AUTHENTIC ASSESSMENT FRAMEWORK

A QUICK REFERENCE

IBRAHIM A. HALLOUN
www.halloun.net

Educational
Research
Centre

CONTENTS:

www.educationalrc.org

P. O. Box 11-9232; Beirut, Lebanon

Phoenix Series

Educational research in the last three decades has systematically shown that students of all levels often complete and pass their various courses, and especially their science courses, without necessarily understanding what their courses are all about (Bransford et al., 1999; Halloun, 1986, 2001b, 2008; Halloun & Hestenes, 1985a, 1998; Helm & Novak, 1983; Novak, 1994; NSB, 1990–2008). Researchers and concerned educators worldwide have argued and shown that this failure is due, to a large extent, to prevalent forms of assessment that encourage students to learn course materials by rote for the sole purpose of passing classroom and large-scale exams (Bransford, Brown & Cocking, 1999; Hake, 2002; NRC, 2001, 2002, 2003; Osborne & Dillon, 2008).

The traditional view has it that assessment is an end by itself whereby students are assigned particular scores or grades for sanction or graduation purposes. Prevalent forms of assessment, hereby referred to as *traditional assessment*(s), do not integrate assessment with learning and instruction, and they often fail to provide reliable, comprehensive evidence about student learning, evidence that teachers and other concerned educators can use to make informed and viable decisions about instruction, pedagogy and curriculum development. The educational community at large is increasingly recognizing the need to opt for *authentic assessment* that revokes the traditional view and promotes meaningful and equitable learning. The alternative view integrates assessment with learning and instruction so as to come up with reliable and useful evidence about various aspects of student achievement and empower both students and teachers to efficiently succeed in critical self-evaluation and insightful self-regulation of their respective performance.

This document outlines a generic framework for authentic assessment in the sense just mentioned, and discusses how the framework can be deployed in science courses of all levels. The document presents a quick reference that educators, and especially teachers, can readily use to deploy the framework within and outside the context of science. The presented work builds mostly on the author's personal, worldwide experience in developing assessment tools for a variety of purposes, but especially tools developed over the course of the past 25 years for the evaluation of learning and instruction in science, and especially in secondary school and university physics.

The document comes in four sections followed by an appendix. The first two sections concisely discuss the role of assessment in curriculum development and implementation, and then more specifically in instructional design and practice. The third section outlines the proposed authentic assessment framework. The framework is illustrated in the fourth section with examples from science. The document also includes an appendix on certain epistemological foundations in which the framework is grounded for efficient deployment in science education. These foundations are especially important to understand the theoretical background of section 4.

The document is meant to be briefly read so that the reader may come out with a quick idea of what authentic assessment is all about from our own perspective. Discussion is kept to the necessary minimum in all sections, and especially in the third section where the framework is outlined in the form of bullet points. The reader with little background in assessment is encouraged to skim through the appendix and then the fourth section before reading the third section. Subsequently, once the reader comes across a point about the framework in the third section that s/he feels requires to be contextualized for understanding or clarification, s/he would easily find the appropriate illustration in the fourth section, along with its background in the appendix.

1. Assessment and evaluation in a curriculum

The philosophy of educational assessment cannot be isolated from the philosophy of any particular curriculum. As shown in Figure 1, the nature of assessment must be explicitly specified when a given curriculum is put together. The figure shows that the philosophy of assessment, like that of the program of study and of learning and instruction, is determined by the framework of the curriculum itself, and then by the specifics of curriculum implementation.

A curriculum is usually about a given discipline or set of interconnected or integrated disciplines. A country may, for example, adopt separate curricula for separate scientific disciplines (e.g., physics, chemistry, biology, geology), or it may put together a common curriculum for all scientific disciplines. It may do the same for various languages, or it may even integrate science, language and other disciplines together, and thus go for a more generic and encompassing curriculum.

The curriculum framework embodies and reifies the broad vision and aspirations of education in a given country, while accounting for the particular philosophy and pedagogy pertaining to the respective discipline(s). The framework subsequently sets the general *profile* that students are desired to develop following the completion of their education at specific levels (Fig. 2), along with general norms or standards for ascertaining the extent to which students have actually succeeded in developing the target profile.

The nature of the target profile also varies from country to country, depending on the nature and scope of the curriculum, and also on how specific policymakers want the profile to

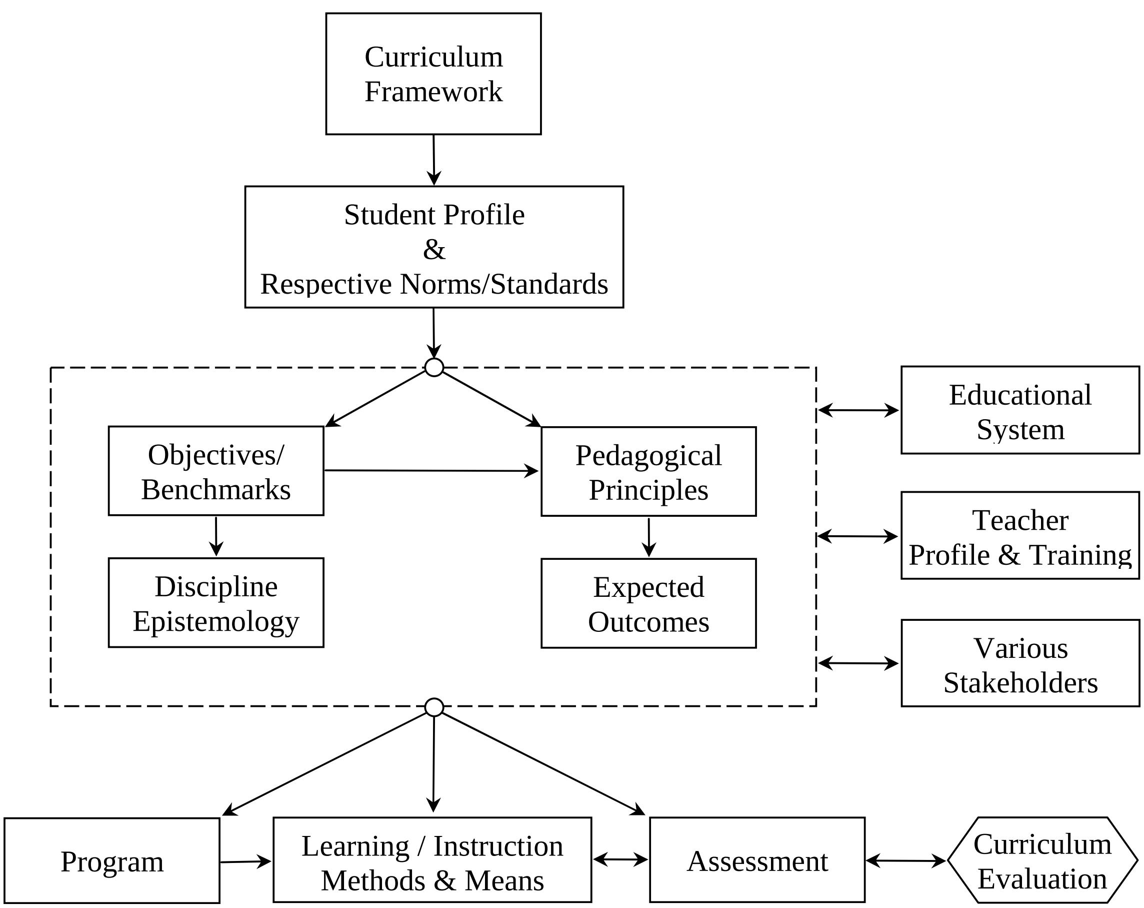

Figure 1. Broad view of a curriculum.

be decreed. On one end of the spectrum stand countries that specify general qualifications which students are hoped to develop. On the other end of the spectrum are countries that spell out in detail measurable outcomes around which the curriculum and the entire educational system would be built. Most other countries stand somewhere in between the two poles of the spectrum, including those countries opting for what is nowadays called standards-based education. Once the target profile is defined, and based on appropriate epistemological and pedagogical foundations, *expected outcomes* are inferred for specific courses. The role of assessment is subsequently to ascertain to what extent *observed outcomes* during curriculum implementation match expected outcomes.

Assessment is thus about student outcomes, about the extent to which individual students develop expected conceptions, processes or dispositions (Fig. 2). Assessment can never directly target educational standards or related objectives. Standards and objectives are *evaluated indirectly* by assessing student achievement of the expected outcomes. In this respect, *evaluation* is to be distinguished from assessment, though it closely relies on the latter. Evaluation is about the process, the system; about the merits of a whole curriculum and especially the merits of its standards and objectives. Assessment is about the product, about student learning, and more specifically about students' *achievement* of the expected outcomes, and, in very limited instance, about students' aptitude to achieve these or other outcomes.

2. Assessment-guided pedagogy

Learning and instruction are still overwhelmingly test-driven virtually all around the world, despite all reform movements that have lately been calling to the contrary (AAAS, 1990; AAC&U, 2002; Bransford et al., 1999; NRC, 1996, 2001, 2002; Osborne & Dillon, 2008). Students, parents and administrators ascertain the value of instruction in terms of student scores on course exams, and especially on large-scale state or exit exams where such exams are instituted. Teachers are thus driven to teach to the test, and to allow students learn by rote

Core-disciplinary dimensions

Content knowledge, drawn from the *episteme* of a given discipline (i.e., from the body of established knowledge shared and accepted by the members of a particular professional community). For example, the episteme of a given scientific community (physicists, chemists, biologists) consists of corroborated scientific theories. A science course is normally about certain *conceptions* (concepts, laws and other theoretical statements that make up certain conceptual models) in a particular scientific theory or set of theories (cf. Appendix).

