EVALUATING SCIENCE AND TECHNOLOGY LEARNING MATERIALS

THE CASE OF THE MODELING CURRICULUM

IBRAHIM A. HALLOUN

CONTENTS

1. Development of STE materials: A basic modus operandi
2. Modeling theory for science and technology education
3. Norms and guidelines of normative evaluation
4. Normative evaluation of modeling materials
5. Evaluation for regulation in action
6. Toward meaningful and equitable learning of science and technology

P. O. Box 2882; Jounieh
www.inco.com.lb/halloun

UNESCO Regional Workshop *Beirut, May 2003*

ABSTRACT

Evaluation of learning materials is an integral part of curriculum development in science and technology education (STE). It is not an end by itself, and it does not mark the end of curriculum development. Evaluation of STE materials is carried out for continuous curriculum regulation. Regulation may extend from minor refinements of some materials and other curriculum components to curriculum reform on entirely new grounds. For reliable outcomes and sustainable impact, evaluation of learning materials needs to be normative. Normative evaluation is conducted with a battery of instruments selected according to a well-defined taxonomy of the sort of knowledge, skills and worldviews students are expected to develop with materials being evaluated. It comes out with quantitative indicators ascertaining in terms of pre-established criteria to what extent students have actually met original expectations. In order to steer curriculum regulation in the right direction, normative evaluation also bears on teacher practice in the classroom and the impact of such practice on student achievement. Normative evaluation is resorted to in the development of the modeling curriculum in physics and other STE fields, at the secondary school and university levels. Respective data are presented and discussed in the context of Newtonian mechanics courses so as to illustrate with practical examples how normative evaluation of learning materials may be conducted in a way to contribute to sustainable reform and meaningful and equitable learning in STE.

Development of new learning materials in any field is an involved process that goes beyond the conception of such materials and their reification in the form of textbooks, equipment, computer software, or the like. The process extends to the evaluation of the materials in question in order not only to determine their value, but more importantly the extent to which they need to be refined or even reconceived altogether. As such, evaluation of learning materials is not an end by itself, and it does not end as long as materials are in use. A curriculum is continuously being put to the test as it is being implemented. This is especially the case with learning materials, the evaluation of which provides significant indices of the curriculum merits and reliable indicators of the direction in which curriculum refinement or reform should be headed. This is true for any educational field. It is more so for science and technology education in an era where virtually all aspects of human life are dominated by the two fields. In this paper, we discuss modes and means for evaluating science and technology education materials, and we illustrate with practical examples and actual data how evaluation outcomes can be used to refine learning materials, all in the context of modern educational theory.

Around the turn of the twentieth century, many an organization around the world has realized the relevance of science and technology education (STE) at the pre-college levels. UNESCO's stand on this matter was among the most influential in developing countries. The organization called for many conferences on the subject and instituted many international and local programs in order to promote STE along the lines of modern educational theory. Project 2000+ was perhaps the most influential in this respect (UNESCO, 1993, 1994). Many STE programs sprung out of this project in dozens of countries around the world. Under the auspices of the STE section at UNESCO, some of these programs were evaluated in the fall of 2001, and appropriate recommendations were made with respect to the general STE policy promoted by UNESCO and the respective programs it sponsors around the world. Some of these recommendations were specifically about evaluation of science and technology learning materials. In this respect, the evaluation committee recommended, among others, that concerned educators (Bibeau, Halloun, May & Reddy, 2002):

- situate their STE curricula in clearly defined educational frameworks;
- develop and evaluate teaching/learning materials accordingly and in the direction of "profound and sustainable" curriculum reform;
- evaluate materials in terms of clearly defined, "reasonable" standards, and ensure that *all* students meet in this respect a "minimal set of STE standards";
- adopt authentic means of student assessment that contribute to the continuous evaluation and enhancement of adopted materials.

This paper discusses how these recommendations may be implemented and what they may lead to when put to practice. The paper comes in six sections. In the first section, general procedures are outlined for the development of science and technology education materials. Discussion concentrates on the fact that materials' evaluation is an integral part of curriculum development, and that evaluation should be meant for continuous curriculum regulation. The entire process should be carried out within the framework of a conveniently chosen educational theory, and this in a way to benefit from potentials and account for constraints that are normally on the ground. Modeling theory, a synopsis of which is presented in the second section of the paper, is advocated as a viable theory for STE. Materials from a physics curriculum developed within the framework of modeling theory serve for illustration in subsequent sections. In the third section is a discussion of the kind of evaluation that STE materials developers need to undertake in order to tell whether things are working in the right direction, i.e., whether they are actually contributing to meaningful, equitable and sustainable reform. This is normative evaluation whereby the value of STE materials is ascertained with particular instruments and with respect to pre-established, and conveniently graded taxonomy and criteria. Taxonomy and criteria pertain to "reasonable" aspects of respectively what instructional modes teachers can assume in the classroom with developed materials and what conceptions, processes, and worldviews students may come out with. Procedures of normative evaluation as implemented in the evaluation of a modeling curriculum are outlined in the fourth section of the paper. Specific outcomes are presented and discussed in the following section so as to illustrate how one can go about evaluating STE materials in a way to regulate a new curriculum and steer it in the right direction. Implications for sustainable reform are then discussed in the last section of the paper.

1. Development of STE materials: A basic modus operandi

Science and technology (S&T) have been the main concern of curriculum reformers around the turn of the past century, and they have been major driving forces behind virtually any curriculum reform movement, for whatever discipline and at any educational level. STE and S&T-driven reforms in other fields were triggered by the colossal growth of S&T knowledge and applications and their increasingly significant impact on virtually every aspect of human life. Calls for scientific or S&T literacy at the level of the general public have become the norm in educational policy around the world, and so has been the drive to look into S&T developments for a better understanding of effective human thoughts and consequently for more effective educational theory.

The state of education, and especially of STE, has reached such an alarming level by the end of the past century that radical reforms in STE policy and curricula have been called for around the world. Various actors in education and in S&T kept heading in virtually opposite directions for the best part of the twentieth century. Educators kept relying until late that century on rules of thumb for conceiving and implementing various educational curricula. Meanwhile, scientists and engineers were developing S&T following systematic norms and rules stemming from well-defined and corroborated scientific theory. Action-research was also becoming the norm in S&T but not in education. This is evaluative research intended to continuously enhance productivity (as well as equity) in many respects. Educators have finally come to realize that, in order to turn things around in STE, and education in general, they need to follow in school practice that has proven to be successful in scientific research and technological developments. In particular, rose the conviction that for meaningful and sustainable reform, STE should be theory-laden and steered by action-research.

In the first respect, educators became more and more convinced (gradually, through practice) that development of learning materials for science and technology, like for any field, has to be grounded in a well-defined educational theory. Educational theory sets the general framework (i.e., underpinnings, norms and guidelines) for making decisions about various aspects of a given curriculum. This includes, in the case of STE, the choice of whether to conceive science and technology in integrated or separate curricula. Such a choice, like the choice of curriculum objectives, is though governed by a number of factors that may be beyond the scope of the chosen theory and/or the control of curriculum developers. These factors include national and local policies about S&T, and especially the feasibility of implementing a particular approach or targeting a specific objective in a given society. Other theory-laden decisions pertain to the kind of profile students should develop under a given curriculum, and to the ways they should go about developing such a profile (i.e., to *what* and *how* should students learn). A *student profile* includes all sorts of *conceptions* (including theories, models and concepts) and *processes* (including related tools, schemes and rules) that any student willing and capable of investing necessary efforts should develop following the completion of the curriculum in question. As a consequence, student general *skills*, *habits* (including learning styles) and *worldviews* (including views about the nature and relevance of S&T), should evolve in specific directions, and this in meaningful and sustainable ways. A synopsis of modeling theory is presented in the next section for illustration. The theory has been implemented in the past two decades in the development of physics curricula at the high school (secondary school) and college (university) levels in a number of countries, and is currently being deployed in other S&T fields at the same and lower educational levels. Examples and data from the modeling curriculum are resorted to for illustration throughout this paper.

