

1. WHAT IS THE PAN-CANADIAN ASSESSMENT PROGRAM?

The Pan-Canadian Assessment Program (PCAP) is the CMEC's most recent commitment to informing Canadians about how well their education systems are meeting the needs of students and society. The information gained from such a pan-Canadian assessment gives the ministers of education a basis for examining the curriculum and other aspects of their school systems.

School curriculum programs vary from jurisdiction to jurisdiction across the country, so comparing results from these varied programs is a complex task. However, young Canadians in the different jurisdictions learn many similar skills in reading, mathematics, and science. PCAP has been designed to determine whether students across Canada reach similar levels of performance in these core disciplines at about the same age and also to complement existing assessments in each jurisdiction so these jurisdictions can have access to comparative, Canada-wide data on the achievement levels attained by 13-year-olds across the country.

Goals

When the ministers of education began planning the development of PCAP in 2003, they set out the following goals for a conceptually new pan-Canadian instrument of assessment designed to:

- inform educational policies as a means of improving approaches to learning
- focus on mathematics, reading, and science, with the possibility of including other domains as the need arises
- reduce the testing burden on schools through a more streamlined administrative process
- provide useful background information using complementary contextual questionnaires for students, teachers, and school administrators
- enable jurisdictions to use both national and international results to validate the results of their own assessment programs and to improve them

The development process

In August 2003, a PCAP working group of experienced and knowledgeable representatives from several jurisdictions and including an external authority on measurement theory, large-scale assessment, and educational policy began the development process. A concept paper was commissioned that would elaborate on issues of structure, development planning, operations, and reporting. Drawing on this concept paper, the working group defined PCAP as a testing program that would:

- be administered at regular intervals
- be administered to students who are 13-year-olds at the start of the school year
- be based on the commonality of all current jurisdictional curricular outcomes across Canada
- assess reading, mathematics, and science
- provide a major assessment of one domain with a minor concentration on the two other domains
- focus on reading as the major domain in the first administration in 2007

For each subject area, a thorough review of curricula, current assessment practices, and research literature was then undertaken, and reports were written to indicate the common expectations among all jurisdictions.

The working groups for bilingual framework development, established for each of the three subject areas, were composed of representatives from several jurisdictions with knowledge and experience in curriculum and assessment for the particular subject. Each working group also had an external expert in the assessment of the particular subject to advise and assist with the development of a framework statement establishing the theory, design, and performance descriptors for each domain. The framework statements were reviewed and accepted by all participating jurisdictions as the basis for test item development.

Bilingual teams for developing the test items were then established; members of these teams were subject area educators selected from all jurisdictions, with a subject assessment expert to supervise. Each subject framework provided a blueprint, with its table of specifications describing the subdomains of each subject area, the types and lengths of texts and questions, the range of difficulty, and the distribution of questions assessing each specific curriculum expectation. Each jurisdiction was also encouraged to submit texts and test-ready materials that they felt were appropriate for the age group and that were not currently in use in their jurisdiction. The results in reading, for example, provided sufficient items for three complete forms for field testing, each 90 minutes in duration.

Texts and questions were developed in both official languages and cross-translated to be equivalent in meaning and difficulty. Jurisdictions reviewed and confirmed the validity of the French-English translations to ensure fair and equitable testing in both languages. All items were reviewed by outside validators and further revised by members of the item development team. These texts and items were then submitted to the framework development working group to be examined in light of the blueprint and to be structured into three comparable field-test forms. Each booklet contained both selected-response and constructed-response items with a range of difficulty accessible to the age group, based on scenarios meaningful to the age group and reflecting Canadian values, culture, and content.

Field testing involved the administration of these temporary forms to a representative sample of students from an appropriate range of jurisdictions in both languages. Approximately 2,000 students in 100 schools across Canada were involved in the field testing. The tests were then scored by teams of educators from the jurisdictions in July 2006. Following analysis of the data from the field tests, each framework development working group reviewed all items and selected the texts and items considered best, from a content and statistical viewpoint, to form two booklets in reading and a booklet consisting of half mathematics and half science, each booklet totalling 90 minutes. The final test booklets were then reviewed and approved by all participating jurisdictions.

Design and development of contextual questionnaires

The accompanying questionnaires for students, teachers, and school administrators were designed to provide jurisdictions with contextual information that would contribute to the interpretation of performance results in the field testing. Such information may also be examined and used by researchers, policy makers, and practitioners to help determine what factors influence learning outcomes.

A questionnaire development group composed of educators and research experts from selected jurisdictions developed a framework to ensure that the questions asked of students, teachers, and school principals were consistent with predetermined theoretical constructs or important research questions. The group reviewed models of questionnaire design found in the three large-scale assessment programs

Texts and questions were developed in both official languages and cross-translated to be equivalent in meaning and difficulty.

(The School Achievement Indicators Program [SAIP], The Trends in International Mathematics and Science Study [TIMSS], and the Programme for International Student Assessment [PISA]), worked to create a shorter, more streamlined model of the questionnaires; and attempted to maximize research value by shaping the questionnaires around selected research issues for the 2007 administration of the test.