Process knowledge, drawn from the *methodology* of knowledge construction and deployment in a given discipline. In the case of science, this knowledge pertains primarily to *model* construction and deployment, along with associated *tools* and *rules* (Appendix). The ultimate target of a given curriculum is to stabilize process knowledge so as to turn it into permanent *skills* in student profiles.

Meta-cognitive dimensions

Learning styles, which primarily are about processes of reflective thinking that help students regulate their own profiles in insightful and meaningful ways.

Emergent trends or *dispositions*, which include habits and attitudes that are characteristic of a given professional community. In the case of science, dispositions include habits and attitudes commonly referred to in recent calls for scientific literacy. They also include respective student views about the nature and relevance of science.

Figure 2. Dimensions of student profiles as anticipated in science curricula. Details of the four dimensions should be specified in the objectives and/or benchmarks of a given science curriculum (Halloun, 2004a/2006, 2008).

just enough material to score well on various exams, so that their administrators remain satisfied with their accomplishments. Assessment is thereby conceived as an end by itself rather than means to a more dignifying end: meaningful and equitable learning of any discipline.

To turn things around, learning and instruction need to be *assessment-guided rather than test-driven*. Classroom assessment should be considered not for merely deciding the fate of students in a given course, but especially for reliably ascertaining the extent to which individual students have meaningfully developed expected outcomes at specific points of instruction, so that appropriate learning activities may be designed and carried out to help students actively engage in self-regulation, and evolve toward the anticipated level of understanding and performance. Large-scale assessment (e.g., mandated local or state exams) should be meant to provide reliable evidence about the extent to which students develop an anticipated profile rather than an indication of students' capacity for rote learning, i.e., their ability to memorize and spontaneously reproduce specific content or problem solving routines discussed in textbooks.

As they currently stand, traditional classroom and large-scale assessments cannot guide learning and instruction. They often fail to provide reliable evidence about student learning, evidence that teachers and other concerned educators can use to make informed and viable decisions about instruction, pedagogy and curriculum development. Research in the last three decades has consistently shown that traditional assessments suffer three major shortcomings. They fail to: (a) reliably ascertain the extent to which students meaningfully achieve expected outcomes, (b) identify progress or evolution paths of individual students throughout the course of instruction, and (c) allow tracking of student evolution along these paths in meaningful ways.

Such shortcomings are primarily due to the fact that the results of traditional assessment are presented in the form of numerical or letter scores that do not necessarily reflect what a given student has actually learned or missed from the expected outcomes. The problem extends to all numerical measures of traditional assessment, whether norm-referenced or criterion-referenced. Traditional norm-referenced assessment results in grades or scores that allow ranking students relative to one another or to some norm group of students (high or low achievers). Criterion-referenced assessment often positions students relative to preset score intervals that do not necessarily reflect a cognitive or epistemological hierarchy of what students were tested about. In both types of assessment, two students may receive the same score (or letter grade) without necessarily achieving, or failing to achieve, the same expected outcomes, and a student may receive a higher score than another without necessarily achieving outcomes of higher cognitive order. The philosophy of all forms of traditional assessment needs to be significantly and feasibly modified so as to curtail the shortcomings mentioned above, and end with long-awaited authentic assessment. As we discuss in sections 3 and 4 below, our research shows that test content can be carefully chosen, and student achievement may be systematically graded and normalized so as not to merely rank students relative to one another or to a preset scale, but to *mark individual students' position on a cognitive evolution path mapped on the epistemology of the discipline* being studied. A student score (or any other measure) would then tell where exactly a student stands on this path, and more importantly, it would reliably tell what a student has actually learned and missed in her/his course or program of study.

Assessment subsequently becomes *authentic* in the sense that it would provide valid, reliable and transparent indicators of what students have actually learned and what they can do with their knowledge, and not just a rough estimation of their capacity to memorize and recall things by rote. Authentic assessment bring about outcomes: (a) that individual students can readily use to evaluate and regulate their own profiles, and (b) upon which teachers can reliably rely to ascertain students' achievement and diagnose their difficulties, so that they may determine how to mediate learning successfully. More specifically, authentic assessment enables teachers (and all concerned stakeholders, especially curriculum developers): (a) discern knowledge that is critical for success and meaningful learning of course materials, (b) ascertain how successfully students develop such knowledge at specific points of instruction and how they evolve in the course of instruction, and (c) interpret the measures of assessment so that informed and constructive decisions can be made about learning and instruction, as well as about various aspects of a given curriculum.

3. Authentic assessment framework

Authentic assessment, like any other educational activity, must be grounded in a well-defined framework that sets the norms, principles and rules for designing, implementing, interpreting, evaluating and refining all possible assessment tasks. The framework in question must, in principle, be the framework of the curriculum whose expected outcomes are the object of assessment (Fig. 1). It may eventually be a complementary framework well-aligned with the curriculum framework. In the following, we outline an authentic assessment framework that is flexible and plastic enough to be easily integrated in, or adapted to, any curriculum.

The proposed framework stems from over 25 years of research in assessment and curriculum development in which the author has been involved worldwide, mostly in science education. It significantly deviates from the traditional view of assessment. Educators used to traditional assessment might, at first sight, dismiss it as unrealistic or unfeasible. Our research and practice suggest that every single aspect of the framework is feasible within the appropriate classroom environment and educational system. Yet we do acknowledge that it may be quite impossible for someone used to traditional assessment to shift entirely to the proposed framework. Such a shift must be done gradually so that the initiated user can fully grasp the foundations of the framework and develop knowledge and skills required for its implementation. It must also be done only following appropriate training in the context of an educational system that would provide necessary support to allow and sustain the conversion to authentic assessment.

Assessment is a **process** primarily for ascertaining to what extent students achieve

expected outcomes spelled out in a given curriculum. It uses specific

tools (i.e., tests or other tasks, and respective means) that yield

observed outcomes (actual student **achievement** in assessment tasks). These outcomes are quantified with objective

measures (e.g., test scores) that are statistically

analyzed in order to determine the

proficiency of individual students, i.e., the extent to which observed outcomes are consonant with expected outcomes, and subsequently help students

regulate their profiles (Fig. 2), teachers regulate their instructional design and practice, and all concerned stakeholders

evaluate the viability of all curricular components.

Figure 3. The language of assessment.

Major tenets of authentic assessment are presented in the following ten subsections in the form of concise bullet points. Critical aspects are illustrated in the context of science courses in the last section of the document, and frequently used terminology is featured in Figure 3. The framework covers a broad range of assessment aspects, theoretical and practical, and it cuts across all disciplines commonly covered in K-16 education. As presented, the framework is somewhat comprehensive, but not exhaustive. Each subsection deserves at least one chapter in an assessment textbook, in order to broadly cover aspects herein discussed, as well as other salient aspects not accounted for in our discussion.

1. Reference, outcomes and goals

1. *Reference*. Educational assessment is grounded in the framework of a given curriculum in order to serve the purposes of this particular curriculum. All theoretical and practical aspects of assessment are set in reference to that curriculum. There is no "one-size fits all" assessment. Yet, authentic assessments all share some common grounds, the best part of which may be found in, or derived from, our proposed framework.
2. *Outcomes*. Assessment targets expected outcomes specified, at least in part, in a given curriculum. Assessment does not target educational standards or curricular objectives, although it serves to partially and indirectly evaluate standards and objectives. Educational standards are commonly defined for a broad range of disciplines (e.g., science or mathematics), and objectives are specified for individual disciplines (e.g., physics or algebra) so as to define components of the profile (expected outcomes) that individual students are anticipated to develop by the end of a given course or set of related courses.
3. *Goals*. Authentic assessment is meant not to sanction or rank students, but to allow various stakeholders to reason from evidence so as to reliably ascertain their accomplishments vis à vis student achievement and evolution in the context of a given curriculum, and make viable inferences for meaningful and equitable learning.

2. Stakeholders and utility

4. *Stakeholders*. Students and teachers are major stakeholders in assessment. Other stakeholders include parents, administrators, curriculum developers, educational researchers, and other educators and policymakers.
5. *Learning utility*. Authentic assessment is an integral part of learning. It helps individual students ascertain their strengths and weaknesses so as to gain increasingly better control over the evaluation, regulation and evolution of their own profiles, and subsequently succeed in meeting the ends of the followed curriculum.
6. *Instructional utility*. Authentic assessment is an integral part of instruction. It helps a teacher ascertain how successful individual students are in achieving the expected outcomes at particular points of instruction, monitor and mediate student evolution, and subsequently make informed and productive decisions about instructional design and practice. It also helps the teacher evaluate and regulate her/his own teaching practice so that learning becomes more and more meaningful and equitable.
7. *Curriculum and systemic utility*. Authentic assessment helps evaluate and refine various components of the respective curriculum, from curriculum framework to program of study and assessment modes (Fig. 1). It also helps concerned administrators and policymakers make informed decisions about the educational system so that it facilitates curriculum deployment in the most efficacious and efficient ways possible.

8. *Pedagogical utility*. Authentic assessment helps educational researchers continuously evaluate and refine educational theory and practice. It also guides design and implementation of various forms of professional development in education, including pre-service and in-service teacher training, action-research, and the works of professional learning communities.

3. Foundations

9. *Context*. Assessment is discipline-based. The discipline may consist of a single field of study (e.g., a given science or a given language) or of an integration of many related fields. Discipline episteme and methodology provide the core-disciplinary knowledge (content and process knowledge) that students need to develop and against which student knowledge state needs to be ascertained (Fig. 2). They also provide the necessary context for student development and assessment of the meta-cognitive dimensions of student profiles (learning styles and dispositions).

