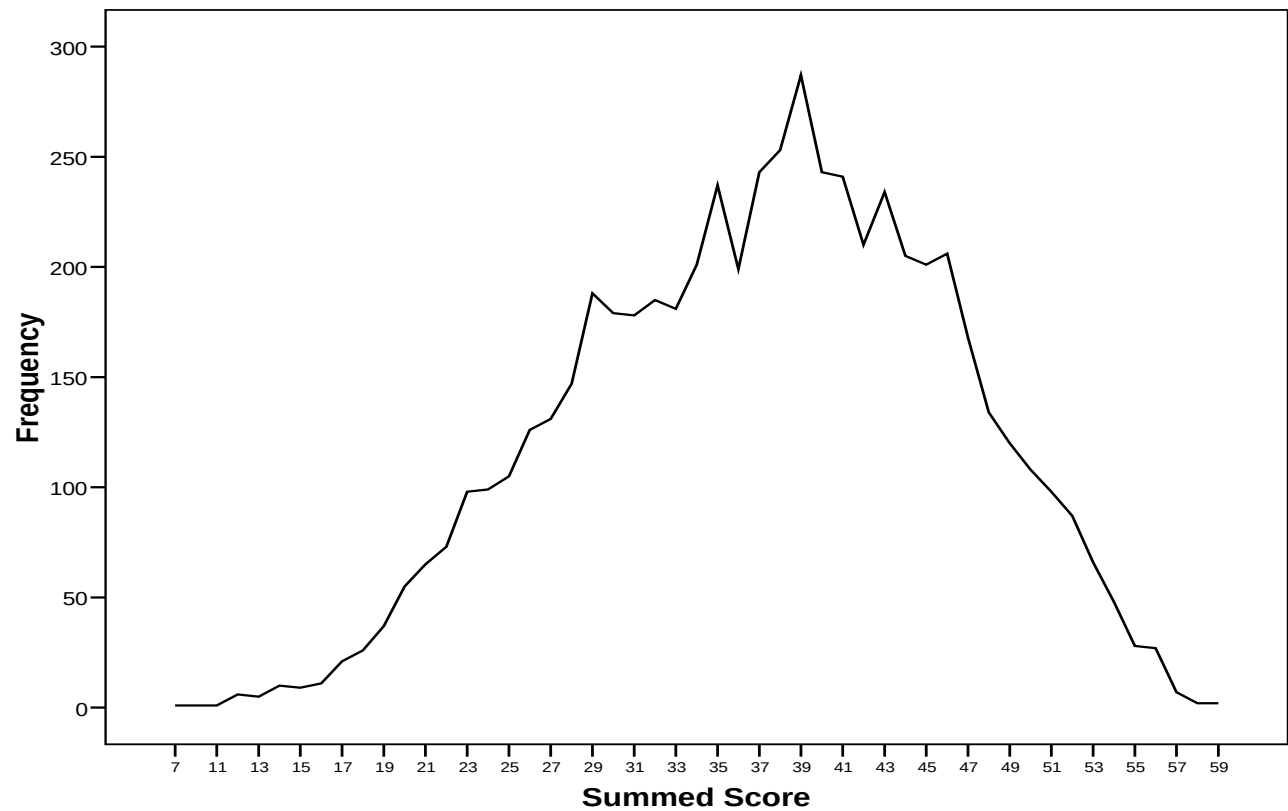


Figure 7.1.11 Score Distribution for Junior Math (English)

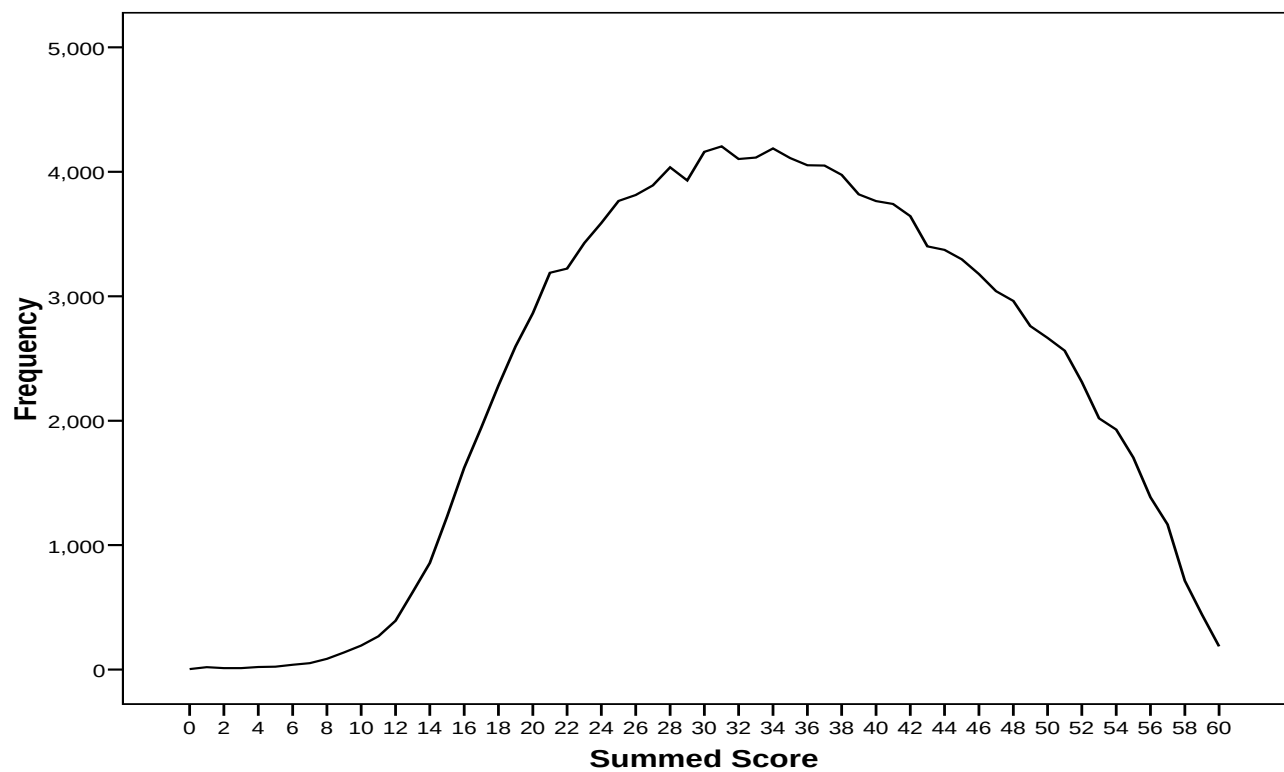
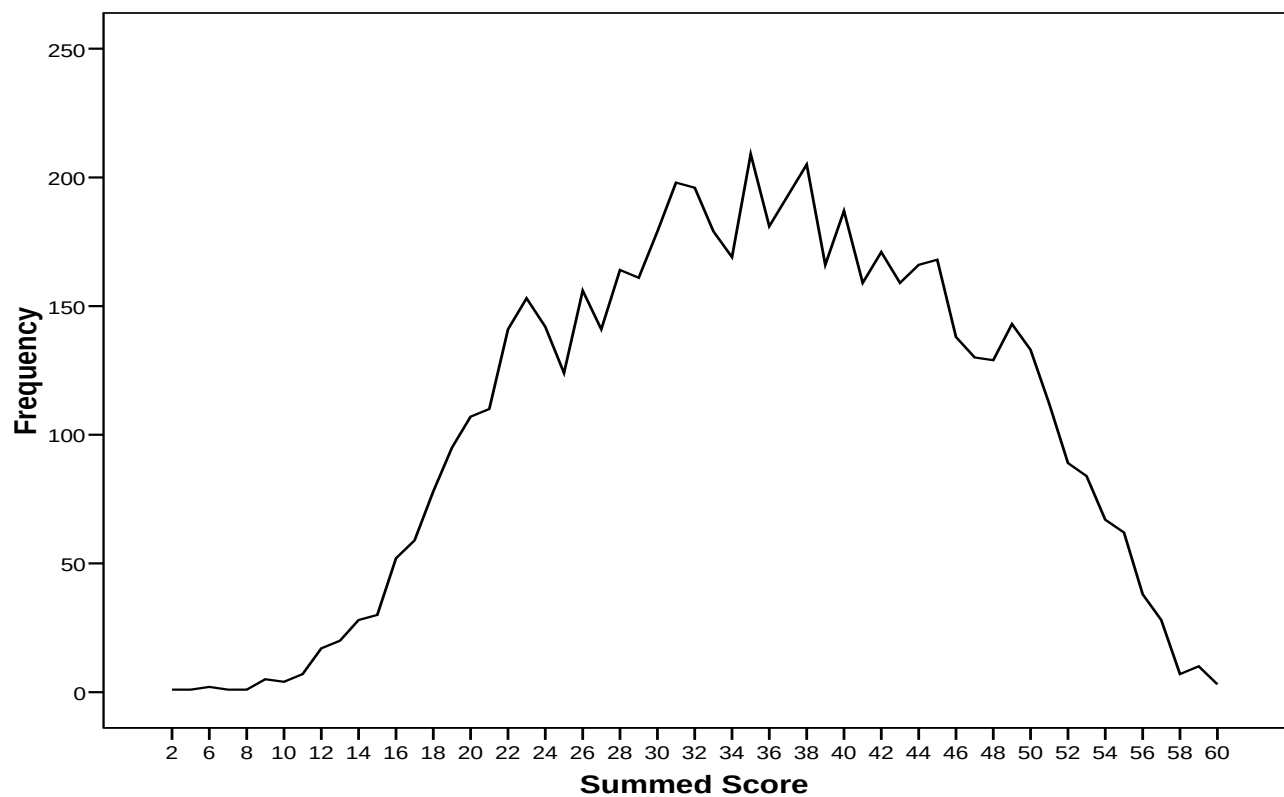


Figure 7.1.12 Score Distribution for Junior Math (French)



Classical Item Statistics and IRT Item Parameters

Table 7.1.1: Item Statistics – Primary Reading (English)

Item Code	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3441209	3.0	C	R	2	0.63	0.45	1.24	–0.17
3441180	1.0	I	R	1	0.80	0.18	0.34	–2.16
3445050	1.0	I	R	4	0.62	0.30	0.58	–0.17
3441191	2.0	I	R	4	0.62	0.37	0.77	–0.13
3443552	2.0	I	R	1	0.66	0.30	0.56	–0.44
3443580	3.0	E	R	4	0.96	0.36	1.21	–2.33
3443577	1.0	C	R	2	0.59	0.31	0.55	–0.02
36005	3.0	C	R	1	0.66	0.45	1.07	–0.34
3442039	2.0	C	R	4*	0.52 (2.08)	0.41	0.38	–0.13
3443776	1.0	C	R	4*	0.48 (1.94)	0.43	0.41	–0.10
3443557	1.0	E	R	4*	0.60 (2.39)	0.48	0.44	–0.95
3445021	1.0	C	R	4*	0.52 (2.07)	0.46	0.43	–0.45
3441530	1.0	I	R	4*	0.56 (2.26)	0.48	0.38	–0.80
3441234	1.0	C	R	4*	0.42 (1.70)	0.45	0.43	0.07
3441230	1.0	E	R	1	0.84	0.34	0.65	–1.65
3442026	3.0	I	R	4	0.87	0.45	1.14	–1.41
3444164	3.0	C	R	4	0.67	0.41	0.87	–0.37
3441481	2.0	C	R	3	0.59	0.24	0.39	0.05
3441233	1.0	I	R	1	0.75	0.44	0.96	–0.75
3442037	3.0	C	R	1	0.65	0.35	0.67	–0.34
3445712	1.0	C	R	4	0.83	0.43	1.02	–1.17
3442028	1.0	I	R	2	0.85	0.43	0.96	–1.42
3442029	1.0	I	R	4	0.77	0.32	0.55	–1.20
3441432	1.0	I	R	3	0.77	0.40	0.80	–0.97
3442545	2.0	C	R	1	0.58	0.34	0.62	0.02
3442044	3.0	I	R	3	0.71	0.52	1.41	–0.52
3443515	3.0	E	R	1	0.84	0.37	0.71	–1.59
3441223	1.0	I	R	3	0.81	0.47	1.17	–1.04
3443861	2.0	I	R	1	0.95	0.37	1.11	–2.29
3441147	1.0	C	R	3	0.48	0.34	0.81	0.55
3441146	3.0	C	R	4	0.68	0.40	0.82	–0.44
3441145	3.0	I	R	3	0.48	0.34	0.78	0.54
3441224	2.0	I	R	4*	0.40 (1.59)	0.38	0.36	0.40
3445113	1.0	C	R	4*	0.56 (1.26)	0.49	0.41	–0.60
3441157	1.0	C	R	4*	0.49 (1.96)	0.51	0.45	0.05
3443855	1.0	C	R	4*	0.58 (2.33)	0.55	0.47	–0.48

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

C = Connections I = Implicit
E = Explicit R = Reading

Table 7.1.2: Item Statistics – Primary Reading (French)

