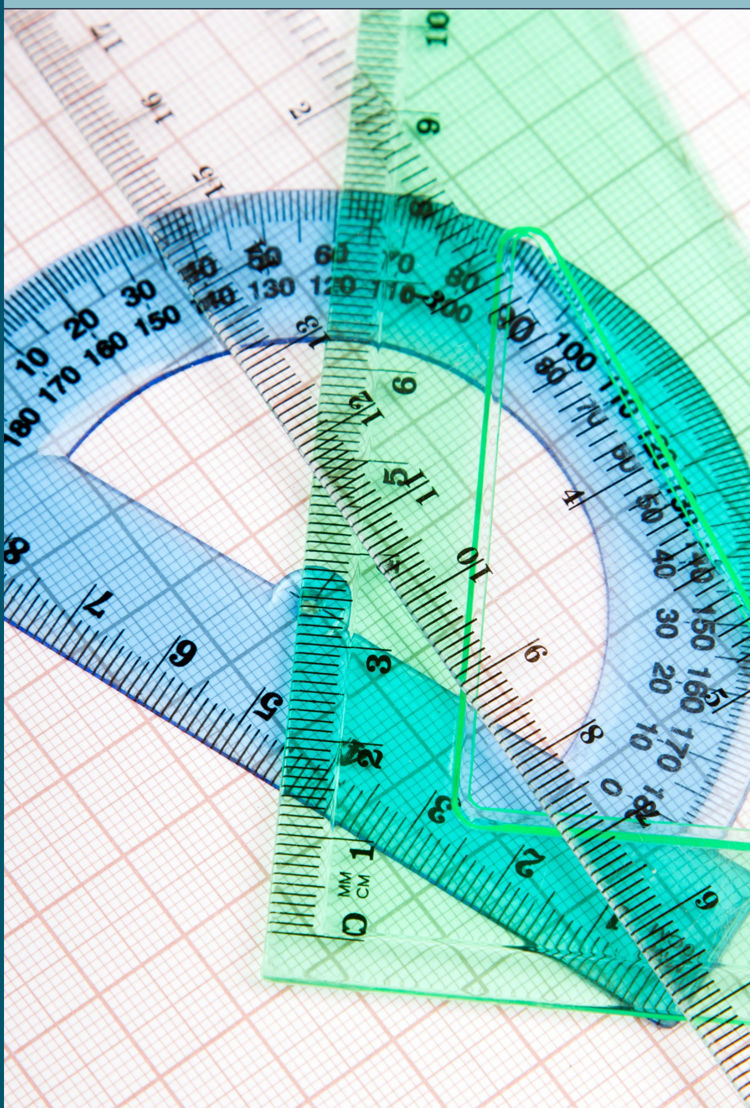
Pan-Canadian Assessment Program

PCAP 2019

Assessment Framework



cmec

Pan-Canadian Assessment Program

PCAP 2019

Assessment Framework

Council of Ministers of Education, Canada

cmec

Council of Ministers of Education, Canada

Conseil des ministres de l'Éducation (Canada)

The Council of Ministers of Education, Canada (CMEC) was formed in 1967 by the provincial and territorial ministers responsible for education to provide a forum in which they could discuss matters of mutual interest, undertake educational initiatives cooperatively, and represent the interests of the provinces and territories with national educational organizations, the federal government, foreign governments, and international organizations. CMEC is the national voice for education in Canada and, through CMEC, the provinces and territories work collectively on common objectives in a broad range of activities in early childhood education, elementary, secondary, and postsecondary levels and in adult learning.

Through the CMEC Secretariat, the Council serves as the organization in which ministries and departments of education undertake cooperatively the activities, projects, and initiatives of particular interest to all provinces and territories. One of the activities on which they cooperate is the development and implementation of pan-Canadian testing based on contemporary research and best practices in the assessment of student achievement in core subjects.

Council of Ministers of Education, Canada
95 St. Clair Avenue West, Suite 1106
Toronto, Ontario M4V 1N6

Telephone: 416-962-8100
Fax: 416-962-2800
E-mail: cmec@cmec.ca
Web: www.cmec.ca

Ce rapport est également disponible en français.

Table of Contents

Chapter 1. Introduction

What is the Pan-Canadian Assessment Program?

The Pan-Canadian Assessment Program (PCAP) is a collaborative project that provides data on student achievement in Canadian provinces and territories.[1] It is part of the ongoing commitment of the Council of Ministers of Education, Canada (CMEC) to inform Canadians about how well their education systems are meeting the needs of students and society. Every three years, close to 30,000 Grade 8/Secondary II[2] students from across Canada are assessed with respect to their achievement of the curricular expectations common to all provinces and territories in three core learning domains: reading, mathematics, and science. The information gained from this pan-Canadian assessment provides ministers of education and other stakeholders with a basis for examining their provincial curriculum and other aspects of their school systems.

School programs and curricula vary from province to province and from territory to territory across the country, so comparing results in these domains is a complex task. However, young Canadians in different provinces and territories 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 to complement existing provincial/territorial assessments with comparative Canada-wide data on the achievement levels attained by Grade 8/Secondary II students.

The PCAP assessment cycle

PCAP assessments are administered every three years to students who are in Grade 8/Secondary II. Each assessment cycle collects achievement data using a cognitive test with a major emphasis on one of the three learning domains—reading, mathematics, or science—and a minor emphasis on the two remaining domains. PCAP also collects a significant range of contextual information (e.g., on demographics, socioeconomic factors, and school teaching and learning conditions) to enhance interpretation of student performance.

Each PCAP assessment includes questions on all three domains, although the focus shifts, as shown in Table 1.1. The repetition of the assessments at regular intervals yields timely data that can be compared across provinces and territories, and over time. For the fifth assessment, in 2019, the focus was on mathematics, as it had been in the second assessment, in 2010, with reading and science as the minor domains.

---

[1] All ten provinces have participated in each PCAP administration. The three territories did not participate in PCAP 2019.

[2] PCAP is administered to students in Secondary II in Quebec and Grade 8 in the rest of Canada.

Table 1.1 PCAP assessment cycle

Domain	Cycle 1			Cycle 2		
	Spring 2007	Spring 2010	Spring 2013	Spring 2016	Spring 2019	Spring 2022
Major	Reading	Mathematics	Science	Reading	Mathematics	Science
Minor	Mathematics	Science	Reading	Mathematics	Science	Reading
Minor	Science	Reading	Mathematics	Science	Reading	Mathematics

PCAP, which replaces an earlier assessment called the Student Achievement Indicators Program (SAIP), is coordinated by CMEC.

The report on the PCAP 2019 assessment results will be released in 2020 and will provide results on Grade 8/Secondary II student achievement in mathematics, science, and reading at both country and provincial levels. Extensive contextual information will be collected from questionnaires completed by students, teachers, and principals. The report with the assessment results will also include one chapter with contextual information for the major domain, mathematics. More extensive contextual information will be published in 2021 as part of the contextual report and should offer insight into some of the factors that may influence student performance in mathematics.

Large-scale assessments and classroom assessments

Large-scale assessments cannot and should not attempt to duplicate or imitate classroom assessments as if they were identical processes. According to curricula across Canada, classroom assessments serve both formative and summative purposes, each informing the other reflexively. However, they are aimed more at helping students take the next steps in learning than at judging the end points of achievement. Multiple modes of assessment, including observation and interviewing, are needed to provide a complete picture of the students' competency. In contrast, large-scale assessments are mainly one-time measures. The content and the administration procedures are standardized to ensure that the results mean the same thing in different contexts.

The difference between classroom assessments and large-scale assessments is based on the need for quite different information: immediate and contextualized data for the former as opposed to rigorously comparable results for the latter. However, both types of assessment are useful at different levels of the education system. Assessments external to schools are used for system accountability and to inform planning for board improvement. They can have a valuable impact on teaching practices and function as a pedagogical resource, provided the education community uses the results in the ways for which they were designed.

Purposes of assessment

PCAP must reflect the changes in our understanding of assessment that have happened since the first administration of SAIP. It is important to understand these changes in order to clarify our understanding of the purpose and limitations of country-wide large-scale assessments.

Although public attention is often focused on the results of large-scale pan-Canadian and international assessments, research suggests that valid and reliable classroom assessments used by teachers in their daily practice provide powerful tools to improve student achievement (Olsen, 2002).

Provincial/territorial examinations and large-scale assessments usually have different purposes than in-class assessment and evaluation practices. They can be used for varying purposes, which can include the following:

- providing information for teachers to improve student achievement;
- selecting students who go on to postsecondary studies;
- certifying graduates;
- fostering accountability for schools, school systems, and provincial/territorial systems here in Canada, and for national systems abroad.

Large-scale assessments such as PCAP may be used to inform education policy, but they are not used to certify and select individual students. Individual students do not have any personal stake in doing well in large-scale assessments; therefore, communication about their purpose and use is critical. Clarity of purpose will help students, teachers, and administrators to understand the importance of these assessments and to take them seriously so that the results accurately reflect the amount of learning that has taken place (Simon & Forgette-Giroux, 2002).

For the purpose of understanding the role of large-scale assessments, it is useful to consider the following three major purposes of assessment that can be used in conjunction with each other to improve student achievement (Earl, 2003): assessment *for* learning, *as* learning, and *of* learning, as described below:

- "Assessment *for* learning" is part of effective planning for teaching and learning. It involves both the teacher and students in a process of continual reflection on and review of progress. When students are given quality feedback, they are empowered to understand how to improve their learning. When teachers are given quality feedback by assessing student understanding, they can reflect and adjust their teaching strategies in response.

- "Assessment *as* learning" actively involves students in their learning processes. Students take responsibility for their learning, constructing meaning for themselves. They develop the ability to determine what they have already learned and decide how to further organize and enhance their learning. Teachers assist this student-driven process by providing opportunities for reflection and critical analysis.

- "Assessment *of* learning" provides a snapshot of student achievement relative to specific curriculum requirements. Assessment *of* learning is often associated with large-scale assessments, and data collected from these assessments are often made public and can be used to inform allocation of resources, monitoring of standards, and approaches to teaching and thus to promote accountability. Assessment *of* learning provides evidence of achievement for public reporting (Stiggins, 2002) and requires precision tools and elements such as tables of specifications, rating scales, and criteria to be used in the development, delivery, grading, and reporting of the assessment tasks.

PCAP is an assessment *of* learning.

Mode of administration

PCAP 2019 marks the beginning of the transition from a paper-based assessment to an online assessment. Students today interact extensively with technology—both in the classroom and in their daily lives. Digital interaction and engagement is now a permanent and ubiquitous part of our society. Accordingly, the move to an online assessment is aligned with current educational and social practices, and supports increased student engagement. To facilitate the transition from a paper-based assessment to an online assessment and also preserve the trend measurement from previous cycles of PCAP, some schools completed the paper-based assessment. Minimal edits were made to how students respond to questions to ensure comparability of achievement across modes (e.g., in multiple-choice questions, distractors were presented with fillable bullets in the online format and circles lettered A to D in the paper format). The mode of administration for each school was randomly allocated by the sampling contractor.

Presentation of PCAP results

Every PCAP report provides data for the three learning domains in the form of mean scores. While overall mean scores, and the comparison of provincial results to the Canadian mean scores, are useful indicators of the performance of education systems overall, they do not provide much information about student learning.

To provide a detailed understanding of what students know, understand, and can do, PCAP has developed useful benchmarks or performance levels that align a range of scores to levels of knowledge and skills measured by PCAP as an assessment of learning. For the major domain, which was mathematics in 2019, PCAP used four performances levels, which provide an overall picture of students' accumulated proficiency at Grade 8/Secondary II. Performance levels are reported for the overall domain as well as by subdomain.

A standard-setting exercise is used to determine the performance levels for the major domain. A group of educators from each province set the "cut scores" for each level using the "bookmark" method (Lewis, Mitzel, Mercado, & Schultz, 2012)—that is, determining the relative difficulty of the full set of assessment instruments and delineating the point along a scale that defines the achievement of each level of success, thus determining the "cut score." Once suitable cut scores are set, student performance within the range of cut scores is refined. These refined descriptors of performance-level results more clearly indicate what students should know and be able to do at each level.

The achievement results in the minor subject domains (reading and science, in 2019) are reported only as overall mean scores. Together, these two minor domains constituted approximately one-third of the assessment. Because the students responded to a smaller subset of items for the two minor subject areas, their results by subdomain and by performance level are not reported.

PCAP results are weighted based on population size—provinces with a larger population have a greater weight. This weighting has implications for the mean scores: because English-language students from Ontario and French-language students from Quebec contribute the greatest number of test results, their average scores are more likely than those of any other population to be closest to the Canadian English mean and Canadian French mean, respectively.

The actual results from students' assessments are called "raw scores." The raw scores are converted to a scale, which has a range of 0 to 1000. These raw scores are standardized, providing a common measurement so that meaningful comparisons can be made of scores obtained from different populations over time and on different versions of a test. The standardized scale used for PCAP assessments places scores on a normal distribution with a midpoint or mean of 500 and a standard deviation of 100. The scale midpoint of 500 is equal to the pan-Canadian average for each subject in the baseline year.[3] The majority of students in Canada—about two thirds—will score between 400 and 600, or within one standard deviation of the mean. This mean can then be used as a reference point that allows the comparison of Canada-wide results.

Reporting by language

English and French versions of the assessment are equivalent. The results obtained from students educated in the francophone school system of their respective provinces are reported as French. The results obtained from students educated in the anglophone school system of their respective provinces are reported as English. Results achieved by French-immersion students who wrote in French will be calculated as part of the anglophone results, since these students are considered to be part of the English-language cohort.

Reporting PCAP achievement over time

[3] The baseline year is the year in which the domain was the major domain assessed (2007 and 2016 for reading, 2010 and 2019 for mathematics, and 2013 for science).

One of the strengths of PCAP is its measurement of changes over time in student performance. The PCAP achievement scales provide a common metric on which provinces can compare students' progress at the Grade 8/Secondary II level in the three core subjects from one assessment year to another. Items that were administered in the baseline years, known as "anchor items," provide the basis for linking the assessment results. Such links enable provinces to have comparable achievement data from 2007, 2010, 2013, 2016, and 2019, and to analyze changes in performance over time.

Applications of PCAP data

PCAP is designed as a system-level assessment to be used primarily by provincial ministries of education to monitor and assess their respective education systems. PCAP data are reported by province (and, where data are available, by territory), by language of the school system, and by gender.

The goal of national (and international) large-scale assessments is to provide reliable information about academic achievement and insight into the contextual factors influencing it. The data from studies such as PCAP provide policy-makers, administrators, teachers, and researchers with meaningful insights into the functioning of education systems and how they might be improved.

It should be noted that PCAP is not designed to report valid results at the student, school, or school board level: its results complement classroom assessment but do not replace it. Although public attention is often focused on the results of large-scale, standardized assessments, research suggests that valid and reliable classroom assessments used by teachers in their daily practice provide powerful tools to improve student achievement (Olsen, 2002). Therefore, it is important to recognize the key roles of both classroom assessments (formative and summative) and larger-scale summative assessments such as PCAP in providing valuable information about student learning. Table 1.2 summarizes the similarities and differences between large-scale assessments like PCAP and classroom assessments.

Table 1.2 Comparison of large-scale and classroom assessment

Large-scale assessments	Classroom assessments
Summative assessment	Program of formative and summative assessments
Standardized procedures, randomly administered	Multiple modes and instances of assessment adapted to student learning needs
Supports analysis of education systems	Supports and assesses the learning of individual students
Fosters system accountability	Provides educators and students with immediate, context-specific feedback on learning
Differentiates by student achievement	Differentiates by student achievement, learning needs, and strengths

Chapter 2. Mathematics Assessment Framework

The development of an assessment of proficiency in mathematics is a result of the provinces' and territories' desire to gather information that would allow them to evaluate their mathematics programs and curricula in relation to pan-Canadian and international assessment results.