As mentioned above, curriculum developers, and especially educators engaged in the development of learning materials of all sorts (textbooks, equipment, computer software, etc.), have not only to ground their work in a well-defined educational theory. They also have to take into account realities on the ground that affect their work (Figure 1). Concerned actors need to consider these realities in their *strengths* and *weaknesses*, take advantage of the *opportunities* they provide and circumvent *threats* they may impose. This is often referred to as the *SWOT* rule of the game in business and industry; the rule applies as much to education as to any other enterprise. The *SWOT* rule applies to all *stakeholders* in the process. These include material developers, publishers, teachers, students, administrators, parents, and other concerned groups and communities. The rule also applies to *systemic structures* that are involved in the process and that include support systems and programs, and to logistics of all levels for developing, field-testing and implementing new materials at various scales, and subsequently refining (or reconceiving) such materials.

Teachers have traditionally been the most marginalized stakeholders in the development of curricula and even of learning materials. They have often been treated as mere traders of finished products or as passive reporters about, or disseminators of, canned

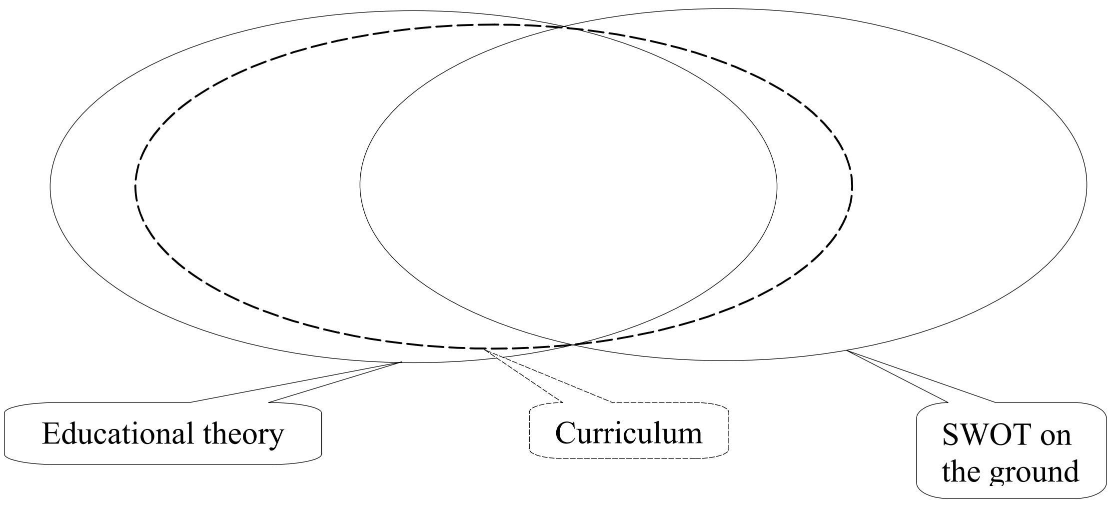

Figure 1: Curriculum developers cannot entirely adhere to their educational theory and account for the *SWOT* rule on the ground. They need to compromise between the two, but in a manner that favors more their theory than ground constraints.

information and goods. Instead, teachers should take an active role at different stages of curriculum and materials development. Perhaps the most important role teachers can and should play in this respect is in *action-research*. As such, teachers should consider themselves, as individuals and groups, as most concerned about continuous *evaluation* of learning materials and subsequently of the respective curriculum, with the intention of contributing to the *regulation* (refinement or even reconception) of materials and curriculum so that they ensure *meaningful* and *equitable* learning experience to all participating students, and thus *sustainable* curriculum changes and/or reform. For action-research to succeed in this direction, systemic structures should be conveniently placed in order and in harmony with the motto of *continuous evaluation for regulation* of learning materials (and respective curriculum).

Evaluation of learning materials is thus not an end by itself, and it does not mark the end of a process (curriculum development). Such an evaluation is meant to *gauge* and direct a curriculum in the direction of meaningful and equitable learning, and of sustainable reform. It is a continuous process that lasts as long as concerned materials are being used. Furthermore, the evaluation in question is not confined theoretically to the educational framework of a given curriculum, and is not restricted in practice to a limited group of outside evaluators nominated by certain authority. Evaluation norms and criteria need to account for the *SWOT* rule on the ground, though not necessarily to the same extent as to the originally adapted theoretical framework. Evaluating bodies should span from external evaluators not originally involved in the conception of the curriculum to various stakeholders in the process and especially concerned teachers. All in all, evaluation of learning materials should answer at least the following *triad* of *A*-questions:

1. What materials have actually been made *available*, and how do they fair between original intentions of curriculum developers (as stemming from the adapted educational theory) and realities on the ground (SWOT)?
2. What is then *anticipated*, of teachers and of students, to accomplish?
3. What has actually been *achieved* on the ground? Who and what made that possible? How do we know?

2. Modeling theory for science and technology education

Modeling theory is a theory of science (and technology) education that this author has been working on for the past two decades. The theory has been implemented mostly in the development of science (and particularly physics) curricula independently of technology curricula, and this at the secondary school and college (university) levels (Halloun, 1984, 1994, 1996, 1998a, 1998b, 2000; 2001a, 2003; Halloun & Hestenes, 1987). Its recent implementation in some universities in integrated S&T curricula has shown its viability in this respect

Modeling theory is grounded in philosophical and cognitive tenets drawn from the philosophy of science and educational research. It calls for science courses to empower students with tools, skills and habits that help them achieve a *paradigmatic evolution* whereby they transcend their naïve realism (or common sense) and evolve into the realm of science. The evolution is *student- centered*, *teacher-mediated*. It is promoted in a learning environment that is structured enough (by the teacher) to be in line with scientific theory and inquiry, yet flexible enough to account for differences in students' initial knowledge state, and to allow for their self-evaluation and self-regulation.

A person's thought and behavior in a given situation are always governed by a particular paradigm or set of paradigms that constitute a paradigmatic profile. A *paradigm* is, for us, an intellectual system that: (a) determines the conditions of a person's conscious experience, and that (b) sets forth standards, rules and guidelines that govern the choice and the course of all actions necessary for the reification and continuous evaluation of the experience in question. A *paradigmatic profile* of an ordinary person consists of a set of paradigms, of the same scope or of closely related scopes, some of which may be in accordance with scientific paradigms while others at odds with such paradigms.

Educational research shows that secondary school and university students enrolled in science courses are often encumbered with *naïve* paradigms that are at odds with scientific paradigms, and that student paradigms are not significantly affected by conventional instruction of lecture and demonstration. Many science educators have argued that this situation cannot be reversed unless science teachers bring into their classrooms whatever *modeling* practices scientists resort to in their research. The pivotal role of *models* and *modeling processes* in learning scientific theory has also been recognized by prominent organizations such as the American Association for the Advancement of Science and the National Research Council (cf. Halloun, 2003 for an extensive list of references).

