

On the evaluation of curriculum reforms

STEFAN THOMAS HOPMANN

The paper considers the current international trend towards standards-based evaluation in a historical and comparative perspective. Based on a systematization of evaluation perspectives and tools, two basic patterns of curriculum control are discussed: process evaluation, and product evaluation. Whereas the first type has dominated the Continental European context for two centuries or more, and the latter type the English-speaking world for most of the 20th century, the last two decades have brought about a situation in which both types are more and more intertwined. The paper discusses the historical pre-conditions, empirical findings, and possible prospects around this development.

Curriculum reforms almost always create great expectations: raising the stakes, getting ‘better’ performance, and the like. Those reforming a curriculum also hope, of course, that a change of goals, contents, and of the ways and means will enhance teaching somehow and in some way. It is no surprise that they want to know whether or not the new curriculum has had the impact ascribed to it; in other words, they want an *evaluation* of its effects—or do they?

Historically this has *not* been the case, at least in most European countries. It is only recently, within the last two decades or so, that curriculum authorities have started to evaluate systematically what happens after a new curriculum is implemented, or how the old curriculum is doing. Before this development, newly-developed curricula came and went, without systematic testing and evaluation. At best, the authorities based their development of new curricula on more or less systematically collected hearsay and more or less educated guesses about what the old curriculum did, and what the new one might do. To understand why this has been the case and why it worked out quite nicely—and why this might be changing now—we have to go far back into the history of schooling.¹

The ways and means of changing and evaluating a curriculum are closely connected to the ways and means of organizing the social control of schooling as a whole, and to the role teachers and the school administrations are ascribed. In the following paper, I want to give a short account, based on

Stefan Thomas Hopmann is Professor of Education at the Norwegian University for Technology and Science (NTNU) at Trondheim and at the Agder University College, N-4604 Kristiansand, Norway; e-mail: stefan.t.hopmann@hia.no. His research deals with historical and comparative issues in curriculum, schooling, information technology, evaluation, etc. His recent publications include a comprehensive collection on *Teaching as Reflective Practice* (with Ian Westbury and Kurt Riquarts, Mahwah, NJ: Lawrence Erlbaum Associates, 2000) and a study (with Katrine Nesje) on curriculum implementation in Norway, *En lærende skole* (Oslo: Cappelen, 2002).

a rather simplified systematization of the tools of evaluation available within these systems, of the two dominant Western traditions of controlling, changing, and evaluating the curriculum of public school systems.² I will start out by giving rough sketches of (1) the basics of evaluation, and (2) institutionalization; I well know that much could be added and that reality is much more blurred than any systematization of main characteristics can reflect. However, drawing on my systematization, I will outline briefly how the two main patterns for evaluating the curriculum came into being. Finally, I will discuss how current developments may lead to an alignment of the two traditions identified. In all of this, I focus on the historically prevalent forms of evaluation *as a social practice*, leaving aside the issue of what else evaluation might be.³

Evaluation

Logically, there are three basic levels in any social practice which can be the focus of some kind of evaluation:

- the *people* involved,
- the *processes* engaged in, or
- the *products* emerging from these processes.

We look primarily at the personal level, the people involved, if we believe that personal characteristics, i.e. abilities, attitudes, etc., are decisive for the quality of what they do. For instance, we might look if the people involved get along with each other, or if there is a mismatch of personalities or skills of some kind—which might have negative consequences. This approach works, to some degree, when we evaluate a few people in a narrow setting largely shaped by those involved rather than by the environments in which they act. However, in larger social systems this approach doesn't lead far enough. If organizations depend overwhelmingly on the personal characteristics of their members (as in charismatic organizations, to use Weber's terms), they tend not to be stable outside the direct reach of these individuals. Any significant change in leadership or staff puts the organization at risk.

Larger social systems, like institutions or organizations, are supposed to function irrespective of the idiosyncrasies of their members. In a factory or an administration, we want to rely on things working out, even if the people involved are replaced by others. If we want to change or improve such an organization or institution, we will look at the *processes* which shape the interactions within that system. In a totally routinized environment, which deals almost exclusively with *well-defined* problems, i.e. if *a*, then *b*, as in a traditional factory, that might be sufficient. All we have to know is if the relations between the *a*'s and the *b*'s regulating the interaction are working properly.

In modern organizations, however, the routines, and virtually all other processes, depend on how the people fit into their places within the organization. For instance, something might happen which requires more than a mechanical response of the type 'if *a*, then *b*'. The modern answer to this problem is to professionalize the relation between people and processes,

i.e. to develop *professions* in which a specific education and a specific set of standards ensure that their members fit into the processes the profession takes care of. This move is especially the case where the process being taken care of is comparatively *ill-defined*, i.e. requiring some kind of situational adjustment and assessment. The decision to ask a professional to do something is based on the trust that a professional background enables him or her to provide such situational judgement and to render appropriate services based on that judgement.

