ASSESSMENT FRAMEWORKS:
Your Guide to Fair and Reliable Assessments

RICHARD JONES

June 20, 2025

Introduction

Most jurisdictions and testing organizations utilize framework documents, while others may not. Among those that do, there is a wide range of content included, often presented in various formats.

An assessment framework is a foundational document that identifies the guidelines, procedures and tools that will be used to guide assessment development, administration, marking/scoring, data analysis and reporting of results. The document outlines how the assessment will be conducted, focusing on what should be assessed and how. It provides a structured guide for designing and implementing assessments, ensuring they are aligned with learning outcomes and intended purposes. This article intends to describe the content of an effective framework document and outline the benefits of developing and publishing such a fundamental guide.

Framework Content[1]

Before embarking on the design and development of large-scale assessments, addressing several issues and making well-informed decisions is crucial. Involving key stakeholder groups at this stage can be extremely advantageous, as their insights and feedback can assist assessment developers in crafting comprehensive plans. Furthermore, engaging stakeholders often fosters broader acceptance of the assessment program. If the resulting document is made publicly available, it will reflect transparency, ultimately helping to promote a greater understanding and acceptance of the initiative. As decisions regarding the core principles of the assessment program are made, it is beneficial to capture them in a document referred to here as a framework document. This foundational resource is an invaluable tool for the team members engaged in the program's design, development, and implementation, and also serves as an important communication device.

The framework document should include information about the purposes of the large-scale assessment program, the key guiding principles that govern the assessment process, and details on how the resulting information will be generated, utilized, and reported. (In contrast, technical reports provide detailed accounts of all assessment processes and procedures.) A comprehensive framework may include the following types of information:

- **Ethical Considerations**

  The assessment program should subscribe to key principles to be referred to during all stages of design and implementation. For instance, to ensure fairness, equity and inclusion, test developers should make assessments as fair as possible for test takers from a variety of backgrounds, including those of different cultures, genders, geographic regions, as well as those with special learning needs. Information about how fairness will be addressed should be included in the document. Statements about these fundamental principles should reference important resources such as Canada's *Principles for Fair Student Assessment Practices for Education in Canada* and the American *Standards for Educational and Psychological Testing.*[2]

- **Assessment Purpose(s)**

  It is essential to clearly define the purpose(s) of a large-scale assessment program before beginning its development, as these purposes will influence all aspects of the program and affect its design. For example, if one program's primary goal is to ensure public accountability for schools, school boards, and provincial or state levels, while another aims to provide formative or diagnostic information to enhance individual student learning, the two assessment programs will differ significantly in various ways. Designing an assessment that effectively meets multiple purposes is difficult, if not unfeasible.

---

[1] Information in this section has been excerpted or adapted from:
Jones, R.M. (2014). *Large-Scale Assessment Issues and Practices: An Introductory Handbook.* La Vergne, TN. Lightning Source (Ingram) Books.

[2] *Principles for Fair Student Assessment Practices for Education in Canada.* (1993). Edmonton, Alberta: Joint Advisory Committee.

American Educational Research Association, the American Psychological Association, & the National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing.* Washington, D.C.: American Education Research Association.

- **Frame of Reference for Interpretation**

  A decision needs to be made regarding whether the assessment will be norm-referenced or criterion-referenced. Norm-referenced assessments compare students' performance to a norm, which is the average or typical performance of a specific group of individuals who have taken the same test. This approach serves as a benchmark for interpreting students' outcomes relative to their peers. In contrast, criterion-referenced assessments compare students' performance against established criteria or specific definitions of success. Recently, there has been a general shift from norm-referenced to criterion-referenced assessments, as educational jurisdictions have emphasized the importance of designing tests that align with specific curriculum learning expectations (content standards or outcomes).

- **Language of Testing**

  A statement regarding the languages used in testing should be included in a framework document. In some jurisdictions, assessments are developed in only one language, while others may require assessments to be administered in two or more languages. For example, many U.S. states are mandated to conduct tests in both English and Spanish, in Canada, most jurisdictions provide assessments in English and French (the two official languages), and in Switzerland, languages available for assessment include English, French, German, Italian and Spanish. The process of developing, administering, scoring, and reporting assessments in multiple languages raises several important questions. One key issue is whether a test in one language will be a direct translation of a test in another language or if tests will be developed independently for each language.

- **Sample Versus Census Tests**

  The purpose of the testing program will significantly influence the decision on whether to conduct a census or a sample of test takers. This decision has important implications for the design of the assessment. For example, if

test takers are randomly selected (sampled), reporting will likely only be possible at broader levels, such as the provincial, state, or national levels. Individual test-taker reports, or reports at the school or school district levels, may not be feasible.