10. *Cognition*. Authentic assessment, like program of study, learning and instruction, is grounded in pedagogy and discipline epistemology (Fig. 1). Discipline epistemology helps convert the specified objectives into measurable educational outcomes, and pedagogy, guided by cognitive science, helps matching outcomes to the cognitive level of students of particular age and intellectual maturity. In retrospect, assessment feeds back into pedagogy by indicating to what extent specified outcomes actually match the cognitive level of the target student population.

11. *Design principles*. Reliable evidence requires that assessment means and methods be designed and implemented following well-defined principles that ensure the viability (validity, reliability, transparency, feasibility, efficiency and fairness) of such means and methods, as well as of all subsequent inferences.

12. *Commensurability and psychometrics*. The success of a given curriculum is determined by the extent to which students develop the target profile, more specifically by the extent to which observed outcomes are commensurable with expected outcomes. Commensurability implies that the two sets of outcomes be comparable in objective and precise measurable ways, i.e., with quantitative measures. Whence the role of psychometrics in documenting, analyzing, interpreting and reporting obtained measures, with appropriate descriptive and inferential statistical models of educational and cognitive measurement.

13. *Limitations*. No assessment can ever fully reveal what a person actually knows or can do. Limitations we are concerned about here come primarily from the main three sides of assessment: examiner (often a teacher), examinee (a student), and medium (assessment tool). From the first two sides come, among others, inherent constraints in student capacity and forms of expression and performance, as well constraints and discrepancies in examiner's and examinee's manipulation and interpretation of assessment tools and outcomes. Such tools are also limited in their transparency, i.e., in their capacity to reveal what they are supposed to reveal, especially when it comes to constructs (student features or traits that we assume to be the cause of an observed outcome and that we cannot directly assess). Assessment thus results in measures of observed outcomes that provide only estimates of a student's profile, estimates whose limits of approximation and precision are function of the adopted design and measurement principles.

4. Scope and sequence

14. *Scope*. A curriculum is conceived to foster the development of a particular student profile (Fig. 1). Authentic assessment comprehensively targets all four dimensions of the profile (Fig. 2), and reveals how students proceed to develop such a profile, and not just to what extent they succeed in achieving expected outcomes. Depending on the curriculum philosophy and afforded opportunities, the four dimensions may be weighed differently so that some dimensions may be targeted more broadly and deeply than others.

15. *Focus.* The scope of assessment (expected student profile) is often so broad that no authentic assessment can ever cover every single outcome specified in the curriculum. Authentic assessment, and especially summative assessment (cf. point 20 below), thus focuses on critical expected outcomes that empower students to have their profiles efficiently and successfully evolve in the desired direction.

16. *Continuum*. Authentic assessment seeks continuous and not discrete, dynamic and not static, measures. It concentrates more on revealing student progress and profile evolution along actual developmental paths – there is a multitude of such paths – than on providing episodic snapshots of individual students' achievement.

17. *Sequence*. Authentic assessment (like learning and instruction) follows reiterative and not linear paths. Following each assessment exercise, students are afforded to go back, reflect on, and insightfully regulate, their current profiles so that they become increasingly commensurable with the target profile. The same assessment exercise needs to be subsequently revisited, in one form or another, to establish evidence of successful regulation.

5. Types

18. *Achievement vs. aptitude assessment.* Achievement measures are actual status measures; they are about what a student have actually learned in a given course. Aptitude measures are provisionary measures; they are about what a student can possibly learn or achieve in the course beyond what achievement measures indicate. For all practical purposes in a given course, authentic assessment relies more on achievement than on aptitude assessment. The latter is relied upon mostly for diagnostic or predictive purposes at the beginning of a given course or course units.

19. *Content vs. performance assessment*. Content assessment is about what students know in a given course (mostly content knowledge and, to a lesser extent, some dispositions in Fig. 2), whereas performance assessment is about how students go about constructing and deploying what they know (process knowledge and learning styles in Fig. 2). Authentic assessment is comprehensive in the sense discussed in point 14 above. To this end, it relies on both content and performance assessment. Unlike content assessment, performance assessment requires observing and measuring students in action. Performance tasks consist of group or individual activities conducted inside or outside the classroom, and observed by the teacher directly or remotely with the use of appropriate technology (cf. points 26, 27 below). Such tasks include demonstrations, case studies, laboratory experiments, and respective whiteboard presentation and discussion.

20. *Internal vs. external (large-scale) assessment*. Authentic assessment may be internal or external. Internal or classroom assessment has a limited scale (a single class or a limited number of similar classes in a given school), and is owned by a particular teacher and her/his students. External assessment is a large-scale assessment conducted by district, state or international agencies, be it public or private, and is owned by the administering

agency (e.g., mandated state exams, International Baccalaureate, standardized tests). Internal and external assessments complement one another for optimizing learning and instruction, and especially in curriculum evaluation and refinement.

21. *Formative vs. summative assessment*. Formative assessment is the most frequent and indicative form of assessment. It is about the details of a given course (e.g., content or process knowledge pertaining to every section in a given chapter). Summative assessment is about the big picture in some units of a given course (assessed in partial exams) or an entire course (assessed, say, in final exams, exit exams, standardized tests). Authentic assessment relies on both types of assessment, though more on formative than summative assessment, for ascertaining student progress and for planning instruction and assigning learning tasks, and on a weighed balance of both types of assessment for deciding the fate of individual students at the end of a given course, or for graduation purposes at any grade level.

22. *Placement vs. diagnostic assessment*. Placement assessment serves to ascertain what a student knows at the beginning of a course, about prerequisites and course content, in order to determine the appropriate learning path for the student and assign him/her to the proper group for team work. Diagnostic assessment serves to ascertain what a student knows and does not know at various points of instruction (e.g., at the beginning of every course unit or activity) in order to decide on the proper remedial learning tasks for the student to overcome her/his difficulties. Authentic assessment relies on both forms of assessment, but especially on diagnostic assessment for refining the learning path of individual students.

There are of course other types of assessment that may be distinguished and contrasted following a variety of classification criteria. For all practical purposes, the types discussed above are the most prevalent and most useful for authentic assessment.

6. Tools

23. *Tools*. Authentic assessment requires a variety of coherent and complementary assessment tools. Various expected outcomes cannot be assessed all the same way, and no single tool can ever fully disclose whether a student has actually achieved a given outcome. Tools include tasks and respective means. Tasks include, but are not limited to, written tests and assignments, ethnographic surveys and interviews, field/observation reports, case studies, and experimental projects. Means include all necessary materials (from exam paper to computers and laboratory equipment) for carrying out a given task.

24. *Closed (or objective) and open (or subjective) tasks.* Closed tasks describe problematic situations and provide students with alternative answers/solutions for choice. Examples include true/false questions, multiple-choice and rating-scale tests. Open tasks leave it to students to come up with their own answer or solution. Open tasks are usually more reliable than closed tasks in all types of authentic assessment. Closed tasks may be sufficiently reliable if they are developed and validated following rules and principles typically followed in the development and implementation of standardized tests.

25. *Written and oral tasks*. Except for performance assessment that is better conducted orally and through think-aloud, all types of authentic assessment discussed above can be conducted in either written or oral form. Written assessment is not limited to paper-and-pencil. It extends to whiteboard presentations and various forms afforded by today's technology, and especially information and computer technology (ICT). It may be conducted locally (e.g., in a given classroom) or remotely (e.g., at home).

26. *Individual and group tasks*. Authentic assessment relies on a balance of individual tasks and teamwork. Individual tasks are especially needed to determine the position of individual students on a given evolution path. Group tasks are especially useful to help individual students efficiently evolve on such a path.

27. *Technology*. Technology, and especially ICT, is used to improve the efficiency and friendliness of assessment, but not necessarily the quality of observed outcomes. For example, ICT helps allocating less time for assessment and more time for active learning and instruction in the classroom, and reducing students' fear and tension.

7. Taxonomy and critical thresholds

28. *Taxonomy*. Expected outcomes in a curriculum are derived from, or grounded in, the episteme of a given discipline. Their choice is governed by: (a) respective epistemology and methodology, and (b) related student cognition and adopted pedagogical principles (Fig. 1). A detailed taxonomy of such outcomes is spelled out along the four dimensions of Figure 2 for every instructional unit in a given course. The taxonomy is followed in all activities throughout instruction, from lesson planning and implementation, to assessing student achievement of the expected outcomes.

29. *Benchmarks*. Due to a variety of practical constraints, especially time constraints, no assessment, and especially no summative assessment, can ever fully cover all expected outcomes or all aspects of any given expected outcome. Authentic assessment, and particularly summative assessment, is thus carried out relative to specific benchmarks, i.e., the most critical elements of expected outcomes, especially those pertaining to the big picture in a given unit or course. Benchmark taxonomy is especially useful in profile and curriculum evaluation.

30. *Normative assessment and proficiency*. Authentic assessment is normative in the sense that it is conducted relative to a pre-established taxonomy and success criteria. The proficiency of a given student is a relative state that defines the extent to which the student has actually achieved an expected outcome or a set of related outcomes, i.e., the commensurability between observed and expected outcomes. The most meaningful proficiency in authentic assessment is the one defined in terms of benchmark taxonomy.

31. *Success criteria*. A match or consonance between an expected outcome and an observed outcome (commensurability) is determined in terms of pre-established norms and criteria. The match in question can never be perfect since no student, especially no secondary school or college student, can fully master any given expected outcome. The criteria set the level at which an observed outcome can be considered as successful achievement of the expected outcome. Such criteria are set in authentic assessment for individual outcomes and benchmarks, as well as for an entire profile, so that obtained evidence can be measured and interpreted objectively, and independently of the special interests and idiosyncrasies of any of the stakeholders, especially teachers and students.

