Summative Assessment

Resource selection and writing
Robert Howe, Consultant in Higher Education Pedagogy and Assessment Specialist

Document retrieval
Andrée Dagenais, Librarian at the CDC

Summary

Summative assessment is almost an everyday topic of discussion. All teachers assess student learning; always in a “summative” way and almost always in a “formative” way. For many, summative assessment is the assessment that “counts,” and formative assessment is the one that “does not count.”

This literature review might seem outdated to some, because almost all the texts that deal with learning assessment are explicitly or implicitly rooted in summative assessment. It is, however, important to go back in order to reconnect with the basic principles of measurement and evaluation. This is what we offer in this selection of documents—references that are all conducive to reflection.

In this Issue

1. Introduction
2. Fundamentals
3. Quebec classics
4. Contemporary guides
5. Articles
6. Further reading
1. Introduction

As in any activity in the humanities, the assessment of learning does not rely on ready-made formulas or solutions. Here as elsewhere, there are complex problems that are never identical and must always be approached with caution. In assessing human traits, a nuanced attitude of caution requires judgment and methodical reflection. Individuals involved in assessment must educate their personal judgment for themselves by, among other things, reading the authors who, although sometimes not so contemporary, think out loud for us in their writings. These are the authors who interest us here. They are the ones who inspire researchers today. They have, indeed, helped to educate the professional judgment of stakeholders in contemporary measurement and assessment in Quebec and elsewhere.

So let’s take a closer look at the fundamental reflection on summative assessment. In these few readings, there is a sampling of the most relevant documentation by the authors who give meaning to the basic concepts used today to name things and make the best decisions in our practice of measurement and summative assessment.

Some of the references that follow are not recent. We have chosen them because they are definitely critical sources of contemporary knowledge in measurement and assessment. They guide practitioners and researchers today just as beacons guide marine pilots. Benjamin Bloom was such a beacon. Over 50 years ago, the author and publisher gave us the famous Bloom’s Taxonomy and, a bit later, his famous Handbook.

If we talk about literacy to evoke the necessary professional and cultural foundations of human activity, we must include these and other resources, as fundamentals of “basic training” in measurement and assessment.

In the context of francophone Quebec, we will name some essential authors who are sometimes forgotten today but who are still worth consulting for their analytical and methodical thinking in the area of assessment.

Happy reading!
2. Fundamentals


This is a classic; a must for pedagogy and assessment. It is a book about the “state of the art” of assessment intended for classroom teachers. Even though it is an older book and the approaches it presents have evolved, in particular in cognitive sciences and in focusing on competency development, its basic principles are still relevant today. The genius of Benjamin Bloom and his co-authors can be seen in the comfort of their moral and professional authority. They are among the pioneers of modern pedagogy. Many of the principles of good practice in assessment have been inspired by the Handbook. It has historical value: it represents the beginning of a pedagogical view of psychometrics, the beginning of the transition from docimological\(^1\) to measurement and assessment—a process that has evolved, especially since the publication of the Handbook. It should be noted that it was this handbook that brought into the classroom the concept of formative assessment, until then used in another context. Nothing could be farther from the truth than to say the handbook has no value today.

Chapter 4 is devoted to summative assessment (summative evaluation at the time), described by the authors based on its distinguishing characteristics, the first of which has to do with the purpose for and uses of summative assessment: generating a grade and certifying the attainment of learning objectives.

The second characteristic relates to the time or times when summative assessment is used in the learning context. This assessment is performed at a time when substantial learning has been attained, at key times during the course or at the end of it. It is then that we ask “Has the student learned?”

The third concerns the overall, holistic characteristic of summative assessment, as opposed to formative assessment, which is more analytical, and more focused on the details and components of learning. Summative assessment involves a far more general judgment on the attainment of substantial, unifying, overall objectives attained as a whole over some substantial part of a course or at the end of it.

In short, summative assessment makes a summary. The French have translated “summative” by “sommative,” and some might confuse the word with the idea of making the “sum” or adding up, which evokes totalling grades from a number of assessments. But in spirit, summative assessment is intended to mean a “summary,” “to sum up” with regard to the essentials, not details.