Item Code	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3462588	B	C	R	2	0.50	0.17	0.32	1.05
3462537	B	I	R	4	0.90	0.29	0.70	-2.17
3462533	C	I	R	2	0.57	0.32	0.60	0.15
3462529	A	I	R	1	0.57	0.38	0.82	0.11
3460766	B	C	R	2	0.45	0.34	0.77	0.77
3460758	A	I	R	3	0.68	0.41	0.86	-0.42
3460784	C	E	R	4	0.73	0.34	0.67	-0.76
36010	C	C	R	4	0.48	0.18	0.35	1.13
3462590	A	C	R	4*	0.49 (1.94)	0.43	0.49	-0.35
3462591	A	C	R	4*	0.44 (1.75)	0.47	0.49	-0.21
3460797	A	E	R	4*	0.63 (2.51)	0.47	0.32	-2.00
3460791	A	C	R	4*	0.52 (2.09)	0.50	0.53	-0.42
3463315	B	I	R	4*	0.44 (1.77)	0.47	0.44	-0.19
3463301	A	C	R	4*	0.62 (2.47)	0.49	0.44	-1.38
3463339	B	C	R	3	0.67	0.39	0.72	-0.43
3470166	C	C	R	4	0.67	0.33	0.59	-0.48
3463038	C	I	R	4	0.48	0.21	0.35	1.05
3463316	A	E	R	2	0.93	0.28	0.71	-2.52
3463036	C	I	R	2	0.72	0.43	0.91	-0.67
3463348	C	I	R	2	0.93	0.27	0.68	-2.55
3470176	C	C	R	4	0.64	0.26	0.40	-0.42
3470171	A	I	R	3	0.70	0.32	0.56	-0.69
3463325	A	I	R	1	0.92	0.29	0.75	-2.27
3481932	C	C	R	4	0.76	0.40	0.79	-0.91
3461610	A	I	R	1	0.62	0.11	0.16	-0.54
36008	C	E	R	4	0.74	0.28	0.47	-1.11
36009	C	C	R	2	0.43	0.23	0.46	1.24
3461613	A	C	R	2	0.49	0.32	0.69	0.58
3461831	B	I	R	3	0.75	0.35	0.64	-0.95
3461832	C	I	R	1	0.61	0.19	0.28	-0.22
3461807	C	C	R	4	0.78	0.36	0.69	-1.16
3461809	C	I	R	4	0.69	0.38	0.77	-0.50
3462264	A	I	R	4*	0.48 (1.91)	0.52	0.55	-0.27
3461866	A	C	R	4*	0.56 (2.23)	0.50	0.58	-0.72
3461838	A	C	R	4*	0.53 (2.10)	0.47	0.53	0.91
3461837	A	C	R	4*	0.54 (2.14)	0.52	0.58	-0.47

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

C = Connections

I = Implicit

E = Explicit

R = Reading

Table 7.1.3: Item Statistics – Junior Reading (English)

Item Code	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3444788	1.0	E	R	4	0.98	0.16	0.57	-4.52
3444790	1.0	I	R	1	0.70	0.21	0.32	-1.11
3444809	3.0	C	R	4	0.62	0.33	0.64	-0.21
3444798	2.0	I	R	1	0.52	0.24	0.39	0.52
3444048	1.0	I	R	4	0.79	0.27	0.47	-1.53
3444089	3.0	I	R	2	0.86	0.29	0.57	-1.95
3444084	2.0	I	R	3	0.66	0.28	0.47	-0.57
3445013	1.0	C	R	2	0.51	0.31	0.65	0.45
3444793	1.0	I	R	4*	0.58 (2.32)	0.43	0.41	-0.82
3444795	1.0	C	R	4*	0.57 (2.29)	0.48	0.47	-0.85
3444077	1.0	C	R	4*	0.42 (1.70)	0.46	0.45	0.21
3444063	1.0	C	R	4*	0.44 (1.75)	0.49	0.53	0.21
3443953	1.0	I	R	4*	0.59 (2.36)	0.45	0.44	-0.90
3443896	1.0	C	R	4*	0.54 (2.15)	0.51	0.50	-0.26
3443878	2.0	I	R	1	0.67	0.38	0.79	-0.43
3443873	1.0	E	R	2	0.78	0.42	0.89	-0.95
3443935	1.0	I	R	1	0.66	0.41	0.80	-0.39
3443984	3.0	E	R	1	0.58	0.39	0.83	0.04
3443876	1.0	E	R	2	0.70	0.29	0.48	-0.79
3443989	3.0	I	R	1	0.82	0.29	0.47	-1.90
3443884	1.0	I	R	1	0.84	0.16	0.24	-3.67
3443943	1.0	I	R	3	0.84	0.28	0.45	-2.19
3443881	1.0	I	R	1	0.82	0.38	0.71	-1.46
3443964	2.0	I	R	2	0.68	0.29	0.45	-0.73
3444078	3.0	C	R	4	0.87	0.35	0.73	-1.83
3444035	1.0	E	R	3	0.80	0.25	0.41	-1.83
3444045	1.0	I	R	2	0.79	0.28	0.46	-1.55
3444895	2.0	C	R	2	0.88	0.31	0.56	-2.29
3444900	1.0	I	R	2	0.54	0.26	0.48	0.32
3444890	1.0	I	R	4	0.88	0.38	0.81	-1.80
3444899	3.0	I	R	3	0.69	0.41	0.85	-0.54
3444082	1.0	C	R	4*	0.62 (2.49)	0.41	0.42	-1.00
3444069	1.0	C	R	4*	0.61 (2.45)	0.43	0.41	-0.84
3444915	1.0	I	R	4*	0.56 (2.22)	0.52	0.51	-0.57
3444903	1.0	C	R	4*	0.58 (2.32)	0.53	0.59	-0.67

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

C = Connections

I = Implicit

E = Explicit

R = Reading

Table 7.1.4: Item Statistics – Junior Reading (French)

Item Code	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3463552	A	E	R	4	0.97	0.21	0.84	-3.09
3463479	A	I	R	3	0.59	0.30	0.53	0.04
3463504	A	I	R	4	0.69	0.15	0.24	-1.21
3463550	C	C	R	4	0.66	0.32	0.56	-0.43
3463646	A	E	R	4	0.92	0.20	0.50	-3.07
3463683	A	I	R	3	0.70	0.25	0.45	-0.81
3463685	B	I	R	3	0.82	0.24	0.43	-1.96
3463642	C	I	R	4	0.58	0.34	0.68	0.07
3463559	A	I	R	4*	0.59 (2.36)	0.46	0.41	-0.79
3463566	A	C	R	4*	0.59 (2.38)	0.55	0.57	-0.79
3463692	A	C	R	4*	0.54 (2.17)	0.54	0.62	-0.62
3463650	A	C	R	4*	0.60 (2.41)	0.55	0.67	-0.96
3469963	A	I	R	4*	0.49 (1.98)	0.39	0.35	-0.86
3463631	A	C	R	4*	0.49 (1.97)	0.45	0.44	-0.36
3463672	A	I	R	2	0.57	0.19	0.32	0.21
3470217	A	E	R	1	0.74	0.34	0.60	-0.91
3463680	A	E	R	4	0.52	0.31	0.54	0.47
3463627	B	I	R	2	0.80	0.33	0.64	-1.27
3482781	A	I	R	4	0.42	0.35	0.77	0.93
3463603	A	E	R	4	0.91	0.32	0.82	-2.02
3481897	A	I	R	1	0.82	0.22	0.36	-2.20
3463630	C	C	R	3	0.46	0.29	0.58	0.83
3482782	A	E	R	3	0.65	0.24	0.36	-0.47
3463695	A	I	R	1	0.95	0.24	0.66	-2.86
3463021	B	I	R	4	0.54	0.31	0.58	0.30
3463023	A	E	R	2	0.73	0.36	0.70	-0.75
3463020	B	C	R	1	0.43	0.33	0.75	0.90
3462626	C	C	R	2	0.50	0.31	0.63	0.55
3463727	C	I	R	4	0.83	0.38	0.79	-1.36
3463728	C	I	R	2	0.60	0.14	0.22	-0.02
3463729	B	C	R	4	0.67	0.31	0.51	-0.53
3463764	A	I	R	1	0.47	0.26	0.51	0.85
3462630	B	C	R	4*	0.64 (2.56)	0.51	0.54	-1.12
3462631	A	C	R	4*	0.59 (2.34)	0.55	0.62	-0.76
3463730	A	C	R	4*	0.61 (2.45)	0.50	0.54	-1.13
3482792	B	I	R	4*	0.51 (2.05)	0.47	0.51	-0.46

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

C = Connections

I = Implicit

E = Explicit

R = Reading

Table 7.1.5: Item Statistics – Primary Writing (English)

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3356057	1.0	W	T	6*	0.58 (3.54)	0.55	0.51	-0.57
3356057	3.0	W	V	4*	0.74 (2.99)	0.59	0.84	-1.76
3353030	1.0	W	T	6*	0.56 (3.40)	0.52	0.48	-0.59
3353030	3.0	W	V	4*	0.69 (2.78)	0.58	0.89	-1.46
EQWE10031	1.0	W	T	6*	0.57 (3.43)	0.58	0.60	-0.62
EQWE10031	3.0	W	V	4*	0.68 (2.72)	0.62	0.97	-1.25
3352934	3.0	W	V	4	0.76	0.34	0.62	-1.08
3352857	3.0	W	V	2	0.87	0.35	0.78	-1.68
3352933	2.0	W	V	2	0.65	0.27	0.45	-0.45
3352963	3.0	W	V	3	0.70	0.23	0.36	-0.98
3352803	3.0	W	V	2	0.76	0.36	0.69	-0.96
3352889	3.0	W	V	2	0.75	0.34	0.64	-0.94
3352858	3.0	W	V	4	0.85	0.35	0.75	-1.57
3352794	2.0	W	V	1	0.70	0.26	0.44	-0.85

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

T = Content V = Conventions W = Writing

Table 7.1.6: Item Statistics – Primary Writing (French)

Item Code	Expectation	Strand	Cognitive skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
18124	A	W	T	6*	0.38 (2.30)	0.44	0.32	0.34
18124	C	W	V	4*	0.63 (2.53)	0.58	0.68	-1.17
18004	A	W	T	6*	0.44 (2.64)	0.49	0.37	0.07
18004	C	W	V	4*	0.67 (2.69)	0.57	0.68	-1.55
OQEF10019	A	W	T	6*	0.49 (2.94)	0.58	0.58	-0.16
OQEF10019	C	W	V	4*	0.64 (2.60)	0.60	0.89	-1.14
3462475	C	W	V	2	0.84	0.29	0.60	-1.68
3462525	C	W	V	1	0.82	0.26	0.45	-1.88
3462548	C	W	V	2	0.67	0.22	0.36	-0.62
3462569	C	W	V	1	0.85	0.30	0.64	-1.65
3462500	C	W	V	2	0.89	0.26	0.56	-2.33
3462521	C	W	V	1	0.82	0.30	0.58	-1.50
3462575	C	W	V	1	0.88	0.25	0.53	-2.29
3462570	C	W	V	4	0.78	0.19	0.31	-1.99

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

T = Content V = Conventions W = Writing

Table 7.1.7: Item Statistics – Junior Writing (English)

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3356065	1.0	W	T	6*	0.67 (4.00)	0.56	0.52	–0.97
3356065	3.0	W	V	4*	0.77 (3.08)	0.56	0.77	–2.07
3352983	1.0	W	T	6*	0.56 (3.38)	0.57	0.53	–0.58
3352983	3.0	W	V	4*	0.71 (2.83)	0.63	0.94	–1.56
EQWE10069	1.0	W	T	6*	0.57 (3.43)	0.60	0.52	–0.41
EQWE10069	3.0	W	V	4*	0.75 (3.01)	0.61	0.86	–1.41
3444571	3.0	W	V	3	0.80	0.30	0.54	–1.50
3444272	3.0	W	V	2	0.68	0.28	0.48	–0.62
3444638	3.0	W	V	1	0.76	0.30	0.51	–1.26
3444777	2.0	W	V	4	0.81	0.33	0.61	–1.45
3444563	3.0	W	V	4	0.75	0.36	0.66	–0.99
3444431	3.0	W	V	3	0.49	0.27	0.53	0.67
3444255	3.0	W	V	2	0.76	0.24	0.39	–1.45
3444211	2.0	W	V	1	0.84	0.33	0.64	–1.65

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

T = Content V = Conventions W = Writing

Table 7.1.8: Item Statistics – Junior Writing (French)

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
36045	A	W	T	6*	0.51 (3.07)	0.52	0.40	–0.68
36045	C	W	V	4*	0.72 (2.86)	0.60	0.80	–2.04
18336	A	W	T	6*	0.55 (3.29)	0.57	0.52	–0.85
18336	C	W	V	4*	0.68 (2.71)	0.63	0.98	–1.50
OQEF10082	A	W	T	6*	0.60 (3.58)	0.59	0.58	–0.87
OQEF10082	C	W	V	4*	0.67 (2.68)	0.65	1.04	–1.19
3462997	C	W	V	4	0.51	0.36	0.73	0.43
3462611	C	W	V	2	0.15	0.16	0.75	2.60
3462994	C	W	V	4	0.72	0.34	0.63	–0.75
3462900	C	W	V	4	0.73	0.17	0.26	–1.65
3462642	C	W	V	4	0.75	0.26	0.44	–1.19
3462991	C	W	V	3	0.55	0.25	0.45	0.33
3463004	C	W	V	3	0.74	0.25	0.43	–1.11
3462836	C	W	V	2	0.73	0.28	0.49	–0.96

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

T = Content V = Conventions W = Writing

Table 7.1.9: Item Statistics – Primary Mathematics (English)