A completely different approach would be to ignore the people and processes and only to focus on end-products, i.e. to look only at the *products* of a social practice. All of us undertake this kind of evaluation each and every day as we buy things based (at best) on a comparison of different products, but without the slightest idea of how the product was manufactured and who was involved in its making. In the modern era, the *market* and its forces are the most powerful tools of such product evaluation and—because of the success of this approach in creating the Western economy—it is often believed that this is the most important way of evaluating all kinds of problem-solving. However, those providing a product also have to look at the process of its production, at least in terms of trading, financial, organizational, personnel, etc., costs, and pay-offs. Otherwise, they run the risk of going bankrupt, or destroying other pre-requisites of their problem-solving, even though the product as such might be what they wanted. Looking at production may force them to also look at the processes and people involved. For well-defined products (like groceries or machines), it is easy to draw this line from the people and processes towards the qualities of the product. However, if the product itself is ill-defined, as is the case with many social services (like health care, education), those buying (i.e. evaluating) a product without knowing its origins cannot be sure that the product as they see it is indeed first and foremost a result of the production line (i.e. institution) offering with it.

Thus, we have a kind of continuous scale of evaluation levels: we might start at one end or the other, or in-between, and choose to focus more on one factor or another depending on what we believe about the social action which is to be evaluated:

- *people*
professionals;
- *processes*
production; and
- *products*.

Looking at the people, processes or products involved in a social action is difficult in cases where what is going on is doubtful, and depends on who is looking: a thief has a different view on the products of a theft than those who owned the product before; a lover may think differently about the outcomes of his love affair than the husband of his beloved. Accordingly, the efficiency of the police or of divorce courts might be evaluated rather differently.

Thus, a second necessary differentiation looks at *who* is doing the evaluation. In principle, there are three possibilities. For instance, looking at people, it may be that

- one is evaluating oneself;
- one is evaluated by an acquaintance or by someone who knows one's work; or
- one is evaluated by someone who is not familiar with what is going on.

The first case, *self-evaluation*, is something we do every day—and is not noticed unless we communicate the results of our self-evaluation. However, in social organizations such self-evaluation might be something which we are obliged to do, and there may be rules for how do it and report the results. Self-evaluation is only adequate if what we do doesn't affect others—if it is only our own business. Otherwise, there will be some kind of external evaluation, either informal or formal, depending on the social action in question.

The most basic form of external evaluation is the one done by someone close to us, who knows who we are and looks at what we are doing. In this case, we talk about *peer evaluation*. We have different kind of peers: those close to us, like colleagues who share in our doing, i.e. 'peers from within'; and we have those who come from the outside and take a fresh look at what we do. They are peers as long as they share something with us, either by being a part of the same process or by having comparable personal qualifications or duties.

If there is absolutely nothing which connects us in this sense to the evaluating person, we may talk about a truly *external evaluation*, e.g. those who buy the products we made or who evaluate us from a kind of 'alien' stance of no familiarity with the persons and processes involved.

These different perspectives can also be seen in a continuum where the different ways of evaluating can be completely detached from each other or intertwined:

- *self-evaluation*
peers from within;
- *peer evaluation*
peers from outside; and
- *external evaluation*.

Which level of evaluation is used, and who does it is different in different types of organization, and has changed throughout history. To simplify somewhat, we can say that the more complex a social action becomes, the more likely it is not to restrict evaluation to self-evaluation by the persons involved but to move towards more complex evaluation approaches, including evaluation of the processes and products, with peers or external observers undertaking the evaluation. For modern public *institutions*, there is good reason to talk about a movement from the personal level of self-evaluation towards external product-control as the predominant type of evaluation (see Figure 1).

Institutionalization

In the tradition of modern sociology, from Max Weber to Talcott Parsons and Niklas Luhmann, the process of *institutionalization* is one main force

<i>Level</i> Type	<i>Self- evaluation</i>	<i>Peer evaluation</i>	<i>External evaluation</i>
Persons			
Processes			
Products			

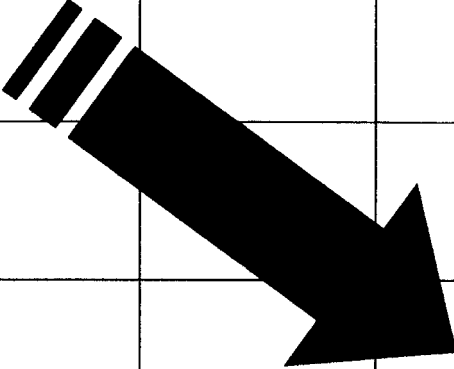


Figure 1. Types and levels of evaluation.

shaping the structures of modern Western societies.⁴ The theories may differ somewhat in what they see as the driving forces behind this process (functional or other types of differentiation, belief systems, etc.), but they agree on the fact that we organize most of our social life with the help of institutionalizations which give our activities more or less stable frames of reference.

Of course, institutionalizations differ widely in their size (from local to global), scope (from single purposes to all matters encompassing multiple societies), and life-span (from short-lived to lives of many centuries). However, for the purposes of this paper it is enough to suggest that institutionalization is the modern way of dealing with social problems *which do not solve themselves*. In this respect, we can draw a continuum indicating the degree to which social problems are defined, from clear-cut *well-defined* problems where we know precisely everything that there is to know about this problem (*if a, then b*), to *ill-defined* problems where we do not know what the problem is and what a possible solution might be.

What is striking in the history of modern societies are the differences in how attention to large-scale social problems has been institutionalized, depending on how well-defined the problem was, or is. As a rule—with exceptions—we can say that ill-defined social problems tend to end up as public matters, while well-defined social problems are mainly left to the people and their private institutionalizations, i.e. to market forces. Thus, the production of well-defined goods is left to private institutions (from farms to companies) if they are available, with public institutions moving in to ensure either the good's availability or distribution in times of shortages. In such

cases, the availability or distribution becomes ill-defined, and public institutions move in to 'relieve' private providers of the risks which occur if the availability or distribution of the problem-solution is unclear.