Context for developing a mathematics assessment framework

The development of the PCAP mathematics assessment framework was informed by a literature review of mathematics assessments and test design (CMEC, 2005a) and common pan-Canadian curricular components.[4] There are two types of assumptions to be considered in the preparation of this framework: assumptions about the learning of the domain assessed (in this case, mathematics) and about the assessment itself.

Domain assumptions

- The learning of mathematics is a process in which students link and build on previously learned concepts and procedures so that their understanding and knowledge of mathematics become deeper and more refined, as does their ability to apply the mathematics they have learned.
- Students learn more complex concepts and procedures by connecting to existing knowledge through meaningful experiences.
- A well-articulated mathematics curriculum places concepts and procedures along a continuum. Student mastery of these concepts and procedures and the level of development of competencies in them will also be placed along this continuum.
- Although the domain is divided into separate subdomains (strands of the curriculum) the content of the subdomains is often interwoven and dependent, as are associated concepts and procedures. The evaluation of certain content and associated concepts and procedures cannot be done in isolation.

Assessment assumptions

- Although the assessment is administered to Grade 8/Secondary II students, it also assesses the concepts and procedures learned in earlier grades.
- PCAP is not a student assessment but a program assessment.
- Provinces and territories want enough information at the domain level to reliably identify strengths and weaknesses in their programs.

[4] For updated mathematics curricula, please visit the official websites of the provinces and territories.

Large-scale assessments in mathematics

The mathematics assessment framework is based on three major mathematics assessments, namely, the former School Achievement Indicators Program (SAIP), the Programme for International Student Assessment (PISA), and the Trends in Mathematics and Science Survey (TIMSS).

It is also closely aligned with provinces' and territories' own curricula, which generally have been guided by National Council of Teachers of Mathematics (NCTM) standards as articulated in the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics, 2000) and *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence* (NCTM, 2006).

School Achievement Indicators Program (SAIP)

The SAIP mathematics content component was designed to evaluate levels attained by 13- and 16-year-old students in numbers and operations, algebra and functions, measurement and geometry, data management and statistics, and problem solving. SAIP reported on students' overall performance in mathematics (CMEC, 2002).

Programme for International Student Assessment (PISA)

PISA assesses 15-year-old students' performance in mathematical literacy. In 2003, it broadly defined mathematical literacy as "an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments, and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned, and reflective citizen" (OECD, 2003, p. 15). In its 2014 report, PISA's definition of mathematical literacy was refined into "(a)n individuals' capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals in recognising the role that mathematics plays in the world and to make the well-founded judgements and decisions needed by constructive, engaged and reflective citizens" (OECD, 2014, p. 28). PISA reports on students' performance in mathematics overall, as well as producing separate scales for space and shape, change and relationships, quantity, and uncertainty and data.

In 2018, PISA was administered as a computer-based assessment in the 80 participating countries. Two reasons for having a computer-based mathematics assessment were described in the report on the 2012 PISA results:

> First, computer-based items can be more interactive, authentic and engaging than paper-based items. They can be presented in new formats (e.g., drag-and-drop), include real-world data (such as a large, sortable data set), and use colour, graphics, and movement to aid comprehension. Students may be presented with a moving stimulus or representations of three-dimensional

objects that can be rotated, or have more flexible access to relevant information. New item formats can expand response types beyond verbal and written, giving a more rounded picture of mathematical literacy (Stacey and Wiliam, 2013).

Second, computers have become essential tools for representing, visualizing, exploring, and experimenting with all kinds of mathematical objects, phenomena, and processes, not to mention for realizing all types of computations—at home, at school, and at work. In the workplace, mathematical literacy and the use of computer technology are inextricably linked (Hoyles et al., 2002).

The design of the computer-based assessment ensured that mathematical reasoning and processes take precedence over mastery of using the computer as a tool. Each computer-based item involves three aspects:

- the mathematical demand (as for paper-based items);
- the general knowledge and skills related to the information and communications technologies (ICT) that are required (e.g., using a keyboard and mouse and knowing common conventions, such as arrows to move forward). These are intentionally kept to a minimum; and
- competencies related to the interaction of mathematics and ICT, such as making a pie chart from data using a simple "wizard," or planning and implementing a sorting strategy to locate and collect desired data in a spreadsheet. (OECD, 2014, p. 491)

The PISA 2012 report also indicated that:

In general, there is a high degree of consistency in student performance on items delivered on paper and by computer. However, there are important exceptions.

In the field of mathematics, one participant (Shanghai-China) saw a large difference, of around 50 score points, in favour of the paper-based format. Three other countries and economies showed substantial differences in the same direction: Poland (28-point difference), Chinese Taipei (22-point difference), and Israel (20-point difference). Conversely, there are also countries for which computer delivery of the assessment appears to have been advantageous. The largest difference, of about 30 score points, was seen in Brazil. Colombia also saw a difference of about 20 points in the same direction. The United States, the Slovak Republic, and Italy also saw marked, albeit smaller, differences in favour of the computer delivery of the assessment. Across OECD countries, the performance advantage of the computer-based assessment is slightly higher for boys than for girls.

> Further analyses are needed to explore the extent to which these differences are driven by the different nature of the tasks, by the differences in the mode of delivery, or by student familiarity with computers. (OECD, 2014, p. 491)

Trends in Mathematics and Science Survey (TIMSS)

The TIMSS mathematics assessment framework is structured around mathematical content and cognitive processes. Numbers, algebra, measurement, geometry, and data are the five subdomains covered by the assessment. The four cognitive processes identified are: knowing facts and procedures, using concepts, solving routine problems, and reasoning. These subdomains and cognitive processes assess a student's ability to draw upon relevant mathematical knowledge and efficient and accurate computational skills; link mathematical facts to make extensions beyond current knowledge; use mathematics to solve problems based on familiar settings; and apply logical and systematic thinking to unfamiliar situations. TIMSS reports on the students' performance overall in mathematics, as well as on each one of the subdomains and cognitive domains. In Canada, TIMSS assesses the performance of students in Grade 4 (9-year-olds) and Grade 8/Secondary II (13-year-olds) in mathematics and science.

National Council of Teachers of Mathematics (NCTM) standards

NCTM presents 10 standards, five of which relate to content and five to process standards, in the *Principles and Standards for School Mathematics* (NCTM, 2000) and *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence* (NCTM, 2006). The content standards are: numbers and operations, algebra, geometry, measurement, and data analysis; the five process standards relate to problem solving, reasoning and proof, communication, connections, and representations. Each jurisdiction then defines mathematics to suit the needs of its population and ministerial or departmental philosophy.

All Canadian provinces and territories use the NCTM documents as a starting point or guide for the development of their mathematics programs. In the western provinces and territories, these documents form the basis for *The Common Curriculum Framework for K–9 Mathematics* as they do for the Atlantic provinces' *Foundation for the Atlantic Canada Mathematics Curriculum*. Ontario and Quebec also consult the NCTM documents when constructing and revising their mathematics curricula.

Mathematics within the provinces and territories

Mathematics curricula within the various provinces and territories in Canada are structured around the NCTM content strands (referred to as subdomains in the PCAP mathematics assessment framework) and processes that specify the conceptual and procedural knowledge that students should acquire in school mathematics. They provide a comprehensive foundation for all students to reason and communicate mathematically and use mathematical knowledge and skills effectively in postsecondary education, the workplace, and daily life.

The content strands across provinces and territories are generally defined as: numbers and operations, patterns and relations, geometry and measurement, and data management and probability. Each province and territory defines a number of mathematical processes deemed to be essential to the effective study of mathematics. These generally include problem solving, reasoning, making connections within and outside the discipline, representing, and communicating. The processes reflect the manner through which students acquire and apply mathematical knowledge and skills and are interwoven throughout the content strands.

In recent years, much attention has been focused on 21st-century skills. These are usually described as those skills that individuals will have to master to succeed in the 21st century. They include creativity and innovation; critical thinking and problem solving; communication and collaboration; information literacy; media literacy; information and communications technology (ICT) literacy; flexibility and adaptability; initiative and self-direction; social and cross-cultural skills; productivity and accountability; and leadership and responsibility. They are seen as cross-curricular competencies that are present in mathematics, science, language arts, social studies, geography, and the arts.

Working definition of mathematics

Mathematics can be defined in a variety of ways. *The Report of the Expert Panel on Student Success in Ontario* (Ontario Ministry of Education, 2004) states that mathematics "is a fundamental human endeavour that empowers individuals to describe, analyse, and understand the world we live in" (Ontario Ministry of Education, 2004, p. 9). Most dictionaries define mathematics as "the study of the measurement, properties, and relationships of quantities and sets, using numbers and symbols" (*The Free Dictionary*) or "the abstract science of number, quantity, and space studied in its own right or as applied to other disciplines such as physics, engineering, etc." (*Concise Oxford Dictionary of Current English*, 1990). The publication *Everybody Counts* (National Research Council, 1989) describes mathematics as "the science of pattern and order." This very simple definition of mathematics challenges a common view of mathematics as a discipline dominated by computation and rules without reasons and instead makes one think of mathematics as a science of things that have a pattern or regularity and logical order. Mathematics is finding and exploring this regularity or order and then making sense of it (Van de walle, 2004).

For the purpose of the PCAP assessment, mathematics is broadly defined as the study of patterns and relationships and as a discipline involving conceptual understanding, procedural knowledge, and processes.

The domain is divided into four strands or subdomains:

1. numbers and operations (properties, equivalent representations, and magnitude);
2. geometry and measurement (properties of 2-D figures and 3-D shapes, relative position, transformations, and measurement);
3. patterns and relationships (patterns, algebraic equations and expressions, and linear relations); and

4. data management and probability (data collection and analysis, experimental and theoretical probability)

The four subdomains incorporate several processes or 21st-century skills, such as:

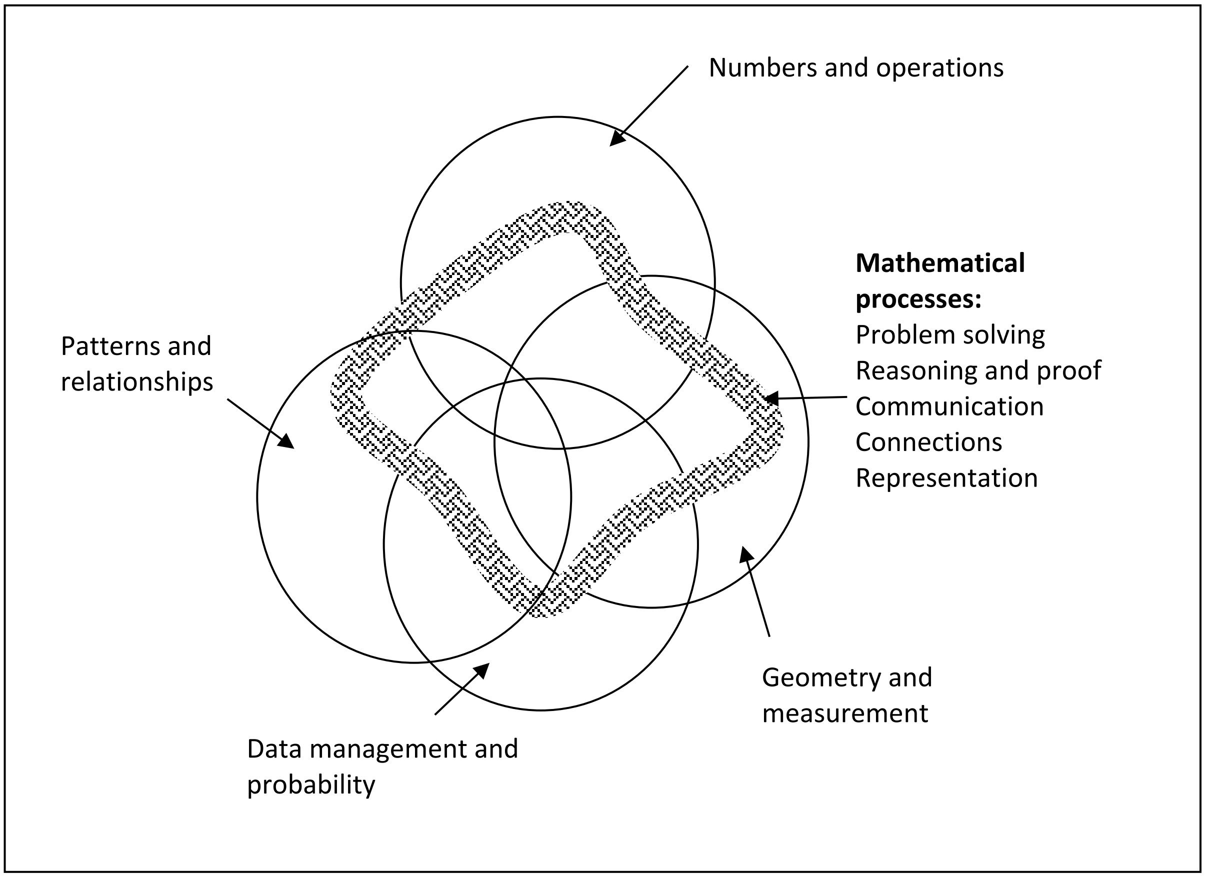
- critical thinking and problem solving
- creativity and innovation
- communication and collaboration
- information and communications technology (ICT) literacy
- flexibility and adaptability
- initiative and self-direction

The limitations of a large-scale assessment reduce the number of processes or skills that can be reliably assessed. Therefore only five NCTM process standards, which are a subset of the 21st-century skills set described above, have been selected for this assessment. They are:

- problem solving
- reasoning and proof
- communication
- connections
- representation

The subdomains are traditional groupings of conceptual and procedural knowledge as outlined in this framework, and the processes are present in all subdomains. As illustrated in Figure 2.1, the concepts and procedures of the subdomains intersect, while the processes are interwoven through all subdomains.

Figure 2.1 PCAP mathematics assessment framework



Assessment design

The majority of the items in the PCAP mathematics assessment are presented in groups within simple and relevant scenarios. The scenarios are distributed across four booklets to ensure each booklet is made up of items spanning all four subdomains and five processes. Booklets are designed so that a student would need approximately 30 minutes to complete each group of items.

The groups of items may contain selected-response and constructed-response items. The number of items per group may vary slightly, depending on the distribution of item types in the group. No group would contain only one type of item.

When mathematics is the primary domain, as it was in PCAP 2019, the student completes approximately two groups of assessment items. When mathematics is a minor domain, selected concepts and procedures in mathematics that cross over all strands—for example, proportionality—are chosen as the focus of the assessment. Anchor items are distributed across booklets to support accurate estimations of students' performance.

The assessment should be accessible to all participating students; therefore, the reading level and vocabulary used is consistent with what can be expected of Canadian Grade 8/Secondary II students. As well, information in the items is represented in a variety of modes (e.g., graphical, tabular, symbolical, written).

Specific considerations

1. *Use of calculators*: This assessment does not focus on students' ability to perform calculations but rather on their ability to choose the appropriate operation, demonstrate their understandings, and assess the relevance of their answer in a given situation. All students should, therefore, be allowed to use a calculator, preferably of the type they would normally use in their mathematics class. A calculator is also built into the online assessment. The decision to use or not to use a calculator should be the student's. Using or not using a calculator should have no effect on the student's performance on this assessment, and this must be a consideration in item design.
2. *Use of manipulatives:* The use of manipulatives (concrete objects) as teaching tools is encouraged in all provinces and territories, and they should be found in all schools. They should help and support students in developing a better understanding of concepts as they go from concrete to abstract representations. The assessment will be designed so that manipulatives are not supplied or required to perform the assessment tasks, but they will be permitted if the student requests them. They will be limited to what is normally available to the students in their mathematics class.