Item Code	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3459511	PV2	AP	P	4	0.66	0.30	0.52	–0.45
3459114	NV2	AP	N	2	0.91	0.25	0.60	–2.49
3459363	NV3	PS	N	1	0.46	0.30	0.66	0.80
3459309	NV3	KU	N	3	0.80	0.37	0.73	–1.21
17135	NV3	KU	N	2	0.77	0.32	0.57	–1.21
3459515	PV2	PS	P	1	0.38	0.13	0.38	2.10
3459405	DV3	AP	D	3	0.70	0.38	0.71	–0.57
3459028	MV1	PS	M	2	0.53	0.31	0.68	0.41
3459018	MV1	AP	M	3	0.47	0.29	0.59	0.81
3459406	GV1	KU	G	1	0.74	0.34	0.62	–0.88
3459421	GV2	KU	G	4	0.54	0.29	0.54	0.35
3459059	MV2	AP	M	3	0.57	0.41	0.90	0.11
3459032	MV2	KU	M	1	0.82	0.35	0.70	–1.40
3459509	PV2	KU	P	3	0.76	0.31	0.57	–1.10
3459502	PV1	AP	P	2	0.68	0.36	0.68	–0.45
3459004	MV1	KU	M	3	0.74	0.26	0.44	–1.14
3459040	MV2	AP	M	4	0.19	0.24	0.93	1.84
3459029	NV1	PS	N	3	0.46	0.38	1.03	0.62
3459019	NV1	KU	N	2	0.73	0.36	0.67	–0.83
3459487	GV3	AP	G	1	0.46	0.25	0.53	0.92
3459410	GV1	AP	G	2	0.77	0.36	0.71	–1.01
3459396	DV2	KU	D	1	0.58	0.28	0.46	0.03
3459339	NV3	AP	N	1	0.68	0.37	0.75	–0.44
3459076	MV2	KU	M	1	0.61	0.42	0.90	–0.10
3459027	MV1	AP	M	2	0.62	0.39	0.72	–0.15
3459395	DV1	AP	D	3	0.45	0.39	1.13	0.64
3459480	PV1	KU	P	4	0.61	0.37	0.69	–0.11
3459496	PV1	KU	P	3	0.76	0.37	0.73	–0.93
3459457	DV3	PS	D	4*	0.58 (2.32)	0.52	0.42	–1.51
3459370	NV1	AP	N	4*	0.67 (2.69)	0.55	0.39	–1.46
3459494	GV3	AP	G	4*	0.59 (2.34)	0.44	0.38	–0.80
3459348	MV2	AP	M	4*	0.78 (3.11)	0.50	0.35	–2.52
3459419	DV2	AP	D	4*	0.60 (2.40)	0.47	0.41	–1.05
3459545	PV1	AP	P	4*	0.60 (2.39)	0.55	0.40	–1.37
3459477	GV2	PS	G	4*	0.41 (1.64)	0.33	0.29	–0.16
17054	NV3	PS	N	4*	0.66 (2.64)	0.58	0.49	–1.25

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

AP = Application

KU = Knowledge and Understanding

N = Number Sense and Numeration

M = Measurement

PS = Problem Solving

P = Patterning and Algebra

D = Data Management and Probability

G = Geometry and Spatial Sense

Table 7.1.10: Item Statistics – Primary Mathematics (French)

Item Code	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3463878	NV4	AP	N	3	0.93	0.17	0.42	–3.61
3463841	NV1	KU	N	1	0.89	0.31	0.79	–1.78
3463782	NV2	PS	N	3	0.55	0.43	1.11	0.16
3463881	NV1	AP	N	4	0.66	0.31	0.57	–0.39
3463842	NV3	AP	N	4	0.83	0.34	0.80	–1.34
3463798	MV1	AP	M	3	0.77	0.36	0.74	–0.95
3462508	GV1	PS	G	4	0.49	0.30	0.62	0.60
3463753	DV1	PS	D	2	0.72	0.31	0.58	–0.81
3462513	GV2	AP	G	3	0.87	0.27	0.59	–2.05
3463787	MV3	KU	M	4	0.49	0.25	0.51	0.75
3463874	PV2	KU	P	3	0.76	0.36	0.77	–0.89
3463796	MV2	AP	M	2	0.45	0.42	1.20	0.58
3463679	DV1	KU	D	4	0.60	0.36	0.72	–0.04
3463864	PV1	KU	P	4	0.81	0.30	0.58	–1.44
3463885	NV3	KU	N	4	0.85	0.28	0.56	–1.84
3463827	NV3	PS	N	4	0.81	0.37	0.88	–1.15
3463888	NV2	AP	N	2	0.57	0.32	0.69	0.15
3463865	PV3	AP	P	2	0.65	0.28	0.49	–0.34
3463694	DV2	AP	D	3	0.74	0.28	0.50	–1.02
3462499	GV1	AP	G	4	0.68	0.05	0.09	–2.78
3463867	PV2	PS	P	2	0.62	0.43	0.93	–0.16
3463802	MV2	KU	M	3	0.81	0.34	0.74	–1.22
3463797	MV1	PS	M	2	0.48	0.32	0.81	0.62
3463791	DV1	KU	M	3	0.55	0.32	0.68	0.26
3463654	GV1	AP	D	2	0.54	0.30	0.60	0.31
3462490	NV4	KU	G	2	0.57	0.12	0.18	0.48
3463889	PV3	KU	N	3	0.68	0.41	0.88	–0.41
3463890	NV3	PS	N	4*	0.53 (2.13)	0.47	0.33	–0.82
3463353	GV1	PS	G	4*	0.64 (2.54)	0.34	0.19	–2.91
3463877	PV1	PS	P	4*	0.43 (1.71)	0.33	0.21	–1.74
3463779	DV1	AP	D	4*	0.53 (2.14)	0.38	0.33	–0.87
3463818	NV1	AP	N	4*	0.67 (2.67)	0.46	0.29	–2.15
3463360	GV1	AP	G	4*	0.76 (3.05)	0.24	0.16	–4.06
3463804	MV3	PS	M	4*	0.51 (2.06)	0.44	0.29	–1.31
3463769	DV1	PS	D	4*	0.64 (2.57)	0.47	0.32	–1.70

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

AP = Application

KU = Knowledge and Understanding

N = Number Sense and Numeration

M = Measurement

PS = Problem Solving

P = Patterning and Algebra

D = Data Management and Probability

G = Geometry and Spatial Sense

Table 7.1.11: Item Statistics – Junior Mathematics (English)

Item Code	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3459401	PV1	AP	P	3	0.68	0.35	0.60	−0.62
3459671	PV2	AP	P	3	0.83	0.32	0.66	−1.49
3459725	MV2	PS	M	4	0.26	0.44	1.68	1.23
3459741	MV1	KU	M	2	0.69	0.35	0.63	−0.57
3478223	NV2	AP	N	3	0.60	0.36	0.66	−0.09
3459755	NV3	KU	N	3	0.73	0.42	0.93	−0.69
3459740	DV3	AP	D	2	0.82	0.36	0.80	−1.32
3469904	DV2	KU	D	2	0.56	0.48	1.12	0.07
3458687	GV1	KU	G	3	0.80	0.33	0.67	−1.29
3459412	GV3	PS	G	2	0.55	0.32	0.60	0.27
3459039	MV2	KU	M	3	0.43	0.38	1.02	0.72
3459397	MV1	AP	M	4	0.62	0.39	0.73	−0.17
3458668	PV2	KU	P	3	0.54	0.13	0.20	0.78
3458670	NV1	KU	N	3	0.52	0.34	0.72	0.40
3458660	NV1	PS	N	1	0.61	0.29	0.57	−0.06
3459760	GV3	AP	G	3	0.72	0.34	0.61	−0.84
3458981	GV2	PS	G	3	0.60	0.33	0.60	−0.05
3459407	PV1	KU	P	2	0.66	0.44	0.93	−0.35
3459756	PV1	PS	P	2	0.52	0.40	0.78	0.31
3459431	MV2	AP	M	4	0.28	0.40	1.31	1.28
3459418	MV2	PS	M	2	0.42	0.19	0.42	1.44
3458733	DV1	AP	D	3	0.74	0.34	0.62	−0.94
3459034	PV2	AP	P	3	0.84	0.40	1.00	−1.28
3469901	DV2	AP	D	2	0.61	0.33	0.57	−0.18
3459759	DV3	PS	D	3	0.63	0.30	0.50	−0.30
3459764	NV2	AP	N	3	0.61	0.46	0.94	−0.12
3459014	NV3	PS	N	3	0.35	0.45	1.41	0.92
3459682	MV2	KU	M	4	0.57	0.33	0.60	0.11
3459544	DV3	AP	D	4*	0.63 (2.53)	0.54	0.42	−1.32
3481990	NV3	PS	N	4*	0.69 (2.75)	0.61	0.49	−1.50
3459445	GV1	AP	G	4*	0.60 (2.38)	0.48	0.36	−0.88
3459773	MV2	PS	M	4*	0.44 (1.78)	0.61	0.52	−0.13
3459455	PV2	PS	P	4*	0.67 (2.69)	0.61	0.49	−1.22
3459654	DV2	PS	D	4*	0.48 (1.92)	0.63	0.53	−0.24
3459561	NV1	AP	N	4*	0.59 (2.35)	0.61	0.48	−1.07
3482402	GV3	PS	G	4*	0.45 (1.75)	0.57	0.57	0.06

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

AP = Application

KU = Knowledge and Understanding

N = Number Sense and Numeration

M = Measurement

PS = Problem Solving

P = Patterning and Algebra

D = Data Management and Probability

G = Geometry and Spatial Sense

Table 7.1.12: Item Statistics – Junior Mathematics (French)

Item Code	Expectation	Cognitive Skill	Strand	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Slope	Location
3464140	PV1	AP	P	3	0.84	0.36	0.87	−1.31
3463981	DV2	KU	D	2	0.74	0.36	0.72	−0.77
3463911	MV3	AP	M	2	0.43	0.24	0.65	1.05
3464084	GV2	PS	G	1	0.69	0.35	0.65	−0.53
3464121	NV3	AP	N	3	0.47	0.29	0.61	0.78
3463897	NV2	KU	N	2	0.75	0.42	0.94	−0.72
3464150	PV2	AP	P	4	0.39	0.23	0.52	1.44
3463902	MV1	PS	M	4	0.66	0.32	0.56	−0.41
3463920	NV1	KU	N	3	0.64	0.32	0.61	−0.30
3463977	DV1	AP	D	2	0.42	0.40	1.05	0.76
3464077	GV1	PS	G	3	0.51	0.25	0.46	0.64
3463907	MV2	PS	M	4	0.32	0.26	1.06	1.40
3463982	DV2	AP	D	4	0.70	0.40	0.77	−0.57
3464082	GV2	KU	G	2	0.81	0.26	0.49	−1.61
3464142	PV1	KU	P	2	0.75	0.41	0.88	−0.76
3463893	NV2	AP	N	1	0.67	0.35	0.66	−0.44
3463901	MV1	AP	M	2	0.54	0.37	0.79	0.29
3463979	DV2	PS	D	2	0.36	0.21	0.53	1.75
3464079	GV1	AP	G	4	0.58	0.45	1.05	0.08
3463905	MV1	KU	M	3	0.64	0.43	0.91	−0.20
3464066	NV1	AP	N	2	0.47	0.53	1.40	0.40
3464151	PV2	PS	P	1	0.74	0.45	1.01	−0.70
3463892	NV3	PS	N	4	0.49	0.40	0.85	0.50
3463903	MV1	KU	M	3	0.63	0.33	0.59	−0.23
3464087	GV2	KU	G	1	0.56	0.40	0.85	0.19
3463975	DV1	PS	D	2	0.53	0.42	0.95	0.28
3463906	MV2	KU	M	1	0.44	0.31	0.77	0.86
3464078	GV1	KU	G	3	0.60	0.39	0.78	−0.03
3463899	NV1	AP	N	4*	0.61 (2.45)	0.59	0.48	−1.09
3463916	MV1	AP	M	4*	0.64 (2.54)	0.57	0.42	−1.21
3464091	GV1	AP	G	4*	0.52 (2.06)	0.49	0.35	−1.34
3464159	PV2	AP	P	4*	0.59 (2.35)	0.48	0.42	−1.27
3464127	NV3	PS	N	4*	0.65 (2.59)	0.55	0.41	−1.28
3464092	GV2	AP	G	4*	0.66 (2.63)	0.51	0.37	−1.52
3464157	PV1	PS	P	4*	0.45 (1.82)	0.48	0.37	−0.76
3463987	DV1	PS	D	4*	0.67 (2.67)	0.41	0.29	−1.95