32. *Critical thresholds.* Such thresholds define critical levels of proficiency that individual students need to attain before they can be introduced to new materials in a given course. The most critical of these thresholds are what we call the basic threshold and the mastery threshold. The basic threshold corresponds to the minimum standards of meaningful understanding that any student should meet, irrespective of the initial competence level and interests of the student. This critical threshold embraces, in any discipline, so-called primary outcomes. These are the most elementary, but the most fundamental expected outcomes in a given course, both from cognitive and epistemological perspectives. No

student can achieve any higher-order outcome without mastering first all primary outcomes. Teachers' role in mediating learning is most critical below the basic threshold. The extent and importance of teacher's mediation gradually diminishes until students reach the mastery threshold. Most students can virtually proceed on their own in achieving expected outcomes that fall beyond the latter threshold.

8. Measurement

33. *Indicators*. An indicator is a quantifiable evidence of student proficiency. Given the complexity of an anticipated student profile and respective outcomes, no single indicator can ever measure the proficiency of individual students at any point of instruction. Numerous indicators are used in authentic assessment to measure student proficiency in each of the four dimensions distinguished in Figure 2, especially in formative assessment.

34. *Measures.* A measure is a value (often a numerical score) assigned to a given indicator on an appropriate scale. This scale may be nominal (e.g., true or false indicators), ordinal (rating-scale indicators), interval (numerical- or letter-graded indicators), or ratio-type. Authentic assessment relies on all sorts of measure, but especially on interval measures. Though they are traditionally treated as ratio-type, numerical scores actually are interval measures (they have no real zero, and a double score does not indicate a twice better achievement). Appropriate transformations can sometimes be carried out, with extreme caution, to convert any of the first three types of measures into ratio-type for statistical analysis purposes.

35. *Graders*. Authentic tasks, including those of classroom assessment, may be graded by stakeholders other than teacher, especially students. Student self grading and peer grading of various assessment tasks are crucial in promoting critical thinking and self-evaluation and self-regulation.

36. *Matrices and evolution maps.* For efficient interpretation and statistical manipulation, measures corresponding to a given dimension in student profiles are documented in well-defined matrices. Such matrices may be readily interpreted in authentic assessment to track the evolution of individual students' proficiency throughout the course of instruction, or graphically converted into evolution maps to this end.

37. *Criterion-referenced and norm-referenced measures*. Individual and matrix measures of authentic assessment can be interpreted and deployed either as criterion-referenced or norm-referenced measures. Either way, these measures do not loose their primary utility in positioning individual students' proficiency on a given evolution path, and especially relative to respective critical thresholds.

38. *Viability*. Indicators and respective measures of authentic assessment are chosen and scaled so as to provide relatively objective, valid, reliable, transparent, feasible, efficient and fair evidence about student proficiency, within certain limits of approximation and precision that should be well kept in mind during the entire assessment process, and especially while interpreting observed outcomes and respective measures.

9. Data collection and analysis

39. *Documentation*. Multiple and transparent means are used for documenting and reporting assessment tasks and measures. Means include, but are not limited to, task templates, spreadsheets, student evolution logs, and portfolios.

40. *Analysis*. Assessment measures are analyzed (and interpreted) based on all five foundations of authentic assessment delineated in sub-section 3, and against taxonomy and criteria discussed in sub-section 7. Classical and modern statistical models are used in this respect, with keen attention to the assumptions underlying any of these models.

41. *Feedback*. Teachers provide students, directly or through ICT, with timely and explicit feedback about their performance on assessment tasks. Feedback includes a detailed report about observed outcomes, along with guidelines on how to ascertain these outcomes against expected outcomes (self-evaluation) and how to proceed in self-regulation.

10. Interpretation

42. *Classroom implications*. Internal authentic assessment is used for planning, revising and orienting instruction, and especially for assigning remedial activities for individuals and groups of students with non-satisfactory proficiency. It is also used, in conjunction with external assessment, to evaluate and regulate instructional practice through continuous action-research conducted by individual and groups of teachers giving the same course.

43. *Large-scale implications*. Authentic assessment is used, in all its forms, to evaluate curriculum, pedagogy, and educational theory. It is also used to guide various aspects of professional development, from pre-service to in-service teacher training, and from specific training workshops to the works of professional learning communities.

44. *Transfer*. Assessment interpretation is not confined to the target course or discipline. Students' ability to transfer what they achieve in a given course or discipline to other courses or disciplines is a major indicator of meaningful learning. Longitudinal documentation, analysis and interpretation of assessment measures are thus critical in evaluating curriculum, pedagogy and educational theory.

4. Authentic assessment in science education: An illustration from classical mechanics

The framework presented above is generic in the sense that it applies to any course at any educational level. In this section, we briefly illustrate how major practical tenets of the framework can be deployed in the case of science education. In particular, we discuss how internal, formative and summative assessment of achievement can be conducted in a typical secondary school or college physics course dealing with the Newtonian theory of mechanics. We concentrate our discussion on the taxonomy and critical thresholds because of their importance not only in assessment, but throughout the learning process, from planning lessons and carrying out learning activities to assessing student achievement.

Our discussion is situated in a pedagogical and epistemological framework that answers common calls for reform in science education worldwide. It is derived from our work on modeling theory in science education, and is outlined in the appendix.

For all practical purposes, taxonomy is perhaps the most critical aspect of assessment. Without a clearly spelled out taxonomy, teachers and students alike would be lost throughout the educational process, and assessment would fail to portray a reliable picture of student proficiency. In the following, we provide a partial taxonomy of the core-disciplinary dimensions pertaining to the Newtonian theory of mechanics. The taxonomy is about the most important three levels in the conceptual hierarchy of any scientific theory: concept, model and theory (Appendix).

Cognitive and epistemological research has shown that there is a common pattern in the structure of all scientific concepts, and in the way these concepts are constructed and deployed in the scientific community. The same also goes for all scientific models or theories. Concepts' structural pattern is captured in our work in what we call the *content concept schema*, and the respective construction and deployment pattern, in what we call the *process concept schema*. The dual concept schema (content and process) is a generic template that allows any teacher or concerned educator to lay out explicitly the core-disciplinary taxonomy of any scientific concept, whether for instructional design and implementation or for assessment. The schema is discussed in the appendix, and is illustrated below with the case of the Newtonian concept of force.

The content concept schema basically says that a student can build any scientific concept by answering four major questions, pertaining respectively to the four dimensions of a schema, scope, expression, organization and quantification (Fig. 4):

1. What does the concept represent in the real world, and what is its function?
2. What is the concept called, and how is it featured in scientific discourse and episteme?
3. What is the concept type, and how is it related to other concepts in a given scientific theory?
4. How is the concept measured?

Missing any of the four questions (schematic dimension) results in student failure to meaningfully understand what a scientific concept is all about. All four dimensions need to be covered in both instruction and assessment (mostly formative assessment).

Figure 5 complements Figure 4 by laying out the process taxonomy of the force concept. Processes shown in this figure are specific to the concept of force. There are other generic processes that apply to virtually any scientific endeavor. These are shown in Figure 7.

Scientific models are of a higher cognitive and epistemological complexity. They make extensive use of scientific concepts, and provide them with the appropriate context for meaningful understanding. Students gain the meaning of scientific concepts only when they deploy them in model construction and deployment. Models are also at the core of scientific theory. They are even what scientific theory is all about (Appendix). Without learning explicitly how to construct and deploy scientific models, students fail to develop skills of scientific inquiry and meaningful understanding of any scientific conception, from concepts, laws and other theoretical statements, to scientific theory.

Figures 6 and 7 lay out respectively the dimensions of content and process taxonomy of any scientific model (following the dual model schema discussed in the Appendix). Figure 6 further illustrates the content taxonomy in the case of the uniformly accelerated particle model of Newtonian theory.

Figures 4 through 7 present taxonomies that are especially useful in formative assessment. No summative assessment can afford following such taxonomies in their details. In fact, such details are not all needed for summative assessment, whether within or at the end of a given course. Summative assessment is about the big picture in a given course, and more specifically about a given scientific theory in a science course. A useful content taxonomy for such assessment is provided in Figure 8. This taxonomy corresponds to the two most critical models in the Newtonian theory of mechanics, models that set what we consider the *basic threshold* of meaningful understanding.

As discussed in tenet 32 of our assessment framework, the content of any course, and especially any science course, can be divided into three levels of increasing cognitive and epistemological complexity (Appendix). These levels are separated by two critical thresholds that we call, respectively, the basic threshold and the mastery threshold. Summative assessment needs to particularly target these two thresholds. Figure 8 delineates a taxonomy for all content knowledge that falls just below the critical threshold in a physics course dealing with Newtonian mechanics. Our research and experience has long shown that secondary school and college students fail to understand anything about Newtonian theory unless they master first all elements in the taxonomy shown in this figure. They particularly cannot develop any of the higher-level content material unless they first reach the critical threshold delineated in the figure. A teacher thus cannot move on in instruction beyond this threshold unless students develop first the content of Figure 8. That is why summative assessment is important, even most critical, at the level of the basic threshold.

Our research has also shown that a taxonomy similar to the one shown in Figure 8 allows teachers and other concerned stakeholders to trace a typical evolution path that they can model in instructional design and curriculum development, and against which they can ascertain actual student evolution paths, and subsequently the effectiveness of instruction and curriculum. We have also devised standardized tests that set the critical and mastery thresholds at special levels of performance, and more specifically at specific scores on such tests. Ample details can be found elsewhere (Halloun, 2004b, and references therein; www.halloun.net).

The concept of force in Newtonian theory

1. **Scope**

   What does the concept ***represent*** *in the real world, and what is its* ***function****?*

   The concept of force represents agent-object *interaction*. A force of particular characteristics is associated with a particular kind of interaction.

   No physical body can act on itself. An interaction takes place between at least two bodies, an « agent » (acting body) and an « object » (body acted upon) whose state is being investigated.