What the assessment measures

Specific conceptual and procedural knowledge being assessed

<u>Numbers and operations</u>

The student shows evidence that he or she can:

- demonstrate an understanding of the inverse relationship between perfect squares and square roots, multiplication and division, and addition and subtraction;
- find the exact square root of numbers that are perfect squares and the approximate square root of numbers that are not perfect squares;
- demonstrate an understanding of and find factors for numbers less than 100;
- find prime factorization of composite numbers and use it to find least common multiples of numbers less than 100;
- order and compare positive fractions and positive and negative decimals;
- generate equivalent expressions for percentages, fractions, and decimals;
- represent rational numbers with diagrams and on a number line;
- explain and apply the order of operations for decimals, fractions, and integers;

- demonstrate an understanding of the four operations (+, –, ×, ÷) on positive fractions, negative and positive decimals (× and ÷ decimals limited to two-digit multipliers and one-digit divisors);
- demonstrate an understanding of the four operations with integers;
- select appropriate operations to solve problems involving rational numbers (except negative fractions) set in contextual situations;
- describe ways to estimate sums, differences, products, and quotients of positive fractions and decimals;
- apply the commutative, associative, and distributive properties, and order of operations to evaluate mathematical expressions;
- demonstrate an understanding of percentages greater than or equal to 0%;
- demonstrate understanding of proportional relationships using per cent, ratio, and rate;
- use ratio and proportionality to solve problems involving percentages that arise from real-life contexts, such as discount, interest, taxes, tips, and per cent increase and decrease;
- recognize a proportional relationship from context, table of values, and graph and use to solve contextual problems;
- solve problems using proportional reasoning in the different subdomains, e.g., numbers and operations, geometry, probability.

<u>Geometry and measurement</u>

The student shows evidence that he or she can:

- compare and classify 2-D geometric polygons using appropriate geometric vocabulary and properties, such as line symmetry, angles, and sides;
- apply the relationships for the sum of the angles of a triangle to find the measures of missing angles and solve other problems;
- demonstrate an understanding of congruence of polygons;
- draw and describe the image of a combination of translations, rotations, and/or reflections on a 2-D shape (not on coordinate plane);
- identify and plot points in the four quadrants of a Cartesian plane using integral ordered pairs;
- demonstrate an understanding of the relationships among radii, diameter, and circumference of circles and use these relationships to solve problems;
- calculate the measures of the circumference and area of a circle and use the calculations to solve contextual problems;
- calculate the perimeter and the area of triangles, rectangles, and parallelograms and use the calculations to solve contextual problems;
- calculate the surface area of right prisms and use the calculations to solve contextual problems;
- identify, use, and convert among SI units to measure, estimate, and solve problems that relate to length and area.

Patterns and relationships

The student shows evidence that he or she can:

- represent linear patterns and relationships using words, drawings, tables, graphs, algebraic expressions, and equations;
- make connections among various representations of linear relationships (words, drawings, tables, graphs, algebraic expressions, and equations);
- use different representations of linear patterns and relationships to make generalizations, predict unknown values, and solve problems;
- demonstrate an understanding of the different meanings and uses of variables as a place holder, in rules, in formulae, as changing quantities, and as dependent and independent variables;
- translate statements describing mathematical relationships into one or more algebraic expressions or equations in a variety of contexts;
- evaluate algebraic expressions given the value of the variable within the set of rational numbers (except negative fractions);
- show that two or more expressions are equivalent by using properties such as commutative, associative, distributive, and order of operations;
- show that two equations are equivalent by using properties of equality; order of operations; and commutative, associative, and distributive properties;
- distinguish between algebraic expressions and algebraic equations;
- solve linear equations using the most appropriate method (concrete, inspection, trial and error, and algebraic) involving a one-variable term, $\frac{ax}{b} + c = d$, where a = 1, $b \neq 0$, and c and d are rational numbers, for integral solutions and to verify solutions;
- use linear equations to solve problems involving proportion and measurement problems (area, perimeter, unknown angles of polygons).

Data management and probability

The student shows evidence that he or she can:

- collect data:
  - formulate questions for investigation
  - select, justify, and use appropriate methods of collecting data (primary and secondary data; categorical, discrete, continuous data; sampling)
- organize and display data:
  - organize data into intervals
  - select, use, and justify an appropriate representation for displaying relationships among collected data (including circle, line, and bar graphs)

- analyze data:
    - make inferences and convincing arguments about a problem being investigated based on an interpretation and analysis of charts, tables, and graphs used to display given or collected data
    - evaluate data interpretations that are based on graphs, tables, and charts
- understand measures of central tendency:
    - describe a set of data and solve problems using mean and range
    - compare different populations using the mean and range
    - determine the effects of variation in data on measures of central tendency (outliers, gaps, clusters)
- understand probability concepts:
    - identify all possible outcomes of two independent events using tree diagrams, area models, tables, or lists
    - determine probability of a single or two independent events, and describe using fractions, decimals or percentages
    - use the probability of a single or two independent events to make predictions about a population
    - compare theoretical and experimental probabilities of a single and two independent events in appropriate contexts

In addition, five processes—problem solving, reasoning and proof, communication, connections, and representation—highlight ways of acquiring and using the content knowledge outlined in the above subdomains.

<u>Problem solving</u>

The student shows evidence that he or she can:

- solve multi-step problems presented in context that require using and making connections among mathematical concepts, procedures, and processes;
- solve multi-step problems presented in context that show evidence of understanding the problem, making a plan, carrying out the plan, estimating and evaluating the solution for reasonableness;
- explain the process used to solve a problem and verify the reasonableness of solutions by using numbers, words, pictures/diagrams, symbols and equations, estimation;
- apply a variety of problem-solving strategies to solve problems, such as drawing a picture or diagram, looking for a pattern, using "guess and check," making a table, working a simpler problem, or working backwards.

Reasoning and proof

The student shows evidence that he or she can:

- analyze a problem, make and assess conjectures, justify conclusions, and plan and construct an organized mathematical argument by applying logical reasoning (inductive, deductive) and mathematical knowledge;
- make and test generalizations from patterns and relationships using logical reasoning;
- use counter-examples to evaluate conjectures;
- evaluate mathematical arguments;
- select and use appropriately various types of reasoning (algebraic, geometric, proportional, probabilistic, statistical, quantitative) to solve problems presented in context.

Communication

The student shows evidence that he or she can:

- communicate mathematical ideas and solutions clearly and precisely to others using appropriate everyday and mathematical language, units of measurement, and a variety of representations (written, graphical, numerical, and algebraic);
- formulate clear and complete arguments using a variety of representations (written, graphical, numerical, and algebraic) to justify conjectures and solutions to problem situations;
- use symbolic language of mathematics correctly.

Connections

The student shows evidence that he or she can:

- recognize and connect mathematical concepts and procedures to contexts outside of mathematics, such as other curricular areas, personal life, current events, sports, technology, arts and culture, media;
- make connections between different representations (written, graphical, numerical, and algebraic) of mathematical ideas.

Representation

The student shows evidence that he or she can:

- create and use a variety of representations (written, graphical, numerical, and algebraic) to organize, record, and communicate mathematical ideas;
- connect, compare, and translate among different mathematical representations;
- select and apply the appropriate representations to solve problems.

Cognitive levels

The cognitive demands were defined by the reasoning required by the student to correctly answer an item, thus referring to the complexity of mental processing that must occur to answer a question, perform a task, or generate a solution. The three categories of cognitive demands are identified as low, moderate, and high.

Cognitive Level I (low)

The items at this level ask the student to:

- recall information (facts, procedures, definitions);
- identify properties;
- recognize an equivalent representation;
- perform a specific or routine procedure;
- solve a one-step (word) problem;
- retrieve information from a table or graph;
- identify a simple number or geometric pattern;
- draw or measure simple geometric figures;
- recognize an example of a concept;
- compute a sum/difference/product/quotient; and
- convert among different representations of a number (fraction, decimal, per cent).

For this level of item, the student is required to solve problems that have been determined to have a relatively low cognitive demand. Typically, a student at this level is able to retrieve information from a graph or solve previously learned routine problems and solve problems that require mostly recall and recognition.

Cognitive Level II (moderate)

The items at this level ask the student to:

- apply properties to evaluate an expression or find a measurement or solve a problem;
- represent a situation mathematically in more than one way;
- select, use, and interpret different representations depending on the situation;
- solve a contextual problem involving the use of more than one mathematical concept or procedure;
- retrieve information from a graph or table or geometric figure and use this information to solve a problem requiring multiple steps;
- extend a number or geometric pattern;
- formulate a routine problem given data and conditions;
- compare geometric figures or statements;
- compare two sets of data using the mean and range of each set;
- organize a set of data and construct an appropriate display;

- justify a solution to a problem with one solution; and
- interpret a simple argument.

Cognitive Level III (high)

The items, at this level, ask the student to:

- analyze properties;
- describe how different representations can be used for different purposes;
- perform procedures having multiple steps and multiple decision points;
- solve an unfamiliar problem;
- generalize a pattern and write the rule algebraically;
- formulate an original problem given a situation;
- analyze a deductive argument;
- justify a solution to a problem with multiple solutions;
- analyze similarities and differences between procedures and concepts;
- describe, compare, and contrast solution methods;
- interpret data from a series of data displays;
- formulate a mathematical model for a complex situation; and
- analyze the assumptions made in a mathematical model.

Assessment specifications

The following tables describe the percentage distribution of items by subdomain and by cognitive demand. For a valid comparison over time, anchor items will be selected to adequately represent each of the subdomains.

Table 2.1 Distribution of subdomains in PCAP Mathematics

Subdomain	
Numbers and operations	30–40%
Geometry and measurement	20–30%
Patterns and relationships	10–20%
Data management and probability	20–30%

Table 2.2 Distribution of cognitive demands in PCAP Mathematics

Level	Categories of cognitive demand	
I	Low cognitive demand	15–25%
II	Moderate cognitive demand	50–60%
III	High cognitive demand	15–25%

Performance-level descriptors

Performance-level descriptors can be used to demonstrate achievement in mathematics. The PCAP 2019 assessment is designed to report on how provincial performances measure up to the expected level of achievement on two factors: cognitive demand and degree of difficulty of the items. The cognitive demands are defined by the level of reasoning required by the student to correctly answer an item, from low demand to high demand, while the levels of difficulty are determined by a statistical determination based on the collective performance of the students on the assessment. See Table 2.3 for the operationalization of the descriptors for the four levels of performance.

Reporting this level of specificity will support provinces and territories in developing, adopting, and adapting education policies and programs so as to focus on continuous improvement. As cited in Crocker (2005), "It will also enable provinces and territories to improve their own assessments and to validate their results by comparing them to both national and international results" (p. 1).

Table 2.3 Performance-level descriptors

Level	Performance-level descriptors
1	Students at this level were able to: • recognize previously learned information (e.g. geometric shapes) • retrieve information from graphs, tables or diagrams • calculate percentages and translate between percentage and decimal notation • compare and order numbers, including decimal representations • solve previously learned routine problems with explicit instructions in the stem • solve problems with one step calculation, including problems with several one step calculations • identify single transformations (e.g., reflections)
2	Students at this level were able to: • recall facts, definitions or terms (e.g., parallel, perpendicular, range) • carry out calculations involving one or more operations, including operations of different types • use provided formulae • compare and order numbers, including fractional representations • identify the algebraic expression or equation for a given context • solve problems involving probability • solve problems that require proportional reasoning, including ratios • calculate straightforward perimeter and area in a non problem-solving context • evaluate a variable expression • retrieve information from tables, diagrams or graphs and apply it to solve a problem • solve problems that are clearly defined as to what is required

3	Students at this level were able to: • apply mathematical concepts to non-routine or unfamiliar situations • interpret information from tables, diagrams or graphs • generate the algebraic expression or equation for a given context • solve problems requiring algebraic and spatial reasoning (e.g., carry out multiple transformations, manipulate variable equations) • solve problems using relevant information and/or hidden assumptions • select appropriate strategies to solve a problem • make use of logic to support solutions • describe the relationship between quantities
4	Students at this level were able to: • solve problems that require complex reasoning at the analysis and synthesis levels • use appropriate and efficient strategies to solve problems • generalize patterns and write an algebraic rule • communicate mathematics clearly by explaining and justifying complete solutions • combine information from different mathematical domains to solve a problem (e.g., solve a problem requiring both algebraic reasoning and spatial sense) • make connections between a variety of representations in order to solve a problem

Understanding the performance levels in mathematics

For the PCAP mathematics assessment, the four performance levels were illustrated by test items and examples of student work showing the levels assigned to them and explanations for the scores in the PCAP 2010 public report (CMEC, 2011, pp. 19–20). A more comprehensive set of sample items will be available in a forthcoming issue of *Assessment Matters!*

Chapter 3. Science Assessment Framework

The PCAP science assessment framework delineates the conceptual framework for the science component of PCAP (PCAP Science). It is informed by the provincial and territorial science curricula of the participating populations (CMEC, 2005b).

This framework lays out a theoretical foundation based on current research and sound practices in the field of science education. It builds upon two other CMEC initiatives in Canadian science education: SAIP Science assessments and the *Common Framework of Science Learning Outcomes K to 12* (CMEC, 1997a). It provides a working definition for scientific literacy upon which assessment items are designed.

Context for developing a science framework

In 1984, the Science Council of Canada published a report entitled *Science for Every Student: Educating Canadians for Tomorrow's World*. Recommendations were organized around three general areas: science education for all, redirecting science education, and monitoring science education. The report endorsed the concept of science for all and described the importance to Canada of having its citizens acquire a good working knowledge of science concepts and develop inquiry skills to apply these concepts to the world around them. "Science education must be the basis for informed participation in a technological society, a part of a continuing process of education, a preparation for the world of work, and a means for students' personal development" (Science Council of Canada, 1984, p. 18).

Upon the release of *Science for Every Student*, science curriculum development in Canada's provinces and territories began to emphasize the importance of developing a scientifically literate population while continuing to encourage and support students who demonstrate a strong interest in the sciences and in possibly pursuing science-related postsecondary studies and careers.

In 1996, CMEC administered SAIP Science I as an "assessment of scientific literacy" (CMEC, 1996). The assessment items for SAIP Science were intended as an opportunity to ask students to relate their understanding of science to real-life situations that were familiar to them. "Students' knowledge of science concepts and their application to society around them, as well as the understanding of the nature of science, were measured by responses to multiple-choice and constructed-response questions. Questions were presented in groups within simple and common scenarios that required the application of knowledge to situations familiar to young people" (CMEC, 1996, p. 9).

SAIP Science I and SAIP Science II (administered in 1996 and 1999, respectively) included a practical task component that required students to demonstrate their ability to apply scientific inquiry and problem-solving skills to simple hands-on tasks. The practical task component was not administered in SAIP Science III (conducted in 2004).[5]

SAIP scoring was based on criteria outlined in the chapter entitled "SAIP Science Assessment Framework and Criteria" of the *SAIP Report on Science I Assessment* (CMEC, 1997b). Student achievement was assessed according to questions that addressed:

- knowledge and concepts of science:
    i. physical sciences – chemistry
    ii. physical sciences – physics
    iii. life sciences – biology
    iv. Earth and space sciences
- the nature of science
- the relationship of science to technology and societal issues

Learning science as a school subject involves more than learning about conceptual knowledge related to science or the skills required for scientific inquiry. It requires understanding that science is a human endeavour that uses processes to produce evidence-based knowledge and arguments to propose explanations about the natural world. These explanations may change over time as scientists strive to provide verifiable and reliable evidence that is defensible.

In 1997, CMEC published the *Common Framework of Science Learning Outcomes K to 12* as part of the Pan-Canadian Protocol for Collaboration on School Curriculum. The intent of this document was to provide direction for curriculum developers across Canada and to harmonize science learning when revising science curricula for their particular provinces and territories (CMEC, 1997a).

The common framework built upon the work of the Science Council of Canada and stated the following vision for scientific literacy in Canada:

> The framework is guided by the vision that all Canadian students, regardless of gender or cultural background, will have an opportunity to develop scientific literacy. Scientific literacy is an evolving combination of the science-related attitudes, skills, and knowledge students need to develop inquiry, problem-solving, and decision-making abilities, to become lifelong learners, and to maintain a sense of wonder about the world around them.
>
> Diverse learning experiences based on the framework will provide students with many opportunities to explore, analyse, evaluate, synthesize, appreciate, and understand the interrelationships among science, technology, society, and

[5] See also CMEC (1997b, 2000, and 2005a) reports for more details on SAIP I, II, and III assessments.