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

AP = Application

KU = Knowledge and Understanding

N = Number Sense and Numeration

M = Measurement

PS = Problem Solving

P = Patterning and Algebra

D = Data Management and Probability

G = Geometry and Spatial Sense

Grade 9 Assessment of Mathematics

**Figure 7.1.13 Score Distribution for Grade 9 Applied Mathematics,
Winter Administration (English)**

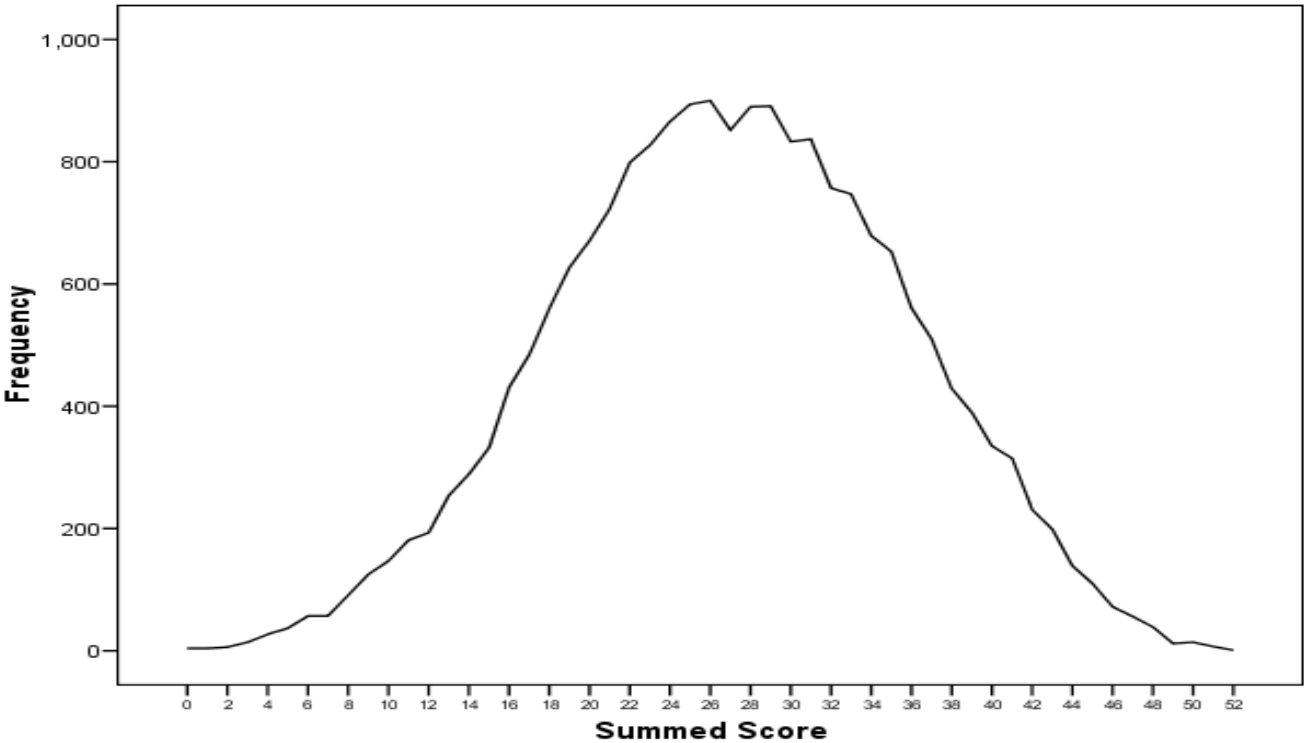


**Figure 7.1.14 Score Distribution for Grade 9 Applied Mathematics,
Spring Administration (English)**

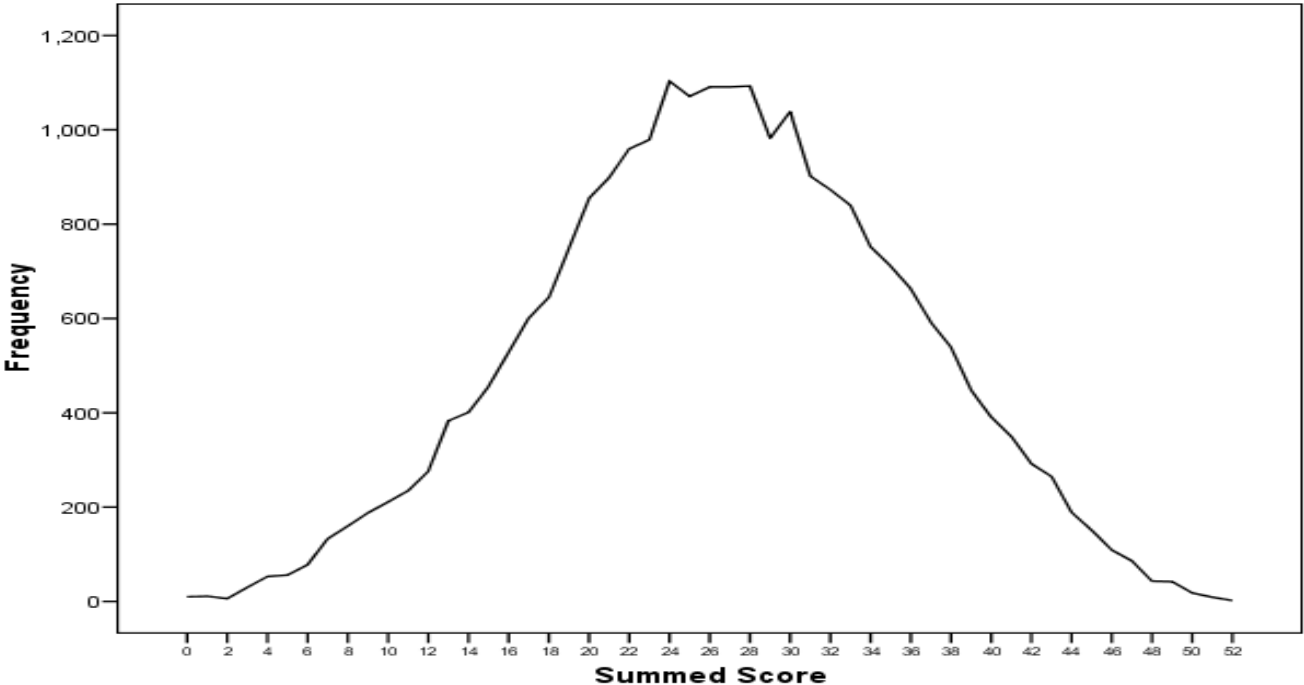


**Figure 7.1.15 Score Distribution for Grade 9 Academic Mathematics,
Winter Administration (English)**

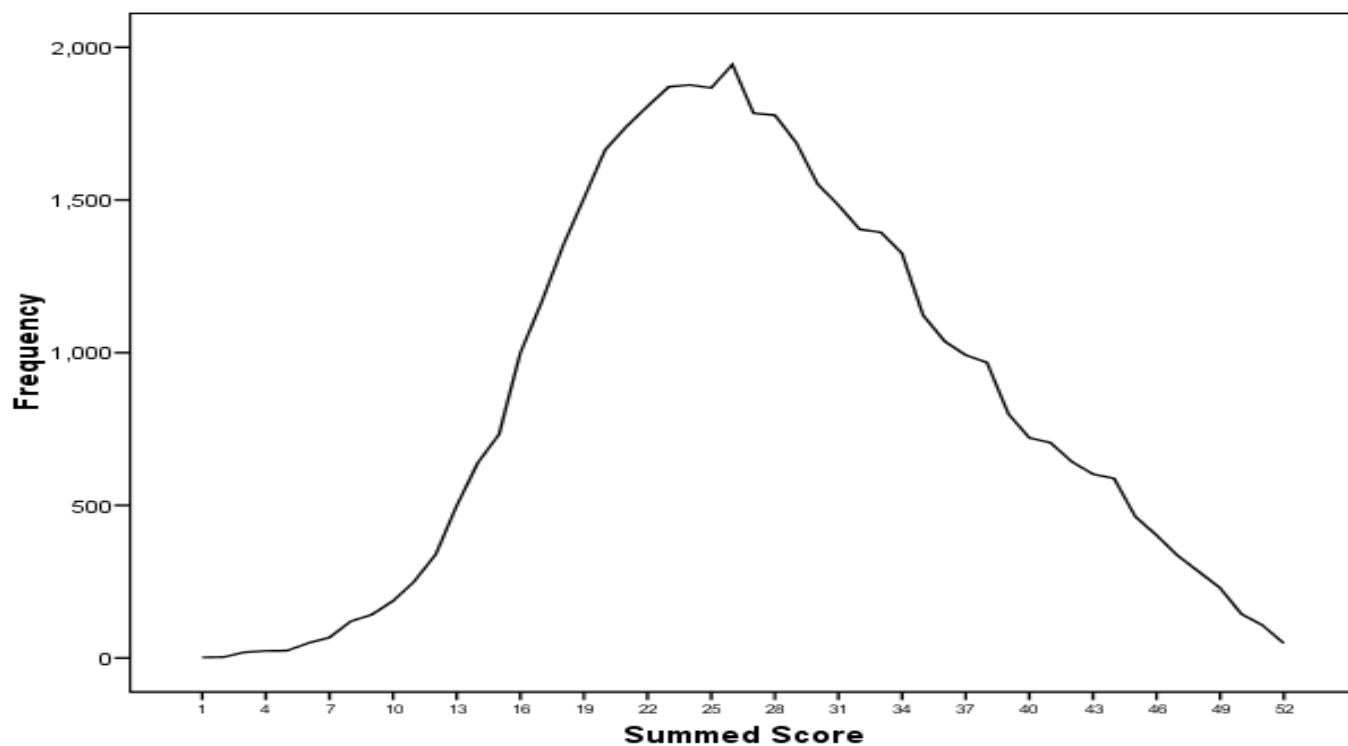


**Figure 7.1.16 Score Distribution for Grade 9 Academic Mathematics,
Spring Administration (English)**

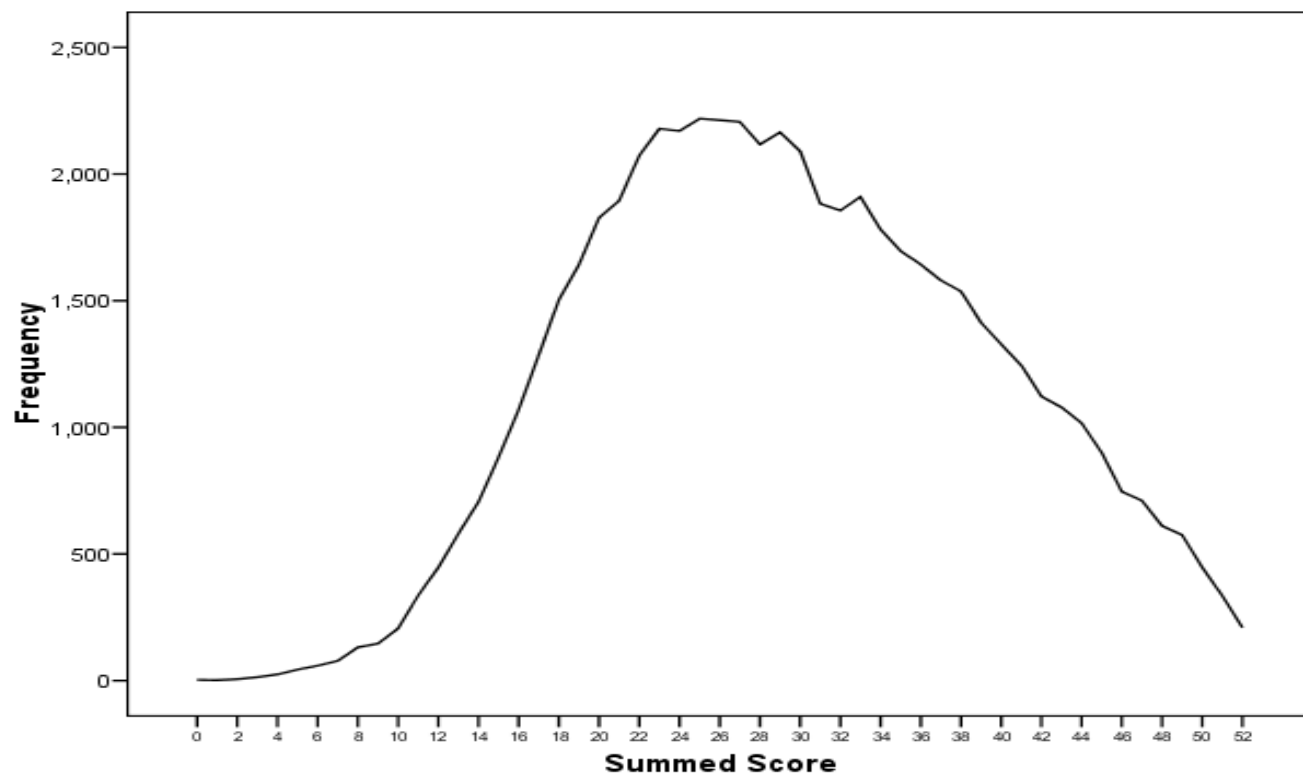


**Figure 7.1.17 Score Distribution for Grade 9 Applied Mathematics,
Winter Administration (French)**

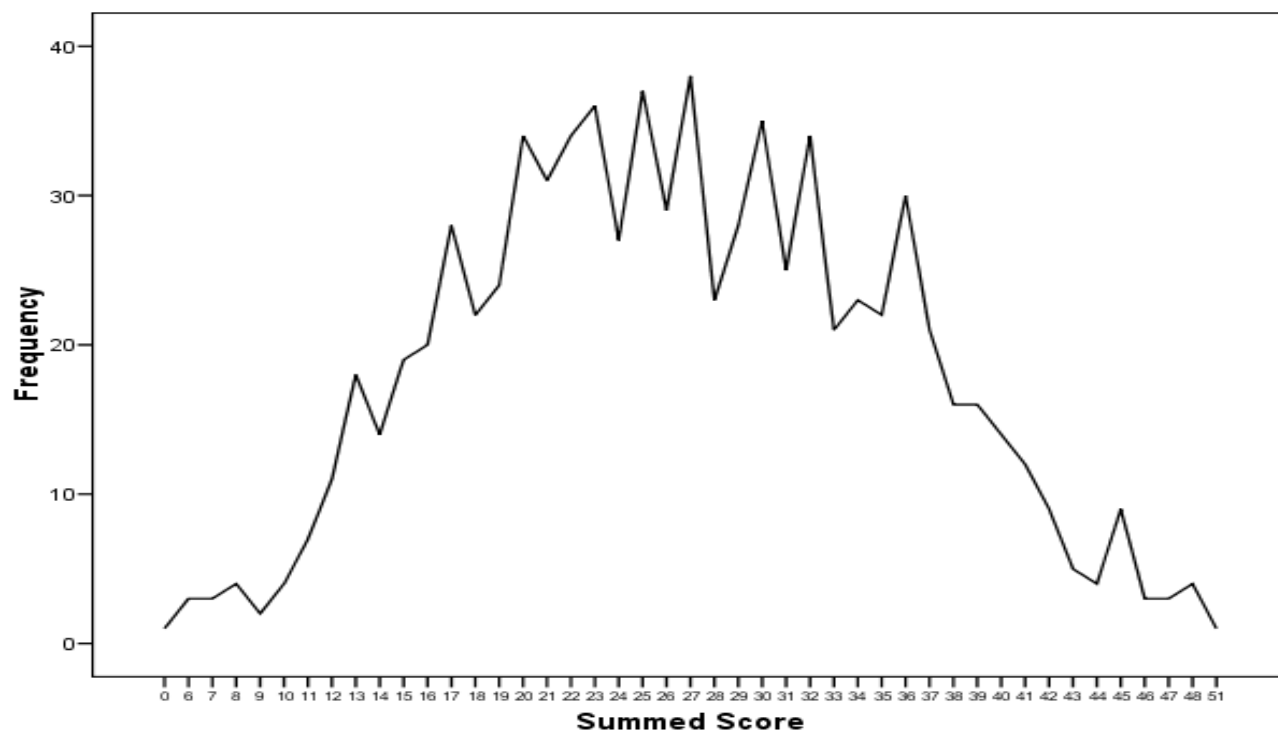


**Figure 7.1.18 Score Distribution for Grade 9 Applied Mathematics,
Spring Administration (French)**

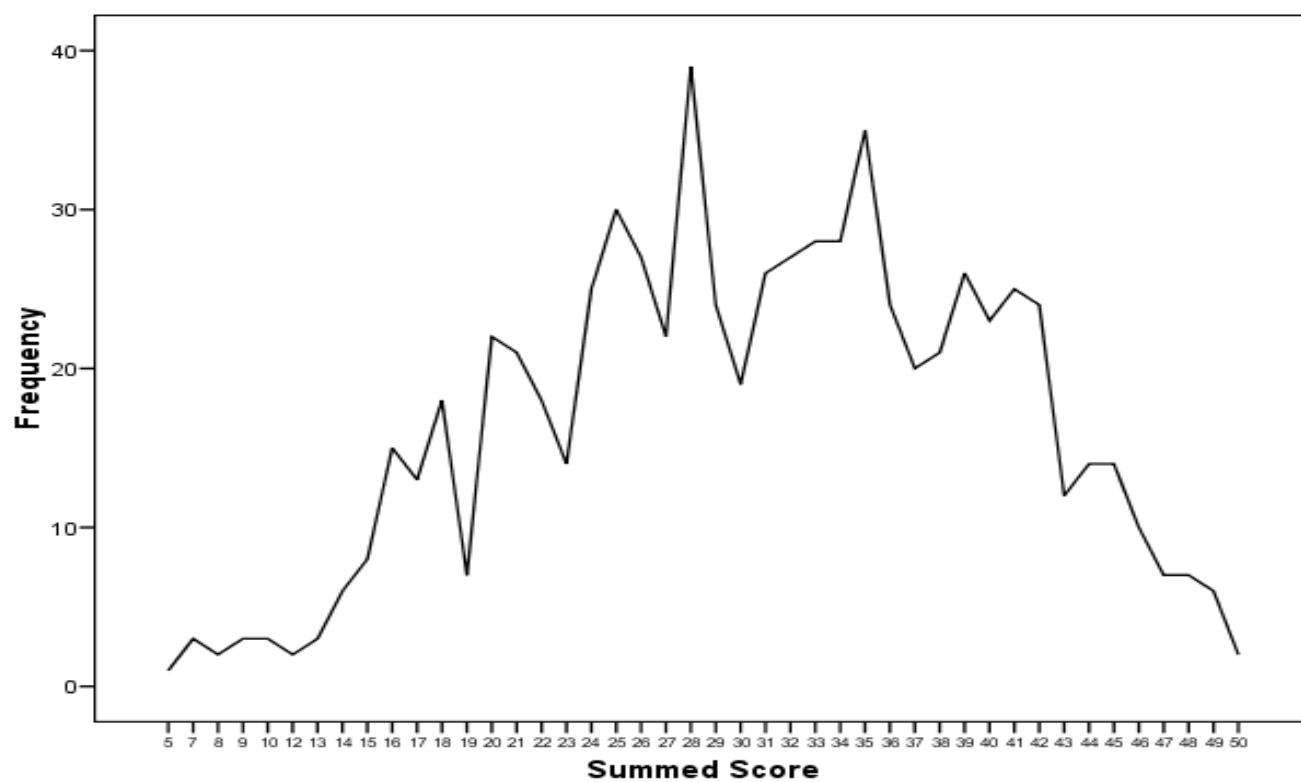


**Figure 7.1.19 Score Distribution for Grade 9 Academic Mathematics,
Winter Administration (French)**

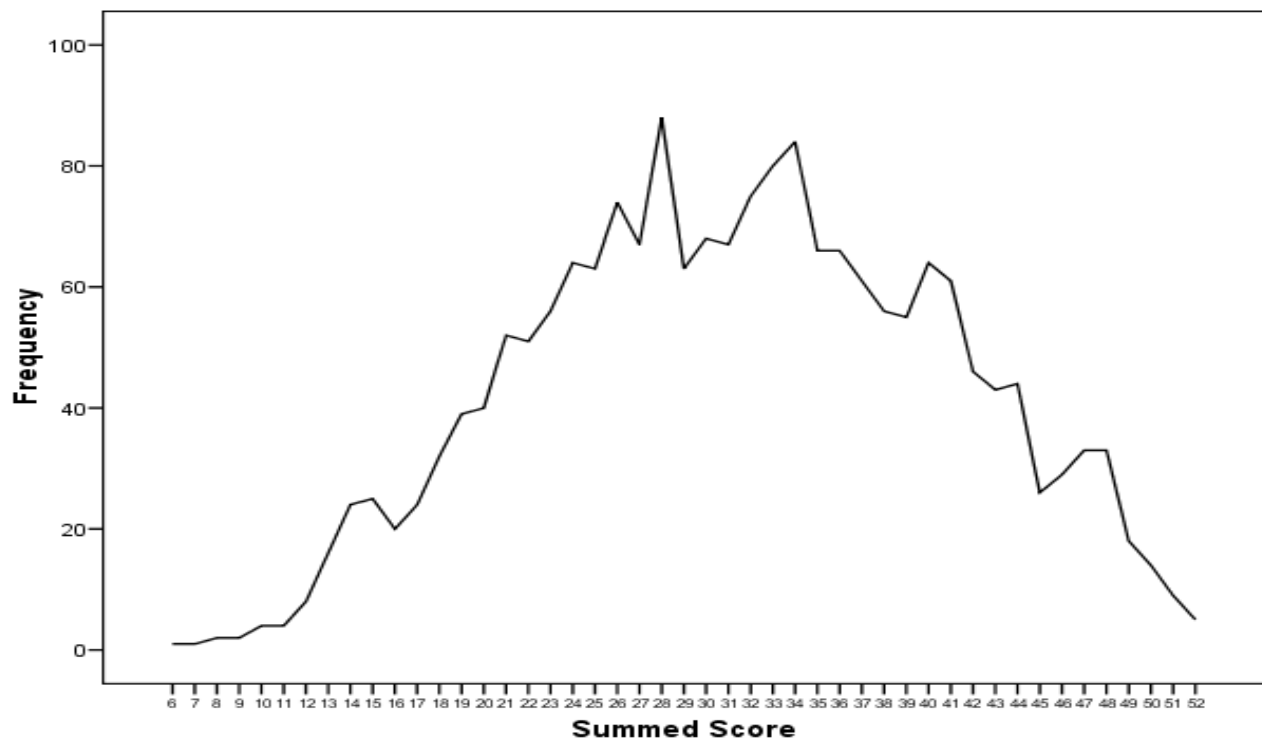


**Figure 7.1.20 Score Distribution for Grade 9 Academic Mathematics,
Spring Administration (French)**

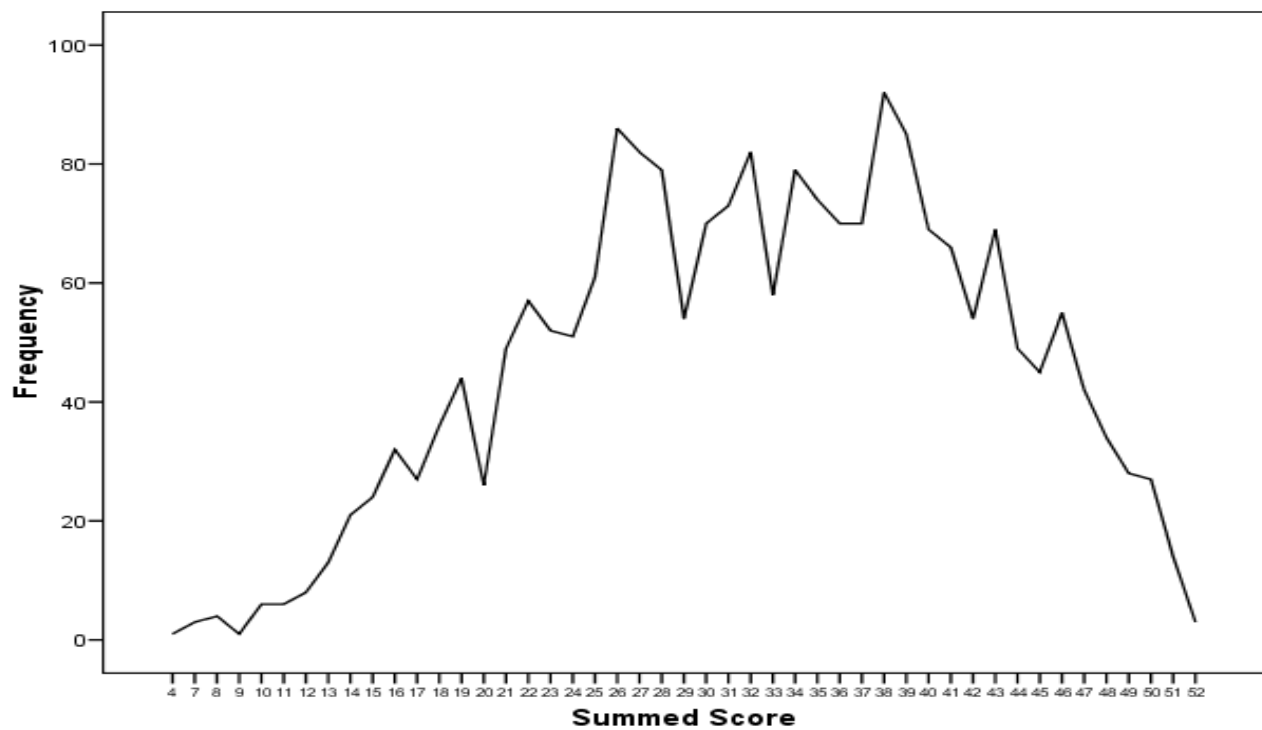


Table 7.1.13: Item Statistics – Grade 9 Applied Mathematics Program, Winter Administration (English)