In this chapter, the concept of a “table of specifications” is proposed for the first time. This matrix (remember the year is 1971) has an x-axis and y-axis relating teaching content to objectives to be attained (behaviours) for each element of content. Obviously, at the time, the priority focus was not yet on competencies to be developed. The main focus then was on content, although the pedagogical discourse of Bloom and others ultimately lead to the development of competencies (skills and abilities).

\(^1\)It was in 1920 that Frenchman Henri Piéron and a colleague proposed the word *docimologie* (docimology) to describe the science of examinations or the science of testing.
The very general table of specifications lists the main content categories and, from there, outlines the characteristics that should eventually allow the evaluator to witness student learning. Traditional instruments such as multiple-choice questionnaires were primarily used. Bloom offers the table of specifications as an analytical tool promoting coherence between what we want to teach and what we will assess at the end.

The next step in preparing a summative assessment instrument is to assemble the tasks or questions in a coherent structure with instructions and guidance to enable students to demonstrate their competency. Finally, a grading strategy, a scale, an answer key, is developed from which a numerical or other symbol or grade will be derived for decision making or for certification.

By reading through the literature of the psychometric culture of the time, we can see the foundation for the function of summative assessment. The most common intended uses of “tests” or evaluation tools were to assign grades, certify skills and abilities, predict success in subsequent courses, serve as the initiation point of instruction in a subsequent course, provide feedback to students, and provide comparisons of outcomes of different groups.

In 1971, we were in an era of normative assessment. However, with thoughtful and informed reading, we can easily transpose onto our current criterion-referenced assessment context the recommended principles for care to be taken in the selection of questions or tasks to put to students to generate meaningful and credible grades. Of course, for many years now, we have no longer referred to discrimination or difficulty indices. But the representativeness of questions about what is being measured must be based on an educated and structured professional judgment, using a meticulously created and validated grid (table of specifications\(^2\)). Chapter 4 presents the structure and use of this grid.

In addition to generating a grade, summative assessment certifies that the student possesses, “at least at that time,” certain skills, knowledge, and abilities. This is the source from which Scallon will later describe the concept of certification assessment. In this function, summative assessment should infer that the student can undertake the study of deeper texts, has the foundation to start a more advanced course, has the skills to start solving problems in this lab, etc. It involves the level attained in learning. This means that tasks or questions in the assessment tool have been carefully chosen in order to justify such inferences. There is an assumption here that a known level of performance exists, above which students can do the specified assignment and below which they cannot. Defining this level, this success threshold, requires professional judgment.

Later in this fourth chapter, Bloom and his colleagues discuss the technical means for giving the prescribed test questions and tasks the measurable qualities of validity and reliability in a spirit of fairness and credibility.

---
\(^2\) In 1996, a PERFORMA task force called Pôle de l’Est (Eastern Pole) published a document entitled *Processus de planification d’un cours centré sur le développement d’une compétence*. In Chapter 5 of this text, the group created, characterized, and recommended a planning grid that it called a synoptic view (*vue synoptique*). The table is based on the same concerns as Bloom’s Table of Specifications and seeks to be relevant in the search for consistency in what is taught and what will be evaluated. See: [http://www.cdc.qc.ca/performa/720834-deshais-et-al-planification-dev-competence-rimouski-performa-1996.pdf](http://www.cdc.qc.ca/performa/720834-deshais-et-al-planification-dev-competence-rimouski-performa-1996.pdf)
Like all previous editions, this fourth edition of *Educational Measurement* is a comprehensive survey of all contemporary knowledge in measurement and assessment. This reference book, written by mainstream experts in assessment, is without a doubt an essential work in this field. Here, the leading American authors in measurement and assessment have contributed chapters dedicated to all aspects of the field, several of which do not concern higher education in Quebec—especially if you consider the important role played by standardized tests in the United States. Every topic is addressed: concepts, methods, and research as well as historical, legal, pedagogical, and prospective issues.