> the environment that will affect their personal lives, their careers, and their future. (CMEC, 1997a, p. 4)

The four foundation statements in the common framework delineated four critical aspects of students' scientific literacy upon which the pan-Canadian document is organized. Although presented separately, they were intended to be interrelated.

- *Foundation 1: Science, technology, society, and the environment (STSE)*—Students will develop an understanding of the nature of science and technology, of the relationships between science and technology, and of the social and environmental contexts of science and technology.
- *Foundation 2: Skills*—Students will develop the skills required for scientific and technological inquiry, for solving problems, for communicating scientific ideas and results, for working collaboratively, and for making informed decisions.
- *Foundation 3: Knowledge*—Students will construct knowledge and understandings of concepts in life sciences, physical sciences, and Earth and space sciences and apply these understandings to interpret, integrate, and extend their knowledge.
- *Foundation 4: Attitudes*—Students will be encouraged to develop attitudes that support the responsible acquisition and application of scientific and technological knowledge to the mutual benefit of self, society, and the environment.

PCAP builds upon the earlier work of SAIP and reflects the changes in Canadian science curricula since SAIP was administered in 1996. It also reflects our evolving understanding of effective assessment instruments since SAIP.

A literature review of Canadian Grade 8/Secondary II science curricula conducted in preparation for PCAP (CMEC, 2005c) clearly identifies scientific literacy as the goal of science education in all Canadian provinces and territories. This framework provides a working definition of scientific literacy for PCAP Science that underpins the design of this PCAP assessment component.

The PCAP science assessment framework:

- describes the competencies and subdomains of PCAP Science;
- recommends using contexts that provide opportunities for students to demonstrate their use of science-related attitudes, skills, and knowledge;
- describes the types and characteristics of the assessment items;
- contains tables of specifications to guide item development;
- discusses scoring and reporting scales; and
- includes an appendix with samples of five PCAP Science assessment units.

This framework also takes into account findings from large-scale international assessments.

Large-scale assessments in science

There are two major international assessments for science in which many Canadian provinces participate: PISA, conducted by the OECD; and TIMSS, administered by the International Association for the Evaluation of Educational Achievement (IEA).

PISA science assessment

PISA is an international assessment of mathematics, reading, science, problem solving, and financial literacy of 15-year-old students. It defines the domain of science as encompassing both “knowledge *of* science” and “knowledge *about* science” (OECD, 2013). Knowledge *of* science refers to knowledge of the natural world across the major disciplines of physics, chemistry, biological sciences, Earth and space sciences, and science-based technology. Knowledge *about* science refers to knowledge of the means (scientific inquiry) and goals (scientific explanations) of science.

TIMSS

TIMSS is an assessment of intended science curriculum content for various grade levels. TIMSS items are developed through an analysis of curriculum policies, textbooks, and other curriculum materials in use in participating countries.

The following table is a comparison of PCAP, PISA, and TIMSS science assessments.

Table 3.1 Comparison of PCAP, PISA, and TIMSS science assessments

PCAP	PISA	TIMSS
National assessment	International assessment	International assessment
Grade 8/Secondary II	15-year-olds	Grade 4 and Grade 8/Secondary II
three-year cycle (science as major domain, 2013)	three-year cycle (science as major domain, 2006 and 2015)	four-year cycle

Why scientific literacy?

There is general consensus that scientific literacy in an important goal for school science (Bybee, McCrae, & Laurie, 2009; Heinsen, 2011; Osborne, 2007; Roberts, 2007, 2011). This notion is reflected in the science curriculum documents not only of Canadian provinces and territories but in the documents of other countries too, such as the US National Research Council’s *A Framework for K–12 Science Education: Practices, Crost-Cutting Concepts and Core Ideas* (National Research Council, 2012) and *The Australian Curriculum: Science*, that country’s science curriculum from Kindergarten to Year 10 (ACARA, 2012). Much has been written about what should be included as part of a science curriculum in order to promote a scientifically literate population (Bybee, 1997; Fensham, 2000; Hodson, 2002).

Scientific literacy reflects the emphasis of “science for all” and is inclusive of both those who choose to pursue further study in science and those who choose other careers and interests that are not specific to science. Both science and technology are creative human endeavours with a long history in all cultures of the world. The intent of scientific literacy is to appreciate the nature of science and technology, the relationships between them, and their social and environmental contexts. Scientific literacy pertains to the application of science and how it helps or hinders humankind. It involves social issues and careers.

Scientific literacy also involves using knowledge to critically assess information, and it is important for Canadians to be able to make informed decisions about science-related issues that society faces, which can include:

- the usefulness of science to society;
- the negative effects or unintended consequences of science;
- scientific principles that could enable scientific research or result in the development of new or improved technologies;
- issues related to science, taking into account personal, community, and environmental factors;
- social issues; and
- careers.

Almost daily, we are bombarded with science-related issues that affect our environment, our health, our food, and our economy. A scientifically literate person may be better able to draw appropriate conclusions from the evidence and information that is provided by others and to distinguish personal opinion from evidence-based statements. He or she may also be better prepared to distinguish the kinds of questions and problems that can be solved by science and technology from those that cannot be answered in these ways.

Defining scientific literacy

Although recognized as a goal of school science, the term “scientific literacy” continues to elude a clear definition (Osborne, 2007; Roberts, 2007, 2011). Hodson (2006) suggests that there are some commonalities and that a reasonable definition of scientific literacy should include:

- a general understanding of some of the fundamental ideas, principles, and theories of science;
- some knowledge of the ways in which scientific knowledge is generated, validated, and disseminated;
- some ability to interpret scientific data and evaluate their validity and reliability;
- a critical understanding of the aims and goals for science and technology, including their historical roots and the values they embody;

- an appreciation of the interrelationships among science, technology, society and the environment; and
- an interest in science and the capacity to update and acquire new scientific knowledge and technological knowledge in the future. (p. 294)

Roberts (2007, 2011) describes two visions of scientific literacy, with Vision I focusing on science looking within itself as well as its products and processes, and Vision II focusing on the situations in which science plays a role in society and everyday life.

PCAP definition of scientific literacy

PCAP Science defines "scientific literacy" as:

> a student's evolving competencies of understanding the nature of science using science-related attitudes, skills, and knowledge to conduct inquiries, to solve problems, and to reason scientifically in order to understand and make evidence-based decisions about science-related issues.

This definition is amplified in the following paragraphs to ensure the clarity of its intent.

… scientific literacy …

The PCAP definition of scientific literacy includes more than information recall. Using the term "scientific literacy" rather than "science" highlights the importance that PCAP places on assessing an understanding of the nature of science and the use of scientific knowledge and skills within societal and environmental contexts. The definition also acknowledges that the disposition to use scientific knowledge and skills is mediated by a student's attitudes toward science and the importance of engaging in science-related issues as a reflective citizen.

… evolving competencies of …

Scientific literacy is a continuously evolving process and is part of being a lifelong learner. The PCAP definition of scientific literacy recognizes that students continue to evolve and develop competencies as they move from grade to grade and mature into adulthood. The term "competency" is used to articulate the importance of students being able to identify questions or issues to pursue science knowledge that will inform the question or issue; to seek answers to practical problems requiring the application of their science knowledge in new ways; and *to reason scientifically* when making decisions based on an understanding of the relationships among science, technology, society, and the environment when engaging with science-related issues.

… understanding the nature of science …

A key aspect of scientific literacy is an understanding of the nature of science as a human endeavour. Some important characteristics of science include the type of questions posed and the approaches to data collection; the obligation to make connections to current and historical knowledge; the reporting of methods and procedures used in obtaining evidence; the use of logical, evidence-based arguments and explanations; addressing issues of relevance, reproducibility, validity, integrity, and accuracy; the tentative nature of knowledge claims; and an openness to skeptical review. Thus, science is always evolving, and new knowledge and theories supersede existing ones.

… using science-related attitudes, skills, and knowledge …

Being scientifically literate implies an understanding of the importance of understanding science and its role in and its interrelationships with technology, society, and the environment in order to make informed, evidence-based decisions on which to base one's actions. This requires applying science knowledge to science-based issues. Skills such as questioning and planning, data collecting, interpreting, and communicating, as well as attitudes such as an interest in and awareness of science-related issues, respect for scientific inquiry, and a sense of stewardship are brought to bear in a variety of science-related contexts.

… to conduct inquiries …

Carrying out scientific inquiries requires combining an understanding of how scientific studies are undertaken and the use of content knowledge, research skills, knowledge of the nature of science, and science-related attitudes to gather verifiable evidence that supports explanations of natural phenomena. It is understood that students must often acquire knowledge that is new to them, not necessarily through their own scientific investigations, but through libraries, the Internet, and other resources. Students should recognize important characteristics of scientific investigations and the types of answers one can reasonably expect from science.

… to solve problems …

While scientific inquiry involves answering questions, solving problems involves searching for solutions to practical problems. For PCAP, this includes applying science knowledge in solving problems, identifying criteria, and evaluating solutions.

... and to reason scientifically ...

Reasoning scientifically involves using evidence to draw conclusions or develop and use models. It includes the ability to identify relationships, analyze numerical and pictorial information, and understand the basis for and limitations of models.

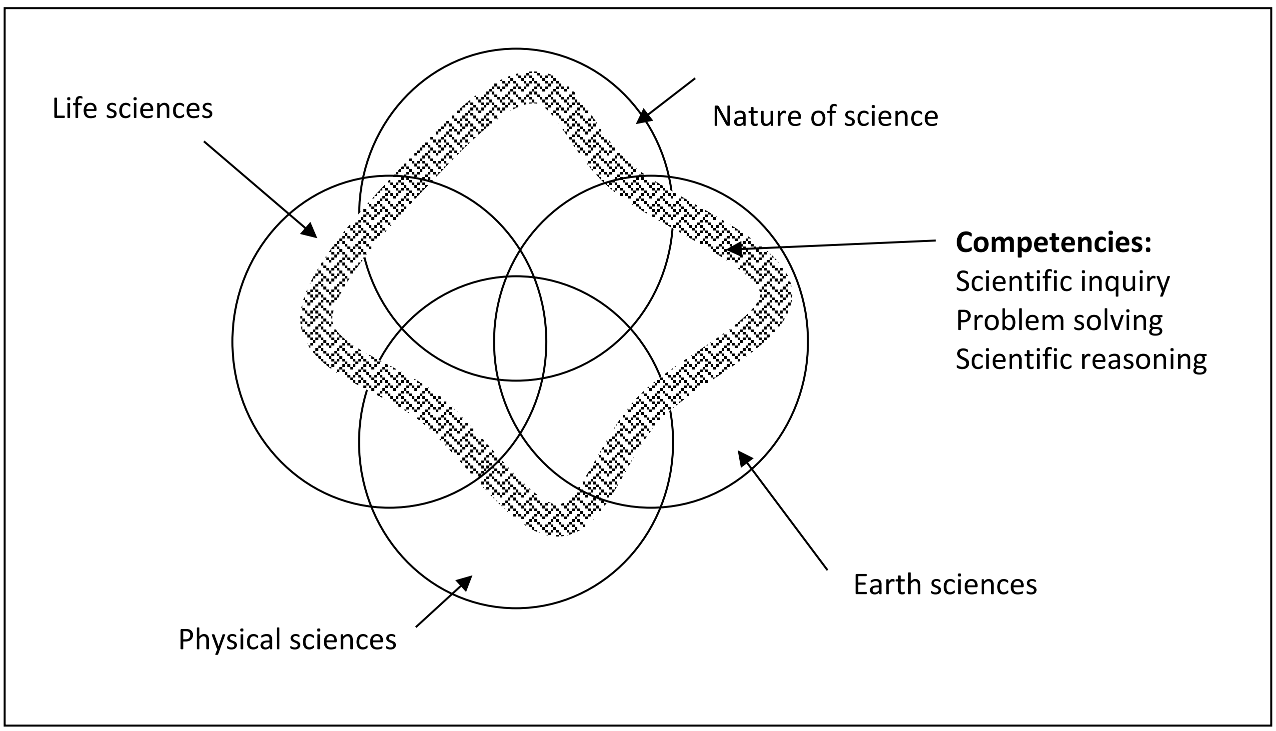
… in order to understand and make evidence-based decisions about science-related issues

Making evidence-based decisions implies knowing, selecting, and evaluating information and data critically using scientific reasoning. Science-related issues are omnipresent and vary in complexity, often including different perspectives such as political, economic, health, and public safety. Students need to recognize that, for many issues, there may not be sufficient information to make valid evidence-based decisions, rendering it necessary to be cautious in the interpretation and the communication of the decisions.

Organization of the domain of science

For PCAP assessment purposes, the domain of science is divided into three competencies, four subdomains, and attitudes within a given context. Figure 3.1 articulates the organization of PCAP Science as a primary domain for assessment. It reflects the intended science curricula for students in Canadian provinces and territories,[6] as well as the foundation statements in the pan-Canadian *Common Framework of Science Learning Outcomes K to 12* (CMEC, 1997a).

Figure 3.1 PCAP science assessment framework



[6] For updated science curricula, please visit the official websites of provinces and territories.

Competencies

An understanding of science is important for young people to be able to participate in and understand that science and technology affects their lives both in the present and in the future. Scientific literacy is developed when students are engaged in demonstrating the competencies of scientific inquiry, problem solving, and scientific reasoning. PCAP Science places a priority on being able to assess these competencies.

Scientific inquiry: Understanding how inquiries are conducted in science to provide evidence-based explanations of natural phenomena

Scientific inquiry requires students to address or develop questions about the nature of things, involving broad explorations as well as focused investigations (CMEC, 1997a). It is from the perspective of the student in that they focus on the “why” and “how” of science.

The PCAP assessment of students’ ability to use scientific practices provides evidence that they can:

- formulate hypotheses;
- make observations;
- design and conduct investigations;
- organize and communicate information;
- analyze and interpret data (e.g., using graphs and tables);
- apply the results of scientific investigations;
- select alternative conclusions in relation to the evidence presented;
- provide reasons for conclusions based on the evidence provided; and
- identify assumptions made in reaching their conclusion.

Problem solving: Using scientific knowledge and skills to solve problems in social and environmental contexts

Problem solving requires students to seek answers to practical problems requiring the application of their science knowledge in new ways (CMEC, 1997a). Students demonstrate this competency by applying their knowledge of science, their skills, and their understanding of the nature of science to solve science-related problems. It is part of the problem process that includes problem finding and problem shaping, where “problem” is defined as the state of desire to reach a definite goal.

The PCAP assessment of students’ ability to solve problems provides evidence that they can:

- define the problem;
- formulate questions;
- communicate the goals related to the problem;

- solve problems by recognizing scientific ideas;
- select appropriate solutions in relation to an identified problem;
- verify and interpret results (communicate, reflect);
- generalize solutions (recognize and apply science in contexts not typically thought of as scientific);
- provide reasons for the solution and how it meets the criteria to solve the problem;
- identify assumptions made in solving the problem; and
- show an awareness of sustainable development and stewardship when addressing problems.

Scientific reasoning: Being able to reason scientifically and make connections by applying scientific knowledge and skills to make decisions and address issues involving science, technology, society, and the environment

Scientific reasoning involves comparison, rationalization, or reasoning from the student in relation to an existing theory or frame of reference. Students demonstrate this competency by applying their knowledge of science, their skills, and their understanding of the nature of science to make informed, evidence-based decisions. They draw conclusions or make comparisons to an existing frame of reference or perspective. Students identify questions or issues and pursue science knowledge that will inform the question or issue.