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Location	Slope
11826	PN10	N	KU	1	0.79	0.3	-1.44	0.58
24576	PN04	N	AP	2	0.39	0.17	1.78	0.38
25044	PR07	R	KU	4	0.82	0.20	-2.50	0.34
25285	PR04	R	AP	4	0.78	0.29	-1.50	0.52
25028	PM06	M	PS	3	0.41	0.32	0.86	0.84
13664	PM10	M	KU	4	0.65	0.29	-0.50	0.49
24520	PN03	N	AP	4*	0.61 (2.45)	0.48	-1.01	0.27
13704	PR12	R	AP	4*	0.52 (2.09)	0.44	-0.32	0.27
13706	PM09	M	AP	4*	0.39 (1.55)	0.46	0.47	0.29
10531	PN10	N	KU	4	0.43	0.23	1.33	0.40
24527	PN06	N	PS	2	0.46	0.15	1.53	0.26
24779	PR07	R	AP	3	0.69	0.23	-1.17	0.34
13731	PR10	R	AP	4	0.42	0.19	1.54	0.34
14036	PR13	R	PS	2	0.88	0.29	-2.15	0.61
24661	PM08	M	AP	3	0.74	0.37	-0.93	0.76
24526	PN06	N	KU	2	0.75	0.25	-1.77	0.35
24638	PN06	N	PS	4	0.11	0.15	2.58	1.10
25127	PR13	R	KU	2	0.50	0.30	0.41	0.55
24588	PR14	R	AP	3	0.44	0.24	1.07	0.47
10639	PR13	R	PS	1	0.48	0.30	0.51	0.71
10645	PM09	M	AP	1	0.54	0.30	0.16	0.61
9248	PN11	N	AP	4*	0.30 (1.21)	0.41	0.89	0.26
13693	PR07	R	AP	4*	0.72 (2.87)	0.44	-2.44	0.26
13697	PR18	R	PS	4*	0.46 (1.82)	0.53	-0.17	0.33
9768	PM06	M	PS	4*	0.45 (1.80)	0.60	0.07	0.47
25035	PN13	N	AP	3	0.50	0.30	0.58	0.54
14026	PR09	R	KU	1	0.80	0.29	-2.05	0.44
10332	PR04	R	AP	3	0.65	0.21	-0.90	0.26
13722	PR11	R	PS	3	0.31	0.26	1.46	0.95
24609	PM06	M	PS	3	0.31	0.25	1.36	1.15
24351	PM08	M	KU	2	0.61	0.38	-0.20	0.76

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

AP = Application

KU = Knowledge and Understanding

N = Number Sense and Algebra

PS = Problem Solving

R = Linear Relations

M = Measurement and Geometry

Table 7.1.14: Item Statistics – Grade 9 Applied Mathematics Program, Spring Administration (English)