Whether a social problem, or the availability of a problem-solution, is seen as ill-defined or well-defined depends, of course, on the viewpoint of those governing a problem or its environment (e.g. society). What a problem is is not a feature embedded in some kind of 'nature' of the problem itself but in the *definitions* of the problem and its solutions which prevail. In short, it is a question of social power. Thus, in one society the distribution of certain goods may be seen as highly ill-defined, e.g. highly controversial, whereas the same distribution in another society may not create a need for public institutions, because the situation is evaluated as being well-defined, i.e. solved, within the predominant frame of problem-understanding.

These rather abstract definitions come to life if we think about specific social problems. In a society, where the education of the young is not seen as being problematic, there is no need for institutionalization. Institutionalization of education only occurs if there are doubts about the adequacy of 'natural growth' into adult life as a preparation for the future. And, if education is seen as problematic, the solution depends on whether or not the problem is seen as well-defined. If young people 'need' education, but those rearing them are seen as knowing what to do, there is no need for public involvement. It is left to the devices of families, etc. to set up the necessary frame for solving the problem of education—either within the family (or whatever the primary group for child-rearing might be) or within institutions, based on the agreement of those involved. The process leading to other types of problem-institutionalization emerges only if there is doubt about the adequacy of privately-provided procedures for child-rearing. Thus, it is not by chance that almost all educational writers, from Plato to Luther to those writing school laws and regulations today, base the need of public education on the argument that privately-provided child-rearing is never, or at least at present, not adequate.

However, taking care of an ill-defined problem by institutionalization does not necessarily make the problem well-defined. It is, rather, a way of limiting the risk of ill-defined problems by placing them within better-defined structures. It is no surprise, therefore, that most institutionalizations of ill-defined social problems are accompanied by discussions on how the institution should treat the problem. Thus, from the origins of public schooling there have been huge differences in the views on the form of solution of the educational problem by this or that institution (see, e.g. Aristotle, *Politics*, ch. 8, as an early example). The modern answer to this problem was the invention of the *professions*: by leaving the problem solving and the running of institutions to professionals, it could be assumed that the problem would find the best possible solution, if the professionals concerned could be considered the best people available to deal with it. In this sense, the invention of the teacher as a professional is like the similar inventions of the medical, military, and legal professions.

Professions stabilize the interaction of *people* and *processes*. They do so by laying out standards for professional education (and the subsequent service-

rendering) which culminate in rites of passage (i.e. exams, internships, evaluations, etc.) either to enter or to leave (i.e. be thrown out of) the profession.

These standards and the measures to enforce them can also be divided along the line of well-defined to ill-defined: well-defined, respected professions are often allowed to define and defend the standards of professional education and service themselves, whereas less well-defined professions run the risk of their standards and membership rules becoming a non-professional, i.e. public, matter.

In addition, one can define the *production* of problem-solutions within these institutions—which *processes* were expected to deliver which *products*. When, for example, professionals were (and are) seen as having competence for their tasks, it was (and is) natural for them to assert a need for the profession itself to undertake the definition of the *processes*—as the body arguably knowing most about the range of possible problem-solutions. This claim was often extended: only the *product* of their involvement, i.e. the ‘aims of education’, should be defined externally. Of course this is not easy to do: an ill-defined problem is an ill-defined problem because we don’t know exactly how to solve it, i.e. what the product of the problem-solving process should be. The converse is also true: it is difficult to come up with process-definitions if the product to be achieved is difficult to define. The most common solution to such problems has been a kind of power-sharing: the public (or whoever represents it) defines the structures of the institution and its professions (for courts, armies, hospitals, schools, etc.), but leaves the definition and control of the day-to-day rendering of services to the professionals involved.

It is easy to demonstrate historically how the level of problem-definition and the degree to which the professionals were trusted have shaped the process of institutionalization.⁵ Thus, the predominant form of pre-modern schooling in Europe was a kind of *contract school* where parents made contracts with the schoolmaster for what their children (of course typically only boys) were supposed to learn in school: how much reading, writing, arithmetic, etc. Parents did not interfere in the ways these outcomes were achieved; instead they looked for another professional if they judged the quality of the results to be poor. Such locally-controlled contract schools were the prevalent type of schooling in many parts of western and northern Europe until the late-18th century. And, such schools function well if the problematic part of education is seen in terms of learning contents and skills which are difficult to teach in the home. The second most successful type of school institutions were, of course, the schools run by professionals with the aim of fostering new professionals like, for example, monastery and cathedral schools, and such emerging institutions of higher education as the mediaeval universities.

The picture changed slowly and gradually as less well-defined problems moved into the field, with, for instance, the need of the emerging states to secure the loyalty of their inhabitants or as churches sought to secure the commitment of their constituencies by other means than pure force, i.e. by measures of personal inclusion. Whereas the distribution of the parts and pieces of knowing can be easily defined, it is difficult to specify means for

fostering and measuring attitudes or ways of being. The argument (as stated, for example, by Luther) continues to be the inability of the primary child-rearing environments (i.e. in most cases the family) to provide an appropriate education. However, the new task of the schools goes beyond the deployment of knowledge and skills to the enforcement of habits. Faced with such new problems, the definition of what is needed could no longer be left to the primary environments or the emerging professions alone but became a public matter. The development of public institutions of mass schooling followed, along with process regulations (i.e. school laws, etc.), and different kinds of product control (i.e. by means of confirmation and other examinations).