   The concept of force is *explanatory*. It explains the *change of state* of a given object, more specifically the change of its linear momentum (or velocity).

   No change of state takes place in the absence of any (net) interaction. A change of state reflects the presence of unbalanced interactions, and may be *explained* with the force descriptor.

2. **Expression**

   What is the concept ***called,*** *and how is it* ***featured*** *in scientific discourse and episteme?*

   Force is the name of the concept, and no other name can be used to call this concept. As a vectorial concept, it is often represented by an arrow in appropriate force (vector) diagrams, and denoted symbolically by a bold letter in algebraic representations (mathematical equations).

3. **Organization**

   What is the concept ***type****, and how is it* ***related*** *to other concepts in a given scientific theory?*

   The concept of force is a *prime interaction* descriptor. It does not derive from any other concept, though other concepts may derive from it (e.g., the derived concept of work). It is implicitly defined through Newton's four laws of dynamics.

   The concept of force is related to: (a) the *intrinsic* properties of respective object and agent through *interaction laws* such as Newton's law of universal gravitation, (b) its *effect* on the object to which it is applied through *causal laws* (e.g., Newton's second law of dynamics, often expressed in the form $\mathbf{F} = m\mathbf{a}$), laws that relate the interaction concept of force to object descriptors.

4. **Quantification**

   How is the concept ***measured****?*

   Force is a *vectorial* concept (as opposed to *scalar*, like the concepts of mass or temperature), and hence its measurement requires the specification of a direction, a magnitude and a unit which is the Newton (N) in SI.

   Force is an *extensive*, *additive* and *ratio-type* concept (as opposed respectively to *intensive*, *non-additive* and *interval-type concept*, like temperature).

   A force is *indirectly* measured physically; there are no direct means (or physical probes) for comparing a given force to a standard force in the same way, say, the length of an object is physically measured by comparing it to the graduation of a ruler. A force is always measured through its effect on a given object, like stretching or compressing a spring.

   The effect of a force on an object is instantaneous and lasts as long as the force is exerted on this particular object. It stops only when the force is no longer exerted on the object.

   No two forces can be added together (through Newton's fourth law or law of composition) unless they are exerted simultaneously on the same object.

   The dimension of a force may be given symbolically by: $[Force] = \dfrac{[Mass][Length]}{[Time]^2}$

   The characteristics of a force are invariant under Galilean transformations, i.e., when changing inertial reference systems.

Figure 4. A partial content taxonomy pertaining to the Newtonian concept of force in secondary school and college physics courses.

Constructing and deploying the concept of force in Newtonian theory

1. *Semantic processes.*:
- Recognize that the concept of force represents an interaction between at least two real bodies.
- Distinguish between agents and objects. An agent is a distinct body that exerts a force on an object.
- Recognize that particular agents exert forces of particular characteristics.
- Recognize that the concept of force is a vectorial concept and establish the correspondence between a vector (mathematical representation) and the force (physical reality) that it represents.

2. *Syntactic processes.*:
- Realize that an agent and an object exert equal and opposite forces the magnitude of which may be expressed in terms of intrinsic parameters with appropriate interaction laws (e.g., gravitational and electrostatic interaction laws).
- Realize the independence of different forces exerted by different agents on the same object.
- Realize that forces may be added together only when acting on the same object.
- Realize that Newton's Second law ($\mathbf{F} = m\mathbf{a}$) relates a cause (F) to an effect (a), and thus that the equality between the two terms of the given equation represents a causal implication and not an identity.

3. *Inferential processes*:
- Recognize that the state of motion may be defined with the concept of velocity (or momentum) and not with the concept of position in a given reference system.
- Realize that the state of motion changes whenever the direction or magnitude of an object's velocity changes, and that the concept of acceleration may be used to quantify such a change of state.
- Realize that only a change of state as described above needs to be explained, and that the concept of force may be used to explain such a change.
- Alternatively, realize that an object is subject to certain force(s) when its velocity changes.

4. *Evaluative processes*:
- Realize what data need to be collected to determine whether an object is subject to certain forces.
- Be capable of analyzing data and recognizing when outcomes reliably indicate that forces are exerted on an object.
- Recognize the limits of approximation and precision associated with the evaluation of particular forces.
- Be capable of estimating certain forces and establishing upper and lower bounds on such estimations.

5. *Mathematical processes*:
- Representing forces with appropriate vectors (force diagrams), following appropriate semantic rules.
- Operating with force-vectors following appropriate syntactic rules (e.g. breaking a force-vector into components or adding force-vectors in an appropriate coordinate system).
- Coordination of various mathematical representations of a given force.
- Establishing the correspondence between force diagrams and motion diagrams.

Figure 5. A partial process taxonomy pertaining to the Newtonian concept of force in secondary school and college physics courses.

Uniformly accelerated particle model

1. **Domain**

   What ***pattern*** *does the model* ***represent*** *in the real world?*

   All physical systems that are in *translation* with *constant acceleration* in inertial reference systems. A given system interacts with one or many physical agents that exert on it a net *constant force*.

2. **Composition**

   What ***concepts*** *does the model consist of?*

   Objects: *One* particle representing a system in consideration, and depicted by a geometric point in a coordinate system representing the reference frame where the motion takes place.

   Environment: Two types of agents are distinguished in Newtonian mechanics, those of interaction at-a-distance (or long-range interaction), and those of contact interaction.

   Object descriptors refer to *intrinsic* and *state properties* of the particle-like system. Only one intrinsic property is accounted for in any particle model: the mass of an object. State properties are the kinematical properties of the object. They include position, displacement, velocity, acceleration, kinetic energy, etc. In a coordinate system, state properties are depicted in a *motion map* consisting of a trajectory on which can be shown position, velocity and acceleration vectors at instants of interest.

   Interaction descriptors include, among others, the concept of force. Forces exerted on the particle-like system by its agents are often depicted by arrows in a *force diagram*. The net force is evaluated using the superposition principle (Newton's 4th law or law of composition).

3. **Structure**

   How are concepts ***related*** *to each other, and what is the subsequent* ***function*** *of the model?*

   Topology facet: The geometric structure, if any in this model, is often restricted to the relative position of the particle-like object and long-range agents.

   State facet: The translation of the particle is described with *state laws* (often called equations of motion) involving only kinematical concepts, such as:

   $\mathbf{a} = \text{constant} \qquad \Delta\mathbf{v} = \mathbf{a}t \qquad \Delta\mathbf{r} = \mathbf{v}_o t + \frac{1}{2}\mathbf{a}t^2 \qquad \Delta\mathbf{v}^2 = 2\mathbf{a}\Delta\mathbf{r}$

   These laws can be depicted with appropriate *graphs*, *relational diagrams* and/or *motion maps*.

   Interaction facet: Object-agent interaction is expressed with appropriate interaction laws, like the Newtonian law of universal gravitation, or the Coulomb law of electrostatic interaction.

   Causal facet: Change of state, e.g., change in the particle's velocity, is explained by *causal laws* such as Newton's 2nd law (dynamical law) or the Work-Energy theorem (conservation law).

   The model may be kinematical (descriptive function), and/or dynamical (explanatory function).

4. **Organization**

   How is the model ***integrated*** *in its theory?*

   Newtonian Theory relates the uniformly accelerated particle model to:

   - other models within the family of particle models (Fig. 2).
   - rigid body models and other families of models (e.g. fluid models) within the theory.

   Rules are established within the theory to combine this model with other basic models in order to study physical systems undergoing more complex motions. For example, this model can be combined with the model of bound particle in uniform circular motion in Figure 2 to come up with the *emergent model* of a particle in circular, uniformly accelerated translation.

Figure 6. A partial content taxonomy pertaining to the Newtonian model of a uniformly accelerated particle model in secondary school and college physics courses.

Constructing and deploying scientific models

- *Exploratory analysis*. This process is a complex analysis process required to analyze physical situations that one needs to explore (describe, explain, predict or post-dict). It extends from the delineation and setting the boundaries of individual systems in the situation (system / pattern recognition) to the discrimination between primary and secondary constituents and properties of the system (discriminatory analysis).
- *Model adduction*. This process allows one to decide when is it appropriate to use a particular scientific model to explore a given situation, or come up with certain innovations about the physical pattern that the model represents. Model adduction relies heavily on one's understanding of the domain of the model, its function and its organization.
- *Mathematical modeling*. This process is about transforming a conceptual scientific model (or a physical situation represented by the model) into a mathematical model that can be efficiently manipulated in model analysis. Mathematical modeling relies heavily on semantic rules discussed in the process concept schema below, and on one's understanding of the rules of mathematical depictions.
- *Model analysis*. This process is about "running" a scientific model (or, often its mathematical counterpart) to explore a given physical situation. Model analysis relies heavily on syntactic processes discussed in the process schema below, and on one's understanding of the structure of the model and of various mathematical operators used in the process.
- *Model reification*. This is about using a scientific model to control or modify an existing reality, or to invent a new one that manifests the pattern represented by the model. This process relies on engineering and technology.
- *Model evaluation*. This process takes place throughout model construction or deployment, i.e., in conjunction with the other processes listed above. Every step taken through any of those processes need to be evaluated in order to ensure not only the viability of the step in question, but also the viability of the processed model for the situation in which it is used, and to subsequently consolidate the model or refine it in the context of the scientific theory to which it belongs.

Generic inquiry processes.
In addition to the above model-specific processes, there are other generic processes needed in concept or model construction and deployment, or any other form of scientific inquiry. These include: semantics and syntax, critical thinking (including discriminatory analysis and evaluation), analogical and metaphoric reasoning, conjecturing (making assumptions about a particular situation), inference making, analysis and synthesis (in the broad sense), extrapolation, and mathematical processing and interpretation.