A *scientific theory* consists, for us, of: (a) a set of models or families of models, and (b) a set of particular rules and theoretical statements that govern model construction and deployment and that relate models to one another and to specific patterns in the real world. A *scientific model* is a conceptual system mapped onto a specific *pattern* in the structure and/or behavior of some physical systems so as to allow us: (a) describe, explain, and predict (or postdict) the pattern in question, and, eventually (b) control or change physical realities exhibiting the pattern, and (c) reify the pattern in new realities. Among the models of a theory is a subset of *basic models* the understanding of which is set at a paradigmatic *threshold* that a student needs to cross so that s/he could meaningfully learn the scientific theory in question. A *basic model* is one that is simple enough to facilitate student understanding of fundamental tenets and conceptions (concepts, laws, etc.) of the respective theory, and development of fundamental tools, skills and habits of scientific inquiry. Yet it is generic enough to serve in the construction of more complex models in the theory.

Modeling processes are pivotal in scientific inquiry. These processes include: (a) model construction and validation in the context of real world situations, and (b) model adduction and analysis for solving empirical or rational problems. In their endeavor, scientists rely, often implicitly, on various tools, the most important of which are modeling schemata.

From a scientific perspective, a *modeling schema* serves, on the one hand, as an organizational template for structuring models or related conceptions in a compact and coherent way. It offers, on the other, a set of well-defined rules for evaluating and deploying scientific models or their building blocks. From a pedagogical perspective, a modeling schema is indispensable for students to retrace scientists' steps meaningfully and efficiently. It also provides teachers with reliable means for planning instruction, and for assessing student learning and teaching practice.

Under modeling instruction, a student is thus supposed to develop a *profile* of the following characteristics:

- *Conceptions* consisting, in a given course, of a particular scientific theory organized, in a middle-out structure, around a set of basic models in the context of which are developed all lower-level conceptions (laws, concepts, etc.).

- *Tools* consisting of modeling schemata and various other tools (especially mathematical) for model construction and deployment.
- *Schemes* including those associated with various modes of scientific processes, especially modeling inquiry, and those associated with reflective learning styles.
- *Rules* associated with all the above, including semantic and syntactic rules associated with various conceptions, and rules of engagement associated with various tools and schemes of inquiry and with scientific discourse and negotiations.
- *Emergent skills, habits and worldviews* that are characteristic of the scientific community, especially those skills and habits pertaining to model-laden theory and paradigm, and respective views pertaining to the nature and relevance of science (and technology).

Students develop their profiles in five-stage *learning cycles* (exploration, model adduction, model formulation, model deployment, paradigmatic synthesis). Each cycle is devoted to the development of a specific model and the corresponding rules of engagement. All along a cycle, individual students ascertain their ideas in light of empirical and rational evidence at their disposal. They often negotiate these ideas with their peers under the supervision of the instructor, and this in a way that helps them detect and resolve any incommensurability with science that might emerge in the process.

Modeling theory has been corroborated so far within the context of secondary school and university physics courses, especially in U.S.A. and Lebanon. It is now being extrapolated into other scientific fields and educational levels. Two decades of research and development have shown that the theory actually fosters the paradigmatic evolution we are calling for, and that it brings about an equitable learning experience that narrows significantly the traditional gap between high and low competence students (Halloun, 1984, 1994, 1996, 1998a, 2001a, 2003; Halloun and Hestenes, 1987; Wells, Hestenes & Swackhamer, 1995).

3. Norms and guidelines of normative evaluation

Evaluation of S&T learning materials needs to be a *normative* process in order to contribute to meaningful and sustainable reform. As such, answers to the first question about available materials (in the a-triad of § 1) must come by comparison to clearly defined norms and standards that are set in terms of the chosen educational theory and the SWOT rule on the ground. Similar norms and standards should be set in answer to the second triad question, and this *before* a curriculum is implemented and learning materials are developed, so that the third question be answered explicitly in terms of these pre-established benchmarks. In other words, it is imperative that benchmarks be clearly identified ahead of time, as a curriculum is being conceived and before learning materials get developed and used, so that the evaluation process does not get biased by what goes on on the ground, especially in the classroom. For *evidence* (or counterevidence) to be reliable, it must be *criterial*, i.e., established against *objective criteria* that are not tainted by the special interests and idiosyncrasies of any of the stakeholders, especially end-of-the-line material users (teachers, students, administrators).

Two major steps need to be taken in normative evaluation before criterial evidence can be sought for STE materials:

1. Establish a detailed *taxonomy* of conceptions, processes and worldviews that would make up the profile that students are anticipated to develop following the completion of a curriculum.
2. Set *criteria* that establish whether individual students have actually developed each element of the anticipated profile, and to what extent they have done so.

In parallel, similar taxonomy and criteria need to be set for teachers' practice in the classroom in order to subsequently assess whether teachers have actually done everything that is necessary for students to develop the anticipated profile. Curriculum and material developers need then to specify which items in the taxonomy each of the materials is supposed to foster and to what level. Furthermore, and in order for the evaluation process to focus on realistic aspects and result in meaningful outcomes, it is better that taxonomy and criteria be *graded*, i.e., that they be about ordered levels of student competency. At least two levels or *thresholds* need to be identified in this respect:

1. *Basic threshold*. This is the most fundamental level. It corresponds to the *minimum standards* of meaningful understanding that *any* student should meet, irrespective of the initial competence level and interests of the student.
2. *Mastery* or *critical threshold*. This is the highest threshold that students need to cross in order to master all fundamental conceptions and processes in a given S&T course. In an ideal and truly equitable situation, all students willing to invest necessary efforts should be capable of reaching this threshold. In traditional classroom settings, critical threshold usually corresponds to the level of understanding of supposed A-students.

Students' failure to reach the basic threshold would be a doom indicator for the evaluated materials and curriculum (provided that materials have originally been considered as suitable, and that teachers have actually used them as anticipated). In S&T courses, this threshold corresponds to conceptions and processes that are most elementary in the scientific theory that is the object of a given course, and that serve in the construction of the most elementary models of the theory. For example, in classical (Newtonian) mechanics, and according to modeling theory, basic threshold corresponds to concepts and principles needed for the construction of the free particle model and the uniformly accelerated particle model. The first model corresponds to physical objects at rest or in linear uniform translation under no net force in a given inertial reference system. The second model corresponds to physical objects that are in linear or parabolic translation under a net constant force in such a reference system. Critical threshold in classical mechanics corresponds to conceptions and processes that are indispensable for the construction and deployment of the two models just mentioned, as well as of at least the model of a particle in uniform circular motion and the harmonic oscillator model.

Answers to all three questions in the A-triad (§ 1) are usually sought by internal and external panels of experts. *Internal panels* consist of people who have been directly involved with material development and implementation, teachers and publishers included. *External panels* consist of outside people who have not been involved in the process in any way. It is preferable that internal and external panels set their evaluation strategies separately and independently of one another before they get together to coordinate their efforts. It is even preferable, if the situation allows it, that each panel conducts its own evaluation on the ground independently of the other panel. In the end, the two panels would get together to compare and analyze their outcomes. In the entire process, both types of panels need to assess things in terms of the same taxonomy and criteria originally set by curriculum developers. However, by the end of the evaluation process, they may recommend some changes in these respects.

Evaluation panels usually rely entirely on their own judgment, based on qualitative evaluation, and without recourse to quantitative indicators, in order to answer the first two questions of the triad, the ones about the nature of available materials and anticipation. As such, panel judgment may be reliable and sufficient with respect to these two questions, provided that evaluation is made in terms of pre-established taxonomy and criteria. However, qualitative evaluation provides only circumstantial (and often anecdotal) evidence when it comes to achievement as addressed in the third question of the triad. Reliable evidence (or counterevidence) about achievement can only be sought through the use of appropriate instruments that measure teacher practice and student profile evolution as quantitatively as possible, and that provide outcomes that can be subject to all sorts of inferential statistics, including correlation between teacher practice and student achievement. A battery of instruments described below is often needed in this respect so that various aspects be adequately assessed of student profile as well as of teacher practice. Outcomes would not only answer questions about teacher practice and student achievement (independently and in relation to one another), but they would also serve to reconsider various aspects of the curriculum, including learning materials, and thus refine (or even overhaul) the curriculum accordingly and revisit the A-triad of questions with refined taxonomy and criteria. In this respect, the triad of questions constitutes a *closed cycle* that evaluators and other stakeholders need to go through periodically, over and over again, as long as the curriculum under evaluation is in effect (Figure 2).