An ill-defined problem remains ill-defined, even when institutionalized. It is how the problem gets dealt with that receives *placement* within society—and the acceptance of the new placement as a better definition of the problem is not guaranteed. Thus, the emerging public school system did not replace the existing types of privately-held or -controlled instruction, i.e. contract schools and the like, at least not immediately; the process of change took up to 300 years after the Reformation, or was never finished—depending on where in Europe and under what circumstances the developments took place. The expansion of non-public school systems continued in many areas in Europe until the mid-18th century, depending on the economic and political sustainability of private solutions over church-run or state-controlled schools. Establishing that one solution is more powerful, i.e. better-defined than another, takes place by way of forms of *evaluation*, from evaluating the people and processes involved to product evaluation by supply and demand, once again depending on how well-defined the problem in view is considered to be by whom.

The evaluation of institutions and the institutionalizing of evaluation

As I have suggested, social problem-solving has to be somewhat independent of personal idiosyncrasies in order to function on a larger scale and over the long run. Thus, the professions and specified production lines connected with them, i.e. institutions, were invented. Therefore, the main focus for the evaluation of institutions becomes either the processes or the products involved; the evaluation of the involved people (insofar as they are accepted as professionals) tends to become subordinated and restricted to the question of whether the processes or the products provided by individual professionals are in accordance with professional and institutional standards. Whether the process or the product perspective dominates depends on how well defined the profession and/or the problem-solving (production) is, with the possible mix stretching from strong professions dealing with weakly-defined problems (like handicraft) to weak professions dealing with better-defined problems (like manufacturing) (see table 1).

However, every other combination is possible, producing specific types of professions and forms of institutionalization as well as different approaches to evaluation: Well-defined problems dealt with by well-defined

Table 1. Professions and problems.

	Problems		
	Ill-defined ←————→ Well-defined		
Well-defined Professions Ill-defined			

professions will less likely create a demand for external evaluation than ill-defined problems treated by ill-defined professions. The weaker a profession is considered to be, or the more difficult the assessment of process qualities seems to be, the more likely evaluation will look to find at least some well-defined features of the product to measure—while strong professions enjoy substantial freedom to define both the processes and products themselves, etc.. Thus, the evaluation of institutions spans a whole range of different approaches.

Traditions of process control

This interaction between professions and their institutions proved to be decisive for the struggle between the private and the public school sector and initiated cycles of forms of evaluation which mixed different approaches to evaluation or switched from one type to another depending on:

- the interaction of school evaluation with other frames of public problem-solving; and
- the power of the professions working within the institutions.

The first of these issues is well illustrated by the introduction of the visitation within Protestant (Lutheran) schooling, initiated by Phillip Melanchthon’s detailed instructions on visitation, i.e. a form of inspection, and by the first school laws of the Reformation outlining how schools should proceed. The visitation added to the then-prevalent forms of entrance and exit evaluation (mostly done by *internal* examinations), a new dimension of *external* control, focusing on *processes* of inclusion instead of *products*. Thus, a typical visitation report would, for instance, conclude with the statement that the students showed considerable abilities in the three Rs—which was what contract schools already provided—but that the instruction or the students were obviously lacking the religious spirit—which was what the church and the state were looking for. Also connected to the visitation approach was the requirement that teachers in public schools needed the approval of the respective authorities, a first step in the process towards a unified profession.

However, the strategies of both process evaluation and control only worked within the still-weak public sector, and could not, therefore,

contribute very much to winning the competition with the private sector. Thus, a new move in the evaluation cycle followed, the introduction of confirmation in Protestant states (a step which Luther had refused to take, maintaining his teaching of a division between the two worlds). By means of a first step connecting certain civil rights (such as marriage) to confirmation, and then, in a second step, connecting confirmation to participation in at least some years of schooling, parents were forced to send their children to school, even if they did not see an educational rationale in this step (at least when compared to what could be learned in the privately-controlled areas of education). This intertwining proved to be so successful that, in those areas where the confirmation requirement was first introduced (as early as 17th century in some parts of Denmark) it only took about two or three generations before the school system included almost all available children. Thus, the next move in the evaluation cycle was added: *external product* control in addition to *external process* evaluation, performed by assessing the students' knowledge in externally-defined exams and confirmations.

However, this was still not enough to break the prevalence of non-public schooling. The break only came about with help from the professionals running the schools, i.e. the teachers. It is an historical irony that one of the most powerful forces in shaping the teachers' move towards a unified profession was a counter-movement to the Lutheran view of schooling and confirmation, i.e. Pietism. One of Pietism's basic criticisms of the then-model of schooling was that teaching only the catechism and other more or less religious content was not enough to turn children into truly Christian adults. Only instruction which brought the faith to life could turn knowing into belief. For such instruction, it was not enough that the teachers knew what they were supposed to teach; in addition they should be able to 'stage' instruction in a faith-filled way as *praxis pietatis*, as a living Christianity.