Figure 7. A partial process taxonomy pertaining to scientific models in secondary school and college physics courses.

1. Law of Inertia (Newton's 1st law)

The state of inertia of physical bodies is characterized with a constant velocity (that is not necessarily zero) in Galilean reference systems.

No external cause, and more specifically no interaction, is needed to maintain such a state.

2. Interaction and Force

No physical body can act on itself. An interaction takes place between at least two bodies, an « agent » (acting body) and an « object » (body acted upon) whose kinematical and/or dynamical state is being investigated.

The concept of force represents agent-object interaction. A force of particular characteristics is associated with a particular kind of interaction. These characteristics are not affected by the kinematical state of the object (current or past), or by the object interaction with other agents. In particular, motion does not imply force (impetus), and the force exerted by a given agent on an object does not build up or get used up because of the motion of the object.

The force acting on an object lasts as long as the interaction with the respective agent is taking place. It vanishes at the instant the interaction is brought to an end. The same goes for the force effect on the object.

3. Law of Interaction (Newton's 3rd law)

Agent and object exert simultaneous forces on one another. The two exchanged forces are equal and opposite, irrespective of the physical or kinematical properties of either body.

4. Law of Cause and Effect (Newton's 2nd law)

An object must interact with at least one agent in order to change its state of inertia, and more specifically to change the direction or the magnitude of its velocity.

The concept of acceleration represents the effect of interaction between agent and object. Acceleration and not velocity of object is proportional to the exerted force and inversely proportional to the object mass, and this irrespective of the nature of interaction.

5. Law of Composition (Newton's 4th law) / Superposition Principle

Many forces can be composed only if exerted simultaneously on the same object.

Simultaneous interaction of a given object with many agents is identical in cause and effect: (a) to the absence of any interaction when the sum of all forces acting on the object is zero, or, otherwise, (b) to its interaction with a single agent that exerts on it a force equal to the vectorial sum of all forces exerted by the original agents.

The kinematical state of the object may be determined by the superposition of motions that it would have undergone, during the same period, under each dynamical state separately.

6. State Laws

The kinematical state of a given object, from a particular moment onward, depends on the velocity of the object at this moment and its interaction with all influential agents. This state is independent of prior motion of either object or agents.

Under the action of a constant force, an object maintains a uniformly accelerated motion following: (a) a linear trajectory when its initial velocity (at the time the force starts acting) is either zero or pointing in the (same or opposite) direction of the force, or (b) a parabolic trajectory when this is not the case with the velocity.

The velocity of a uniformly accelerating object changes in proportion to the duration of motion and not to the distance traveled. For a given acceleration, duration of motion and velocity change are independent of the object mass. When the object slows down until a point where it turns around in the opposite direction, the object does not stop at this point; motion in both directions is symmetric and it takes place all along with the same acceleration.

Whatever their motion in a given reference system, two objects that occupy the same position at a given time do not have necessarily the same speed at this time. However, two objects may have the same acceleration when they move with different velocities.

Figure 8. A partial *content* taxonomy associated with the *basic threshold* in Newtonian theory of mechanics.

APPENDIX

SCHEMA-BASED PEDAGOGY
OF
MODEL-CENTERED SCIENTIFIC EPISTEME

Middle-out, model-centered scientific episteme

Traditional instruction fails its mission partly because it is test-driven, but mostly because, in the first place, it does not empower students for meaningful and equitable learning of course materials. The case is especially true in science where students are driven to learn by rote loose bundles of theoretical statements and problem solving routines. Research has constantly shown that students often complete and pass their science courses without necessarily understanding what science is all about (Bransford et al., 1999; Cobern, 1995; Hake, 2002; Halloun, 1986, 2001b, 2008; Halloun & Hestenes, 1985b, 1998; Novak, 1994; NRC, 2001; NSB, 1990-2008). Prominent reform programs are calling for science curricula to reverse the situation by aligning course materials with scientific episteme and practice in ways to help students meaningfully understand the structure of scientific knowledge and develop skills of scientific inquiry (AACU, 2002; AAAS, 1993; NCEE, 1983; NRC, 1996, 2002; NSTA, 1995; Osborne & Dillon, 2008).

A scientific *episteme* is a body of conceptual knowledge accepted by a community of scientists. It consists primarily of a set of scientific theories that have been corroborated in the real world, i.e., theories whose viability (validity and reliability) has been established through sufficient evidence in the real world. A scientific theory is, for us, a conceptual system that consists of: (a) a set of models or families of models, and (b) a set of generic rules and theoretical statements (axioms, laws, etc.) that are particular to the theory in question and that govern model construction and deployment. The theory is corroborated indirectly when its models fulfill, to certain extents, specific functions in the real world. A scientific model *represents*, in some respects and to a certain extent, a particular *pattern* in the structure and/or behavior of real-world systems, and serves specific *functions* regarding the pattern in question. These functions may be *exploratory*, i.e., about pattern description, explanation, post-diction or prediction, or *innovative*, i.e., about pattern reification through control or change of existing physical realities (systems or phenomena), and/or invention of new realities like in technology (Halloun, 2001a; 2004a/2006, 2007).

Prominent educational reform programs agree that the construction and deployment of *scientific conceptual models* are *critical* for reaching high levels of scientific literacy or proficiency, and that science courses need to help students understand the pivotal role of models in scientific episteme and of modeling in scientific inquiry (AAAS, 1993; Johsua & Dupin, 1989; Justi & Gilbert, 2002; NAGB, 2004; NCTM, 1989, 1991; NRC, 1996; OECD, 2003). Our work on modeling theory and assessment highly resonates with this position (Halloun, 2000, 2001a, 2004b/2006, 2007). Our work also resonates with cognitive research that shows the importance of models and modeling in all sorts of human endeavor (Bower & Morrow, 1990; Gentner & Stevens, 1983; Giere, 1992; Johnson-Laird, 1983; Lakoff, 1987; Lakoff & Johnson, 1980), and that implies that models occupy the basic level in *middle-out* organization of human knowledge (Lakoff, 1987).

Many cognitive scientists have shown that, in accordance with the theory of prototypes and basic-level categories of Eleanor Rosch, "categories are not merely organized in a hierarchy from the most general to the most specific, but are also organized so that the categories that are cognitively basic are 'in the middle' of a general-to-specific hierarchy… Categories are not organized just in terms of simple taxonomic hierarchies. Instead, categories 'in the middle' of a hierarchy are the most basic, relative to a variety of psychological criteria" (Lakoff, 1987, pp. 13 and 56). For example, "dog" is "in the middle" of a hierarchy between "animal" and "retriever", just as "chair" is between "furniture" and "rocker" (Fig. 9). Categories in the middle are *basic* in the sense that: (a) they ensure in the best way possible a cohesive structure of human knowledge of any type, and that (b) they constitute the most accessible, efficient and reliable building blocks in knowledge construction and deployment.

Rosch and other cognitive psychologists concentrated their work on *prototypical* or *familiarity criteria* in their "general-to-specific hierarchy". We believe, and our work has begun to show, that the middle-out hierarchy is generic, and that it extends to all sorts of classification criteria. For instance, this hierarchy still holds when *intrinsic structural criteria* are adopted with respect to either the physical world (as seen from a human perspective) or the conceptual world as indicated in Figure 9. In this perspective, models occupy the middle of conceptual hierarchy, between theory and concept in any scientific episteme, just like solar systems occupy the middle of physical hierarchy between a galaxy and its planets.

The model-centered, middle-out structure of scientific theory ensures theory coherence and consistency from an epistemological perspective, and it facilitates people's development of scientific knowledge from a cognitive perspective. A scientific model is to theory and concept what an atom is to matter and elementary particles. Each elementary particle is essential in the structure of matter, but its importance cannot be conceived independently of its interaction with other particles inside an atom. It's the atom and not elementary particles that give us a coherent and meaningful picture of matter, and it's the atom that displays best the role of each elementary particle in matter structure. Now, Bohr's model of the atom is essential for understanding hydrogen-like atoms, and is often referred to as a "model" in physical science textbooks. However, the word model is seldom used in reference to other scientific models, which would give students the false impression that Bohr's model may be about the only scientific "model". Furthermore, various concepts and laws are often presented episodically, one after another in a given chapter, without relating them to one another in the context of

Categories Hierarchy (according to Eleanor Rosch & George Lakoff)

SUPERORDINATE	Animal	Furniture
BASIC LEVEL	Dog	Chair
SUBORDINATE	Retriever	Rocker

Furniture ← Chair → Rocker

***Our own extrapolation to science** (Halloun, 2007):*

Real World Structural Hierarchy:

SUPERORDINATE	Matter	Galaxy
BASIC LEVEL	Atom	Solar System
SUBORDINATE	Elementary particle	Planet

Conceptual Hierarchy in a Scientific Theory:

SUPERORDINATE	Theory
BASIC LEVEL	Model
SUBORDINATE	Concept

Model Hierarchy:

SUPERORDINATE	Emergent model
BASIC LEVEL	Basic model
SUBORDINATE	Subsidiary model

Figure 9. Middle-out hierarchies.

The term "hierarchy" takes here a new meaning; it is no longer restricted to a linear, top-down or bottom-up, order.

appropriate models, whether implicitly or explicitly. Students are thus deprived of the opportunity to develop a coherent, model-based, picture of scientific theory and episteme, and they end up with a piecemeal, fragmented picture of the world. To get a feel of this picture, imagine what your knowledge about physical realities would look like, should you have learned at school that matter consists of elementary particles and should no mention about the atom was ever made.