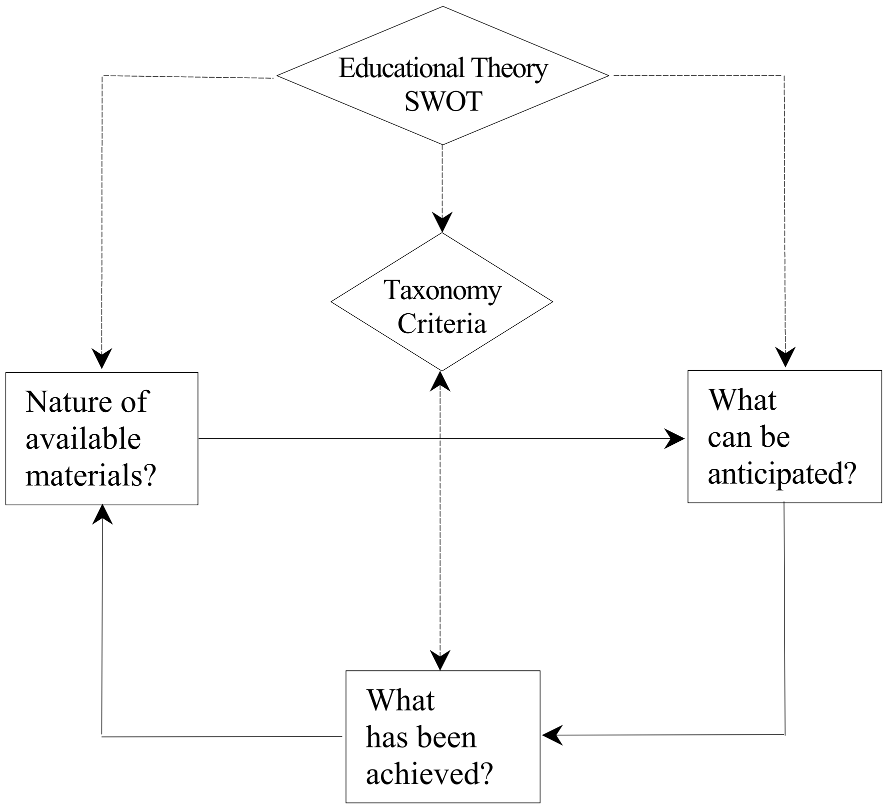

Figure 2: The triad of evaluation questions constitutes a closed cycle that evaluators keep going through periodically as long as the curriculum under study is in effect. Educational theory and the SWOT rule are accounted for as materials are being developed and in setting all sorts of expectations about teachers and students. Taxonomy and criteria that govern the process may be refined as a consequence of the evaluation exercise.

The battery of evaluation instruments includes:

1. For assessing teacher practice and viewpoints:

- *Surveys*, including paper-and-pencil surveys and interviews with teachers using the materials being evaluated.
- *Observation grids* prepared in accordance with the taxonomy and criteria set for teacher practice. Evaluators visit classes armed with the grids in which they note their observations about what takes place in the classroom. Class visits may sometimes be supplemented or even substituted with video-tapes of class sessions.
- *Teaching logs*, whereby teachers record regularly and systematically their own activities and comments about their own practice.
- *Listservs* maintained for groups of teachers sharing similar interests, and whereby peer discussions take place about what works and what not in the classroom. Such a tool has many potentials and capabilities including immediate feedback by peers and support staff (including material developers), which helps concerned teachers reconsider matters of interest in a timely manner.

2. For assessing student profiles:

- *Conceptual surveys*, including paper-and-pencil surveys and interviews with concerned students. Examples of such surveys are discussed in the following section.
- *Observation grids* similar to the ones used for teacher practice and covering all sorts of student activities in the classroom. Different grids may be used for different activities and student products.
- *Student logs*, whereby teachers record regularly and systematically what every student and/or groups of students achieve of the respective taxonomy in every teaching unit.
- *Formal course assessment*, including notes and grades on homeworks, exams and all sorts of assignment.

Conceptual surveys administered to students are perhaps the most reliable among all these instruments, especially when these surveys are developed and standardized by outside evaluators, and when they allow quantitative assessment of student achievement. Evaluators though need to ensure that such instruments are developed in accordance with a taxonomy that matches the one they are concerned with, and that instruments' criteria match the targeted benchmarks. *Standardized instruments* offer a special advantage. Corresponding criteria are set not only intrinsically in terms of the scientific theory the understanding of which is under investigation, but, more importantly, following inferential analysis of actual outcomes obtained elsewhere. This would facilitate grading the criteria as mentioned above (say, between basic and critical thresholds) and setting thresholds at objective and realistic levels.

4. Normative evaluation of modeling materials

Normative evaluation as outlined above was implemented in the development of modeling theory, and particularly in the development of learning materials within the framework of the theory in question. Major aspects of the process are described in this section so as to illustrate how one can go about conducting such an evaluation in the direction of continuous curriculum refinement.

Development of modeling theory practically began in 1981 as part of this author's doctoral dissertation at Arizona State University (Halloun, 1984). The theory was first developed in the context of introductory college physics courses, then it was gradually deployed into secondary school physics courses. Now that the theory has been duly corroborated for physics curricula, efforts are under way to deploy it into science, technology and mathematics curricula of all educational levels. The modeling curriculum in physics was first implemented in limited settings in USA and Lebanon. Now, it has become a national curriculum in USA where it was rated an *exemplary* curriculum by the Department of Education for the last couple of years, and the only curriculum of the kind last year. The modeling curriculum is currently being implemented to various degrees in about twenty countries around the world.

In its most efficient form, the modeling physics curriculum is implemented at the secondary school and college levels in small classes of about thirty students each (scale-up projects are currently in place in some universities). Students always work collaboratively in groups of three students each. The typical classroom includes eight to ten workstations. Each group is assigned a particular workstation where most, if not all, learning activities take place. A workstation is equipped with necessary laboratory equipment, preferably of MBL type (Microcomputer-Based-Laboratory). In the absence of appropriate modeling software that has not been developed yet, MBL can enhance learning logistics (reduce efforts and time invested by teachers and students), but not necessary the quality of conceptual understanding. Thus, in the current status of things, traditional laboratory equipment and MBL may bring students up to the same quality of learning (i.e., the same level of paradigmatic evolution), but MBL does so more efficiently than the other.

Learning activities are designed with student participation following specific guidelines provided by the teacher. Manuals of modeling activities are also available for secondary school courses*. A manual is divided into units, each devoted to the construction and deployment of a specific model (Halloun, 2001, 2003). Students actively construct a given model through appropriate experiments conducted on their workstations and/or outside the classroom in real world settings. Then they deploy it in various activities including traditional paper-and-pencil problems.

Formal evaluation of the modeling curriculum takes place in accordance with the norms and guidelines described in the previous section. Individual teachers implementing the curriculum are asked to do the same in order to assess the effectiveness of their own instruction and participate in action-research leading to continuously more meaningful and equitable learning experience across the board. To these ends, teachers and evaluators make systematic use of batteries of instruments similar to the ones described above (§ 3). However, standardized instruments of conceptual understanding are the most relied upon for evaluating the modeling curriculum in terms of student achievement, and particularly for assessing whether students have achieved to an acceptable level the paradigmatic evolution promoted by the curriculum. Let us illustrate the process in the case of classical (Newtonian) mechanics.