The consequence of this idea was the invention of *seminaries* for teachers, that is schools where future teachers could learn both the contents *and* the ways and means of staging a living faith. From the mid-18th century onwards, pedagogical supplementary courses were also initiated at universities for those who were to become teachers. As a result of a common education, and as a result knowing others working within teaching, it became possible for teachers to develop the self-understanding of a new specialized profession, and to act accordingly. The pace of this professionalization process was remarkably different regionally, as was the pace of school development as a whole. In the early-19th century in parts of Denmark and Prussia, for instance, almost all teachers had completed some kind of seminary education, which had become a formal pre-requisite for becoming a teacher. In other school systems, where the public sector still struggled with positioning itself (as in England), this professionalizing transformation was not completed until well into the 20th century (or is still in the making).

The concept of the normal school as an exemplary stage is closely connected to the insight that successful instruction depends on a specific staging. In such a school or seminary, the how's and why's of the staging could be made visible, and from this the idea of transferring the script of the staged lessons by means of detailed curriculum guidelines added a very

strong tool of *external process control* to the evaluation tool-kit. However, the most successful version of this approach did not come from the Pietist movement, or from other seminaries, but from Jean-Baptiste de la Salle, a French Catholic, from Joseph Bell, an Anglican, and dissenters like Joseph Lancaster, the English Quaker, working in far-less-developed school environments. The so-called *monitorial instruction* which they developed and which, in a number of different versions, had enormous success throughout Europe and US in the late-18th and early-19th century, provided the model for this development.

The main idea of monitorialism was a very detailed prescription of the progression of learning and of the teaching environment, thus permitting students to teach each other and thereby allowing the teacher to take care of large groups of students (up to thousands, it was said). The monitorial movement started out as a surrogate for more traditional schooling in places where the public school sector was not yet developed in order to provide for mass schooling. However, it had its greatest success in areas where there were comparatively well-established school systems with well-educated teachers (as in, e.g. Denmark or Sweden), because the system provided the emerging teaching profession with its own language and set of tools (whereas the Pietist concept copied the education and language of the clergy). For an emerging profession, the monitorial approach was a useful tool in unifying the teaching profession and in making its specific competencies visible.

For professionals, it seems obvious that the definition of the processes and products of their work should be left to themselves: they know most about their problems and the possible solutions to those problems. This interest coincided with the need of the state to gain greater control over the school sector as a whole. Thus, both the state and the emerging teaching profession had an interest in diminishing the influence of local authorities over the schools, either to secure the dominance of the larger public school system or their independence from local expectations—or both. Monitorial instruction could help to achieving this; however, it severely limited the teacher's leeway to shape his own instruction. It left the teachers as the not-very-independent working tools of a teaching system. In the late-18th and early-19th century, this unsatisfactory situation led to the introduction of a new type of control over and evaluation of schools: the freedom of teachers as to teach according to their own standards, but within centralized guidelines which provided frames of reference for their institutions and for the processes within them.

This new structure was copied from the most effective model of organizing work in post-Napoleonic Prussia, *licensing*. The state provided general curriculum guidelines (syllabi or *Lehrplan*) outlining what to teach combined with prescriptions for who could teach, e.g. those who had passed the required teacher examinations, but left the how-to-do-it to what was called the pedagogical or methodological 'freedom' of the local teaching staff or school.⁶ With this system in place, the state could now rely on the support of the teachers when enforcing its school policies: the teaching profession received a stable frame of reference—the school as a public institution—which left its members considerable freedom and professional

independence in their day-to-day activities. From around the 1820s on, this system of curriculum control diffused through most of continental Europe, and it still constitutes the framework for steering the curriculum.

This open system of process control enhanced the independence of the teaching profession, which then turned against all other forms of external school evaluation and control, denouncing them as not being professionally grounded. And, over time, these external controls either lost their connection to the school sector (in the way that confirmation did in the 19th century) or they were emptied of content (as the regulations for final examinations were in the 20th century). The definition of the specifics of the products (what was considered as an adequate student achievement) was, thus, up to the schools and teachers. Passing the final internal exams of one type of school became enough to gain access to the following stages (e.g. the transition from secondary schools to universities), irrespective of what students actually knew. (To that point, such assessment had been subject of external outcome and/or entrance examinations.) Another very visible side-effect of this was the general introduction of marks or grades instead of content-related characterizations as a main way of evaluating students (this shift—at first developed within the Catholic school system—occurred in most parts of the European world in the course of the 19th century).

The teaching profession gained still more influence over the definition of the institutional framework by means of its involvement in the production of these frameworks by, e.g. the commissions developing the curriculum guidelines, and by taking over more and more positions—both in school administration as well as in the political systems dealing with schooling. It took about a century to achieve this position, but today in Continental Europe there are virtually no parliamentary committees or branches of administration dealing with schooling in which the majority of those involved are not active or former teachers. This gives the teaching profession a unique intertwined fabric of *professional self-control* and *self-evaluation* on all levels of the public school sector.

