

## **FSA 2008 Technical Notes**

The notes below provide information about the technical quality of FSA tests: test development, test administration and scoring, as well as data analysis.

### **1. Overview of FSA 2008**

FSA is a set of three tests in reading, writing and numeracy. FSA measures important skills in the provincial curriculum. It measures skills students have gained in several school years, not just in a single year.

The primary purpose of FSA is to help schools, school planning councils, school districts and the province evaluate how foundation skills are being addressed and make plans for improvement. The secondary purpose is to give parents, teachers and principals information about individual student performance.

FSA helps answer important questions such as “Are students learning vital skills they will need later?” “Is student achievement improving over time?” “Are there any trends in student performance at the school, district or provincial levels?” and “How are specific groups of students doing?”

With limited exceptions, all Grade 4 and Grade 7 British Columbia students are expected to participate in FSA.

FSA results can show students, parents and teachers where students are doing well and skills they need to work on. FSA results complement classroom work and report card information.

### **2. Developing the FSA 2008 Reading, Writing and Numeracy Tests and Scoring Rubrics**

BC teachers are contracted by the Ministry to develop the FSA tests. Teachers are involved in all aspects from designing the [table of specifications](#), writing items, field-testing the items, reviewing the tests and developing the [FSA scoring rubrics](#). [FSA sample questions](#) are available for students and teachers to familiarize themselves with the format of the tests.

### **3. Technical Test Review**

Technical test reviewers include teachers, curriculum specialists, and assessment consultants. Members review items against the following criteria: grade-level appropriateness; accuracy; language; bias and sensitivity.

### **4. Item Types**

Reading and numeracy scores are based on multiple-choice and written-response questions; writing scores are based on two writing tasks.

The reading component includes 39 multiple-choice questions (1 mark each) and one written-response question (4 marks), for a total of 43 marks. The writing component includes two written tasks, one shorter (4 marks) and one longer (8 marks), for a total of 12 marks. The numeracy component includes 40 multiple-choice questions (1 mark each) and two written-response questions (4 marks each), for a total of 48 marks.

## **5. Administration of FSA tests**

The FSA tests were administered to Grades 4 and 7 students during February 4 to 15, 2008. Schools were able to schedule testing sessions at any time during two weeks of this period, provided they followed the guidelines detailed in the test [administration manual](#). Schools were provided with an extended administration window.

## **6. Item Analysis and Review**

Classical item statistics are produced for both field test and actual administered items. Analyses include verification that scoring keys and rubrics are correct, determination of item difficulty and discrimination, and computation of item non-response rates (items skipped and items not reached).

## **7. Scoring FSA**

Scoring of the multiple-choice questions is automated. The written-response section of the assessments was scored locally by teachers, using provincial scoring guides and examples of student work.

For reading and numeracy, all students' scores were scaled and equated. Students' raw scores (i.e., total number of points) were translated to scaled scores. The scales were developed such that they ranged from 200 through 800, with a mean of 500. Scaling converts raw points from one scale to another. For writing, students' raw scores were used to calculate numbers and proportions of students at each level of performance, and average scores.

All FSA reading and numeracy items were calibrated using Item Response Theory (IRT) for equating purposes. This calibration allows for the difficulty of items to be weighted, based on actual student responses.

IRT equating will also allow for more precise comparisons for results from one year to the next. Selected FSA 2008 items will be used as anchor items to equate to future FSA results, allowing for comparisons over time with 2008 results as a new baseline.

Due to the changes made to the FSA tests, new cut scores have been established through a standard-setting method, called the bookmark method, where teachers provided the Ministry with recommended cut scores using historical data as well as the FSA 2008 results. Based on teachers' recommendations, the Ministry set two cut scores for each component at both grades: the cut between *not yet meeting* and *meeting*, and the cut between *meeting* and *exceeding expectations*.

Numbers and proportions of students in the school, district and provincial reports are calculated using the new cut scores. Therefore, FSA 2008 results represent new baselines. Comparisons of student performance with previous FSA results should not be made.