The PCAP assessment of students' ability to reason scientifically provides evidence that they can:

- recognize patterns;
- develop plausible arguments;
- verify conclusions;
- judge the validity of arguments;
- construct valid arguments and explanations from evidence;
- connect scientific ideas and thereby build one on another to produce a coherent whole;
- use reasoning in order to make an informed decision for a particular issue in relation to the evidence;
- use reasoning in order to understand a science-related issue;
- provide reasons for the decision based on the evidence provided;
- identify assumptions and limitations of the chosen decision for that issue;
- develop and use models;
- show respect and support for evidence-based knowledge; and
- display interest in and awareness of science-related issues.

For each competency, students are assessed on their understanding and ability to critique the practices and processes related to these competencies.

Subdomains

The four subdomains targeted by PCAP Science are aligned to pan-Canadian science curricula and the *Common Framework of Science Learning Outcomes K to 12* (CMEC, 1997a). They include: nature of science, life sciences, physical sciences, and Earth sciences.

<u>Nature of science</u>

PCAP defines the nature of science as involving an understanding of the nature of scientific knowledge and processes by which that knowledge develops. Science provides a way of thinking and learning about the biological and physical world based on observation, experimentation, and evidence. Science builds upon past discoveries. Theories and knowledge are continually tested, modified, and improved as new knowledge and theories supersede existing ones. Scientific debate on new observations and hypotheses is used to challenge, share, and evaluate data through peer interaction and dissemination of information through written publications and presentations. "School education can develop the abilities to relate evidence to conclusions and distinguish opinion from evidence-based statements and so feed into the public understanding of science" (Fensham & Harlen, 1999, p. 762).

The PCAP assessment of students' understanding of the nature of science provides evidence that they can:

- understand how collecting evidence, finding relationships, and proposing explanations relate to the development of scientific knowledge;
- distinguish between processes and terminology that are scientific and those that are not;
- describe the processes of scientific inquiry and problem solving in evidence-based decision making;
- distinguish between qualitative and quantitative data;
- identify characteristics of measurement (e.g., replicability, variation, accuracy/precision in equipment and procedures);
- distinguish between various types of scientific explanations (e.g., hypothesis, theory, model, law);
- give examples of scientific principles that have resulted in the development of technologies; and
- demonstrate scientific literacy with respect to issues related to the nature of science.

The subdomains of life sciences, physical sciences, and Earth sciences are assessed using the following descriptors:[7]

[7] Please note that, although these descriptors reflect the commonalities of pan-Canadian curricula, they are not intended to be an exhaustive list.

Life sciences

- Explain and compare processes that are responsible for the maintenance of an organism's life.
- Describe the characteristics and needs of living things.
- Distinguish between cells and cell components.
- Describe the function and interdependence of systems related to inputs and outputs of energy, nutrients, and waste.
- Demonstrate scientific literacy with respect to issues related to life sciences.

Physical sciences

- Describe the properties and components of matter and explain interactions between those components [e.g., states of matter (i.e., solids, liquids, and gases); properties and changes of matter; particle theory; mass and volume].
- Demonstrate scientific literacy with respect to issues related to physical sciences.

Earth sciences

- Explain how water is a resource for society.
- Explain patterns of change and their effects on water resources on Earth (e.g., water distribution; weather; weathering and erosion; effect of water on regional climates).
- Demonstrate scientific literacy with respect to issues related to Earth sciences.

NOTE: Although the interrelationships between science and technology are an important part of developing scientific literacy, it must be emphasized and made clear that PCAP Science is not designed to assess the technological literacy of students writing this assessment.

Attitudes

Attitudes toward science determine students' interest to pursuing scientific careers (Osborne, Simon, & Collins, 2003). Since the creation of new scientific knowledge is essential for economic growth, students' attitudes toward science are a subject of societal concern and debate in many countries (OECD, 2006a).

To analyze students' attitudes, PCAP Science assesses:

- interest in and awareness of science-related issues;
- respect and support for evidence-based knowledge; and
- awareness of sustainable development and stewardship.

Assessment design

General design of the assessment

PCAP Science is organized into "assessment units" that provide a developmentally appropriate context for which specific questions (assessment items) will follow. Assessment units comprise an opening situation usually followed by three to six assessment items assessing both a competency and a subdomain.

Any text assumes that students have a degree of reading literacy. In PCAP Science, text selections are at a level that is accessible to the vast majority of Grade 8/Secondary II students. Questions that predominately assess reading or mathematics are avoided. The vocabulary is consistent with the level of understanding that can be expected of Canadian students at this level.

Since PCAP Science is an assessment of scientific literacy, each assessment item is coded to one of the three competencies and one of the four subdomains. Attitude items are embedded in the units within specific contexts.

Contexts

Each assessment unit has a context that is interesting and relevant to Grade 8/Secondary II students and relates to the science, technology, society, and environment (STSE) component of Canadian science education. Health, sports, media, the environment, and consumerism or consumption are possible areas of application relevant to Grade 8/Secondary II students where science and technology has an effect on their lives. Developers of the assessment items ensure that the contexts are developmentally appropriate and not culturally or geographically dependent.

The selection of contexts is mindful of the scientific competencies, understanding, and attitudes that students have acquired by the end of Grade 8/Secondary II. In the majority of Canadian education systems, this grade marks a transition period in the curriculum.

PCAP Science recognizes that as students advance from grade to grade, their ability to use science-related attitudes, skills, and knowledge builds over time. Students who are in Grade 8/Secondary II are at the stage of early adolescence and experience dramatic changes in physical, intellectual, social, and emotional growth. Their social and environmental experiences are more personal and local, although they are highly curious and can relate to real-life problems and situations. It must be recognized that, in this period of early adolescence, Grade 8/Secondary II students are idealistic, have a strong sense of fairness, and are reflective and introspective in thoughts and feelings. They confront moral and ethical questions head-on and have a willingness to learn new things they consider useful (Forte & Schurr, 1993).

The contexts chosen for PCAP science assessment units were intended to captivate the interests of Canadian Grade 8/Secondary II students and thereby increase their motivation to participate in writing the assessment.

PCAP Science item developers identify appropriate and relevant contexts for Canadian Grade 8/Secondary II students. A question they asked themselves was "What is important for Grade 8/Secondary II students to know, value, and be able to do with respect to understanding science within a situated context?" Contexts must be relevant to students' interests and lives and need to be sensitive to linguistic and cultural differences. The context for the assessment unit is introduced through an opening situation and could be in the form of a brief narrative and include tables, charts, graphs, or diagrams.

Contextualized embedded attitude items

The vast majority of Canadian provinces and territories include the development of positive attitudes as an important component to be embedded within science teaching and learning. This importance must be mirrored in PCAP Science. PCAP Science gathers data about students' attitudes using both contextualized embedded attitude items and a student questionnaire (when science is the major domain). Gathering data about students' attitudes both in and out of context will provide data on whether attitudes vary between these two types of methodologies and how this affects achievement. Hidi and Berndorff (1998) argue that situational interest can have an important effect on both cognitive and motivational functioning; however, investigations of its role remain "haphazard and scattered." By using both contextualized attitude items and a student questionnaire, PCAP Science could provide data to further this area of research.

The framework for PISA 2006 (OECD, 2006b) describes two types of contextualized embedded attitude items that are worthy of including in PCAP Science. These are "match-the-opinion" items and items consisting of a set of Likert-style responses.

"Match-the-opinion" items require students to choose the opinion that best matches their own from a list of four ordered opinions about an issue. Each given opinion represents a different level of commitment toward evidence-based knowledge and toward sustainable development and stewardship.

In a Likert-style item, students are asked to indicate their agreement to specific statements. Different scales may be used to operationalize statements in a given context. In PCAP Science, the criteria used to define "interest in science" have response scales that use a format that indicates interest (e.g., high interest, medium interest, low interest, no interest) rather than agreement (e.g., strongly agree, agree, disagree, strongly disagree), which tends to favour responses thought to be "socially desirable." Likert-style items are efficient and minimize demands on student response time.

PCAP Science contains sufficient attitude items to prepare a reliable scale; however, responses to the attitude items will not typically be included in the overall score for scientific literacy. Nevertheless, they will provide an important component of profiling student scientific literacy.

Several, but not all, assessment units contain contextualized embedded attitude items.

Challenges for assessment unit developers

The overall purpose of PCAP Science is to measure scientific literacy, and assessment items therefore focus on critical components (competencies and subdomains) that contribute to scientific literacy. This is different from traditional test items in science, which may have a greater emphasis on knowledge recall and application. All items must be clearly mapped to both a competency and a subdomain. PCAP Science items provide students with the opportunity to demonstrate both competency and knowledge regarding the practices of science.

Limitations of the assessment tasks

Although the design of this framework has been consistent with the intent of science curricula across Canada, PCAP Science is not a comprehensive assessment that includes every aspect of content knowledge, skills, and attitudes that are in every science curriculum for Canadian Grade 8/Secondary II students.

PCAP Science does not have a performance-based practical component. Performance-based tasks usually require observation and the completion of a product or practical task. The time constraints for PCAP Science are such that the test is written by students in 90 minutes, with each booklet containing assessment items for both the primary domain (mathematics) and the minor domains (science and reading). These time constraints, as well as financial considerations, do not allow for practical items. Teamwork and cooperative skills identified as important in Canadian science curriculum documents are not evaluated in this domain.

Assessment specifications

A table of specifications is a guide for assessment that indicates the emphasis placed on the measurement of students' understandings within various learning domains and reflects the degree of curricular commonality among Canadian provinces and territories. Table 3.2 summarizes the percentages devoted to each competency and subdomain in the assessment.

Table 3.2 Distribution of competencies and subdomains in PCAP Science

Competencies	
Scientific inquiry	30–40%
Problem solving	15–25%
Scientific reasoning	40–50%

Subdomains	
Nature of science	20–30%
Life sciences	20–30%
Physical sciences	20–30%
Earth sciences	15–25%

Understanding the performance levels in science

To meet the intended goal of PCAP Science to be an assessment of scientific literacy, the development of numeric scales of student achievement is required. The process of arriving at these scales is iterative and draws upon past experiences of assessing science achievement; it is also informed by the research into the cognitive development of science. Reporting scales need to be revisited each time science is the primary domain.

For the PCAP 2013 science assessment (O'Grady & Houme, 2014), the four performance levels were illustrated by test items and examples of student work showing the levels assigned to them and explanations for the scores. A comprehensive set of sample items is available in issue 8 of *Assessment Matters! PCAP 2013: Science resources for teachers*.

Chapter 4. Reading Literacy Assessment Framework

This chapter delineates the conceptual framework of the reading component of PCAP. It is informed by the curriculum objectives, goals, and outcomes of the participating provinces and territories.[8] As well, it reflects current research findings and best practices in the field of literacy development and the learning of reading.

In Canada, all curricula seek to develop student literacy in the broadest sense of the word, including the ability to understand, critically analyze, and create a variety of forms of communication (i.e., oral, written, visual, digital, and multimedia). These curricula recognize that reading is a cross-curricular skill necessary in all school subjects, as well as a life skill with applications beyond the classroom. This particular PCAP framework design was shaped by careful attention to Canadian curriculum guidelines for those classes and grades that serve Grade 8/Secondary II students. Consequently, it reflects provincial and territorial language-arts curricula, of which literacy is an integral component.

The framework lays out a theoretical foundation based on sound research and practice. It establishes a practical blueprint for the test and defines reading literacy and its elements. It describes the subdomains of this reading literacy assessment and identifies the types of texts and the characteristics of the items. The test design, including tables of specification, is provided, along with rationales for the various elements and descriptions of performance levels.

Theoretical background for reading comprehension

Our understanding of the reading process has evolved over time, leading to a number of different theories and models in linguistic and cognitive fields (see Ruddel & Unrau, 2004; Rayner & Reichle, 2010; and Snowling & Hulme, 2005, for reviews). One of the most influential theories today remains the theory of van Dijk and Kintsch (1983), which was extended to the so-called “construction-integration” model (Kintsch, 1988, 1998). This model describes all steps of reading comprehension, from decoding words to constructing coherent textual representation. According to researchers, the process of comprehension operates at two levels:

- a textbase—corresponding to a knowledge structure derived from information within the text; and
- a situation model—consisting of propositions that link the text to world knowledge and personal experiences.

Thus, the textbase represents the actual meaning of a text, and the situation model refers to the situation described by the text. The information that is derived directly from the textbase is usually insufficient for full comprehension. Therefore, in addition to the text, the situation model involves prior knowledge stored in long-term memory. Integration of this knowledge

---

[8] For updated reading curricula, please visit the official websites of the provinces and territories of Canada.

helps fill in the gaps in the text and create a mental representation that is complete and coherent. Because background knowledge differs from one person to another, reading of the same text results in the construction of different situation models. Featuring personal associations, inferences, and personal experiences, such models would be subjective and unique.

The role of background knowledge in reading is also emphasized in schema theories that postulate the existence of abstract memory structures, such as frames (Minsky, 1975), scripts (Schank & Abelson, 1977), plans (Schank, 1982), or simply schemata (Anderson, 1984; Schallert, 1982, 1991). Schemata represent a sort of template that allows us to remember and recognize the information. In the context of reading comprehension, they enable us not only to recognize different text types (e.g., novels, detective stories, news articles, research articles, recipes, etc.), but to process and recall the texts as well. It is hypothesized that a text schema is activated in memory after reading the first paragraph and guides us through the rest of the text (Wallace, 1992). In education, schema theories have led to an awareness of why certain textbooks are difficult for learners and how the activation of appropriate text schemata in memory could improve teaching and learning (Harrison, 2004).

The importance of the reader's prior knowledge is also stressed in studies related to student vocabulary. Indeed, a high level of reading comprehension is not possible without adequate vocabulary (Biemiller, 2006). Many studies now conclude that vocabulary is a significant predictor of reading ability (Blachowicz et al., 2013; Scarborough, 2002; Storch & Whitehurst, 2002). For instance, Biemiller and Slonim (2001) found that children who were behind in vocabulary knowledge in Grade 3 remained behind throughout their entire schooling.

Overall, successful reading relies on the trio of author, text, and reader. Theories in both the psychology of reading and English literature are now concentrating on the latter: the reader and his or her knowledge (Harrison, 2004). Indeed, any text is incomplete without the reader's contribution, as it is the reader who makes meaning of it, brings his or her own experience to it, and resolves any inconsistencies in it.

A definition of reading literacy

While earlier PCAP assessments focused solely on the process of reading, PCAP 2016 and 2019 combined two terms: reading and literacy. Adding the term "literacy" broadens the meaning of the ability to read to include skills that will be relevant throughout life for attaining individual and societal goals (Mullis et al., 2009; Organisation for Economic Co-operation and Development, 2013; Smith et al., 2000).

For PCAP, reading literacy is defined as the ability to construct meaning from texts through understanding, interpreting, and responding personally and critically to text content in order to make sense of the world and participate in society. It also includes metacognitive competencies that allow for awareness and application of different reading strategies appropriate to a given context. Reading literacy effectively involves the interaction of reader, text, purpose, and context before, during, and after reading.

The reader

In order to make meaning of a text, the reader must make a connection between what is in the text and what he or she knows or brings to the text. The reader's personal experiences, real or vicarious, allow greater or lesser access to the content and forms of what he or she reads.

Students have varying degrees of:

- knowledge of and about language and texts;
- facility with language strategies;
- knowledge of the way language works in print and in the digital world.

Each bullet is elaborated below:

1. Knowledge of language refers to vocabulary, syntax, punctuation, text structures, and rhetorical devices.
2. Facility with language strategies includes those used before, during, and after reading, such as accessing prior knowledge of content and form or type of text; making predictions; making connections; asking questions during reading and building mental images; determining key ideas and noting important supporting details; using "fix-up" strategies when meaning fails; making inferences; synthesizing; assessing the validity of content; making comparisons with other sources of information; summarizing; and the like.
3. Knowledge of the way language works in print and in the digital world may include the ways in which linear text (or hypertext), formatting practices, visual additions, and the general structuring of text on the page (or web page) affect the construction of meaning in the text. These elements have become more significant in contemporary websites and in promotional texts in particular.