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Location	Slope
11826	PN10	N	KU	1	0.79	0.29	-1.44	0.58
24548	PN05	N	AP	2	0.81	0.32	-1.30	0.81
25044	PR07	R	KU	4	0.83	0.21	-2.50	0.34
13647	PR01	R	AP	4	0.65	0.33	-0.43	0.65
25028	PM06	M	PS	3	0.42	0.31	0.86	0.84
24598	PM08	M	KU	3	0.75	0.25	-1.53	0.38
13762	PN05	N	AP	4*	0.54 (2.16)	0.49	-0.86	0.29
13705	PR12	R	AP	4*	0.37 (1.48)	0.52	0.47	0.38
13715	PM09	M	AP	4*	0.46 (1.86)	0.56	-0.18	0.37
10531	PN10	N	KU	4	0.43	0.22	1.33	0.40
24534	PN05	N	PS	3	0.39	0.13	2.14	0.33
24779	PR07	R	AP	3	0.71	0.23	-1.17	0.34
10294	PR14	R	AP	1	0.39	0.29	1.14	0.70
14036	PR13	R	PS	2	0.87	0.30	-2.15	0.61
11819	PM06	M	AP	4	0.42	0.26	1.09	0.56
24532	PN02	N	KU	1	0.56	0.23	0.16	0.36
13672	PN06	N	PS	3	0.39	0.24	1.31	0.60
25127	PR13	R	KU	2	0.52	0.31	0.41	0.55
13648	PR09	R	PS	4	0.62	0.34	-0.30	0.61
10639	PR13	R	PS	1	0.49	0.32	0.51	0.71
10645	PM09	M	AP	1	0.56	0.33	0.16	0.61
13708	PN13	N	AP	4*	0.28 (1.13)	0.40	1.15	0.22
12079	PR07	R	AP	4*	0.66 (2.65)	0.49	-1.62	0.23
11566	PR17	R	PS	4*	0.49 (1.97)	0.51	-0.07	0.27
13687	PM06	M	PS	4*	0.33 (1.34)	0.51	0.42	0.38
25035	PN13	N	AP	3	0.47	0.29	0.58	0.54
14026	PR09	R	KU	1	0.81	0.30	-2.05	0.44
10332	PR04	R	AP	3	0.65	0.20	-0.90	0.26
13601	PR12	R	AP	4	0.40	0.31	0.94	0.76
24569	PM06	M	PS	4	0.40	0.16	1.80	0.36
13662	PM10	M	KU	4	0.75	0.36	-1.13	0.63

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

AP = Application

KU = Knowledge and Understanding

N = Number Sense and Algebra

PS = Problem Solving

R = Linear Relations

M = Measurement and Geometry

Table 7.1.15: Item Statistics – Grade 9 Academic Mathematics Program, Winter Administration (English)

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Location	Slope
13612	DN11	N	KU	2	0.71	0.25	-0.97	0.44
24836	DR05	R	KU	3	0.83	0.30	-1.56	0.70
13634	DR01	R	PS	4	0.37	0.27	1.27	0.69
10703	DG08	G	AP	1	0.71	0.37	-0.65	0.87
25132	DM12	M	AP	4	0.63	0.41	-0.25	1.01
24624	DM11	M	PS	3	0.48	0.35	0.53	0.76
566	DR10	R	PS	4*	0.45 (1.8)	0.45	-0.51	0.35
13683	DG06	G	PS	4*	0.61 (2.44)	0.52	-0.69	0.41
24430	DM12	M	PS	4*	0.44 (1.77)	0.49	-0.19	0.30
24602	DN06	N	AP	4	0.50	0.35	0.41	0.75
13718	DN03	N	AP	1	0.52	0.27	0.45	0.51
10340	DR13	R	AP	3	0.61	0.28	-0.29	0.52
10535	DG01	G	KU	4	0.68	0.26	-0.71	0.47
14069	DG09	G	PS	4	0.31	0.34	1.26	1.20
24655	DM11	M	KU	2	0.78	0.26	-1.42	0.49
24590	DN01	N	KU	4	0.42	0.19	1.44	0.40
14063	DN10	N	PS	3	0.34	0.40	0.95	1.17
13663	DR09	R	AP	1	0.40	0.15	2.18	0.30
24841	DG10	G	KU	2	0.71	0.35	-0.77	0.71
10729	DG05	G	AP	1	0.39	0.34	0.77	1.02
13627	DM05	M	PS	2	0.44	0.23	1.10	0.49
24518	DN06	N	AP	4*	0.63 (2.51)	0.55	-1.07	0.45
13758	DR08	R	AP	4*	0.37 (1.47)	0.52	0.18	0.36
13686	DG12	G	PS	4*	0.43 (1.74)	0.49	-0.07	0.35
13717	DM07	M	AP	4*	0.62 (2.46)	0.52	-0.94	0.38
10883	DR10	R	KU	4	0.86	0.31	-1.64	0.74
13610	DR07	R	PS	3	0.50	0.34	0.44	0.72
13745	DG07	G	PS	2	0.63	0.34	-0.31	0.66
13670	DG03	G	AP	2	0.48	0.34	0.51	0.80
25071	DM11	M	AP	4	0.51	0.35	0.39	0.69
24819	DM12	M	KU	4	0.84	0.36	-1.26	0.94

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

AP = Application

KU = Knowledge and Understanding

N = Number Sense and Algebra

G = Analytic Geometry

PS = Problem Solving

R = Linear Relations

M = Measurement and Geometry

**Table 7.1.16: Item Statistics – Grade 9 Academic Mathematics Program,
Spring Administration (English)**

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Location	Slope
24581	DN04	N	KU	1	0.49	0.27	0.77	0.60
24836	DR05	R	KU	3	0.86	0.29	-1.56	0.70
13729	DR08	R	PS	3	0.51	0.32	0.51	0.72
10703	DG08	G	AP	1	0.74	0.40	-0.65	0.87
25132	DM12	M	AP	4	0.67	0.42	-0.25	1.01
24551	DM07	M	KU	2	0.71	0.29	-0.64	0.56
13681	DR09	R	AP	4*	0.50 (1.98)	0.52	-0.36	0.35
13707	DG09	G	PS	4*	0.39 (1.57)	0.62	0.44	0.59
11573	DM08	M	PS	4*	0.55 (2.21)	0.57	-0.63	0.40
24559	DN06	N	AP	3	0.65	0.32	-0.27	0.61
13638	DN03	N	AP	2	0.81	0.23	-1.78	0.43
10340	DR13	R	AP	3	0.66	0.28	-0.29	0.52
13755	DG11	G	PS	4	0.44	0.25	1.09	0.55
13605	DG03	G	AP	3	0.42	0.34	0.89	0.96
23552	DM11	M	PS	4	0.50	0.32	0.58	0.64
13628	DN10	N	KU	2	0.75	0.31	-0.95	0.59
14063	DN10	N	PS	3	0.41	0.39	0.95	1.17
13606	DR07	R	AP	4	0.83	0.29	-1.53	0.60
24841	DG10	G	KU	2	0.74	0.34	-0.77	0.71
10729	DG05	G	AP	1	0.46	0.38	0.77	1.02
13753	DM11	M	PS	3	0.47	0.29	0.79	0.63
9220	DN10	N	AP	4*	0.47 (1.89)	0.57	-0.40	0.42
6705	DR11	R	PS	4*	0.71 (2.85)	0.52	-1.51	0.44
9782	DG05	G	AP	4*	0.65 (2.59)	0.52	-0.99	0.34
13699	DM12	M	PS	4*	0.44 (1.77)	0.58	0.03	0.41
10883	DR10	R	KU	4	0.86	0.32	-1.64	0.74
13738	DR04	R	PS	4	0.43	0.23	1.24	0.51
24664	DG01	G	KU	1	0.53	0.29	0.49	0.53
13650	DG07	G	PS	2	0.44	0.38	0.76	0.96
25071	DM11	M	AP	4	0.53	0.35	0.39	0.69
24819	DM12	M	KU	4	0.84	0.36	-1.26	0.94

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

AP = Application

KU = Knowledge and Understanding

N = Number Sense and Algebra

G = Analytic Geometry

PS = Problem Solving

R = Linear Relations

M = Measurement and Geometry

**Table 7.1.17: Item Statistics – Grade 9 Applied Mathematics Program,
Winter Administration (French)**

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Location	Slope
13524	PN02	N	KU	4	0.65	0.24	-0.55	0.41
13527	PN02	N	AP	3	0.47	0.16	1.30	0.27
13819	PN08	N	KU	2	0.69	0.22	-0.78	0.50
18526	PR04	R	KU	2	0.67	0.35	-0.55	0.65
13529	PR05	R	AP	3	0.47	0.23	0.75	0.51
13457	PR14	R	PS	2	0.60	0.27	-0.11	0.47
13517	PM07	M	PS	4*	0.42 (1.67)	0.42	0.32	0.24
13536	PR01	R	AP	4*	0.58 (2.34)	0.40	-1.65	0.27
13537	PR09	R	AP	4*	0.45 (1.80)	0.38	0.71	0.23
18557	PR10	R	KU	3	0.58	0.28	-0.29	0.57
13472	PR09	R	AP	3	0.36	0.18	2.28	0.36
23871	PM17	M	AP	2	0.65	0.26	-0.48	0.51
23809	PM08	M	PS	3	0.45	0.33	0.59	0.81
10589	PM01	M	PS	3	0.50	0.30	0.48	0.62
23815	PM06	M	KU	2	0.47	0.23	0.57	0.56
19113	PN08	N	KU	2	0.69	0.32	-0.71	0.53
13479	PN06	N	KU	1	0.40	0.33	0.86	0.78
18547	PN12	N	AP	2	0.60	0.34	-0.22	0.70
13528	PN09	N	PS	3	0.46	0.17	1.10	0.35
13823	PR11	R	KU	3	0.54	0.25	0.22	0.44
18530	PR04	R	PS	3	0.75	0.22	-1.06	0.51
13492	PM03	M	AP	4*	0.54 (2.16)	0.50	-0.57	0.29
24071	PN12	N	PS	4*	0.41 (1.66)	0.52	0.05	0.33
8965	PR14	R	PS	4*	0.42 (1.67)	0.51	-0.04	0.32
12294	PN01	N	AP	4*	0.58 (2.33)	0.58	-0.77	0.42
23835	PR12	R	AP	2	0.60	0.26	-0.17	0.48
23890	PM01	M	AP	2	0.61	0.31	-0.22	0.70
18539	PM06	M	AP	2	0.46	0.22	0.48	0.53
11728	PM17	M	PS	3	0.42	0.28	0.90	0.72
13477	PM15	M	PS	4	0.21	0.22	1.80	0.88
24224	PM17	M	AP	3	0.70	0.33	-0.77	0.62

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

N = Number Sense and Algebra

M = Measurement and Geometry

AP = Application

R = Linear Relations

KU = Knowledge and Understanding

PS = Problem Solving

**Table 7.1.18: Item Statistics – Grade 9 Applied Mathematics Program,
Spring Administration (French)**

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Location	Slope
13483	PN01	N	KU	4	0.83	0.17	-1.73	0.56
13817	PN02	N	AP	2	0.71	0.16	-0.70	0.57
13816	PN06	N	KU	2	0.52	0.21	0.40	0.54
23889	PR04	R	KU	3	0.66	0.26	-0.30	0.91
10766	PR02	R	AP	2	0.29	0.12	2.82	0.53
13485	PR14	R	AP	2	0.74	0.22	-0.89	0.61
13539	PM02	M	AP	4*	0.52 (2.10)	0.48	-0.54	0.29
13535	PN13	N	PS	4*	0.64 (2.57)	0.48	-0.99	0.28
13827	PR04	R	AP	4*	0.76 (3.06)	0.41	-1.88	0.24
18557	PR10	R	KU	3	0.71	0.24	-0.29	0.57
18816	PR09	R	AP	3	0.58	0.22	0.20	0.56
23871	PM17	M	AP	2	0.68	0.21	-0.48	0.51
23809	PM08	M	PS	3	0.47	0.27	0.59	0.81
10589	PM01	M	PS	3	0.54	0.18	0.48	0.62
23815	PM06	M	KU	2	0.58	0.20	0.57	0.56
19113	PN08	N	KU	2	0.68	0.19	-0.71	0.53
13479	PN06	N	KU	1	0.43	0.23	0.86	0.78
18547	PN12	N	AP	2	0.67	0.23	-0.22	0.70
13822	PN09	N	PS	2	0.41	0.16	1.22	0.57
18597	PR12	R	KU	1	0.48	0.25	0.61	0.79
18530	PR04	R	PS	3	0.72	0.27	-1.06	0.51
13496	PN02	N	AP	4*	0.52 (2.07)	0.50	-0.29	0.32
13538	PR15	R	PS	4*	0.63 (2.53)	0.50	-0.74	0.32
11060	PR09	R	AP	4*	0.64 (2.57)	0.55	-0.96	0.49
13825	PM09	M	PS	4*	0.58 (2.30)	0.47	-1.29	0.23
13459	PR11	R	PS	2	0.49	0.10	1.75	0.17
23890	PM01	M	AP	2	0.65	0.20	-0.22	0.70
18539	PM06	M	AP	2	0.57	0.24	0.48	0.53
13476	PM15	M	PS	4	0.28	0.11	1.73	1.16
13534	PM17	M	PS	1	0.34	0.22	1.21	1.13
13532	PM17	M	KU	4	0.65	0.17	-0.24	0.57