Using initial drafts, a separate group (the chair of the questionnaire development working group and two reading experts) expanded the reading component of the questionnaire. This working group held briefing sessions with the chair of the reading working group and the CMEC coordinator, Education Data and Research, who suggested some areas of interest derived from the most recent round of consultations on the Pan-Canadian Education Research Agenda

(PCERA). It was determined that the main research focus would be on teaching and learning reading strategies. Additional areas of interest included the methods and uses of assessment and the ways in which special needs students are accommodated in schools and classrooms.

Features of the administration of PCAP reading 2007

In the spring of 2007, the test was administered to a random sample of schools and students, representing the national cohort of 13-year-olds and of the jurisdictions. Booklets were randomly assigned to students.

The sampling process resulted in approximately 30,000 13-year-old students writing the test.

Sampling

The sampling process refers to the way in which the schools and students were selected to write the assessment. It is necessary to select a large enough number of participants to allow for adequate representation of the population's performance (the word "population" refers to all eligible students within a jurisdiction and/or a linguistic group). This assessment adopted the following two-step stratified sampling process in the selection of participants:

- the random selection of schools from each jurisdiction, drawn from a complete list of publicly funded schools provided by the jurisdiction
- the random selection of students, drawn from a list of all eligible 13-year-olds within each school

In the case where numbers were smaller than the desired size, all schools and/or all students meeting the criteria within the jurisdiction were selected. This method ensured that we had an adequate number of participants to allow for reporting on their achievement as if all students within the jurisdiction had participated.

The sampling process resulted in approximately 30,000 13-year-old students writing the test. Approximately 20,000 wrote the reading segment, the primary domain, and about 10,000 wrote the mathematics and science segment, which represented the secondary domains. Approximately 15,000 wrote the reading segment in English and 5,000 wrote in French. For mathematics and science, the numbers were 7,500 in English and 2,500 in French.

Reporting results by language

The results obtained from students educated in the French system of their respective jurisdictions are reported as French. The results obtained from students educated in the English system of their respective jurisdictions are reported as English. In most jurisdictions, the results achieved by French immersion students who wrote in French are calculated as part of the English results. However, in Manitoba, the results achieved by French immersion students are calculated as part of the French results. All French and English students were expected to write for 90 minutes, with breaks deemed appropriate by the assessment administrator. Then they completed the contextual questionnaire at the back of their test booklet.

Participation

Each school received the assessment handbook that outlined the purposes of the assessment, the organization and administration requirements, and suggestions to encourage as full participation as possible. These suggestions included a common administration script to ensure that all students encountered the testing process in a similar manner and provided guidelines for accommodating special needs students. PCAP testing is intended to be as inclusive as possible in order to provide a complete picture of the range of performance for the age group. The students who were excused from participation were nevertheless recorded for statistical purposes; they included those with highly limited abilities in any one of the domains, those who would be adversely affected by the test, and those whose parents requested that their children be excused.

Participation rates

In large-scale assessments, participation rates are calculated in a variety of ways and are used to guide school administrators as they determine whether the number of students who completed the assessment falls within the norm established for all schools. In the case of PCAP, a formula for this purpose is provided to the test administrators, thereby ensuring that all schools use the same guidelines and that the set minimum of participating students is uniformly applied. Using this formula, the PCAP student participation rate was over 85%.

Schools were encouraged to prepare and motivate students for the test, aiming for as much positive participation and engagement in the process as possible by teachers, students, and parents. The materials provided included information pamphlets for parents and students; the school handbook also included sample questions in reading that illustrated the types of demands and the descriptions of achievement levels for each question provided.

Schools were also asked to have the teacher questionnaire completed by all the language arts teachers of the participating students in the school and to have the school questionnaire completed by the school principal. All three questionnaires (student, teacher and school) were linked to student results, and unique identifiers were used to preserve confidentiality.

Scoring the student response booklets

The scoring was conducted concurrently in both languages in one location over a three-week period. After all student booklets had been submitted from the jurisdictions, the booklets were scrambled into bundles of 10 so that any single bundle contained booklets from several jurisdictions. The scoring administration team, the table leaders, and the scorers themselves came from several jurisdictions. The whole scoring process included:

- parallel training of both table leaders and scorers in each subject area
- a bilingual committee with responsibility for reviewing all instruments and selecting anchor papers to ensure comparability at every level

- twice daily rater-reliability checks, in which all scorers marked the same student work in order to track the consistency of scoring on an immediate basis
- double scoring, in which 300 of each of the 3 booklets were returned to the scoring bundles to be re-scored, providing an overall inter-rater reliability score

Structure of this report

Detailed performance results were presented in the PCAP-13 2007 report released in April 2008 (CMEC, 2008). A brief summary of these results is reported in the next chapter.

The main focus of this report is on the questionnaire results and specifically on variables associated with achievement. The report is divided into chapters corresponding to major clusters of variables which, according to previous research and theory, may influence achievement. For example, these include demographic characteristics; attitudes; student out-of-school activities and behaviours; allocation and use of time; teaching and learning strategies; and finally, assessment practices.

In each chapter, questionnaire results are first presented descriptively by jurisdiction and language. This is followed by a two-stage analytic process. First, the reading performance levels and mean reading scores are compared for students across categories on each of the variables of interest. The mean comparisons are used to determine whether the variable is significantly associated with achievement. Second, the relationships between questionnaire variables and achievement are examined through a multilevel regression modelling process designed to allow the effects of a single variable to be examined while controlling for other variables in the model.