Chapter 17 of this fundamental book is dedicated to classroom assessment, that is to say the assessment designed and administered by teachers for their students. Interested in and driven by then-recent research in cognitive psychology, Lorrie Shepard wrote this still extremely inspiring chapter on the connections between pedagogy and the assessment of learning. In the American context, there is large-scale use of commercial standardized tests. This chapter looks at the links between, on one hand, formative and summative assessment and, on the other hand, the large-scale administration of external standardized tests. From the outset, based on research, the author denounces the damaging effects of standardized tests on the quality of teaching and learning.

In terms of its functions, we all know that summative assessment is for certifying that students have learned. The French literature also designates it as “certification assessment” because in this function, the evaluative judgment will place the student on one side or the other of a threshold. This threshold corresponds to what is designated as a success or failure for each of the objectives of the learning activities. What most characterizes summative assessment, far more than the administrative act of certification and recognition by a college or university, is that this certification is accompanied by a grade or score that places the student on a certain scale, sometimes as letters, sometimes as numbers, out of 20 or out of 100 (as in the Quebec college network), or using different symbols depending on cultures and conventions. According to Shepard, there are certain dangers or negative effects to grades that vary for each individual. In her text, she describes this phenomenon as threats to the quality of teaching and learning.

Shepard invokes research, including that of Susan Brookhart, demonstrating that even if teachers sincerely seek to be fair and transparent with their students on what will be evaluated and the meaning of the grade, much variability is observed in the practices of assigning grades. For example, some teachers will introduce into a grade variables external to learning, including participation, motivation, attendance, assignment completion, and effort. These classroom practices occur although experts disapprove of polluting grades with these kinds of irrelevant variables. These variables would compromise the validity of the grade as an indicator of the attainment of objectives.

What is most striking in this chapter is that, unlike most textbooks on assessment, Shepard first addresses formative assessment because, in her opinion, classroom assessment should primarily improve learning. Then, in an original approach, she presents summative assessment second by highlighting its negative effects, especially in its function of assigning grades for a report card.
mentions threats to the pedagogical processes of teaching and learning, processes entrenched in the teacher-student relationship. She writes of corruption and paradoxes that can result from combining formative assessment with summative assessment. Think, for example, of the development of valuable cross-curricular competencies acquired through teamwork or the project approach that would be seriously undermined by a spirit of competition among students concerned about their standardized grade average (cote R) or other consequences of individual grades, including their eligibility for another program or some institutions. Consider also the paradox of our entire school system, which values criterion-referenced assessment while all colleges calculate the class average in a traditional normative way. These are paradoxes based on intentions and consequences that have nothing to do with formative and summative assessment.

Shepard also demonstrates that summative assessment does not only have negative side effects and is not only an unpleasant obligation for teachers. Summative assessment makes students study. Thanks to it, students study more, learn more, and are more motivated. They revise contents and integrate them in a personal way, especially influenced by the quality and level of tasks or questions prescribed in their summative assessment. What is evaluated is studied. But even more, Shepard affirms that in a cognitive perspective, the ideal pedagogical system would be one in which formative and summative assessments would be mutually aligned with the learning objectives. Thus, summative assessment is used to mark the achievements and to confirm the attainment of objectives following learning activities nurtured by formative assessment.

Shepard remains focused on the need for coherence between what is assessed in formative assessment and what is assessed in summative assessment. In conclusion, she emphasizes that the two functions of formative and summative assessment can and must be aligned and be mutually relevant if they focus on competency development and not based on social norms.

This chapter by Lorrie Shepard is fundamental and deserves to be highlighted here. Pedagogical counsellors and resource persons involved in the assessment of learning should study and be inspired by it.

3. Quebec classics


Morissette’s book could be described as a classic in assessment. This text covers all topics related to the assessment of learning: planning, from the course outline to the drafting of learning objectives, traditional tools (exam questions), laboratory observation, assessment of attitudes, exam results, the validity and reliability of assessments, and institutional assessment policies. Of course, this text predates the college renewal measures and, above all, the evolution of tools, such as rubric scoring grids, for evaluating complex tasks. However, the manual remains
adequate and relevant because its foundations and its pedagogical approach are and continue to be transposable to contemporary discourse on assessment in the context of a focus on the development of complex competencies.