- **Defining the Construct**

  When developing a large-scale assessment program, an early task is to identify the subject matter and the specific age or grade of the test takers. Often, these decisions are mandated by provincial, state, national, or international authorities. Once these foundational decisions are made, it is crucial to define the constructs—often referred to as domains—that will be measured, as well as to establish connections to the curriculum and current research. For example, it is essential to clarify what is meant by terms such as "literacy," "reading," "writing," and "mathematics." The concept of literacy serves as a good illustration, as it can be defined in various ways, including functional literacy, information literacy, media literacy, visual literacy, global literacy, and technological literacy. A clear definition and description of the construct should be included in a framework document.

- **Alignment with Learning Expectations**

  Once the construct has been defined, decisions must be made regarding the specific content to be assessed. Due to the inherent limitations of large-scale assessments, not all learning expectations can be effectively measured. For example, having students identify areas for improvement in their writing based on feedback from teachers and peers is often not feasible in large-scale assessments, mainly because of time constraints and the need for consistent administration across different jurisdictions. As a result, certain learning expectations that are difficult to measure in large-scale assessments may be more appropriately assessed at the classroom level. Test developers, educators, and test takers must have access to information that outlines which curriculum learning expectations will be assessed and which will not. Many jurisdictions and testing

organizations provide tables that detail this type of information by subject area.

- **Content and Timing**

  In addition to the connections to the curriculum, educators and test takers need to understand the composition and timing of the tests. The framework document will outline the number of items or tasks of various types (such as selected-response and constructed-response) that each test taker will encounter. This includes field-test items, if applicable, and clarifies whether they will contribute to the test taker's score. The document will also specify the number of booklets or sessions involved. Also, it is important to indicate whether the test is timed (meaning there is a strict time limit) or untimed (allowing for additional time within reasonable and practical constraints). Furthermore, the document should contain information on when testing will occur, such as at the beginning, end, or during the school/calendar year.

- **Item Development**

  The framework document should clearly outline the approach to developing test items and questions. For example, will the items be created internally by the testing organization or by an assessment service provider? Will educators participate in the item-writing process? It is important to emphasize that the quality of the items used in the assessment is the foundation of a successful assessment program.

- **Field Testing**

  While the framework document may not delve deeply into the specifics of field testing, it is essential to include a statement regarding the conduct of field testing for assessment items. Most large-scale assessment programs incorporate some form of field testing, which can involve embedding field-test items into assessments or administering separate blocks of field-test items either simultaneously with the main assessment or at a different time.

- **Accommodations**

  All test takers should have an opportunity to demonstrate their knowledge and skills. Therefore, large-scale assessment programs face the challenge of fairly accommodating individuals who are at different stages of language acquisition or have special education needs. The framework document should clearly outline any allowable accommodations. (Accommodations refer to adjustments made to the assessment environment or format that enable test takers with special education needs to demonstrate their learning without altering the content or expectations of the assessment. This differs from modifications, which involve changing the assessment itself, such as altering the content or reducing the difficulty to suit a test taker's needs.) Some accommodations, such as the use of headphones, a quiet setting, alternative locations, and additional time, may be available to all test takers. Those needing special consideration can access different presentation formats, such as Braille, large print, interpretation/sign language services, audio formats, and assistive technology. Response formats may also include assistive technology options, like voice-to-text, as well as verbatim scribing.

- **Alternate Assessments**

  Some jurisdictions offer alternative assessment methods to evaluate the learning progress of the small number of test takers who cannot participate in the standard testing process due to their special education needs. Examples of these alternative assessment formats include portfolios, projects, oral assessments, and specially designed courses. Information regarding the availability of alternate assessments should be described in the framework document.

- **Scoring/Marking**

  A framework document should outline, at least in general terms, how different test items will be scored. Selected-response item types are typically scored using automated technology, while constructed responses are usually scored manually. This can be done in paper format

or online using scoring system technology (although automated scoring, using artificial intelligence [AI], for constructed responses is showing promise). Scoring may occur locally, centrally, or remotely at the testing site through online systems. To ensure consistent scoring of constructed responses, it is essential to establish a set of scoring rules. For each constructed-response item or task, scoring rubrics (also known as scoring guides in some jurisdictions) must be developed. Including sample rubrics in framework documents is useful for promoting an appreciation of the scoring process.

- **Reporting**

The framework document should outline the types of reports that will result from the assessment. For example, will there be individual reports for each test-taker? Will the results contribute to a test-taker's overall grade, and if so, what percentage will they account for? Additionally, will there be reports at the school or school district level? What information will these reports include, and how will they be distributed? The reporting methods should align with the assessment's objectives and reflect the intended interpretation of scores and results.