The text

Definitions of "text" have evolved over time in parallel with changes in technological culture and society. In the modern world, the notion of "text" has expanded and is now used to describe *any language event* (see, for instance, the *Foundation for the Atlantic Canada English Language Arts Curriculum, K–12*[9]). In this context, communication that uses words, graphics, sounds, and/or images in print, oral, visual, or digital form to present information and ideas can be considered a text. This expanded concept of text takes into account the diverse range of language forms with which people interact and from which they construct meaning.

Students must engage with a variety of print and digital texts, such as those generally considered fiction, non-fiction, or a combination of the two. Examples could include: short

---

[9] Retrieved from http://www.ed.gov.nl.ca/edu/k12/curriculum/documents/foundation/

stories, poetry, novels, plays, video clips, pamphlets, labels, instructions, magazine articles, editorials, websites, or online exchanges. Within that range, texts have different degrees of complexity in terms of structure, vocabulary, syntax, organization, ideas, rhetorical devices, and subject matter. The form or type of a particular text plays a part in determining students' success in accessing it.

The reader's purpose

The purpose of the reading activity affects the reader's construction of meaning. Students read texts for a variety of purposes, ranging from the pleasure they get from the text's content and style to the practical information or point of view they acquire from engaging with it. The student's purpose for reading a particular text also influences the strategies and stance he or she takes. Texts of any type may be read for many different purposes. Whereas particular forms or types of text are often considered aesthetic or pragmatic in intention, the reader's purpose may differ from that intent. For example, students of social studies may be required to read a novel or access a website to develop knowledge of a particular culture, era, or event.

The context

Context is important in any reading act because it affects the stance the reader takes toward the text. Context refers specifically to the physical, emotional, social, and institutional environment at the time of reading. It includes where, when, and why the student is reading. One of the challenges of large-scale assessment, for example, is that it is inescapably a testing situation, which, in turn, influences the state of mind brought to the reading act. Pre-reading prompts in this test offer some sense of context beyond the testing situation.

As well, context refers more broadly to the world view of the reader. Any meaning constructed by a reader is a reflection of the social and cultural environment in which the reader lives and reads (Bruffée, 1986; Emerson, 1983; Gee, 1996; Heath, 1983; UNESCO, 2011). Peers, family, and community values affect the stance readers take as they engage with text. This interrelationship is described for print media by Johnston and Costello (2005):

> Although we often think of literacy as a set of all-purpose skills and strategies to be learned, it is more complex, more local, more personal, and more social than that. Becoming literate involves developing identities, relationships, dispositions, and values as much as acquiring strategies for working with print. (p. 256)

The interaction

Contemporary concepts of reading recognize that the process of reading involves the *interaction* of reader, text, purpose, and context before, during, and after reading. The interaction is critical for print media (Binkley & Linnakylä, 1997; Bruner, 1990) and even more important for digital media, where the sociocultural contexts are more complex (Legros & Crinon, 2002). There is also recognition that reading is not a finite set of discrete skills, knowledge, and concepts. Rather, it is a process of continuous growth in which readers constantly expand the boundaries of their understanding, interpretation and response to texts. In doing so, they refine the fluency of their integrated reading processes (Paris, 2005).

Subdomains of the assessment

In light of the interactive process of reader, text, purpose, and context, this assessment of reading literacy considers the reader's engagement with text and response to it. Curricula across Canada identify the following major aspects of reading literacy:

- understanding texts;
- interpreting texts;
- responding personally and critically to texts.

These three subdomains are parallel to Gray's (1960) distinction between "reading the lines," "reading between the lines," and "reading beyond the lines"—terms commonly used by Canadian teachers. In each of these categories, there will be different levels of complexity and difficulty. A few examples of types of questions are given below each subdomain description. This is not an exhaustive list and does not represent the full scope of the assessment.

Understanding texts ("reading the lines")

Understanding, or "reading the lines," refers to the process of constructing meaning based on the information directly included in the text. Students use a variety of appropriate strategies to confirm meaning in a variety of familiar and unfamiliar texts. They identify and use concrete and abstract vocabulary, stated conclusions, principal ideas, important details, and/or some aspects of the style and structure of the text. They also make straightforward inferences that are text-based and require very little effort from skilled readers (e.g., determining the referent of a pronoun, describing the link between two characters, etc.).

Students may demonstrate their ability to understand by:

- identifying principal ideas and differentiating them from secondary ideas;
- locating important details;
- using knowledge of vocabulary and cueing systems to make meaning in both familiar and unfamiliar contexts; and/or
- recognizing aspects of style, organization, links between elements, and/or complexity in the text.

Examples of questions targeting understanding include:

1. What were the key ideas in the information you read? Why are they important?
2. What word or phrase best describes the character?
3. Identify the ways that the main character uses to accomplish his or her mission.

4. Put the events of the story in order.
5. Select the graphic that best illustrates the main idea.

Interpreting texts ("reading between the lines")

In order to construct a coherent representation of the text, students need to develop an understanding of the relationships of discrete elements to the whole, or "read between the lines." Readers use symbols, patterns, text features, and other elements to analyze the story in narrative texts, the general idea in information texts, and the arguments in persuasive texts. They make high-level inferences, synthesize information, and draw conclusions about the broader meaning and intent of the text; that is, they consider relationships among elements and ideas in the text to construct deeper meaning and discern more significant implications.

Students may demonstrate their ability to interpret by:

- communicating a broader perspective and/or meaning of the text by recognizing relationships and integrating elements;
- identifying and supporting a thesis with references to details, events, symbols, patterns, and/or text features;
- making logical inferences referring to relevant textual details;
- analyzing and synthesizing elements of the text;
- relying on the text to inform meaning, draw conclusions, and/or connect aspects of the text to each other;
- relating visual elements (diagrams, graphics, photographs, etc.) to the text;
- establishing links between elements of the text and elements from complementary texts; and/or
- explaining how authors use various techniques to create meaning and achieve different purposes (e.g., symbolisms, text features).

Examples of questions targeting interpretation include:

1. Explain how the main character changed from the beginning to the end of the story and the events that led to those changes.
2. Why did the author italicize the four words in this article?
3. Explain how the two points of view in this interview are similar.
4. Using the three texts, explain why the speakers' attitudes are different from each other.
5. What ideas are common between the table, the graphics, and the text?

Responding personally and critically to texts ("reading beyond the lines")

In responding personally and critically to texts, readers go beyond basic comprehension, or "read beyond the lines." They may engage with the text in any number of ways, such as making personal connections between aspects of the text and their own prior experiences, knowledge, values, or points of view; responding emotionally to central ideas or aspects of the text; and taking evaluative stances about the quality or value of the text, possibly in relation to other texts and/or social or cultural factors.

Canadian curricula in reading generally distinguish between personal and critical responses.

<u>Personal response</u>

In personal responses, readers reflect on their own experiences in light of the text and/or identify with aspects of the text. They elaborate personal connections and reactions to the text by providing extended explanations, examples, and supporting arguments from their own experience and knowledge.

The reader may respond personally to the text by:

- identifying parallels and/or disconnections between his or her own prior experiences and elements of the text;
- expressing personal implications and insights;
- making connections supported by his or her own prior experiences, the text, examples, explanations, or thoughtful justifications; and/or
- using evidence (specific details, examples, citations) from the text and his or her own experience to explain his or her understanding of the argument.

Examples of questions targeting personal response include:

1. Which character's attitude most closely resembles your own? In what way?
2. After viewing the ad, would you consider donating to the program? Why or why not?
3. Do you feel empathetic toward the main character? Why or why not?
4. Does reading about another point of view make you think about this issue differently? Explain.

<u>Critical response</u>

In critical responses, readers stand apart from the text, considering the text as an artifact or object and evaluating its quality and/or appropriateness to the world at large. Readers evaluate content, elements of style, or the author's stance. They reflect on the choice of content, sources, quality, accuracy, or relevance of information, relationships, and ideas. Readers support their responses by providing specific, appropriate details and evidence from the text and other sources about issues, themes, characterization, and elements of style.

The reader may respond critically to the text by:

- evaluating elements of the text, based on social, cultural, and/or literary constructs;
- evaluating the quality, sources, accuracy, or relevance of issues, themes, and/or elements of style presented in the text;
- supporting his or her response with details, examples, explanations, or justifications;
- supporting his or her response with reference to the author's style (e.g., voice, stance, organization, structure);
- evaluating elements of the text (e.g., character development, believability, credibility, bias, stereotypes, intrigue);
- recognizing ways in which print media contain bias;
- comparing the text with other sources; and/or
- identifying contradictions and ambiguities within the text and/or with a broader world view.

Examples of questions targeting critical response include:

1. Give examples of how the author's arguments were supported by credible evidence.
2. What do the types of sources in the reference list tell you about the author's biases?
3. Would you trust the information in this newspaper? Explain.
4. Are the feelings of the characters justified?
5. Explain how and why this reading selection changed your mind about the question.

Text types and forms

Texts come in a variety of types or forms that students read for practical or pragmatic purposes: continuous and non-continuous, print or digital, literary or informational, academic or recreational. These texts may include articles, instructions, websites, and other media texts with graphics and other visuals.

The PCAP 2019 assessment includes a range of text types and forms of varying levels of difficulty. These are broadly identified as fiction or non-fiction, recognizing that texts frequently mix forms. (While digital forms are still outside the scope of PCAP 2019, representations of information are diversified as much as possible to move in this direction).

Fiction

Fiction texts usually have a strong narrative aspect, including elements such as character, setting, conflict, plot, theme, and style. Most frequently, students are expected to engage with fiction texts primarily for literary and aesthetic purposes.

Literary reading involves two levels of text processing: (1) extracting information and (2) experiencing it as literature. The two levels can sometimes interfere with each another, and aesthetic experience can be lost if too much factual analysis is involved (Rosenblatt, 1980). Thus, even if literal comprehension of the text ("reading the lines") must usually precede experiencing it as literature (Church & Bereiter, 1983; Peskin, 1998), the degree to which this comprehension is necessary is open to question. For this reason, achievement tests should not call for pointless and unproductive information finding and inference. The challenge of PCAP is to get to the depth of literary comprehension and to test the most vital aspect of it—understanding a story in terms of people's motivations, goals, and social relations.

Non-fiction

Non-fiction texts, such as expository material (textbooks, essays, lab reports, newspaper articles), generally have a different structure from fiction. For example, expository texts explain information, ideas, or a perspective through definition, sequence, categorization, comparison, contrast, enumeration, process, problem/solution, description, or cause/effect. Some non-fiction texts, however, do include narrative elements.

Non-fiction texts also include those written to argue a particular perspective and those written to persuade the reader to take some particular stand or action (persuasion/argument). These texts may include advertisements, editorials, letters to the editor, and speeches. Frequently, they also include visual components.

When assessing the reading of opinion or argument, it is important to distinguish two components: (1) understanding the argument and (2) forming a judgment on the issue. The understanding part is a more direct reflection of reading competence, and it is also far more readily testable. Forming one's own opinion is influenced by a number of factors, of which understanding the opposing arguments is only one, but there are other factors that carry much more weight. Indeed, contemporary research indicates that people arrive at opinions and judgments quickly and spontaneously, bringing in evidence and logic after the fact to justify them (Kahneman, 2011; Stanovich & West, 2000). From a literacy perspective, the route to more rational opinion and judgment lies not in thinking-skill training but in promoting fuller comprehension. Therefore, this assessment will focus on *comprehending* the argument rather than evaluating it or forming one's own opinion. While these last are important in civic life, testing them is fraught with quite possibly insurmountable difficulties.

Assessment specifications

The weighting for the three subdomains in reading literacy to be assessed by PCAP 2019 are shown in Table 4.1.

Table 4.1 Distribution of subdomains in PCAP Reading

Subdomain	Percentages
Understanding texts	35–45%
Interpreting texts	25–35%
Responding to texts	25–35%

Sample texts and test items

An example of a PCAP reading unit can be found in the PCAP 2007 public report (CMEC, 2008, pp. 13–17). Sample questions accompanied by student responses show the types of knowledge and skills demonstrated by students at each level of performance as well as the explanations for the score. A more comprehensive set of sample items is available in issue 14 of *Assessment Matters! PCAP 2016: How is Grade 8 reading literacy assessed in PCAP?*[10]

[10] *Assessment Matters!* is a series of articles and research notes available on the CMEC website, at https://cmec.ca/459/Overview.html

Chapter 5. Questionnaire Framework

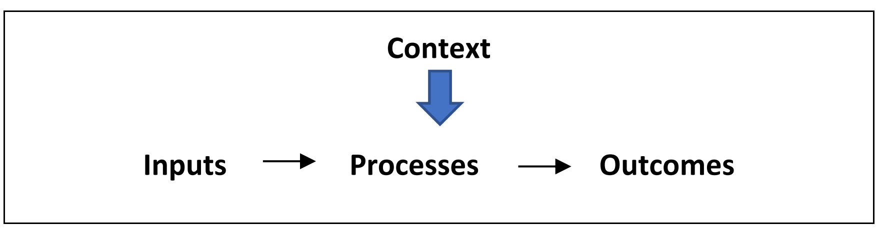
Context for developing a questionnaire framework

PCAP adopted the SAIP practice of administering questionnaires to students, teachers, and school principals. For 2019, the primary domain is mathematics, which means that greater emphasis will be placed on questions relevant to mathematics. Teacher questionnaires will thus target English- or French-language mathematics teachers. All questionnaires include specific questions about mathematics behaviours and strategies as well as more generic questions on student, teacher, and school characteristics.

The core model used in almost all large-scale assessment systems is based on a concept of educational productivity developed, either implicitly or explicitly, within a human-capital theory framework. The model is built around the straightforward concept that education is intended to achieve certain desired **outcomes**, specifically academic achievement, and that these outcomes are influenced by **inputs** and by the **processes** engendered by these inputs. Essentially, this model is a linear causal one, which may be depicted as

Inputs → **Processes** → **Outcomes**

It is generally recognized that education operates in an overall **context** determined by demographic factors, social and economic conditions, infrastructure, and other broad characteristics of the society in which the enterprise operates. The model is therefore better expressed in the following form, where context factors are thought of as overarching and influencing all of the others.



The importance of context is evidenced by the fact that inputs are strongly influenced by the resources a society is prepared to devote to education and by macro-level policies such as the public/private mix of schools or the years of compulsory attendance. The processes used depend on the resources and organization of the school system, the training of teachers, and other contextual factors. Finally, what counts as valued outcomes is influenced by society's perspective on the goals of education. In most countries, a high value is placed on achievement in core academic subjects—hence the emphasis on reading, mathematics, and science in most large-scale assessments.

While a direction of causality is implied by the model, it must be recognized that large-scale assessments yield only correlational data. Much of the research derived from large-scale databases thus requires relatively sophisticated statistical processes designed to help identify possible causal links from the many correlations that exist. Nevertheless, given the immense number of factors that may contribute to achievement, even the most complex models are insufficient to yield clear causal patterns.

Typically, educational productivity studies consider **educational resources** as inputs and either **attainment** (years of education) or **achievement** (knowledge and skills acquired) as outcomes. Contextual factors are considered in some research as extraneous variables which must be controlled in the explicit mathematical models used and which in other studies are the main independent variables of interest. Large-scale assessments have typically used achievement in core subjects as the outcomes but have taken a much more comprehensive approach to examining context, inputs, and processes. In particular, a large field of research exists around the process outcomes portion of the above model.

The SAIP model

The core model for the SAIP questionnaires was derived from the well-known Wang, Haertel, and Walberg (1993) synthesis of factors associated with school learning. The model is empirically based but has no explicit underlying theoretical framework. The concept of a proximal-distal continuum of influences is used to explain the evidence, which indicates that factors that touch more closely on the lives of teachers and students in the classroom are expected to be more influential than state or district policies. More specifically, the Wang, Haertel, and Walberg synthesis indicates that the five strongest influences on achievement are:

1. classroom management
2. metacognitive processes
3. cognitive processes
4. home environment/parental support
5. student/teacher social interactions

Motivation, peer-group influences, quantity of instruction, classroom climate, and other proximal variables are also important.