Meaningful paradigmatic evolution in Newtonian mechanics is set in terms of a set of basic models. The set includes four to five models, depending on the course level. These basic models are outlined in Figure 3. An anticipated student profile includes conceptions

---

* To inquire about modeling materials, please visit: http://modeling.asu.edu/

and processes required for the construction and deployment of these models, as well as generic worldviews about the nature of physics and its relevance in everyday life. A number of standardized instruments are used in order to assess to what extent a student has developed the anticipated profile. Among these instruments are the following:

- The Mechanics Diagnostic Test or MDT (Halloun, 1986, 2001a; Halloun & Hestenes, 1985) or its advanced version, the Force Concept Inventory or FCI (Halloun & Hestenes, 1995; Hestenes, Wells & Swackhamer, 1992). These are paper-and-pencil instruments designed to assess students' understanding of fundamental *conceptions* required for the construction and deployment of the set of basic models in Figure 3. Conceptions include the concepts of force, position, velocity and acceleration and their relations to one another as established by state laws and Newton's laws of dynamics. An item in these instruments consists of a multiple-choice question about a particular situation in everyday life. Answering a given question requires only qualitative understanding of assessed conceptions and no quantitative manipulation of any sort of mathematical representations.
- The Mechanics Baseline Test or MBT (Hestenes & Wells, 1992). This is a multiple choice, paper-and-pencil instrument that assesses student understanding of the fundamental *processes* required for the construction and deployment of the models in questions. More specifically, MDT targets required mathematical representations and operations.

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 3: Basic particle models in Newtonian theory of classical mechanics, with an outline of the translational pattern that each model represents in inertial reference systems (Halloun, 2003).

- The Views About Science Survey or VASS (Halloun, 2001a&b; Halloun & Hestenes, 1998). This is also a paper-and-pencil instrument. Its items are formulated following Halloun's *Contrasting Alternatives* rating scale. VASS is intended to assess student views about the structure and validity of scientific theory, scientific methodology, reflective thinking (learning styles), learnability and personal relevance of physics (or science, depending on the form of VASS used).

Each of these instruments has been administered to thousands of secondary school and university students around the world, students enrolled in physics courses (and other science courses for VASS) that are taught following a variety of instructional strategies. Strategies extend from conventional instruction of lecture and demonstration to all sorts of student-centered courses that follow modern educational theory. All three instruments are duly validated and standardized following graded criteria as discussed above.

Student profile anticipated following a given course of classical mechanics is graded in taxonomy and criteria between two fundamental levels corresponding to the basic and the critical thresholds. As discussed in the previous section, and according to modeling theory, the *basic threshold* of paradigmatic evolution corresponds to conceptions and processes required for the construction of the first two models in Figure 3 (free particle and uniformly accelerated particle). The *critical threshold* corresponds to the first four models in this figure. To each of these thresholds corresponds a particular set of items and a particular score on the FCI, MDT, and MBT. Similar thresholds and characteristic scores are associated with student worldviews as assessed in VASS.

Practice of teachers following modeling instruction is also assessed with an appropriate battery of instruments. Like in the case of students, the most relied upon instrument is a Likert-type, paper-and-pencil survey developed by the modeling curriculum evaluators. This teacher survey, administered at the end of a semester or year, is intended to ascertain to what extent individual teachers implement particular aspects of modeling instruction in their classes. These aspects pertain to modeling tenets, model-centered course, modeling tools and associated rules, modeling/inquiry schemes, student-centered interactive engagement, scientific discourse, fostered learning styles, and action-research. A quantitative score is associated with a teacher answer on a particular item, and a sum of scores is calculated for every teacher. Item lumping and score grading on the teacher survey are done so as to reflect overall teacher practice in two major respects. One pertains to course content and structure, with emphasis on models and related processes. The other pertains to mediated learning and various aspects of learning cycles.

Development of the modeling curriculum began in the eighties, in a limited number of college then high school mechanics classes, in USA, and then in Lebanon. At that time, instructors made use of available textbooks and laboratory equipment. Textbooks were supplemented with handouts making explicit the model-based structure of Newtonian theory, and traditional laboratory equipments were set up to suit the needs of modeling instruction to the extent that is possible. In the mid nineties, development of the modeling curriculum went on in USA on a larger scale. A broad band of high school and college teachers got gradually involved in the process, and high school and introductory college physics courses got gradually all covered in the process. A few years later, the modeling curriculum was widely disseminated within and outside the U.S. In the following section, we report on evaluation data pertaining to selected samples who participated in the development and dissemination of the modeling curriculum in classical mechanics. Data are presented and discussed so as to illustrate with actual examples how the instruments presented above were used in normative evaluation of the modeling curriculum.

5. Evaluation for regulation in action

Use of all student instruments and interpretation of respective data are governed by identical or similar norms and criteria. That is why we concentrate in what follows on only one student instrument, the FCI, and the way it has been used in the evaluation and development of the physics modeling curriculum at the high school level. Examples of FCI data are provided and interpreted in relation to data pertaining to teacher practice at certain stages of curriculum development. The object is to illustrate how data can be collected, displayed, analyzed and interpreted so as to reveal as best as possible weaknesses and strengths in a new curriculum, and thus refine the curriculum accordingly.

Reported data pertain to two samples of secondary school classes. The first sample consists of classes of about fifty high school teachers who participated in the early broad dissemination of the modeling curriculum, and who were involved in the development of the first modeling manuals (and choice of appropriate laboratory setups). The second sample consists of classes of ten secondary school teachers who joined in later in the process, at a time where the modeling curriculum began to take its current shape.

At an early stage of the modeling materials development, about fifty in-service high school teachers were trained at length to implement modeling theory in teaching classical mechanics at the high school level. During their first year of training, these teachers participated in the development and field testing of a learning manual for classical mechanics. The manual includes guidelines for MBL and other activities devoted for the construction and deployment of the five models in Figure 3. Like all others participating in the development and implementation of the modeling curriculum, these teachers administered to their own students all three student instruments as pretest and posttest, the year preceding their training, and regularly afterwards, during and after training. In parallel, teachers themselves were administered the teacher survey at the end of every academic year in order to monitor changes in their teaching practice throughout training and some time afterwards.

Figure 4 shows FCI data for students of these teachers collected at the beginning (pretest) and end (posttest) of the first year of training. Data are displayed in the form of three clouds or clusters (mostly to preserve confidentiality). The boundaries of each cluster actually consist of a best-fit envelope that can be wrapped around a particular cluster of data points. Each point originally represents the posttest average plotted against the corresponding pretest average of FCI scores of students taking classical mechanics with a particular teacher. The three clusters or groups are distinguished with the extent to which teachers implemented modeling instruction in their classes, as revealed with answers on an early version of the teacher survey. Group 1 refers to teachers who claimed to be most systematic in this respect (20% of participants), and group 3 to those (46%) who claimed to be least systematic (virtually sticking to conventional instruction). Group 2 (34% of participants) stands in the middle between the two extreme groups. Needless to say that by the first year of trial, none of participating teachers had already mastered the modeling approach, and modeling materials were just being piloted and were still far from their current state.