Models in a given scientific theory are, for us, also categorized in a middle-out hierarchy as shown at the bottom of Figure 9. In the middle of model hierarchy are basic models. A *basic model* is one that is simple enough to facilitate student learning of fundamental tenets and conceptions (concepts, laws, etc.) of the respective theory and of fundamental tools and skills of scientific inquiry. A basic model is yet generic enough to serve in the construction of more complex models in the theory. The set of basic models in Newtonian theory are given in Figure 10 for illustration. A student needs to understand the entire set of basic models so that s/he could meaningfully learn the theory in question, and realize a meaningful paradigmatic evolution.

At the subordinate level of model categories are subsidiary models. A *subsidiary model* is a simplified basic model, a particular case which students may usually be most familiar with, and that can serve as a stepping-stone for the comprehensive construction of the basic model in question. For example, the model of a particle in free fall (objects falling in vacuum in the absence of any force except for gravity) is a subsidiary model in Newtonian theory. It serves for the progressive construction of the uniformly accelerated particle model (Fig. 10). At the superordinate level of model categories are emergent models. An *emergent model* is one that may be constructed by putting a given basic model together with one or more other models in

Free particle

Physical objects subject to no net force ($\Sigma \boldsymbol{F}_i = 0$), and thus maintaining constant velocity in any inertial reference system ($\boldsymbol{a} = 0$, $\boldsymbol{v}$ = constant).

Uniformly accelerated particle

Physical objects in linear or parabolic translation with constant acceleration ($\boldsymbol{a}$ constant) under a net constant force ($\Sigma \boldsymbol{F}_i$ = constant).

Bound particle in harmonic oscillation

Physical objects undergoing periodic back and forth translation (sinusoidal $\boldsymbol{a}$ function) under a net force that is proportional to their displacement from a center of force ($\Sigma \boldsymbol{F}_i \propto \Delta \boldsymbol{r}$). This model is often called simple harmonic oscillator.

Bound particle in uniform circular motion

Physical objects in uniform circular translation ($a = v^2/r$) under a net centripetal force ($\Sigma \boldsymbol{F}_i \propto \boldsymbol{r}/r^2$) of constant magnitude.

Particle under impulsive interaction

Physical objects whose linear momentum changes significantly, and almost instantaneously, like in the case of collision, under a variable net force ($\Sigma \boldsymbol{F}_i = \boldsymbol{f}(t)$) exerted for a very short period.

Figure 10. Basic particle models in Newtonian theory of classical mechanics, with an outline of the translational pattern that each model represents in inertial reference systems.

Particle models refer to physical objects the internal structure of which can be ignored when they are in *translation* without rotation or precession, in a specific reference system. Each *basic* particle model is made up of a single, dimensionless object: a particle.

order to represent a pattern that cannot be represented by either model separately. The model of a bound particle in uniformly accelerated circular motion is an example of emergent models. It emerges from combining two basic models in Newtonian theory shown in Figure 10, the uniformly accelerated particle model and the bound particle in uniform circular motion.

A science course, especially at the secondary school and college (introductory university) levels, is usually about a particular scientific theory, and sometimes about a set of interconnected theories. The course content can be organized around a number of models that may be graded into categories of increasing structural and functional, and thus *epistemic*, complexity. Each category characterizes a cognitive evolution level that students need to attain at a certain point of instruction. Our research suggests that models of a given theory, and thus course content, can be pedagogically classified into three categories of increasing epistemic complexity (Fig. 10). The first category includes *primary models*. These are *simple basic models* relative to which students usually have the richest repertoire of subsidiary models, and thus in the context of which students can begin to develop the most fundamental conceptions of the theory (generic concepts, laws and other theoretical statements). The second category includes the rest of, and more complex, *basic models*. The third category includes *emergent models*. For example, the five particle models of Newtonian mechanics outlined in Figure 11 are basic models typically targeted in secondary school and introductory university physics courses. The first two models in this figure, the free particle model and the uniformly accelerated particle model, make up the category of primary models Emergent models in classical mechanics courses usually include the model of a particle in uniformly accelerated circular motion, models of particles in elliptical motion and other types of motion with variable acceleration.

The three model categories are organized and graded in such a way that students cannot meaningfully learn any model in a given category before learning all models in the lower category. The three categories are thus separated by *critical* demarcation lines. They are critical in the sense that at the level of each line is set a *threshold* of understanding that students need to meet before crossing into the upper category. Two *critical thresholds* can

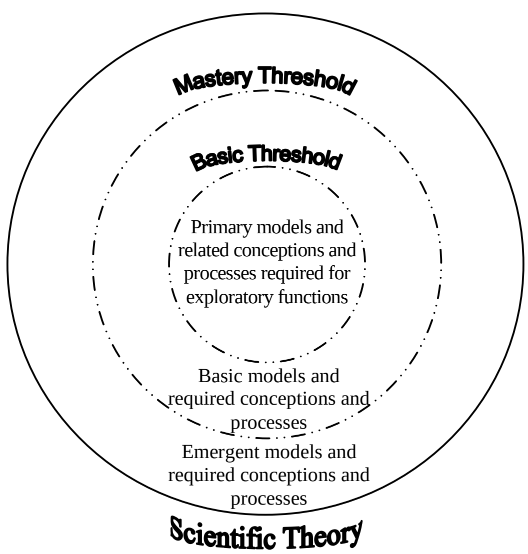

Figure 11. Model-based content categorization and cognitive evolution in a science course.
Evolution from one stage to another is not possible unless students meaningfully attain a particular critical threshold.

thus be set in any given course: the *basic threshold* between primary models and the rest of basic models, and the *mastery threshold* between basic and emergent models. As we discuss below, *formative assessment* in any given course needs to be carried out in specific ways within each category, and *summative assessment* needs to concentrate on the critical thresholds so as to determine the opportune moment for moving in the course of instruction from one category to another (Fig. 11). Both forms of assessment can be efficiently carried out using modeling schemata discussed in the following section.

Modeling schemata for instructional design and assessment

For efficient learning, teachers need to put together and forth coherent and systematic lesson plans that are flexible enough to accommodate both scientific rigor and pedagogical concerns. They especially need to integrate assessment with instruction so as to determine the most effective ways for mediating meaningful and equitable learning of course materials. To these ends, teachers need to be empowered with appropriate tools to lay out course content and design learning activities and assessments that cover comprehensively all fundamental aspects of course materials. The most important of these tools are content and process schemata.

A *schema* is, for us, a generic organizational template for spelling out: (a) all salient features that a student needs to know about a given conception (content schema), or (b) all processes that the student needs to master for meaningful construction and deployment of the conception (process schema). In our modeling theory, we consider that patterns extend from the physical world to the conceptual world, and that there is a pattern in the structure (and related processes) of all scientific conceptions of a given type, be it concepts, laws or any other type of theoretical statements (axioms, definitions, etc.), or models. We further consider that two conceptual patterns are most critical to realize the middle-out, model-centered perspective of scientific theory and episteme (Fig. 9). These are the pattern underlying the structure (or related processes) of all scientific concepts, and the pattern underlying the structure (or related processes) of all scientific models. With these two patterns we associate respectively the concept schema and the model schema. The two schemata are generic in the sense that, for content or processes, there is only one concept schema to cover all aspects of any scientific concept, and only one model schema to cover all aspects of any conceptual model in any given scientific theory. Furthermore, the two schemata are *modeling schemata* in the sense that they promote construction and deployment of concepts and models from the perspective of modeling theory in science education (Halloun, 2001a, 2004a/2006, 2007).

Content schemata

The two modeling schemata are discussed elsewhere with ample details from a content perspective (Halloun, 2000, 2001a, 2004a/2006). We briefly review them in this subsection. As illustrated in the body of the document, the schemata are especially useful in laying out taxonomies of different levels of complexity and details in any science course. Practical examples were given there on their utility in the context of perhaps the most familiar scientific theory of them all, the Newtonian theory of classical mechanics (Figs. 4, 6).

The *content model schema* is a four-dimensional template for putting together any scientific model, at least those models that are the object of study in secondary school and college science. Two of the four dimensions, composition and structure, set the ontology and function of the model, and the other two, domain and organization, set its scope, all in terms of the scientific theory that the model belongs to, and by correspondence to physical realities displaying the modeled pattern. The content model schema is illustrated in Figure 6 with the uniformly accelerated particle model of Newtonian theory.

The *domain* of a scientific model specifies what physical realities (systems and/or phenomena) the model represents in the real world, and under what conditions. The realities in question, called model *referents*, display the physical pattern that the model represents.

Model *composition* consists of concepts representing *primary* constituents and respective properties of physical systems, i.e., only those constituents and properties that are salient to the pattern. Concepts of interest to us at this level are mainly of two types: object-concepts and property-concepts. *Object-concepts* (or conceptual objects) represent physical bodies that significantly contribute to the making of the pattern represented by the model. These may be *objects* that enter in the make up of each physical system of interest, or *agents* in the environment of the system, i.e., physical bodies outside the system that interact significantly with objects inside. *Property-concepts* (or *descriptors*) represent primary physical properties of objects and agents, and of their mutual interaction.

Model composition is meant to discern between primary and secondary aspects of a pattern, i.e. between those aspects that need to be accounted for in the modeling process and those that may be ignored within the considered limits of precision and approximation. In model composition, primary object and property concepts are only listed and not related to one another. Model structure spells out relevant relationships among primary features of the pattern represented by the model, and set the function of the model.

Model *structure* can be defined along four sub-dimensions, or facets, each dealing with a specific aspect of model referents in relation to pattern formation. These are: (a) the topology facet that lays out the *geometric* structure of various objects represented in the model, (b) the state facet that *describes how* each object behaves, (c) the interaction facet that specifies how various objects *interact* with each other, and (d) the cause-effect or causal facet that *explains why* objects behave the way they do. Each facet is distinguished conceptually by the nature of descriptors involved and the ways they are related in space and time. Various relationships are expressed in an appropriate reference system relative to which the pattern is conveniently identified. Such relationships come primarily in the form of laws that set the distinctive descriptive and/or explanatory *function* of the model.