The factors with the least influence are:

1. district demographics
2. school policies
3. state-level policies
4. school demographics
5. program demographics

In general, the SAIP results are consistent with this model. In particular, classroom management variables, such as disciplinary climate and full use of time, have consistently been positively associated with achievement. Variables related to broad school policies and sources of influence on school activities and programs have tended not to be correlated with achievement.

The IEA model (TIMSS)

Early assessments conducted by the International Association for the Assessment of Educational Achievement (IEA) used curriculum as the starting point. Inputs, processes, and outcomes were represented by the concepts of the **intended, implemented, and achieved curriculum**, respectively. The intended curriculum is that represented by state-level curriculum documents and textbooks. The implemented curriculum is that actually taught in the classroom. This facet has often been used synonymously with the idea of **opportunity to learn**. The achieved curriculum is, of course, that which is measured by the assessment instruments.

This framework was retained for the long-term IEA study known as the Trends in International Mathematics and Science Study (TIMSS). However, TIMSS has broadened the traditional IEA approach by also starting with an elaborated version of the inputs-processes-outcomes model.

The PISA framework

The PISA assessment framework is built around a broad concept of **literacy** in the core subject areas. PISA thus has no explicit curriculum focus and, unlike TIMSS, is not based on analysis of curriculum documents. The PISA questionnaires are based loosely on Carroll's (1963) model of school learning, which places learning within a time framework. More explicitly, Carroll treats learning as a function of the ratio of time spent to time needed. Time spent and time needed are thought of in terms of more explicit factors such as available time, student ability, motivation, and quality of instruction. Questions about these constructs are included in the PISA questionnaires.

The early PISA questionnaire frameworks expanded on the time model by introducing variables derived from various syntheses, including the work of Scheerens and Bosker (1997). A key feature of the PISA framework is a "latent variable" approach in which important underlying constructs (such as socioeconomic status, student motivations, classroom processes, school climate) are represented by "index variables" derived from weighted combinations of the observed variables. Most of the research based on the PISA databases has used these index variables. A similar approach was attempted in SAIP, and index variables have been included in recent data files. However, these were derived after the fact and were not an inherent feature of the design. The SAIP index variables seem to have been less useful for research than those of PISA.

The draft PISA 2006 framework explicitly rejected the idea of developing a new model in favour of expanding on existing models. The PISA 2006 model is presented in the document *Contextual Framework for PISA 2006* (OECD, 2006c, pp. 10–14). An important new development for 2006 was the identification of seven **research areas** around which the questionnaires would be framed. This imposed a further dimension on the selection of items, beyond those inherent in the conceptual model. In practice, these research areas are not much different from those that have already been investigated in various PISA-based research projects. The exceptions are a couple of areas specific to the PISA 2006 major domain of science, namely **scientific literacy**

and the environment and **scientific attainment and the labour market**. The identification of possible research areas is also inherent in the PCAP design.

Limitations of existing models in the context of large-scale assessments

Questionnaires accompanying most large-scale assessment programs provide descriptive-comparative information on education systems. This is of considerable value in its own right and serves many of the same purposes as comparative achievement information by informing policy-makers about what is happening in their own and other systems. However, the questionnaires have also been designed to generate research around the productivity model. IEA studies, particularly TIMSS, as well as PISA have generated comprehensive databases that have stimulated a large amount of research over the past decade. Indeed, secondary analysis of large-scale assessment databases seems to have become a new research genre. In Canada, this type of research is also beginning to emerge and is explicitly part of the PCAP design.

While it is too early to expect any new synthesis to have emerged from all of this research, it seems fair to say that the ongoing work has yielded results that are reasonably consistent with earlier research and syntheses around educational inputs, processes, and outcomes. In particular, the research seems to support the concept of **structured teaching** as contributing to higher achievement as well as supporting the major features of Carroll's time model. Broad policy and resource variables do not seem to have much impact.

Work has also reinforced the importance of student characteristics and home environments and drawn attention to the **socioeconomic gradient**, or the magnitude of the difference in achievement across levels of socioeconomic status (SES). This gradient is seen as a measure of the relative equality or inequality of education systems. The goals of increasing average achievement and reducing disparities in achievement, especially those due to SES, have also become part of the discourse stimulated by this research.

Nevertheless, as research based on these databases becomes more intensive, the limitations of large-scale assessment and the associated questionnaire variables have become more obvious. Some of the main limitations are as follows:

- The results of large-scale assessments are characterized by a large number of small correlations. Isolating the effects of single variables or underlying traits thus requires complex models, with many variables. Such models are difficult to interpret and to convey to a policy-making audience.

- The correlational nature of large-scale assessments precludes the testing of true causal hypotheses. The statistical models typically used (e.g., multi-level modelling and structural equation modelling) may best be regarded as quasi-causal models in that they are capable of controlling some but not all extraneous variables.

- It is also difficult to develop firm policy recommendations from these correlational patterns, because no effect is decisive and because the associations are not usually strong enough to warrant significant investment of new resources on a particular approach to school organization, curriculum, or teaching.

- The associations between student characteristics and home-background variables and outcomes tend to be stronger than those between teaching and learning factors and outcomes. The implication drawn by some is that learning is thus largely determined by factors beyond the influence of schools and teachers.

- An argument can be made that the reason for higher correlations for SES and lower correlations for teaching and learning variables is that the latter are inadequately measured. In particular, while large-scale assessments typically measure cumulative achievement over as many as 10 or more years of schooling, the typical questionnaires used in such assessments measure, at most, the student's experience in a single year. The implication is that teaching and learning effects are systematically underestimated.

- Similarly, it may be argued that SES and student-characteristic measures (e.g., attitudes, self-concept) are also more stable long-term measures than the measures of teaching and learning and hence might be expected to show higher correlations with achievement.

- Questionnaires typically measure "perceptions" rather than actual events. This likely affects the accuracy of accounts of school and classroom activities more than the accuracy of measures of student characteristics.

- The questionnaires used in most large-scale assessments are of an "omnibus" nature, attempting to capture a large number of factors in a compact and easily administered format. The resulting lack of detail, especially about school and classroom practices, leaves many important questions unanswered.

- The previous point suggests that the descriptive-comparative function of large-scale questionnaires may not be entirely compatible with the research function. Answering research questions typically requires more information on a narrower range of items than is possible using an omnibus questionnaire format.

- An important example of the above limitation is found in studies of the impact of resources on outcomes. At best, large-scale assessment questionnaires yield information on a few proxies for resources, such as student-teacher ratios or teacher qualifications, but yield no direct information on resource uses or on the number of available resources that actually reach the students whose achievement is being measured.

- "Opportunity to learn" is a potentially powerful construct that has not been well operationalized in previous assessments. Can we make any progress in improving the definition and measurement of this construct? The major obstacle to doing this is the same as that encountered in SAIP, namely, that teachers of the target students do not necessarily know what their students have learned in earlier years.

Description of the PCAP questionnaire framework

The following principles guided the development of the PCAP questionnaires:

1. Include in the questionnaires some core descriptive data useful for both policy and research (e.g., student SES, school demographics, and teacher qualifications).
2. Other than core data, do not duplicate PISA.
3. Attempt to probe fewer areas in greater depth.
4. Identify policy-relevant issues.
5. Exclude areas shown by SAIP and PISA to be non-productive.
6. Focus on the primary domain in developing questions around teaching and learning strategies and behaviours.
7. Identify a limited number of areas that support the directions identified by the Pan-Canadian Educational Research Agenda (PCERA), even if these do not have obvious links to achievement in the primary domain.

In addition, the limitations imposed by the short-term cross-sectional nature of the data on teaching and learning were examined, and it was agreed that asking questions designed to delve into students' longer-term schooling experience should be attempted.

There was no clear sense that any of the existing frameworks are inherently better than others. In the same way that the PCAP assessment is neither explicitly curriculum-based nor literacy-based, a more eclectic approach to questionnaires, based on identified research priorities and on the need to link the questionnaires to the primary domain, is called for.

Core questions

The core section includes a limited number of questions that could be used mainly for descriptive purposes and as comparison or control variables in research models. Most of these are obvious and are included in almost all large-scale assessment questionnaires.

- Student questions:
  - gender identity
  - Indigenous identity
  - home background
  - SES
  - immigration status
  - home language
  - language of instruction
  - immersion

- Teacher questions:
    - teacher demographics
    - teacher qualifications and assignments
    - teacher professional development

- School questions:
    - school demographics and governance
    - community context
    - composition of the student body

Gender differences

Differences in reading achievement tend to favour females; differences in mathematics and science achievement tend to favour males but to a much smaller extent than the reading differences. One of the purposes of the student questionnaire was to uncover some potential explanations for gender differences in mathematics.

Technology differences

With PCAP 2019 being the first PCAP cycle administered online, is was important to include questions on technology, specifically the number and type of devices students have at home and their frequency of use. As well, students were asked questions about the use of virtual manipulatives and web-based resources in classrooms. This could help uncover potential explanations if differences in achievement were seen between the modes of administration.

Time allocation and use

Time has been a major feature of some other assessments, notably PISA. There is also a strong theoretical and empirical basis for time as a contributor to achievement. There have been some problems with the reliability of time measurement in PISA and SAIP. We would like to find ways to enhance the ability to measure time allocations and time loss by omitting previous variables that have little variance (e.g., length of school year) and by asking some more specific questions about engagement in school, such as questions on:

- time lost;
- time spent on subject areas;
- length of class periods;
- homework assignment and completion;
- out-of-school time relevant to learning; and
- absenteeism.

Inclusive classrooms

The questions in this category were intended to address some of the research and policy issues surrounding how to support students with learning disabilities or other challenges that may inhibit their progress in school. The focus was on students with lower levels of achievement (i.e., the bottom quartile) and especially those with identified disabilities requiring some form of special treatment in the school but who are not excluded from the PCAP assessment by virtue of these disabilities. The broad policy context around this area since PCAP began in 2007 was the strong movement toward inclusion of these students in regular classes. As a result, it was deemed inappropriate by the provincial and territorial coordinators to include questions about non-inclusion in the teacher questionnaire. This topic has been broadened to explore how teachers meet the various needs of the students in their classrooms by identifying instructional strategies and resources that teachers use.

Assessment

Many provinces and territories have responded to concerns about the performance of students and schools by implementing provincial and territorial assessment programs. These take different forms and are of different degrees of maturity in different provinces and territories. Assuming that the underlying goal of this policy direction is to improve and not merely to describe achievement or entrench current levels, there is good reason to examine assessment practices in the provinces and territories and particularly the uses made of these assessments. This area has been pursued in a limited way in SAIP and PISA. The intent in PCAP is to expand the scope of questions about assessment and possibly to tailor the questions to the specific features of provincial and territorial assessments.

Some questions included in the PCAP questionnaires have focused on:

- classroom assessment practices;
- teacher knowledge of assessment principles;
- school and teacher use of external assessments;
- strategies to prepare students for assessment; and
- existence and use of external (e.g., district, provincial) assessments.

Attitudes/motivations

This area is examined in some detail in PISA. Questions and constructs in this area are consistently found to be related to achievement. For the most part, this area can be researched using PISA and there is no need to duplicate what is found in PISA. The basic idea here is that PCAP should include only the minimal number of items needed to permit use of attitudes and motivations as control variables in research on teaching and learning strategies. An exception, because it is not included in PISA, is student attributions of success and failure.

Questions were included in the PCAP student questionnaire on:

- attitudes toward school;
- attitudes toward mathematics;
- attributions of performance in mathematics; and
- confidence in mathematics.

Cross-curricular connections

The student questionnaire briefly explored whether students saw connections between mathematics and other areas of learning.

Student learning strategies

The study of student learning strategies is considered one of the core elements of PCAP. This area is also largely subject-specific.

Teaching strategies

Both the SAIP and PISA questionnaires include lengthy lists of teaching strategies to which students (and teachers, in SAIP) are asked to respond. These include generic questions about disciplinary climate, use of time, and student-teacher interactions, as well as more subject-specific questions. Typically, these questions are about the student's or teacher's experience in a particular class in the year of the survey. Because of the narrow scope, this seems likely to result in systematic underestimation of the effects of teaching.

Rather than simply duplicating the kinds of items found on the SAIP and PISA questionnaires, an attempt has been made in designing the PCAP questionnaires to "reach back" to capture the student's longer-term classroom experience. While this will likely be difficult to do, it can, if successful, contribute to our understanding of students' broader school experience and how this relates to achievement. Topics include:

- teacher perceptions of what contributes to mathematics achievement;
- student perceptions of the types of activities used in their mathematics class; and
- school questions on overall instructional philosophy and the approach to mathematics learning.

Opportunity to learn

The TIMSS concept of opportunity to learn (OTL) stems from the core concepts of the intended, implemented, and achieved curriculum. The SAIP Science teacher questionnaires included an elaborated set of items, based on the test framework, asking teachers to indicate whether particular concepts had been taught. However, these questions did not yield much of value in SAIP and were deemed too complex to be used in PCAP, mainly because much of the information pertained to material taught in earlier grades. The PCAP questionnaire working group thus took the view that not much can be done with this aspect of OTL.

Derived variables

Questionnaire items fall into two broad categories: individual items requiring a single response (e.g., how do you identify yourself?) and item clusters, in which individual items represent some underlying scale (e.g., attitude to school or student learning strategies). Sometimes, the items are developed from some explicit theory (e.g., attribution theory for attributions of success and failure). However, the underlying scale is usually more implicit (e.g., attitude toward school is represented by a series of statements about liking various aspects of school).

In analyzing questionnaires in previous PCAP administrations, factor analysis was performed on various question clusters on the student questionnaire. The main goal of this type of analysis was to reduce the complexity of subsequent analyses by identifying, or making more explicit, a smaller number of derived variables or scales, representing clusters of the individual items. Factors are useful if they reduce the number of separate variables to be analyzed and if they can be interpreted in psychologically meaningful ways. For each derived variable, a "factor score" can be computed for each student. These factor scores may then be used to compare provinces/territories or to examine the effect of attitudes toward school on mathematics performance.

To the extent that the PCAP 2019 questionnaires contain items in common with those in PCAP 2007, 2010, 2013, and 2016, the factor analysis is expected to yield similar results to those in previous administrations. Specifically, the following variables in the PCAP 2019 questionnaires may have counterparts in previous questionnaires:

- attitudes toward school
- attitudes toward mathematics
- attributions of success or failure
- learning strategies
- out-of-school activities
- teaching strategies
- teaching resources, materials, and assignments
- disciplinary climate
- assessment strategies

Chapter 6. Assessment Design

Overview

PCAP is based on common curriculum outcomes across Canada. The assessment is not tied to the curriculum of a particular province or territory but is instead a fair measurement of students' abilities to use their learning skills to solve real-life situations. It measures how well students are doing; it does not attempt to assess approaches to learning.

PCAP 2019 is the fifth cycle of PCAP to be completed, and it again focuses on mathematics, defined through five processes (problem solving, reasoning and proof, communication, connections, and representation) that are interwoven throughout the four subdomains (numbers and operations, geometry and measurement, patterns and relationships, and data management and probability).

Provinces and territories work to ensure that the unique qualities of our country's education systems are taken into account. Factors such as linguistic differences, rural and urban school locations, and cultural influences are all considered in both the assessment itself and related context questionnaires. In addition, the common curricular framework for each subject incorporates an agreed-upon perspective for all provinces and territories that is based upon the latest pedagogical research.

One of the strengths of PCAP is its measurement over time of trends in student achievement in the three core subjects. The PCAP achievement scales provide a common metric on which provinces and territories can compare students' progress at the Grade 8/Secondary II level in the three core subjects from assessment to assessment. The scale midpoint of 500 is equal to the national average for each subject in the baseline year, i.e., the first year in which it was the primary domain (2007 for reading, 2010 for mathematics, and 2013 for science). Items that were administered in the baseline years will provide the basis for linking the assessment results. This will enable provinces and territories to have comparable achievement data from 2007, 2010, 2013, 2016, and 2019 and to plot changes in performance over this twelve-year period.