Five critical lines are shown in Figure 4, three horizontal and two inclined. The bottom horizontal line (posttest average of 50%) corresponds to the *basic threshold* on the FCI, i.e., to the minimum level of understanding required for students to be able to construct the first two basic models of Figure 3. The top line (80%) corresponds to the *critical threshold* or mastery level required for meaningful understanding of all five models. The line between the two (60%) corresponds to what we have called *Newtonian threshold*. This is the *minimum* level students need to attain in order to *begin* understanding *all generic* aspects of the Newtonian theory within the context of the models of Figure 3. The three thresholds have been set after long years of FCI administration and analysis of data pertaining to thousands of students around the world. The two slanting lines correspond to average pretest-posttest gains (g) on the FCI for two groups of students. The gain expression is given by:

$$g = \frac{\text{Posttest\%} - \text{Pretest\%}}{100 - \text{Pretest\%}}$$

The g = .30 line corresponds to the national threshold of interactive engagement (IE) courses in USA. These are high school physics courses where teams of students are actively engaged in reflective inquiry in the classroom. Hake (1998) showed that, average gains of IE classes fall all *above* this threshold, while average gains of classes following conventional instruction of lecture and demonstration fall all *below* the same threshold. He also revealed that the overall average gain for IE courses is about .48 (not shown in Figure 4), and that g = .69 is the maximum average gain ever reported in USA for IE courses.

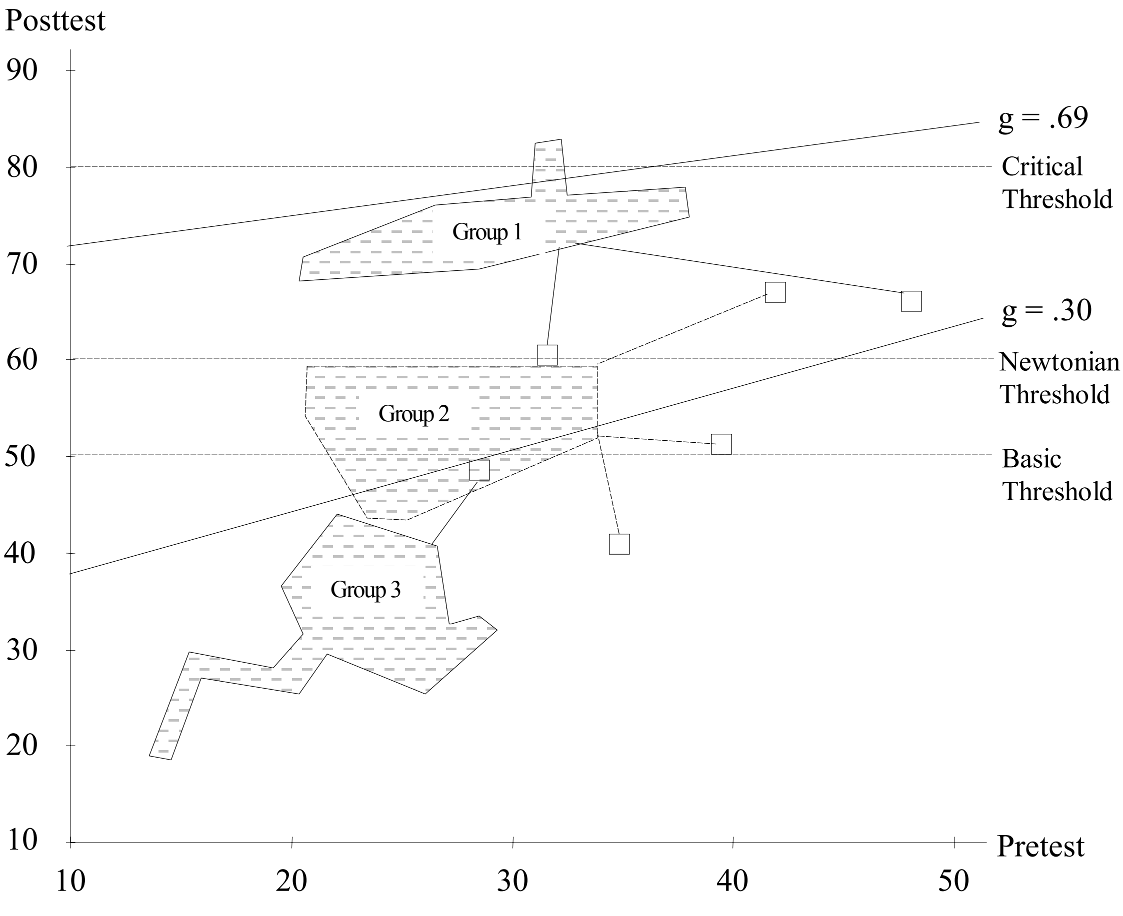

Figure 4: FCI pretest-posttest data (in score percentage) for high school students of three groups of teachers participating in the early stages of modeling materials development. The three groups are ranked from most (Group1) to least (Group 3) systematic in their implementation of various aspects of the modeling curriculum. Individual outliers in each group are represented, like in boxplots, with squares connected to corresponding clusters.

The overall distribution of pretest means in Figure 4 is in line with the international distribution of FCI pretest scores. Pretest class means fall all below the basic threshold (50%) as one would expect. Except for one case in Group 1, pretest means are even all below 40%. Classes in Group 3 appear to be originally at a slight disadvantage with respect to the other two groups; their pretest means are all equal to or below 30%, and only in this group there are two classes whose pretest average is below 20%. Analysis of variance on pretest means revealed no significant differences among the three groups (at the .05 level).

Significant differences are though detected among the three groups on FCI posttest data. Among these differences we note the following:

1. For the same pretest means, classes in Group 1 achieve the highest gains on the posttest, followed by those in group 2 then Group 3. For example, among classes who averaged around 27% on the pretest: (a) all three classes in Group 1 averaged 70% or more on the posttest, (b) the five classes in Group 2 averaged between 45% and 60%, and (c) all three classes in Group 3 averaged below 40%.
2. 80% of classes in Group 1 averaged 70% or more on the posttest (i.e., closer to the critical threshold than the Newtonian threshold), and had pretest-posttest gains averaging well above the IE mean of .48 and close to the .69 ceiling of IE courses. One class in this group reached a posttest average just above the 80% mark of the critical threshold, and had an average gain of .72. Average posttest score and gain in this class were at the time the highest ever reported on the FCI in a high school. Overall, and except for one class in Group 2, posttest means (and gain g for the upper 80% of classes) in Group 1 had no match in any of the other two groups.
3. About 80% of classes in Group 2 averaged between 50% (basic threshold) and 60% (Newtonian threshold) on the posttest. One class in this group had a posttest average of about 68%, and all others between 40% and 50%. Four classes in this group had an average pretest-posttest gain just below the IE threshold of .30, while all others (83% of Group 2) had gains averaging above this threshold.
4. All classes in Group 3 had a posttest average below the basic threshold of 50% and a gain below the IE threshold of .30.

As mentioned previously, participating teachers were in their first year of training on the modeling approach, and learning materials (as well as the teacher survey) were at the early stages of their development. This is what makes the chosen sample so interesting. It comes at the beginning of the large-scale experimental phase of the modeling curriculum, and it includes a wide array of teacher practice in the classroom extending from conventional instruction (including the use of significantly more conventional learning materials than modeling materials) to IE instruction that begins to come close to modeling instruction but that is not quite there yet. As such, the sample is quite heterogeneous in terms of what its members implemented of the modeling curriculum. It allows detailed analysis of what works and what not in this curriculum, with as close as possible to a good control of variables. A specific curriculum dimension can be isolated (to a certain extent), and its impact on student achievement can be conveniently analyzed almost independently of other dimensions.

Analysis of results displayed in Figure 4, and especially of those listed in the four points above reveals the following:

1. By comparison to FCI data pertaining to the students of the same teachers a year before training (not reported in this paper), posttest means and gains are higher for classes of all teachers following their first year of training, even for those in Group 3

who, in all practical aspects, were still following conventional instruction. This already indicates that by and large the modeling curriculum seems to be on the right track.