All these elements together created a situation in which any external evaluation of the outcomes of one or another curriculum guideline in terms of changing teaching procedures or learning products did not make much sense. What happened within the frameworks could be different from place to place, depending on the local school cultures and the capacity of local teacher groups. If a new curriculum framework had any direct impact, it was on the semantics of schooling, i.e. on the one hand on the administrative and public discourse on schooling and, on the other hand, on teachers' argumentation about why this or that could be considered as being within the framework. However, a new curriculum rarely required substantial changes in the traditional forms of teaching. Thus, the knowledge of how to create a bridge, by means of argument, i.e. Didaktik, between institutional frameworks and locally chosen activities became the centrepiece of the professional education of teachers in most European countries.⁷ As one US observer reported with astonishment at the end of the 19th century: 'Ask these teachers about whatever they did, and they will come up with a comprehensive didactical explanation' (Prince 1892: 7).

Traditions of product control

The development of assessment and teaching took a completely different direction in countries where either the public school sector or the teaching profession failed to secure the same prominence as it did in much of continental Europe. The necessary pre-requisites for introducing the licensing principle were simply lacking. Different paths of development occurred within such systems, although most shared common features:

- The school sector remained divided among many private and public providers, following different sets of regulation;
- No unified concept of teacher education and of teaching standards was achieved; and
- No unified institutional frame for schooling based on a common curriculum emerged,
- which made it impossible for the teaching profession to gain the same status and influence as it had within the 'Didaktik system';
- and made it impossible to move from external control and evaluation towards internal control and evaluation, thereby leaving
- considerable influence with external product controls, either as a completely external control (as in externally-designed testing at the end of school stages) or as transition controls, e.g. entrance examinations for subsequent school stages.

The history of schooling in the former British empire provides many examples of this story, not least in England and Wales itself (before the change in the control structures in the late 1980s). However, the most prominent example of this is seen in the US, where neither a unified national school system nor national system of teacher education existed. The administration of schools was seen as a local affair, with the states moving in slowly as regional supervisors. Decisions about who was/might be appropriate as a teacher were left to the same local and regional authorities. And, throughout the 19th century, US schooling was organized in a way that is similar to the contract school tradition, with school districts later moving into the position of collective contractors. Even today, schools are primarily seen as an instrument of the local community's provision of methods for the distribution of knowledge and ability which go beyond the capabilities of the average families. Those who can afford it can opt out of the public system whenever they like (using private schools, home schooling, etc.). Process controls (of the kind used in the larger East Coast cities' hugely successful 19th-century monitorial instruction) never gained national status. Nor could process controls contribute to the construction of an independent teaching profession in that the necessary pre-requisite, a unified teacher education at least on the seminary level, was lacking. All this came rather naturally in that the process of nation-building followed a decentralized (federalist) and non-denominational pattern, without the driving force of the combination of state or nation and church striving for inclusion by means of schooling.

When, at the end of the 19th century, the need to enhance the standards of US public education was finally accepted as a national issue, the one tool available as a result of this history was shared standards of product control—as with the Committee of Ten which defined college and university entrance requirements in the 1890s. All subsequent initiatives to follow up the product approach by process controls, e.g. a national curriculum, have failed. As a result, the 20th century was characterized by an explosive multiplication of external product controls in the form of all kinds of testing and assessment, developed and administered by local or state agencies or private providers, or in a few cases (like the federal government's National Assessment of Educational Progress) on the national level. Not surprisingly (if we take the contract roots of the school system as a whole into account), these product controls focused on the distribution of knowledge and abilities.

However, external product controls, when disconnected from process controls, inevitably encounter the principal difficulty around this approach when it is used to deal with ill-defined problems. It does not acknowledge the degree to which what is measured in fact represents a value that is added by the evaluated institution, i.e. to what degree student achievements are indeed an indicator of the performance of the schools or, rather, reflect factors beyond their reach (like socio-economic circumstances).⁸ In line with the contract tradition, schools are made accountable for the outcomes, whatever their origin might be.

Tracking the impact of new curriculum frameworks does not make much sense in the licensing system in that, given the professional autonomy of teachers, these frameworks are realized in a variety of local activities and outcomes. It is different in product evaluation-based systems. Tracking the impact of changes in the evaluation system makes perfect sense in that the how's and why's of local teaching do not count and the quality of the products of schooling, i.e. student achievement, seems to be well-defined.

It also follows that the teaching profession has different possibilities for showing its professionalism in the different systems (see table 2). In the licensing system, the common core of professionalism is, as I have noted, *Didaktik*, the ability to connect the institutional frame with the not-yet-defined variety of local day-to-day activities and outcomes by means of pedagogical arguments. In the product-centred approach, *efficiency* becomes the core of the profession: the best teacher is the one who gets his students to 'stand up and deliver' at the right time—whatever the cause for their ability to do just that might be.

Changing patterns of evaluation

Thus, one can observe, at the end of the 20th century, two different approaches to curriculum evaluation and control in the Western world:

- the continental licensing or *Didaktik* system, exercising a weak control and evaluation of the processes and almost no external control of the outcomes of schooling; and

Table 2. Varieties of professionalism within the teaching profession.

Differences between process and product control	
Process	Product
<i>Input-based</i> plans and frameworks as main instrument	<i>Output-based</i> output results/standards as main instrument
<i>Different results allowed</i> if the procedures are in accord with the plans/framework	<i>Different procedures allowed</i> if the results meet expectations
<i>The teacher as main addressee</i> pedagogical (didactical) legitimacy	<i>The student as main addressee</i> efficiency of learning
The basic claim of professional expertise	
DIDAKTIK	EFFECTIVE TEACHING

- the product-centred system of external controls as seen, for example, in the US.