In addition to achievement scales for the three domains overall, PCAP 2019 reports relative student performance in relation to the performance-level descriptors defined in the mathematics assessment framework.

PCAP does *not* address individual student performance, nor does it involve comparisons between students, schools, or school boards. PCAP results are not made available to teachers, school boards, regions, or ministries/departments of education to assess students' school performance.

PCAP 2019 student booklet design

For the PCAP 2019 assessment, groups of assessment units were distributed within four booklets. Booklets are designed so that students need approximately 90 minutes to complete all of the items in any one booklet (approximately 60 minutes to be spent on primary-domain items and 30 minutes on minor-domain items). The four booklets were randomly and equally distributed to students within a single class.

To minimize the assessment burden on any one student, each student is presented with only a sample of the items. Following data collection, student responses are placed on common mathematics, science, and reading scales to provide an overall picture of the assessment results in each province and by language and gender. In addition, pairs of booklets containing sets or units of anchor items allow for comparative measurements of student performance from one booklet to another.

Each assessment unit presents a passage or context followed by a series of related items. The contexts chosen for assessment units are intended to captivate the interests of Canadian Grade 8/Secondary II students and thereby increase their motivation to write the assessment. Contexts are introduced with an opening situation that could be in the form of a brief narrative and could include fiction or non-fiction reading passages, tables, charts, graphs, or diagrams. Developers of the assessment items and the Advisory Panel on Test Fairness ensured that the contexts were developmentally appropriate, free of bias, and not culturally or geographically dependent.

Each booklet is composed of sufficient units that together span each of the assessment specifications for the three domains. All the assessment booklets contain a student questionnaire at the end of the booklet.

Texts and questions were developed in both official languages and cross-translated. Items were reviewed by curriculum experts and teachers from different regions in Canada in both French and English to ensure equivalency in meaning and difficulty. Following field testing, differential item functioning (DIF) analysis was conducted to ensure that items selected for the main administration were fair and equitable in both languages.

Characteristics of the items

In measuring any complex and integrated set of skills, it is usually best to include a variety of item types, both to allow all students to respond in the manner that best demonstrates their skill attainment and to measure a greater range of the complex skills involved.

In general, the assessment uses contexts that are complete in themselves, that are short enough to allow a range of text types currently read by the target age group both in and out of class, and that allow for a range of reading demands in a 90-minute time period. A balance of constructed-response and selected-response items allows for an efficient use of student testing time. The percentage of selected-response items is between 70 and 80 per cent and the percentage of constructed responses is between 20 and 30 per cent. Each selected-response item is worth one score point. Constructed-response items generally are worth one, two, or three score points, depending on the nature of the task and the skills required to complete it. In developing assessment items, the choice of item format depends on the competency or subdomain being assessed and the format that best enables the student to demonstrate his or her proficiency.

Selected-response characteristics

The traditional *multiple-choice* format comprises a stem statement and four choices, one of which is correct, while the other three function as distractors. This is the format most familiar to teachers and students. Each item focuses on a single subdomain. Scoring is dichotomous.

True-or-false/yes-or-no/agree-or-disagree items involve a series of statements about which students are asked to draw conclusions and specify whether each is true or false. Scoring is dichotomous.

Constructed-response characteristics

Constructed-response items require students to produce the response. Responses can range from short phrases or two to three sentences to several paragraphs in the case of extended constructed-response items. Constructed-response items may also ask the student to create tables or graphs, sketch diagrams, or design experiments. PCAP includes constructed-response items that are open-ended and measure higher-order cognitive skills and content knowledge.

The inclusion of constructed-response items also reflects good assessment practice in that different assessment formats are required, depending on what students are expected to demonstrate. Constructed-response items allow for partial credit, an important factor when assessing process skills or for items requiring multiple steps.

Releasing material to the public

PCAP 2019 is the fifth in a series of regular three-year studies providing data on trends in mathematics, science, and reading achievement over the twelve-year period from 2007 to 2019. PCAP will be administered again in 2022, and in 2025, and so on into the future. Since the outset of PCAP, a small selection of items has been released to describe the performance scales. Comprehensive sets of sample items for reading, mathematics, and science have been released as part of the publication series *Assessment Matters!*, a series of articles available on the CMEC website. The measurement of trends over time requires that a substantial proportion of the items remain secure; however, as items are released, new items will be developed to take their place.

Contextual questionnaires

Secondary analysis undertaken as part of the contextual report on student achievement in mathematics explores how resources and school and classroom conditions, as well as student characteristics and family circumstances, may impact mathematics achievement in Grade 8/Secondary II students. PCAP administers background questionnaires to students, their mathematics teachers, and their school principals. These questionnaires require about 30 minutes to complete.

Further research

PCAP's design provides for a research phase that follows the release of the public and contextual reports. A series of research articles on more specific topics will follow to provide a broader picture of the interplay between achievement and contextual variables. Wherever possible, comparisons will be made to other large-scale assessment projects in which the provinces and territories participate in order to develop a broader view of education across grades in Canada.

References

Anderson, R. C. (1984). Role of the reader's schema in comprehension, learning, and memory. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 243–258). Hillsdale, NJ: Erlbaum.

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2012). *The Australian curriculum: Science*. Australia: Australian Curriculum, Assessment and Reporting Authority.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2), (pp. 41–51). New York: Guilford Press.

Biemiller, A., & Slonim, N. (2001) Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, *93*, 498–520.

Binkley, M., & Linnakylä, P. (1997). Teaching reading in the United States and Finland. In M. Binkley, K. Rust, and T. Williams (Eds.), *Reading literacy in an international perspective.* Washington, DC: US Department of Education.

Blachowicz, C. L. Z., Fisher, P. J. L., Ogle, D., & Watts-Taffe, S. (2013). *Teaching academic vocabulary, K–8: Effective practices across the curriculum*. New York: Guilford Press.

Bruffée, K. A. (1986). Social construction, language and the authority of knowledge. *College English*, *48(8),* 773–790.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bybee, R. (1997). *Achieving scientific literacy*. Portsmouth, NH: Heinemann.

Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching, 48(8)*, 865–883.

Carroll, J. (1963). A model of school learning. *The Teachers College Record, 64*(8), 723–723.

Church, E., & Bereiter, C. (1983). Reading for style. *Language Arts, 60,* 470–476.

Concise Oxford Dictionary of Current English. (1990). (8th ed.) Oxford: Clarendon Press.

Council of Ministers of Education, Canada (CMEC). (1996). *SAIP Science*. Toronto: Author.

Council of Ministers of Education, Canada (CMEC). (1997a). *Common framework of science learning outcomes K to 12*. Toronto: Author.

Council of Ministers of Education, Canada (CMEC). (1997b). *SAIP 1996: Report on Science I assessment*. Retrieved from http://www.cmec.ca/250/Programs-and-Initiatives/Assessment/Pan-Canadian-Assessment-Program-(PCAP)/SAIP-1996-Science-I/index.html

Council of Ministers of Education, Canada (CMEC). (2000). *SAIP 1999: Report on Science II assessment*. Retrieved from http://www.cmec.ca/Publications/Lists/Publications/Attachments/136/saip1999.science2.en.pdf

Council of Ministers of Education, Canada (CMEC). (2002). *School achievement indicators program (SAIP): Report on Mathematics III assessment*. Toronto: Author.

Council of Ministers of Education, Canada (CMEC) (2005a). *Pan-Canadian assessment program: Literature review of mathematics assessment and test design*. Toronto: Author.

Council of Ministers of Education, Canada (CMEC). (2005b). *SAIP 2004: Report on Science III assessment*. Retrieved from http://www.cmec.ca/Publications/Lists/Publications/Attachments/62/SAIP-Science2004.en.pdf

Council of Ministers of Education, Canada (CMEC). (2005c). The pan-Canadian assessment program: Literature review of science assessment and test design. Toronto: Author (unpublished report).

Council of Ministers of Education, Canada (CMEC). (2008). *PCAP-13 2007: Report on the assessment of 13-year-olds in reading, mathematics, and science*. Toronto: Author.

Council of Ministers of Education, Canada (CMEC). (2011). *PCAP-2010: Report on the pan-Canadian assessment of mathematics, science, and reading*. Toronto: Author.

Crocker, R. (2005). The Pan-Canadian Assessment Program. A Concept Paper, Draft 2. Unpublished manuscript.

Earl, L. M. (2003). *Assessment as learning*. Thousand Oaks, CA: Corwin Press.

Emerson, C. (1983). Outer word and inner speech: Bakhtin, Vygotsky, and the internalization of language. *Critical Inquiry*, *10,* 245–264.

Fensham, P. (2000). Providing suitable content in the “science for all” curriculum. In R. Millar, J. Leach, & J. Osbourne (Eds.), *Improving science education: The contribution of research*. Buckingham, UK: Open University Press.

Fensham, P., & Harlen, W. (1999). School science and public understanding of science. *International Journal of Science Education, 21(7)*, 755–763.

Forte, I., & Schurr, S. (1993). *The definitive middle school guide*. Nashville: Incentive Publications.

Gee, J. (1996). *Social linguistics and literacies: Ideology in discourses* (2nd ed.). London: Falmer Press.

Gray, W. S. (1960). The major aspects of reading. In H. Robinson (Ed.), Sequential development of reading abilities. *Supplementary Educational Monographs, No. 90* (pp. 8–24). Chicago: University of Chicago Press.

Harrison, C. (2004). *Understanding reading development*. Thousand Oaks, CA: Sage Publications.

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge: Cambridge University Press.

Heinsen, L. (2011). Why scientific literacy must be a focus of science education: An argument for the literate citizen. *Alberta Science Education Journal, 42(1)*, 28–32.

Hidi, S., & Berndorff, D. (1998). Situational interest and learning. In L. Hoffmann, A. Krapp, K. A. Renniger, & J. Baumert (Eds.), *Interest and Learning*. Kiel, Germany: Institute for Science Education at the University of Kiel.

Hodson, D. (2002). Some thoughts on scientific literacy: Motives, meanings, and curriculum implications. *Asia-Pacific Forum on Science Learning and Teaching 3(1).* Retrieved from http://www.ied.edu.hk/apfslt/

Hodson, D. (2006). Why we should prioritize learning about science. *Canadian Journal of Science, Mathematics and Technology Education. 6(3),* 293–311.

Hoyles, C., Wolf, A., Molyneux-Hodgson, S., & Kent, P. (2002). *Mathematical skills in the workplace: Final report to the Science, Technology and Mathematics Council*. London: Institute of Education, University of London; Science, Technology and Mathematics Council.

Johnston, P., & Costello, P. (2005). Principles for literacy assessment. *Reading Research Quarterly, 40(*2*),* 256–267.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge: Cambridge University Press.

Legros, D., & Crinon, J. (2002). *Psychologie des apprentissages et multimedia*. Paris: Armand Colin.

Lewis, D. M., Mitzel, H. C., Mercado, R. L, & Schultz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed.). (pp. 225–254). New York: Taylor & Francis.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.

Mullis, I., Martin, M., Kennedy, A. M., Trong, K., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Boston: TIMMS & PIRLS International Study Center, Boston College.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics*. Reston, VA: Author.

National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academies Press.

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

O'Grady, K., & Houme, K. (2014). *PCAP 2013 report on the Pan-Canadian assessment of science, reading, and mathematics.* Toronto: Council of Ministers of Education, Canada. Retrieved from: http://www.cmec.ca/Publications/Lists/Publications/Attachments/337/PCAP-2013-Public-Report-EN.pdf

Olsen, L. (2002, May 22). Up close and personal. *Education Week* online. Retrieved from http://www.edweek.org/ew/articles/2002/05/22/37assess.h21.html?qs=%22up+close+and+personal%22

Ontario Ministry of Education. (2004). *Leading math success, mathematical literacy Grades 7–12: The report of the expert panel on student success in Ontario*. Toronto: Author.

Organisation for Economic Co-operation and Development (OECD). (2003). *PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills, 2003*. Paris: Author.

Organisation for Economic Co-operation and Development (OECD). (2006a). *Evolution of student interest in science and technology studies policy report*. Retrieved from http://www.oecd.org/science/sci-tech/36645825.pdf

Organisation for Economic Co-operation and Development (OECD). (2006b). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: Author.

Organisation for Economic Co-operation and Development (OECD). (2006c, May). *Contextual framework for PISA 2006*. Preliminary version. Paris: Author. Retrieved from https://www.acer.edu.au/files/pisa2006_context_framework.pdf

Organisation for Economic Co-operation and Development (OECD). (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: Author.

Organisation for Economic Co-operation and Development (OECD). (2014, February). *PISA 2012 results: What students know and can do—Student performance in mathematics, reading and science* (Volume I, Revised ed.). Paris: Author.

Osborne, J. (2007). Science education for the twenty-first century. *Eurasia Journal of Mathematics, Science & Technology Education, 3(3)*, 173–184.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*, 1049–1079.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40(*2*),* 184–202.

Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction, 16,* 235–263.

Rayner, K., & Reichle, E. D. (2010). Models of the reading process. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 787–799.

Roberts, D. (2007). Scientific literacy/science literacy. In S. Abell, and L. Lederman (Eds.), *Handbook of research on science education* (pp. 729–780). Hillsdale, NJ: Lawrence Erlbaum.

Roberts, D. (2011). Competing visions of scientific literacy: The influence of a science curriculum policy image. In C. Linder, L. Östmaan, D. Roberts, P-O. Wickman, G. Erickson, and A. MacKinnon (Eds.), *Exploring the landscape of scientific literacy* (pp. 11–27). New York: Routledge.

Rosenblatt, L. (1980). What facts does this poem teach you? *Language Arts, 57,* 386–394.

Ruddell, R. B., & Unrau, N. J. (Eds.). (2004). *Theoretical models and processes of reading* (5th ed.). Newark, DE: International Literacy Association.

Scarborough, H. S. (2002). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York: Guilford Press.

Schallert, D. L. (1982). The significance of knowledge: A synthesis of research related to schema theory. In W. Otto & S. White (Eds.), *Reading expository material* (pp. 13–48). New York: Academic Press.

Schallert, D. L. (1991). The contribution of psychology to the teaching of the language arts. In J. Flood, J. Jensen, D. Lapp, & J. R. Squire (Eds.), *Handbook of research on teaching the English language arts*. New York: Macmillan.

Schank, R.C. (1982). *Dynamic memory*. Cambridge: Cambridge University Press.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.

Science Council of Canada. (1984). *Science for every student: Educating Canadians for tomorrow's world.* Report 36. Ottawa: Author.

Simon, M., & Forgette-Giroux, R. (2002, October 20). Senior school board officials: Perceptions of a national achievement assessment program. *Education Policy Analysis Archives, 10(46).* Retrieved from http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1407&context=coedu_pub

Smith, M. C., Mikulecky, L., Kibby, M. W., & Dreher, M. J. (2000). What will be the demands of literacy in the workplace in the next millennium? *Reading Research Quarterly*, *35(3),* 378–383.

Snowling, M. J., & Hulme, C. (Eds.). (2005). *The science of reading: A handbook.* Malden, MA: Blackwell.

Stacey, K., & Wiliam, D. (2013). Technology and assessment in mathematics. In *Third international handbook of mathematics education* (pp. 721–751). New York: Springer.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences, 23*, 645–665.

Stiggins, R. (2002, June 6). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappa International On-line 83(10).* Retrieved from http://electronicportfolios.org/afl/Stiggins-AssessmentCrisis.pdf

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934–947.

"The Free Dictionary." Retrieved from www.thefreedictionary.com/mathematics

UNESCO. (2011). *Creating and sustaining literate environments*. Bangkok: UNESCO Bangkok Asia and Pacific Regional Bureau for Education. Retrieved from http://www.unescobkk.org/resources/e-library/publications/article/creating-and-sustaining-literate-environments/

Van de walle, J.A. (2004). *Elementary and middle school mathematics: Teaching developmentally* (5th ed.). Boston: Pearson Education.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Wallace, C. (1992). *Reading*. Oxford: Oxford University Press.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63(3)*, 249–294.