What interests us in this book is Chapter 13, entitled Pedagogical Assessment. In addition to clearly describing the very traditional distinctions between formative, summative, and diagnostic assessment, the chapter specifically focuses on the fundamental concept of judgment. The author focuses his reflection on professional judgment and invites evaluators to make use of their professional judgement when evaluating. He beautifully clarifies the nuances between objectivity and subjectivity. What does professional judgment mean? When is it used in the evaluation process? In this chapter, Morissette’s reasoning reads simply and clearly in the excellent pedagogical vision that characterizes him.

This is a brilliant text. To read it is to understand the fundamentals of what we call evaluating and judging. Reading it also makes us, among other things but above all, understand the “why” of summative assessment. Morissette’s book is sometimes forgotten, especially because it is older and because other excellent authors have published more recently. But Les examens de rendement scolaire remains relevant and topical for anyone who wants to understand the fundamentals of the act of evaluating.

An interesting fact: Morissette qualifies as an “exam” any tool that enables us to “measure” learning, including long assignments, performance of all kinds, and written and oral presentations, etc. This therefore includes rubric scoring grids and, of course, traditional paper-and-pencil tests. This generous (or inclusive) view of the word “exam” is plausible from an etymological standpoint. During an assessment, the student is somehow tested when we ask him to demonstrate that he has learned something. For Morissette, to submit students to a questionnaire, or an “authentic” task, or a work assignment to certify learning is to “examine” their learning in order to certify the development of skills.

Chapter 13 gives us other inspiring reflections on the nature of assessment, including excellent distinctions to be made between:

- judgment and opinion;
- judgment and decision;
- normative assessment and criterion-referenced assessment.

Finally, Morissette offers a discussion on the importance of clarifying beforehand the pedagogical objectives (specific terminal or intermediary objectives) upon which the judgment will be made. This is a chapter I do not hesitate to qualify as illuminating, especially in regard to the topic of professional judgment.
SCALLON, Gérard (2004). *L’évaluation des apprentissages dans une approche par compétences*, Saint-Laurent, Éditions du Renouveau Pédagogique. (Available at the CDC, Call number 729607)

Gérard Scallon is the reference Quebec author on summative assessment. The fact that he is cited in many other texts attests to his credibility. This book, dedicated to the evaluation of competencies, deals with numerous aspects of summative assessment including the use of rubrics to guide the judgment of teachers and students, the evaluation of know-how and life skills, the portfolio, and self-assessment.

As a centerpiece of this book, Scallon discusses the issue of integrating assessment in learning. Prior to this work, several other authors also stated that summative assessment and formative assessment should be integrated in and coordinated with learning activities. Scallon reminds the reader of the need to contextualize student learning in a competency-based approach. Learning and summative assessment are based on a variety of situations proposed to students. If the students have failed, either completely or partly, there must be feedback from the teacher to the individual or the group, even if summative assessment is involved (and this stance echoes the recommendation of Lorrie Shepard, presented earlier in this Bulletin). Whatever the learning situation, we cannot pass judgement on learning unless we have followed its progress, which implies that several situations have served as guidelines along the student’s path. Only later will we come to the stage of the final report, known as summative assessment. The acquiring of a competency cannot be inferred only at the end of a learning period. It must have been evaluated continuously or cumulatively in order to guide the teacher’s judgment about its development. It is in this context of coherence between teaching, learning, and assessment, both formative and summative, that we can infer that learning has been reached.


This document, available on the website of Université Laval’s Faculty of Education Sciences, is specifically dedicated to the concept of summative assessment. In it, Scallon describes the history of the concept and shows the functional connections between summative assessment and its certification function. The author describes how Scriven (1967) distinguished summative assessment from formative assessment in the context that interested him at the time (assessment of a product, or a program, for example), and how, in 1971, Bloom and his colleagues adapted these concepts to the classroom context.

In Scriven’s concept, summative assessment consists of a final judgment at the end of a process when everything has been completed. This point of view is not relevant in education and Scallon shows why. In the classroom, we need frequent snapshots of student progress and this justifies a discussion on continuous summative assessment in the learning process. Far from representing an addition of points, continuous summative assessment, inspired by De Landsheere (1974) and others, consists of making regular updates during the learning process, not only at the very end. According to Scallon, Cardinet (1984) speaks of blocks of learning processes. In their summative function, these updates are used to attest to or confirm that such-and-such learning has (or has not) been achieved. In this sense, and this is
one of Scallon’s hobbyhorses, we should be talking about “certification assessment,” although this term has not yet been adopted in Quebec’s assessment culture.