- **Maintaining Consistency**

It is essential to demonstrate that assessments are fair and consistent across different administrations, in terms of both the subject content and the cognitive skills being evaluated. To instill confidence in the assessment results among test takers, educators, stakeholders, and the public, it is important to ensure the consistency of test administration from year to year (or administration to administration), especially if results are to be compared over time. Assessment programs address consistency in various ways. Depending on the specific program, the methods employed may include:

- Utilizing the same or a similar assessment blueprint for each administration

- Maintaining the same test format across different administrations

- Including items or tasks on assessments that have been previously field-tested to equate test-taker performance over time

- Reusing sets of test items from prior administrations

- Employing appropriate equating methods

- Ensuring consistent administration and accommodation procedures

- Instituting monitoring procedures for assessment administration at testing sites

- Implementing consistent, reliable, and valid scoring of test takers' responses

- Ensuring consistent, reliable, and valid data processing and statistical scoring, with proper documentation of procedures

- Conducting independent replication of data analyses

Framework Formats

Effective framework documents contain information about key components of a large-scale assessment; however, they may take different forms. For instance, the Department of Elementary and Secondary Education features a menu of documents related to the Massachusetts Comprehensive Assessment System (MCAS) on its homepage.[3] The website provides information on the following topics:

- Student Participation

- Testing Schedule

- Test Administration Information & Resources

- Accessibility & Accommodations

- Test Design & Development

- Example Test Questions & Practice Tests

- Student Work/Scoring Guides

---

[3] Massachusetts Department of Elementary and Secondary Education. (2025, May 29). *Massachusetts Comprehensive Assessment System.* Retrieved June 11, 2025 from: https://doe.mass.edu/mcas/.

- Assessment Results

- Performance Appeals Process

- MCAS Alternative Assessment

- Technical Reports

Additionally, there are specially tailored resources available for students and parents/guardians.

In Canada, many jurisdictions have developed effective framework documents and other assessment-related resources aimed at specific audiences. Two notable organizations in this context are the Education Quality and Accountability Office (EQAO) and the Council of Ministers of Education, Canada (CMEC).

The EQAO provides framework documents for each of its four provincial student assessments. As an example, you can find a link to one of the assessment frameworks here.[4] Alongside these framework documents, the agency offers additional resources for educators and students, including:

- Released/Sample Questions

- Glossaries/Formula Sheets

- Videos on What to Expect on the Assessment

- Sample/Practice Tests

- Live Webinars (for parents/guardians)

- Additional Resources (such as infographics on EQAO research and highlights of provincial assessment results)

The CMEC's comprehensive framework for the Pan-Canadian Assessment Program (PCAP) can be accessed here.[5]

Conclusion

An assessment framework serves as a blueprint for designing and conducting assessments, recording structured, purposeful, valid, and reliable methods for assessing learner performance. Best practices in large-scale assessments necessitate such a document, as it communicates essential information about the assessment program to test takers and all stakeholders. Furthermore, the creation of an effective framework document fosters transparency in the assessment processes and procedures, which in turn promotes trust and confidence in the validity, reliability, and fairness of the assessment program.

---

[4] Education Quality and Accountability Office. (2025). Assessment of Reading, Writing and Mathematics, Junior Division (Grades 4-6), Framework. Retrieved June 10, 2025 from: https://www.eqao.com/wp-content/uploads/2021/08/junior-division-framework.pdf.

[5] Council of Ministers of Education Canada. (2024). Pan-Canadian Assessment Program: PCAP 2023 Assessment Framework. Retrieved June 10, 2025 from: https://www.cmec.ca/docs/pcap/pcap2023/PCAP-2023_Framework_FINAL_EN.pdf.

About the Author

Richard Jones has extensive experience in the fields of large-scale educational assessment and program evaluation. He has worked in the assessment and evaluation field for more than 35 years. Prior to founding RMJ Assessment, he held senior leadership positions with the Education Quality and Accountability Office (EQAO) in Ontario, as well as the Saskatchewan and British Columbia Ministries of Education. In these roles, he was responsible for initiatives related to student, program and curriculum evaluation; education quality indicators; school and school board improvement planning; school accreditation; and provincial, national and international testing.

Dr. Jones began his career as an educator at the elementary, secondary and post-secondary levels. Subsequently, he was a researcher and senior manager for a multi-national corporation delivering consulting services in the Middle East.

Feel free to reach out to Richard "Rick" at richard.jones@rmjassessment.com (or on LinkedIn) to inquire about best practices in large-scale assessment and program evaluation.