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

N = Number Sense and Algebra

M = Measurement and Geometry

AP = Application

R = Linear Relations

KU = Knowledge and Understanding

PS = Problem Solving

**Table 7.1.19: Item Statistics – Grade 9 Academic Mathematics Program,
Winter Administration (French)**

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Location	Slope
19038	DN03	N	AP	2	0.72	0.23	–1.07	0.43
18570	DN12	N	KU	2	0.82	0.31	–1.31	0.60
13447	DN13	N	PS	4	0.52	0.38	0.28	0.83
23945	DR01	R	AP	3	0.68	0.42	–0.55	0.88
13450	DR09	R	AP	2	0.70	0.33	–0.56	0.61
18951	DR11	R	AP	2	0.78	0.21	–1.73	0.35
12598	DR15	R	PS	4*	0.65 (2.61)	0.53	–1.04	0.39
13503	DN12	N	AP	4*	0.56 (2.24)	0.54	–0.60	0.46
6953	DM20	M	PS	4*	0.56 (2.24)	0.55	–0.48	0.42
13807	DG04	G	KU	4	0.63	0.32	–0.22	0.62
13402	DG09	G	KU	2	0.40	0.4	0.78	0.97
13423	DG07	G	PS	3	0.60	0.29	–0.09	0.60
10478	DM18	M	PS	4	0.58	0.38	0.02	0.84
13410	DM04	M	PS	2	0.40	0.28	1.03	0.70
24311	DM20	M	AP	4	0.75	0.29	–0.91	0.58
13518	DN06	N	KU	2	0.64	0.30	–0.32	0.54
13808	DN12	N	AP	4	0.62	0.27	–0.11	0.47
18580	DR10	R	KU	2	0.69	0.27	–0.72	0.45
10815	DN18	N	AP	3	0.49	0.22	0.78	0.43
10783	DR14	R	PS	4	0.49	0.43	0.39	0.94
13413	DR04	R	PS	1	0.38	0.23	1.36	0.61
13806	DM18	M	PS	4*	0.75 (3.00)	0.46	–2.09	0.34
13505	DR04	R	AP	4*	0.67 (2.68)	0.47	–1.41	0.33
13508	DG07	G	AP	4*	0.55 (2.19)	0.49	–0.26	0.31
13498	DN17	N	AP	4*	0.60 (2.41)	0.44	–0.75	0.28
29	DG04	G	AP	3	0.44	0.33	0.73	0.84
10913	DG09	G	PS	3	0.36	0.37	0.82	0.97
10568	DM15	M	KU	1	0.56	0.32	–0.09	0.69
13811	DM05	M	AP	2	0.41	0.31	0.88	0.87
24193	DM09	M	KU	3	0.62	0.24	–0.32	0.43
18958	DM20	M	KU	4	0.86	0.30	–1.74	0.69

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

N = Number Sense and Algebra

M = Measurement and Geometry

AP = Application

R = Linear Relations

KU = Knowledge and Understanding

PS = Problem Solving

**Table 7.1.20: Item Statistics – Grade 9 Academic Mathematics Program,
Spring Administration (French)**

Item Code	Expectation	Strand	Cognitive Skill	Answer Key/Max. Score	CTT Item Statistics		IRT Item Parameters	
					Difficulty	Corrected Item-Total Correlation	Location	Slope
12288	DN01	N	AP	2	0.72	0.37	–0.65	0.74
18570	DN12	N	KU	2	0.77	0.32	–1.31	0.60
13812	DN20	N	PS	2	0.58	0.28	0.09	0.49
23945	DR01	R	AP	3	0.72	0.39	–0.55	0.88
13450	DR09	R	AP	2	0.68	0.32	–0.56	0.61
18951	DR11	R	AP	2	0.77	0.18	–1.73	0.35
12598	DR15	R	PS	4*	0.65 (2.61)	0.53	–1.04	0.39
13504	DN03	N	AP	4*	0.70 (2.82)	0.54	–1.32	0.43
13500	DN20	N	AP	4*	0.48 (1.94)	0.49	0.01	0.31
10737	DG01	G	AP	2	0.64	0.31	–0.23	0.57
13402	DG09	G	KU	2	0.44	0.36	0.78	0.97
10824	DG04	G	KU	1	0.78	0.21	–1.55	0.38
10478	DM18	M	PS	4	0.60	0.37	0.02	0.84
13801	DM03	M	PS	2	0.62	0.31	–0.12	0.61
24311	DM20	M	AP	4	0.73	0.33	–0.91	0.58
13427	DN05	N	KU	2	0.51	0.30	0.53	0.56
13808	DN12	N	AP	4	0.59	0.28	–0.11	0.47
10738	DR09	R	KU	1	0.65	0.38	–0.24	0.78
10815	DN18	N	AP	3	0.50	0.21	0.78	0.43
10783	DR14	R	PS	4	0.50	0.44	0.39	0.94
23988	DR04	R	PS	4	0.58	0.28	0.15	0.47
13806	DM18	M	PS	4*	0.76 (3.03)	0.46	–2.09	0.34
8994	DR04	R	AP	4*	0.65 (2.60)	0.40	–1.15	0.28
8966	DG03	G	AP	4*	0.51 (2.03)	0.56	–0.39	0.41
13501	DM20	M	PS	4*	0.67 (2.69)	0.62	–0.80	0.47
13814	DG04	G	PS	1	0.52	0.33	0.43	0.69
10913	DG09	G	PS	3	0.46	0.36	0.82	0.97
10568	DM15	M	KU	1	0.66	0.40	–0.09	0.69
10453	DM09	M	AP	3	0.54	0.36	0.37	0.71
24193	DM09	M	KU	3	0.66	0.24	–0.32	0.43
13403	DM20	M	KU	2	0.56	0.18	0.48	0.30

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

Guessing parameter was set to a constant of 0.2 for multiple-choice items

N = Number Sense and Algebra

M = Measurement and Geometry

AP = Application

G = Analytic Geometry

R = Linear Relations

KU = Knowledge and Understanding

PS = Problem Solving

Ontario Secondary School Literacy Test (OSSLT)

Score Distributions

Figure 7.1.21 OSSLT Score Distribution for First-Time Eligible Students (English-Language)

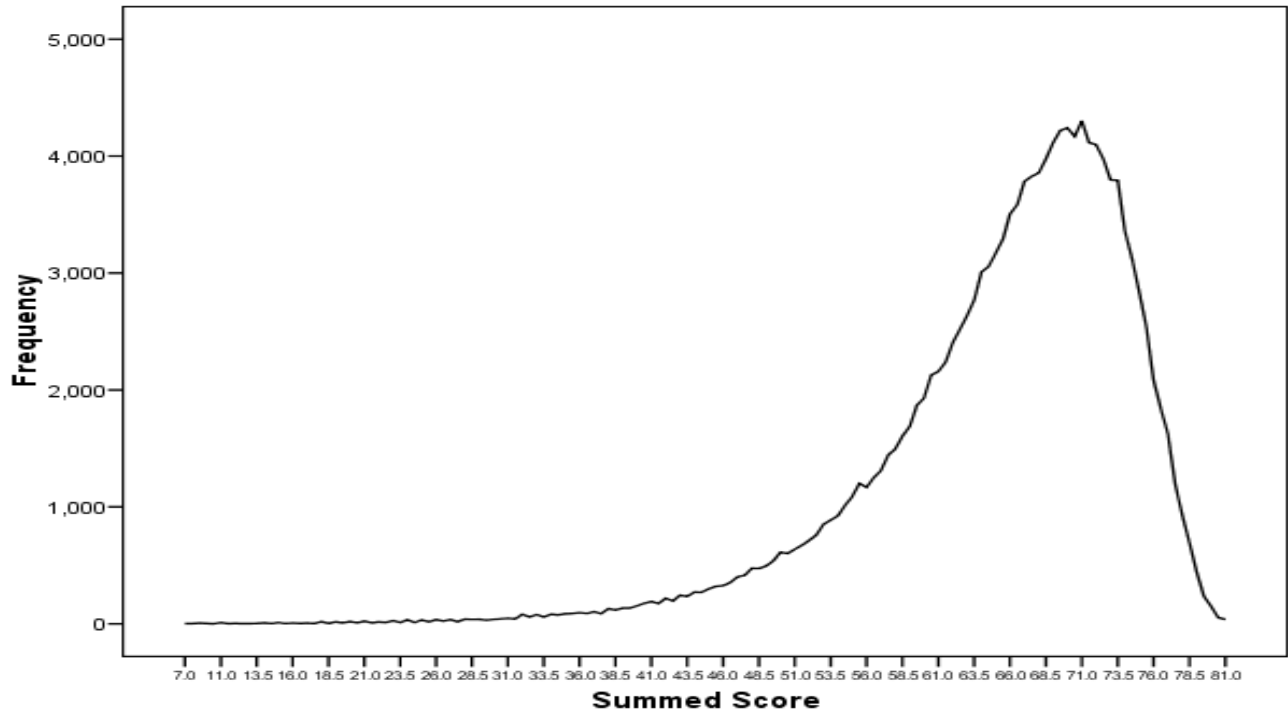


Figure 7.1.22 OSSLT Score Distribution for Previously Eligible Students (English-Language)

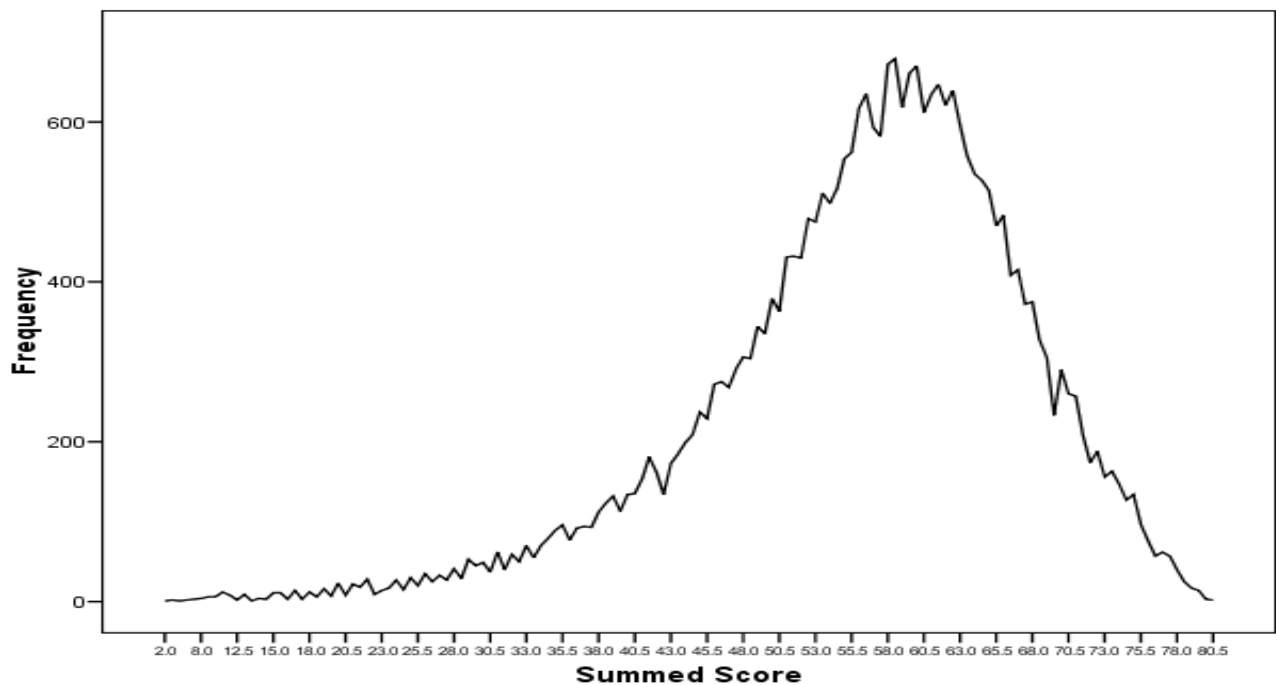


Figure 7.1.23 OSSLT Score Distribution for First-Time Eligible Students (French-Language)

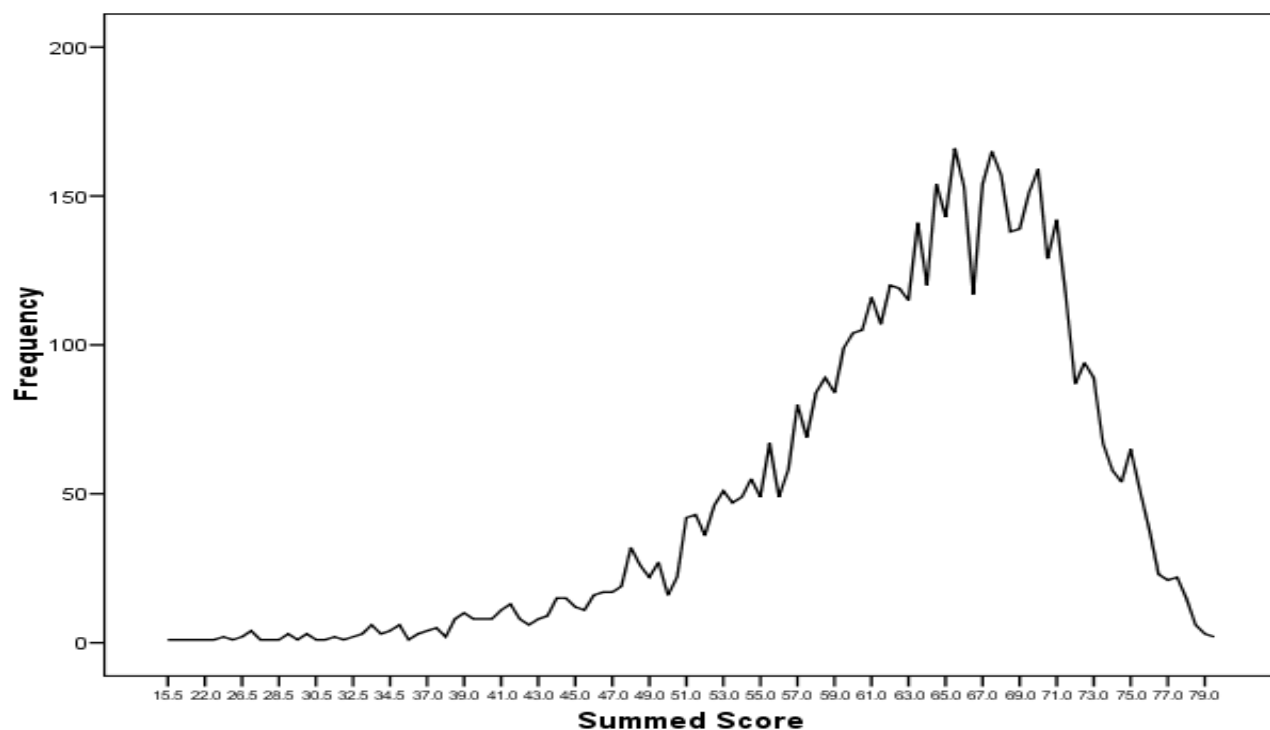


Figure 7.1.24 OSSLT Score Distribution for Previously Eligible Student (French-Language)