2. Posttest and gain results improve gradually from Group 3 to Group 1. This indicates that the more comprehensively teachers follow the modeling approach, i.e., the more dimensions of the modeling curriculum they implement in their classes, the better student achievement.
3. Preliminary analysis of individual teachers' practice revealed that teachers in Group 2 were more systematic in changing their own methodology in the classroom than in changing the content and structure of their courses. They focused on learning cycles, but in the direction of completing course content as available more in conventional textbooks than in the modeling manual that they were supposed to pilot. The same was true, though to a slightly lesser extent, for the two teachers in Group 1 who had the lowest FCI gains.
4. Analysis of Group 1 practice revealed that, and except for the two teachers just mentioned, teachers in this group were trying to balance between instructional methodology and restructuring course content following the modeling manual.
5. Analysis of differences in teacher's practice with respect to content structure revealed that the more systematic teachers were in helping students develop models of well-defined structure, the better student achievement. Specific modeling tools and rules (including MBL related) had proven to be more effective (and sometimes more realistic) than others in this direction.
6. Analysis of student answers on individual items on the FCI revealed a certain hierarchy in conceptual complexity and cognitive requirements across the respective taxonomy. It showed that student understanding of Newtonian theory develops gradually and almost in a specific order through certain elements in the taxonomy, and this irrespective of teacher practice. Basic threshold is crossed only after students develop a specific set of conceptions in Newtonian theory. Newtonian and critical thresholds are reached only after students develop increasingly more comprehensive sets of the theory's conceptions. It was thus revealed that certain clusters of items in the taxonomy are critical for reaching or passing a certain threshold, and that students need to develop conceptions in a certain order within each cluster so that they proceed in their paradigmatic evolution successfully and meaningfully. This, again, was true irrespective of teacher practice. However, the practice of teachers in Group 1 fostered the evolution in question more systematically within and across conceptual clusters than the practice of teachers in Group 2. Teachers in Group 3 were not capable of promoting the evolution to any acceptable level or in any coherent way.
7. Analysis of teachers' implementation of learning cycles revealed that certain phases of a cycle were more critical than others, and that certain refinements were needed at this level. It also revealed the necessity to direct and manage teacher and peer feedback, as well as student discourse, more in certain directions than others.

FCI outcomes were consistent with those obtained with other evaluation instruments. Based on the evaluation exercise conducted with all instruments, the modeling curriculum was refined in many respects, including course content and manual, and MBL setups. The refinement process lasted for a number of years (and is still in process, though it is now more in the direction of exploring new tools than in any other respect). In later stages, teachers from all around the world began implementing the modeling curriculum in secondary school and college physics courses. Figure 5 shows evaluation outcomes for a small group of secondary school classes that joined in later in the process.

Figure 5 pertains to a group of ten secondary school teachers who decided at a certain point to pilot the modeling curriculum in their classical mechanics course. Like all other participants, these teachers administered a number of instruments, including FCI, to their students as pretest and posttest during the first year of their pilot program. Teachers themselves filled out the Modeling Practice Survey (MPS) at the end of the same year. MPS is an upgrade of the teacher practice survey used in Figure 4. It includes 100 Likert questions asking teachers about the extent to which they implement various aspects of the modeling curriculum. A perfect score on the MPS is of 200; it indicates that a teacher claims to implement the modeling curriculum to the letter. A point in Figure 5 shows the score of a particular teacher on the MPS against the average pretest-posttest gain g on the FCI of the students of this particular teacher. The overall distribution of points is quite similar to results obtained elsewhere and at various stages of the development of the modeling curriculum. It indicates a consistent trend in the outcomes: a significant correlation between teacher practice and student achievement, achievement that gets significantly better as teachers account more and more systematically to various dimensions of the modeling curriculum.

Teachers in Figure 5 fall actually in three distinctive groups. The first group consists of the three teachers whose MPS score is equal to or above 140. We will refer to this group as Group 1. The second group consists of the five teachers whose MPS score is above 80 but below 120 (Group 2). The third group consists of the two teachers whose MPS score is

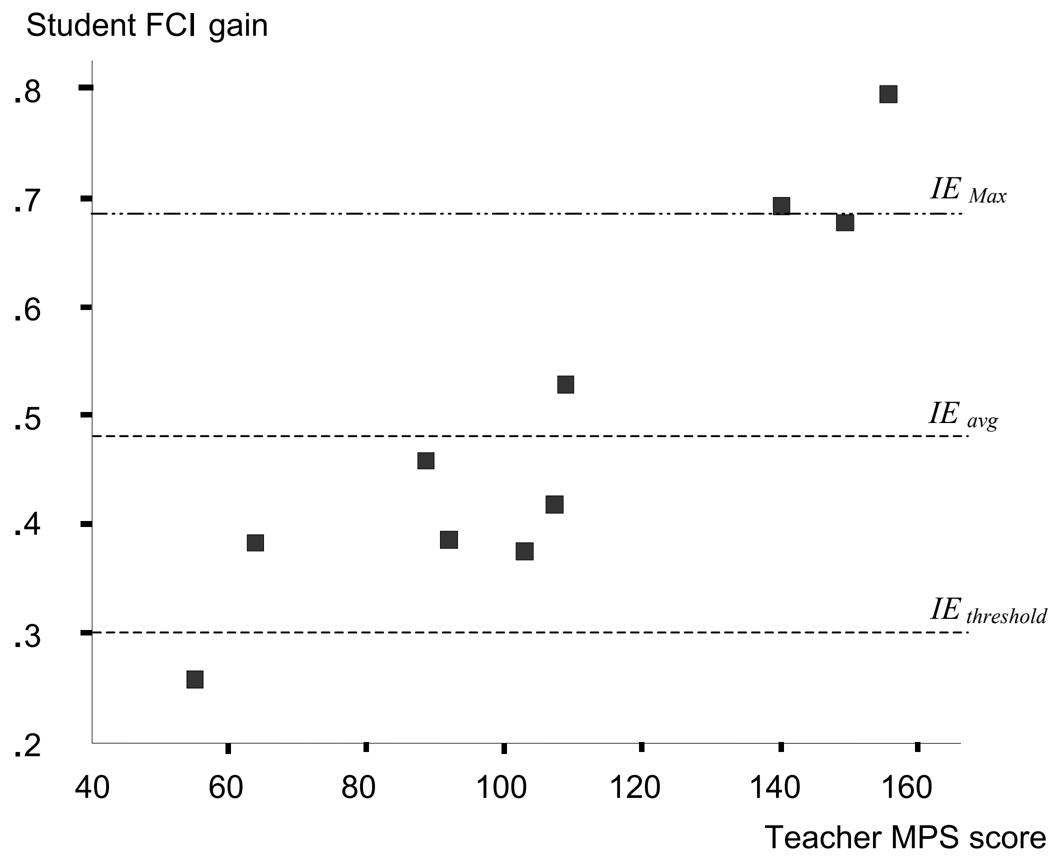

Figure 5: Student FCI gain versus teacher practice as assessed with the Modeling Practice Survey (MPS).

below 80 (Group 3). Close comparison of groups of teachers implicated in Figures 4 and 5 reveals noticeable similarities between groups bearing the same number both in terms of teacher practice and student achievement. Group 1 in both figures consists of teachers who are most systematic in implementing the modeling curriculum. Group 3 in both places consists of teachers who are still favoring conventional instruction. Group 2 is somewhere in the middle between the two, and respective teachers in both figures concentrate more on instructional methodology than on model-centered content. The trend in both figures is typical of all outcomes obtained in the evaluation of the modeling curriculum. It has the following major characteristics:

1. Best results with the modeling curriculum are obtained when all dimensions of the curriculum are accounted for systematically and comprehensively. Selective implementation of some components but not others from each of the two broad dimensions (instructional methodology and course content) may enhance student achievement enough to cross the basic threshold or even the Newtonian threshold, but not enough to reach the critical threshold.
2. When teachers cover comprehensively one broad dimension rather than the other, students, like before, are enabled to cross the basic or even Newtonian threshold but not the critical threshold. This is reflected in Figure 5 where the teacher in Group 3 with the higher MPS score concentrated, though erratically, more on restructuring content than on following the learning cycle, and where all teachers in Group 2 had the balance tipped in the opposite direction.
3. Between the two broad dimensions, the one pertaining to restructuring course content around basic models and emphasizing modeling processes has a significantly higher impact on student achievement than the dimension pertaining to instructional methodology. However, only when the two are conveniently accounted for, students can cross the critical threshold of a given course (and the IE ceiling).
4. Learning materials, other than course manual, and including MBL, have little impact on student achievement if not used in the manner advocated by modeling theory to develop a scientific theory that is the object of a science (or technology) course around basic models and to foster systematic development of modeling schemes and conceptual “tools of the trade”.