## 8. Results Reporting

There are four reports for the FSA 2008:

- A. **Student Summary Report**
- B. **School Summary Report**
- C. **District Summary Report**
- D. **Provincial Report**

With the exception of the Student Summary Report, all reports are available for schools and districts to view or download on the [Ministry website](#). Data files with individual student results are only available for school and district personnel to download from the secure [Ministry secure website](#).

### A. Student Summary Report

The [FSA Student Summary Report](#) includes the student's raw scores on each component. For each component, ranges of raw scores indicate the student performance level: not yet demonstrated meeting expectations, meeting expectations and exceeding expectations. The student performance ranges for FSA 2008 were determined using historical FSA data.

Schools also have access to the list of students enrolled in their school with their scores for the multiple-choice and written-response sections for each component. The list is available on the [Ministry secure website](#) (csv file format).

The FSA Student Summary Report and the student booklets were distributed to parents by March 31, 2008. A template of the report is available in 15 languages.

### B. School Summary Report, C. District Summary Report and D. Provincial Report

The school, the district and the provincial FSA reports each consists of two pages, one for Grade 4 and one for Grade 7.

### Overall Data Display

The reports include the following results:

- *Total number of students* based on February enrolment

For each component,

- *A pie chart* showing *proportions* of students at each level of performance (unknown, not yet meeting, meeting and exceeding)
- *Number and Proportion* of students at each level of performance for the whole population
- *Number and Proportion* of students at each level of performance by sub-groups (male; female; Aboriginal; ESL; and Special Education)

Numbers and proportions of students at each level are based on the new cut scores as recommended by teachers, and may differ from what has been reported on the Student Summary Report. Proportions are calculated using all students as the denominator, including those for whom the FSA performance level is unknown.

- *Average FSA scaled scores* , on a scale of 200 – 800, for reading and numeracy; these averages are based on IRT scaled scores; for writing, the average is based on students’ raw scores.

Average scaled scores and average scores do not include students in the Unknown Performance Level category.

Note: The symbol used to indicate the average scale score for reading may be hidden by the numeracy symbol when the two average scaled scores are similar.

### Details of Data Display

#### Interpreting Average Scaled Scores:

At the school and district levels, confidence intervals for average scaled scores have been estimated using the standard errors of the mean ( $SE = SD / \sqrt{n}$ ), using plus or minus one standard error.

The following table provides the intervals to be used according to the population size of a school or a district.

<b>Population Size (Number of students)</b>	<b>Confidence Intervals</b>
Less than 25	± 33
25 - 35	± 21
36 - 80	± 18
81 - 100	± 12
101 and more	± 10

Example: Using the confidence intervals, the average scaled score for a school with 50 students and an average score of 500, will fall within the range of 483 and 517 (± 1 SE).

<b>Population Size (Number of students)</b>	<b>Confidence Intervals</b>
20 or less	± 0.7
21 – 40	± 0.5
41 – 80	± 0.3
81 or more	± 0.2

Example: Using the confidence intervals, the average score for a school with 50 students and an average score of 6, will fall within the range of 5.7 and 6.3 (± 1 SE).

#### Student Participation

For each component, the number of students in the Performance Level Unknown category represents the number of students who have not responded to at least one question in that component.

The inverse of the proportion of students for whom the performance level is unknown is the participation rate. These rates can be compared to previous FSA administrations.

#### PIRLS Comparison

The *FSA School Summary Report* provides an indication for schools who have performed at or above the International Benchmark on [PIRLS](#). FSA and PIRLS results have been aligned using a common scale. More information is available at [Linking FSA 2008 Grade 4 Reading Results to PIRLS 2006](#).

### 9. Reliability

At the provincial level, reliability coefficients have been calculated for each component at both grades using the Cronbach's  $\alpha$  coefficient:

Grade	Component	Reliability Coefficient
4	Reading	0.88
	Writing	0.57
	Numeracy	0.90
7	Reading	0.89
	Writing	0.61
	Numeracy	0.90

### 10. Validity

To assess the validity of FSA 2008, reviews have been conducted at different stages of the development process to ensure the following:

- item alignment with content standards
- item bias
- sensitivity and content appropriateness
- adherence to the test specifications
- use of multiple item types;
- use of standardized administration procedures, with accommodated options for participation

More information on the design and development process of the assessment instrument is available at [FSA - Design and Development](#)