In the past, doubts about the performance of the school as public institution have been normally dealt with within the trajectories provided by these traditions, either by issuing new curriculum guidelines in the case of the continental European systems or issuing new approaches to assessment in the case of the product-centred systems. Alternative approaches—introducing, for example, assessment tools in licensing systems or curriculum packages or guidelines within assessment systems—have been almost always time-limited local events.

However, in the 1980s and 1990s, both traditions stumbled into a substantial crisis which might well change the traditional patterns. In the US, for instance, the public came to see the whole ‘nation at risk’ because of the shortcomings of the public school system; decades of ‘raising the stakes’ and calling for better performance had obviously not produced adequate outcomes. And, the results of evolving assessment research supported this public view:

- most teachers reported that new assessment strategies had no, or only little, impact on their lesson-planning, teaching, student involvement, student achievement, etc.;
- the format, size, level of detail, etc. of assessment tools had no or very little impact on how students and teachers cope;
- higher stakes, added content, etc. lead to almost nothing or rather the opposite, in short; and
- the main effect of the external product-evaluation tools seemed to be standardization and stress, but neither innovation nor the enhancement of quality.⁹

Not a good bottom line for an evaluation system which had eaten up billions of dollars and caused quite a lot of frustration among teachers, students, and parents! Teachers and students do have these product controls in mind, especially when they are low-achievers facing accountability expectations, even while constantly failing. However, the most natural reaction in such situations was, and is, to narrow the scope of the curriculum to what the product control asks for as stand-up performance—which does not yield a long-term strategy for the improvement of the quality of outcomes as a whole and in the long-run.

On the other side of the Atlantic, the public and their politicians evaluated the situation in a similar way. Decades of reforming the curriculum again and again had obviously not brought about the changes which the curriculum authorities had hoped for. And, the research on the impact of new curricula could be read as neatly supporting this impression:

- most teachers reported that new curriculum guidelines had no or only little impact on their lesson-planning, teaching, their students' involvement, student achievement, etc.;
- the format, size, level of detail, etc. of the guidelines had no, or very little, impact on how students and teachers cope;
- higher stakes, added content, etc. led to almost nothing, or rather the opposite, in short; and
- the main effect of the external process evaluation tools seemed to be legitimation and the distribution of new arguments around the curriculum, but neither innovation nor the enhancement of quality.¹⁰

Not a good bottom line for an evaluation system which had produced literally thousands of guidelines and caused quite a lot of frustration among teachers, students and the public! Of course, teachers do have the guidelines in mind when planning their teaching. They know them and relate to them if they have to defend why what they are doing fits within the guidelines. If a new curriculum presents a completely new challenge, the most reasonable reaction was, and is, to wait until it is replaced by the next generation of guidelines (which happens about every 5–10 years, irrespective of the content area and the school level) and to adapt their didactical argumentation rather than their teaching until the change comes—as they did, e.g., with the 'new math' of the 1960s and 1970s and do today with the 'project method' and other 'new' instructional requirements.

In both the US and Europe, schooling underwent tremendous change in the post-war period: i.e. new contents, new methods, levelling gender differences, broadening the access to higher education, etc. However, in neither system could these changes be reasonably attributed to the tools of control and evaluation in use—nor were the outcomes in terms of achievement or public impressions good enough for those who had to judge the change or be judged by them. Not least, the emerging international comparisons of student achievement (as undertaken by the IEA and others)

seemed to support the diagnosis that schooling in Western societies was sub-standard, which fuelled a tremendous strive to climb the ladder on the comparative evaluation tables.

This common feeling of crisis, a sentiment which was supported by international organizations like OECD and UNESCO, led those responsible on both sides to look across the Atlantic for possible solutions. In the US, in the years since the publication of *A Nation at Risk* (National Commission on Excellence in Education 1983), all states but one have introduced state curriculum guidelines of some kind (i.e. curriculum standards, etc.); 'standards-based reforms' and 'systemic restructuring' are terms found everywhere; more rigorous licensing of teachers and accreditation of teacher education according to national standards has come onto the agenda.¹¹ In short, the tool-kit of *process evaluation* has been introduced as never before.

On the other hand, most western European states have turned to the tool-kit of *product control*, introducing all kinds of assessment and testing to their systems in the hope that this will reassure the public that everything possible to improve the outcomes of schooling is being undertaken. The best-known of these approaches is found, of course, in England and Wales where a formerly locally-controlled school system is now subjected to both an impressive array of national assessment tools and a national curriculum which aims at high standards in every subject for every student. However, other traditionally high-performing countries like Switzerland and Norway have undertaken large-scale evaluations of their school systems. Countries like Sweden and Finland have broadened the number of test and assessment tools they use, combining this approach with a reduction of the traditional centralized process controls. Comparative product assessment has come to these countries in a significant way, and has come to dictate the public discourse on schooling to a hitherto-unknown degree.

It remains to be seen what the outcome of this realignment or rather—taking the internationalization of educational politics into account—the merger of these two traditions will be like. To this point, all we have are guesses. Thus, as we look at the introduction of process tools in the assessment tradition, it seems that these new tools first and foremost serve to co-ordinate the assessment activity, but have no direct impact on local schooling. The 'stand-up-and-deliver' pedagogy of assessment is too strongly ingrained in the day-to-day routines to be thrown out for good by far-away state guidelines. Much will depend on the ability of the teaching profession to use the new tools to re-define its professionalism and to improve its public standing.