6. Toward meaningful and equitable learning of science and technology

Our long experience with the modeling curriculum shows that normative evaluation as described in this paper is a necessary condition: (a) for science and technology education to help students evolve into the scientific realm in meaningful ways and become adequately skilled for S&T practice, (b) for STE to become an equitable enterprise whereby all students willing to invest necessary efforts, and capable of doing so, are enabled to cross certain thresholds, and (c) for curriculum reform to be sustainable. More specifically, evaluation of learning materials must be, on the one hand, an integral part of curriculum evaluation. It must be conducted according to pre-established taxonomy and criteria set primarily in terms of the educational theory within the framework of which curriculum was formulated. On the other hand, normative evaluation should not be an end by itself. It should not be carried out for the mere sake of ascertaining the intrinsic value of materials and curriculum, but for the purpose of continuous regulation of various aspects of the curriculum. Concerned teachers must take an active role as action-researchers in this direction.

S&T learning materials cannot be developed and subsequently evaluated independently of other curriculum components. These include educational tenets, program of instruction (content and structure), learning and teaching strategies, and assessment modes and means. Effectiveness of learning materials is not an intrinsic factor that depends only on the nature of such materials. It especially depends on the way materials are being used in the learning process. Subsequently, it is indispensable that learning materials be evaluated in terms of their effectiveness on student achievement, and this in relation to instructional methodology and student learning styles inside and outside the classroom. This is how we proceeded with the modeling curriculum, as described in part in the previous section.

For the evaluation exercise to be effective, especially in the direction of continuous curriculum regulation, concerned teachers must get involved in systematic action research intended to get their practice increasingly more equitable and student understanding (and skills) more and more meaningful. Unless teachers act as such after they get adequately trained and motivated, curriculum and materials would fail to meet their ends no matter how well they may be originally articulated. This brings us to a critical issue, teacher training and motivation.

Everytime some novelty is introduced in a curriculum, be it materials or else, teachers must be adequately trained so they know how to appropriately adjust their practice. This is true for both pre-service and in-service teachers. It is especially true for the latter, no matter how long an experience they might have. In fact, teaching experience, especially when of conventional type, may be a detriment in the process. Teachers who get used to specific method and materials often develop some inertia and even close-mindedness that prevent them from being capable or even open to adjust to certain novelties on their own. This was in fact the case with Groups 3 in Figures 4 and 5. Our experience suggests that the shift of in-service teachers from conventional to IE or modeling instruction requires long training and particular incentives. Contrary to what some policymakers and educational administrators around the world, and especially the Arab world, actually believe and push for on the ground, innovations in S&T curricula cannot be shoved to in-service teachers in a few days or even weeks of training (if any!).

Such innovations, especially when in line with modern educational theory like modeling theory, require that teachers spend at least one full academic year of special training, followed by years of monitoring and support on the ground. In parallel, teachers need to be offered particular incentives, including but not limited to reduction of contact hours at school and special fringe benefits. Support must be provided, almost on a daily basis in the early implementation of new materials, locally and/or remotely through appropriate support systems and programs. These systems and programs are typically run by educational experts, including master teachers (i.e., teachers who have developed some experience with promoted materials), university professors, material developers and publishers, who are capable of providing timely and effective feedback to all sorts of inquiry put forth by teachers and administrators on the ground. In this respect, particular partnerships between local universities and schools are especially helpful. The object of such partnerships goes beyond formal training of in-service teachers, which must be reserved to universities and teacher training institutions. It extends to all nuts and bolts of a curriculum, and especially to action-research that is better carried out by teams of school teachers and university professors. Teachers bring in their classroom experience and daily concerns, and professors contribute their expertise in educational theory, scientific theory, modern technology, statistical analysis and various aspects of educational research.

All in all, normative evaluation for curriculum regulation is a tedious process that requires understanding and commitment on behalf of many actors, including, but not restricted to, policymakers, administrators, curriculum developers, publishers, teachers, and, of course, students. Long-term training and support systems and programs need to be instituted in partnership with local universities and teacher training institutions. Learning materials need to be developed and evaluated in concert with concerned teachers and primary actors in support systems, and this within the framework of an appropriate educational theory that helps setting ahead of time explicit criteria of success and failure, criteria that account to certain extents to the SWOT rule on the ground. Failure to account to any of the aforementioned would condemn new materials and curriculum no matter how well they may be originally formulated and intended.

References

Bibeau, J. R., Halloun, I., May, J. & Reddy, V. (2002). *Report on the Evaluation of UNESCO's Science and Technology Education (STE) Programme*. Internal report presented to the STE section at UNESCO, Paris.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of physics*, *66* (1), 64-74.

Halloun, I. (2003). *Modeling Theory in Science Education.* Book under review by publisher.

Halloun, I. (2001a). *Apprentissage par Modélisation : La Physique Intelligible.* Beirut, Lebanon: Phoenix series / Librairie du Liban Publishers.

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

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

Halloun, I. (1998a). Schematic concepts for schematic models of the real world. *Science Education, 82* (2), 239-263.

Halloun, I. (1998b). Interactive model-based education: An alternative to outcomes-based education in physics. *South African Journal of Science, 94,* 313-318.

Halloun, I. (1996). Schematic modeling for meaningful learning of physics. *Journal of Research in Science Teaching, 33* (9), 1019-1041.

Halloun, I. (1994). Teaching model construction for solving physics problems. *Recherches Pédagogiques, 19,* 5-17.

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

Halloun, I. (1984). The Use of Models in Teaching Newtonian Mechanics. Unpublished PhD dissertation. Tempe, AZ: Arizona State University.

Halloun, I., & Hestenes, D. (1998). Interpreting VASS dimensions and profiles. *Science & Education, 7* (6), 553-577.

Halloun, I., & Hestenes, D. (1995). Interpreting the Force Concept Inventory. *The Physics Teacher, 33* (8), 502-506.

Halloun, I., & Hestenes, D. (1987). Modeling Instruction in Mechanics. *American Journal of Physics*, *55* (5), 455-462.

Halloun, I., & Hestenes, D. (1985). The initial knowledge state of college physics Students. *American Journal of Physics, 53* (11), 1043-1055.

Hestenes, D., & Wells, M. (1992). A Mechanics Baseline Test. *The Physics Teacher, 30* (3), 159-166.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30* (3), 141-158.

UNESCO. (1994). *Education for All. Summit of nine High-population Countries*. Paris: UNESCO.

UNESCO. (1993). *Project 2000+. Forum International sur la Culture Scientifique et Technologique pour Tous*. Paris: UNESCO.

Wells, M., Hestenes, D., & Swackhamer, G. (1995). A Modeling Method for High School Physics Instruction. *American Journal of Physics, 64* (7), 606-619.