In the dynamics of teaching and learning, we clearly see paths and functional connections between the three major functions of assessment, and Scallon summarizes them well in this document. But what specifically characterizes summative assessment is the certification of competencies. Therefore, like many European authors, including Tourneur (1985), Gerard Scallon openly advocates that we adopt (one day, perhaps) the designation of “certification assessment” in the place of “summative assessment.”

4. Contemporary guides

In Quebec, during the past few years, certain contemporary authors have published a number of texts that demonstrate their thinking and the evolution of their contribution to knowledge about the assessment of student learning. These include master’s degree essays, peer-reviewed articles, and research reports. Here we present the most recent and most integrative of these contributions. The CDC can direct readers to other publications by these authors.

- LEROUX, J.L., dir., et GROUPE DE RECHERCHE EN ÉVALUATION DES APPRENTISSAGES (GRÉAC) (2015). Évaluer les compétences au collégial et à l’université: un guide pratique, Association québécoise de pédagogie collégiale (AQPC) and Chenelière Éducation, 2015, 688 pages. (Available at the CDC, Call number 789012)

This book, published last November, is the most recent publication in the PERFORMA collection of the Association québécoise de pédagogie collégiale (AQPC).

Divided into three main sections, the guide has 20 chapters that address methodologies for guiding assessment practices at colleges and universities from a number of points of view.

Évaluer les compétences au collégial et à l’université states its position from the outset. Rather than focus on the “what” and the “why” of assessment, it specifically addresses the issue of “how” to perform evaluation in higher education using a competency-based approach.

Intended for both experienced teachers and beginners, its chapters provide resources to support quality assessment practices as part of such an approach.

It can be consulted in any order, depending on the reader’s needs and questions, and is organized into three main groups of texts:

1. Context of assessment in higher education
2. Resources for evaluating competencies
3. Challenges related to evaluating competencies
The 688 pages and 20 chapters cannot be summarized here. Nevertheless, we present a brief overview of the rich diversity of the proposed questioning from a sampling of chapters. (Note to the reader: For clarity, the chapter titles have been translated).

Chapter 2: Exercising your professional judgment in higher education. The authors define the concept of professional judgment in evaluation and provide detailed discussion. Who will exercise this judgment? When? Why? What are the characteristics of a rigorous professional judgment? What are the social, ethical, and pedagogical issues of professional judgment in evaluation?

Chapter 6: Designing assessment grids with rubrics. What are the advantages of grids with rubrics? What are the advantages or disadvantages of analytical or holistic grids? What are the rules for setting up and using grids with rubrics?

Chapter 8: Writing questions to assess learning. This chapter describes 10 steps for building a measuring instrument. Some but not all teachers will build an assessment tool for the classroom. If a teacher uses only traditional MCQ assessment instruments, what consequences might this have on student learning and on what is taught? What are the advantages and disadvantages (or limitations) of essay-type (long-answer) questions?

Chapter 13: What guidelines should we use to better guide teamwork? In the context of the individual assessment of work done in a team, what characterizes the major psychological challenges involved? In an appendix, the author describes several formulas for determining the individual’s personal grade in teamwork. Which formulas would likely present ethical problems?

Chapter 15: Ethics in assessment: some points of reference to support teachers in action. In an ethical dilemma, what external or internal guidelines can we use? What is the difference between justice and fairness? How can we educate our professional judgment with regard to ethical issues?

Chapter 19: An approach for the evaluation of attitudes. To identify relevant attitudes to develop in a course or program, what questions can we ask? When assessing an attitude, how do we choose indicators? The authors provide examples of significant indicators of the presence of certain attitudes.

This practical guide focuses on assessing the mastery of competencies and bases this concept on the awareness that the concept of competency is polysemous, having varied meanings. In the book’s introduction (page 17), the reader is warned that the true meaning given to the concept of competency has major repercussions on the evaluation strategies that should be designed to infer competency.

This is an essential guide, written expressly for teachers in higher education.