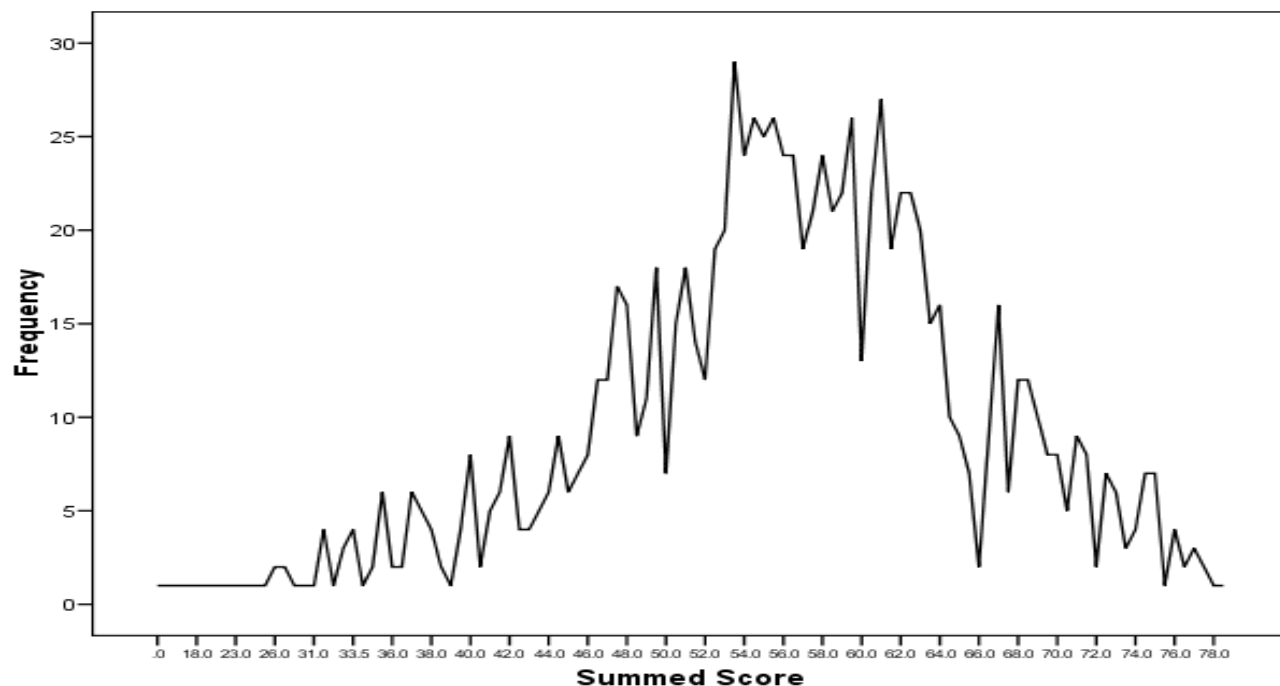


Table 7.1.25: OSSLT Item Statistics (English-Language)

Unique Item ID	Section	Sequence	Strand	Item Type	Answer Key/ Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
						Difficulty	Corrected Item-Total Correlation	Location
ESN2_0024_6104	A	3	R	MC	2	0.72	0.31	-0.87
ESN2_0024_6205	A	4	R	MC	4	0.37	0.08	1.68
ES2_0037_6102	F	1	R	MC	3	0.70	0.31	-0.70
ES2_0037_6105	F	3	R	MC	2	0.87	0.34	-2.08
ES6_0027_6201	I	1	R	MC	2	0.81	0.30	-1.53
ES6_0027_6105	I	5	R	MC	2	0.90	0.29	-2.38
ES9C_0037_6101	L	1	R	MC	1	0.93	0.30	-2.68
ESN2_0024_6101	A	1	R	MC	2	0.89	0.18	-2.20
ESN2_0024_6103	A	2	R	MC	1	0.73	0.27	-0.88
ES2_0037_6104	F	2	R	MC	1	0.84	0.21	-1.66
ES2_0037_6106	F	4	R	MC	4	0.82	0.29	-1.55
ES2_0037_6206	F	6	R	MC	1	0.72	0.25	-0.82
ES6_0027_6103	I	3	R	MC	2	0.78	0.25	-1.21
ES6_0027_6104	I	4	R	MC	1	0.61	0.23	-0.09
ES6_0027_6206	I	6	R	MC	4	0.96	0.35	-3.53
ES6_0027_6209	I	7	R	MC	3	0.91	0.35	-2.47
ES6_0027_6110	I	8	R	MC	3	0.95	0.25	-3.27
ES4_0035_6103	J	1	R	MC	3	0.89	0.31	-2.28
ES4_0035_6105	J	2	R	MC	1	0.88	0.32	-2.09
ES4_0035_6106	J	3	R	MC	2	0.80	0.33	-1.43
ES4_0035_6108	J	5	R	MC	3	0.86	0.35	-1.95
ES9C_0037_4103	L	2	R	MC	2	0.39	0.23	1.22
ES9C_0037_6103	L	3	R	MC	4	0.80	0.35	-1.41
ES9C_0037_6105	L	4	R	MC	2	0.92	0.37	-2.67
ES9C_0037_6207	L	6	R	MC	3	0.94	0.23	-2.91
ES2_0037_6208	F	7	R	OR	3	0.76 (2.27)*	0.39	-1.19
ES4_0035_6209	J	6	R	OR	3	0.87 (2.61)	0.51	-1.76
ESN2_0024_6107	A	5	R	MC	3	0.75	0.40	-1.10
ES2_0037_6107	F	5	R	MC	1	0.86	0.35	-1.89
ES6_0027_6202	I	2	R	MC	2	0.69	0.41	-0.71
ES6_0027_6111	I	9	R	MC	4	0.92	0.25	-2.61
ES4_0035_6107	J	4	R	MC	2	0.67	0.29	-0.48
ES9C_0037_6205	L	5	R	MC	1	0.58	0.23	0.11
ESN2_0024_6109	A	6	R	OR	3	0.69 (2.07)	0.42	-1.01
ES4_0035_6210	J	7	R	OR	3	0.89 (2.68)	0.48	-2.00
WMC_410_002	B	3	W	MC	3	0.95	0.21	-3.17
WMC_430_007	B	4	W	MC	3	0.87	0.23	-1.98

Table 7.1.25: OSSLT Item Statistics (English-Language) (continued)

Unique Item ID	Section	Sequence	Strand	Item Type	Answer Key/ Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
						Difficulty	Corrected Item-Total Correlation	Location
WMC_310_001	H	3	W	MC	2	0.90	0.27	–2.31
WMC_320_010	H	4	W	MC	3	0.79	0.22	–1.26
1W5_049T	D	1	W	SW	3	0.97 (2.90)	0.36	–2.54
1W5_039T	K	1	W	SW	3	0.96 (2.87)	0.34	–2.51
1W3_014T	E	1	W	LW	6	0.72 (4.30)	0.58	–1.05
1W2_010T	G	1	W	LW	6	0.75 (4.50)	0.60	–1.24
WMC_220_005	B	1	W	MC	1	0.61	0.26	–0.14
WMC_220_004	B	2	W	MC	3	0.54	0.20	0.32
WMC_160_004	H	1	W	MC	4	0.71	0.18	–0.71
WMC_150_001	H	2	W	MC	2	0.73	0.20	–0.85
1W5_049V	D	1	W	LW	2	0.98 (1.97)	0.33	–3.09
1W5_039V	K	1	W	LW	2	0.98 (1.97)	0.34	–3.13
1W3_014V	E	1	W	LW	4	0.85 (3.42)	0.59	–1.90
1W2_010V	G	1	W	LW	4	0.85 (3.41)	0.55	–1.96

() Mean score for open-response items (OR, SW, and LW).

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

Note: The slope was set to 0.588 for all items and guessing was set at 0.20 for all multiple-choice items

OR = Open-Response

SW = Short Writing

LW = Long Writing

MC = Multiple-Choice

W = Writing

R = Reading

Table 7.1.26: OSSLT Item Statistics (French-Language)

Unique Item ID	Section	Sequence	Strand	Item Type	Answer Key/Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
						Difficulty	Corrected Item-Total Correlation	Location
FS2N_0069_6201	A	1	R	MC	1	0.81	0.28	-1.43
FS2N_0069_6206	A	3	R	MC	2	0.92	0.20	-2.59
FS2F_0072_6101	F	1	R	MC	2	0.85	0.41	-1.87
FS2F_0072_6105	F	5	R	MC	1	0.97	0.27	-3.67
FS6_0029_6104	I	3	R	MC	3	0.94	0.27	-3.05
FS6_0029_6111	I	9	R	MC	1	0.67	0.34	-0.48
FS9F_0046_6203	L	2	R	MC	3	0.95	0.39	-3.30
FS2N_0069_6203	A	2	R	MC	2	0.79	0.22	-1.21
FS2N_0069_6204	A	4	R	MC	3	0.67	0.33	-0.49
FS2F_0072_6102	F	2	R	MC	1	0.48	0.29	0.62
FS2F_0072_6103	F	3	R	MC	2	0.86	0.32	-1.89
FS2F_0072_6104	F	4	R	MC	2	0.89	0.36	-2.17
FS6_0029_6105	I	4	R	MC	4	0.56	0.16	0.28
FS6_0029_6106	I	5	R	MC	1	0.63	0.16	-0.15
FS6_0029_6208	I	6	R	MC	4	0.29	0.27	1.97
FS6_0029_6108	I	7	R	MC	4	0.75	0.29	-0.98
FS6_0029_6110	I	8	R	MC	3	0.72	0.28	-0.81
FS4_0023_6102	J	1	R	MC	4	0.97	0.11	-3.69
FS4_0023_6103	J	2	R	MC	2	0.59	0.11	0.12
FS4_0023_6108	J	3	R	MC	4	0.98	0.26	-4.54
FS4_0023_6206	J	4	R	MC	4	0.82	0.29	-1.54
FS9F_0046_6202	L	1	R	MC	1	0.91	0.31	-2.41
FS9F_0046_6205	L	4	R	MC	1	0.26	0.25	2.27
FS9F_0046_6106	L	5	R	MC	2	0.97	0.14	-3.70
FS9F_0046_6107	L	6	R	MC	3	0.62	0.15	-0.08
FS2F_0072_6108	F	7	R	OR	3	0.80 (2.39)*	0.24	-1.34
FS4_0023_6109	J	6	R	OR	3	0.81 (2.44)	0.43	-1.98
FS2N_0069_6205	A	5	R	MC	1	0.45	0.16	0.95
FS2F_0072_6107	F	6	R	MC	2	0.90	0.31	-2.30
FS6_0029_6101	I	1	R	MC	2	0.79	0.24	-1.21
FS6_0029_6103	I	2	R	MC	3	0.64	0.22	-0.26
FS4_0023_6107	J	5	R	MC	3	0.90	0.33	-2.31
FS9F_0046_6104	L	3	R	MC	1	0.61	-0.09	0.12
FS2N_0069_6209	A	6	R	OR	3	0.86 (2.59)	0.32	-1.85
FS4_0023_6110	J	7	R	OR	3	0.84 (2.53)	0.44	-2.18
2ÉCM_430_001	B	1	W	MC	2	0.64	0.20	-0.22
2ÉCM_440_001	H	4	W	MC	1	0.89	0.33	-2.28
2ÉCM_320_002	B	2	W	MC	2	0.93	0.30	-2.73

Table 7.1.26: OSSLT Item Statistics (French-Language) (continued)

Unique Item ID	Section	Sequence	Strand	Item Type	Answer Key/Max. Score Code ¹	CTT Item Statistics		IRT Item Parameters
						Difficulty	Corrected Item-Total Correlation	Location
2ÉCM_330_001	B	4	W	MC	3	0.88	0.34	-2.15
2É5_041T	D	1	W	SW	3	0.91 (2.73)	0.36	-2.57
2É5_040T	K	1	W	SW	3	0.82 (2.47)	0.44	-1.37
2É3_014T	E	1	W	LW	6	0.74 (4.44)	0.56	-1.29
2É2_0014T	G	1	W	LW	6	0.64 (3.82)	0.53	-0.68
2ÉCM_144_001	B	3	W	MC	2	0.85	0.20	-1.79
2ÉCM_154_001	H	1	W	MC	4	0.81	0.28	-1.41
2ÉCM_250_003	H	2	W	MC	4	0.46	0.24	0.80
2ÉCM_211_001	H	3	W	MC	2	0.95	0.24	-3.06
2É5_041V	D	1	W	SW	2	0.97 (1.95)	0.27	-3.48
2É5_040V	K	1	W	SW	2	0.94 (1.88)	0.36	-2.04
2É3_014V	E	1	W	LW	4	0.83 (3.32)	0.57	-1.92
2É2_0014V	G	1	W	LW	4	0.78 (3.12)	0.56	-1.52

() Mean score for open-response items (OR, SW, and LW).

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

Note: The slope was set to 0.588 for all items and guessing was set at 0.20 for all multiple-choice items.

OR = Open-Response
MC = Multiple-Choice

SW = Short Writing
W = Writing

LW = Long Writing
R = Reading