Model *organization* situates a given model in the respective scientific theory. It establishes how the model in question differs from, or is similar to, other models in the theory, and how it may be extrapolated in the construction of more complex models.

Concepts are elementary building blocks of models. They gain their significance only when used in model construction, and more specifically in spelling out laws, definitions and other theoretical statements that make up the model structure (Fig. 6). In order to build concepts comprehensively and integrate them coherently into respective models and theory, the model schema is complemented with the concept schema.

The *content concept schema* is a four-dimensional template used for the construction of individual concepts within the context of basic models. The four dimensions are scope, expression, organization and quantification (Halloun, 2000, 2001a, 2004a/2006). They are concisely presented below for property-concepts or descriptors, and were illustrated in Figure 4 with the concept of force in Newtonian theory.

The *scope* of a concept sets the domain and function of the concept. A descriptor represents, to a certain degree and within certain limits, a particular physical property shared by many real world systems or phenomena. It has a domain confined to the represented property and a particular function (descriptive or explanatory) that depends on the nature of the property.

The *expression* of a scientific concept is done objectively so as to name it and feature it in ways to clearly distinguish it from any other concept. Particular *semantics* establish what the concept expression actually delineates in the real world or the rational world of scientific episteme. A mix of verbal, symbolic, iconic, and especially mathematical forms of expression is commonly used to communicate any scientific concept. The mix is necessary to come as close as possible to a comprehensive expression of the concept, since no single form can actually do so alone.

Concept *organization* sets criteria and guidelines for classifying a concept and for relating it to other concepts, all along with appropriate syntax rules. Three concept categories may be distinguished in science: object-concepts, property-concepts or descriptors, and logico-mathematical operators (e.g., equality, addition, derivatives and integrals). Syntax rules set how a given object-concept or, especially, property-concept, can be related to other concepts of its category, mostly using logico-mathematical operators (e.g., state, interaction and causal laws in the structure of a model).

Quantification specifies how a scientific descriptor can be measured. A major distinctive feature of scientific descriptors is that they are measurable according to well-defined laws and rules (otherwise, a descriptor cannot be scientific). The sort of measurement which the descriptor can be subject to can be set accordingly (nominal, ordinal, interval or ratio), along with the means and procedures which one can resort to for determining values of the concept on a certain scale, and by comparison to a certain standard.

Process schemata

The two content schemata discussed above allow teachers to decide what is needed to put together a given concept or model, but not how students should go about constructing or deploying the concept or the model in question. Process schemata are needed to the latter ends. Model construction and deployment require contextual and generic processes. Generic processes extend to all sorts of models in any scientific theory. For those processes, we define a generic *process model schema* in the form presented in Figure 7. Contextual processes pertain to individual concepts and laws in any given model. Typical processes of the sort are presented in Figure 5 which follows the process concept schema.

The *process concept schema* includes processes that are more context dependent than the generic processes listed in Figure 7. As shown in Figure 5, this schema especially includes processes of semantics and syntax. *Semantic* processes are processes required to determine what a concept represents and under what conditions. They rely on one's understanding of the concept scope. *Syntactic* processes are the ones needed to relate one concept to another. They rely on one's understanding of the concept organization. The process concept schema may also incorporate some of the generic processes listed in Figure 7 that have particular importance in concept construction and, especially, in concept deployment. These include inferential, evaluative and mathematical processes. *Inferential* processes are about various sorts of conjectures that one needs to make while using the concept for exploration or innovation purposes. *Evaluative* processes pertain to all sorts of judgment one needs to make about the viability of the concept. *Mathematical* processes pertain to symbolic representations and operations that one can undertake with the concept.

Students need to develop all processes, including generic ones, in the context of specific physical realities that models and concepts represent. One can never expect students to develop any of these processes in the abstract world. Our position in this respect is consistent with the view of many reformists that “processes have to be used in relation to some subject matter; there is no meaning to a content-free process” (OECD, 2003), at least not at the secondary school and college levels.

There are, of course, many other processes required for meaningful understanding of any course materials. Perhaps the most important of which are dialectic and other cognitive processes extending from intrinsic mnemonics for storing information and retrieving it from memory to extrinsic discourse styles for communicating and negotiating one’s own ideas with others. Dialectic processes are concerned with the negotiations that one undertakes while constructing or deploying a given conception within one’s own rational world and between this world and external worlds (whether the empirical world of natural realities or the rational world of an established science). Such processes and related meta-cognitive control factors are beyond the scope of our discussion in this document.

References

American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy, Project 2061*. New York: Oxford University Press.

American Association for the Advancement of Science. (1990). *Science for All Americans. Project 2061*. New York: Oxford University Press.

Association of American Colleges and Universities. (2002). *Greater Expectations: A New Vision for Learning as a Nation Goes to College.* Washington, DC: AAC&U.

Bower, G.H. & Morrow, D.G. (1990). Mental Models in Narrative Comprehension, *Science* 247, 44-48.

Bransford, J.D., Brown, A.L. & Cocking, R.R., Eds. (1999). *How people Learn: Brain, Mind Experience and School.* Washington, DC: National Research Council, Committee on Developments in the Science of Learning.

Cobern, W. W. (1995). Science education as an exercise in foreign affairs. *Science & Education, 4* (3), 287-302.

Gentner, D. & Stevens A. L. (eds.). (1983). *Mental Models.* Lawrence Erlbaum, Hillsdale.

Giere, R.N. (ed.). (1992). *Cognitive Models of Science.* Minnesota Studies in the Philosophy of Science, Vol. XV. University of Minnesota Press, Minneapolis.

Hake, R. R. (2002). Lessons from the physics education reform effort. *Ecology and Society, 5(2),* 28.

Halloun, I. A. (2008). *The Lebanese Secondary School Physics Curriculum and Students' Conceptual Profiles*. (2008). Beirut: Lebanese University.

Halloun, I. A. (2007). Mediated Modeling in Science Education. *Science & Education*. In press.

Halloun, I. A. (2004a/2006). *Modeling Theory in Science Education.* Dordrecht: Kluwer/ Springer.

Halloun, I. A. (2004b). Normative evaluation of mathematics, science and technology curricula: The case of the modeling curriculum. *Proceedings of the 12th annual meeting of the Southern African Association for Research in Mathematics, Science and Technology Education*. Cape Town, South Africa: SAARMSTE.

Halloun, I. A. (2001a). *Apprentissage par Modélisation : La Physique Intelligible.* Beyrouth : Phoenix Series / Librairie du Liban Publishers.

Halloun, I. (2001b). *Student Views About Science: A Comparative Survey.* Beirut, Lebanon: Phoenix series / Educational Research Center, Lebanese University.

Halloun, I. A. (2000). Model-laden inquiry: A prescription for effective physics instruction. *THEMES, 1* (4), 339-355.

Halloun, I. (1986). Le réalisme naif et l'apprentissage de la physique. *Recherches Pédagogiques, 17,* 23-47.

Halloun, I, & Hestenes, D. (1998). Interpreting VASS dimensions and profiles. *Science & Education, 7,* 553-577.

Halloun, I, & Hestenes, D. (1985a). The initial knowledge state of college physics students. *American Journal of Physics, 53,* 1043-1055.

Halloun, I, & Hestenes, D. (1985b). Common sense concepts about motion. *American Journal of Physics, 53,* 1056-1065.

Helm, H. & Novak, J. (eds.): 1983, *Proceedings of the International Seminar: Misconceptions in Science and Mathematics,* Cornell University, Ithaca.

Johnson-Laird, P.N. (1983). *Mental Models.* Cambridge, MA: Cambridge University Press.

Johsua, S. & Dupin, J.J. (1989). *Représentations et modélisations : Le débat scientifique dans la classe et l'apprentissage de la physique*. Berne : Peter Lang.

Justi, R. & Gilbert J. K. (2002). Models and Modelling in Chemical Education. In: J. K. Gilbert, O. de Jong, R. Justi, D. F. Treagust & J. H. van Driel (Eds), *Chemical Education: Towards Research-based Practice,* 47-68. Dordrecht: Kluwer Academic Publishers.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. *What Categories reveal about the Mind*. Chicago, IL: The University of Chicago Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

National Assessment Governing Board. (2004). *Science Framework for the 2005 National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education.

The National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform.* Washington, DC: U.S. Government Printing Office.

National Council of Teachers of Mathematics. (1991). *Professional Standards for Teaching Mathematics.* Reston, VA: NCTM.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: NCTM.

National Research Council. (2003). *Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering and Mathematics.* Washington, DC: The National Academies Press.

National Research Council. (2002). *Scientific Research in Education.* Washington, DC: The National Academies Press.

National Research Council. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment.* Washington, DC: The National Academies Press.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: The National Academies Press.

National Science Board. (1990-2008). *Science & Engineering Indicators.* Washington, DC: National Science Foundation; U.S. Government Printing Office.

National Science Teachers Association. (1995). *Scope, Sequence, and Coordination of Secondary School Science. Volume 3. A High School Framework for National Science Education Standards*. Washington, DC: NSTA.

Novak, J.D. Editor. (1994). *Proceedings of the Third International Seminar. Misconceptions and Educational Strategies in Science and Mathematics*. New York, NY: Cornell University.

Organization for Economic Cooperation and Development. (2003). The PISA 2003 Assessment Framework. Mathematics, Reading, Science and Problem Solving Knowledge and Skill. Available at www.pisa.oecd.org.

Osborne, J. & Dillon, J. (2008). *Science Education in Europe: Critical Reflections.* London: The Nuffield Foundation.