---

3 It should be noted that this chapter refers, among others, to a CDC Bulletin: http://www.cdc.qc.ca/bulletin/eng/bulletin-7-assessing-team-work-sept-2011.pdf
In this recent book, Scallon takes stock of assessment, in particular of summative assessment. Based on the literature on assessment in the 2000s and on research on learning concepts, the author discusses assessment methodologies. The book is thus dedicated to showing how we have evolved from assessing knowledge to assessing increasingly complex skills and abilities. The tools and methods vary, based on what has to be evaluated, and Scallon clearly characterizes these methods.

The author deals in depth with the concept of judgment. Learning objectives are becoming increasingly diversified in order to respond to competency development mandates. During the era when learning partly focused on knowledge, we long measured, quantified, and even summed up observations on students’ answers to paper and pencil tests. Here, Scallon discusses methods that demand a more qualitative approach to evaluation. To evaluate is to judge.

Judgment is an integral part of the assessment process. Because we value the acquisition of high-order abilities, it has become essential to take into account the context of student performance as well as their progress, behaviours, and involvement. This requires adopting multiple information-collection processes and applying professional judgment.

Scallon describes the three stages of judgment: observation, inference, and evaluation. In this context, “inference” consists of estimating an individual’s characteristics based on his or her observed behaviour in a sampling of tasks. In this way, the assessment consists of placing the individual in a performance situation. Three types of information are presented: a scenario (a problem, data, or an issue); a task to be performed; instructions (instructions and requirements to be met during the performance by the student).

From there, the author demonstrates that judgment, in a summative assessment which focuses either on the student’s procedure or on a product, must be based on tools such as assessment grids and scales of judgment that are more qualitative than quantitative. And since high-order competencies are the result of learning over time, Scallon also discusses appropriate methods for assessing student progress. He presents concrete examples of step-by-step progress in learning as well as guidelines or markers of progress towards the development of competency. The book ends with a presentation and discussion of various evaluation procedures including methods for evaluating professionalism.
5. Articles

With the following exception, the Centre de documentation collégiale has not found articles dealing explicitly with summative assessment in its collection.


This article is sometimes cited as proposals for questioning issues on summative assessment. It discusses the concept of a double standard for pass-fail decisions, the value of the final evaluation of a course, the impact of summative assessment on student motivation and on studying, and the presence of the human factor in measurement error.

The article is based on our search for coherence and consistency with the concept of competency-based approach, the concept of minimum competency, and the sometimes exaggerated weight certain IPESAs determine for the summative assessment at the end of the semester.

6. Further reading

The National Council on Measurement in Education (NCME) is a professional organization for those involved in educational assessment and measurement, research and development in the field of assessment as well as psychometrics. This organization, of which the author of this Bulletin is a member, is known particularly through its website: [http://www.ncme.org/NCME](http://www.ncme.org/NCME)

The NCME provides a glossary in which summative assessment is defined from the perspective of its use:

**Summative use of assessments**: Using assessments at the end of an instructional segment to determine the level of students’ achievement of intended learning outcomes or whether learning is complete enough to warrant advancing the student to the next segment in the sequence. This is contrasted with formative use of assessments.
About the Author

A graduate of Université de Montréal, Robert Howe is an assessment specialist. A member of the National Council on Measurement in Education (NCME), he acts as a consultant and trainer for several colleges in the areas of learning assessment, program evaluation, and teaching assessment.

He was a pedagogical counsellor, a trainer in the PERFORMA program (Université de Sherbrooke), and assistant director at Cégep régional de Lanaudière à l’Assomption. He was a lecturer at Université de Montréal and, since 2012, has been a lecturer at Université de Sherbrooke where he teaches assessment in the university’s microprogramme de formation initiale à l’enseignement collégial (MIFIEC).

He recently acted as a consultant in arbitration cases on labour law disputes involving the assessment of staff competencies.

Since 2009, he has written numerous College Documentation Bulletins for the Centre de documentation collégiale (CDC).

Most of the documents referred to in this Bulletin are available online or upon request from the Centre de documentation collégiale (CDC).

For more information about our services or to manage your subscription, please consult our website at www.cdc.qc.ca.