There are some indications that the inclusion of product controls may have a comparatively larger effect on the licensing system. Teacher unions throughout Europe have complained that these tools undermine the professionally-adjusted service-rendering which had been the characteristic of the European system. And the move towards evaluating schools by knowledge and ability distribution contributes to the re-surfacing of the contract school, i.e. the concept of opening the school sector up to market solutions by, for example, calling for more local school 'autonomy' or simply by private competition.

It may be that, as in so many times in the past, the traditions prove to be strong enough to cool the impact of whatever is added to the evaluation tool-kit, and each system will continue to follow the main lines of their institutional history. However, if the merger of the two traditions does achieve all of its intended goals, any comparative evaluation will show inevitably some systems and countries above and some below the international mean. It may be that the one best solution would be to send all the evaluators and their tool-kits off to the shores of Wobegon to evaluate what may be one of the most successful mergers of cultural traditions, i.e. find out how the Norwegian bachelor-farmers of Minnesota manage their educational processes so that all the woman are strong, all the men are good-looking, and—not least—all the children above average.¹²

Notes

1. Traditional accounts of the history of evaluation tend to reduce its scope and content to the Anglo-Saxon component, leaving out the much older roots stemming from, e.g. Europe's 16th-century religious Reformation and Counter-reformation. By proceeding in this way, they miss the fundamental structures and processes which have shaped the differences between, e.g. the US, the UK, and many European countries. Sometimes, this leads to real errors of fact: that, for instance, one William Farish invented quantitative marks to assess student performance in 1792. If that was the case, it was a classical re-invention of a wheel which was widely known in Continental Europe by this time (see, as a typical example, Maddaus and Stufflebeam (2000: 4)).
2. The paper is based on research undertaken within the evaluation of the Norwegian curriculum reform of 1997 (EVA 97 supported by a grant of the Norwegian Research Council) as well as my contributions to a number of other national and international evaluation projects. For examples and further references, see Künzli and Hopmann (1998), Goodson *et al.* (1999), Hopmann (2001), and Nesje and Hopmann (2002).
3. A good account of the various shapes and sizes of policy evaluation is given by Vedung (1997) and Stufflebeam *et al.* (2000). Older, but still useful accounts may be found in Kaufmann *et al.* (1986) and Chelimsky and Shadish (1997). Models of curriculum evaluation are discussed in, e.g. McLaughlin and Phillips (1991), and Jackson (1996).
4. See, e.g. Luhmann (1964, 2000), Schülein (1987), Schmalz-Bruhns (1989), and Leschinsky (1996).
5. The following historical account is based on historical studies which I have conducted in Germany, the US, and the Nordic countries (see, e.g. Hopmann 1988, Haft and Hopmann 1990). A more detailed account will be published in Hopmann (in press).
6. In a very real sense, this system is analogous to the privileges given to holders of drivers' licences: with a licence a driver is free to drive anywhere, although there are restrictions intended to regulate the overall traffic system in order to sustain its overall coherence and order.
7. Didaktik means far more than the English term 'didactics'. Didaktik is the centre-piece of professional teacher education in Continental Europe, covering all subjects relevant for teaching; in the Anglo-Saxon tradition these subjects are assigned to different disciplines such as curriculum theory, educational psychology, teaching methods, etc. (see Hopmann and Riquarts 1995, Westbury *et al.* 2000).
8. There is a never-ending debate both within and outside the evaluation community on what these assessments do indeed assess and to what degree this is an effect of the institutions involved (see, e.g. Teddlie and Reynolds 2000).
9. A useful summary of the pros and cons of assessment is given in Mehrens (1998).

10. For examples, see e.g. Künzli and Hopmann (1998), and Nesje and Hopmann (2002).
11. To deal with the most problematic cases, the Chicago school system, for example, even brought back the old-fashioned ideas of scripted lessons and external school visitation/inspection, the strongest tools of process control.
12. The lake and town of Lake Wobegon, Minnesota, is the subject of a popular series of radio monologues, 'News from Lake Wobegon', that are a feature of the US radio programme, 'Prairie Home Companion' (<http://www.phc.mpr.org>).

References

- CHELIMSKY, E. and SHADISH, W. R. (eds) (1997) *Evaluation for the 21st Century: A Handbook* (Thousand Oaks, CA: Sage).
- GOODSON, I., HOPMANN, S., and RIQUARTS, K. (eds) (1999) *Das Schulfach als Handlungsrahmen. Vergleichende Untersuchungen zur Geschichte und Struktur der Schulfächer* [School subjects as a frame of reference. Comparative studies on the history and structure of the school subjects] *Bildung und Erziehung*, Beiheft 7 (Cologne, Germany: Böhlau).
- HAFT, H. and HOPMANN, S. (eds) (1990) *Case Studies in Curriculum Administration History* (London: Falmer).
- HOPMANN, S. (1988) *Lehrplanarbeit als Verwaltungshandeln* (Kiel, Germany: Institut für die Pädagogik der Naturwissenschaften).
- HOPMANN, S. (2001) Von der gutbürgerlichen Küche zu McDonald's: Beabsichtigte und unbeabsichtigte Folgen der Internationalisierung der Erwartungen an Schule und Unterricht. In E. Keiner (ed.), *Evaluation (in) den Erziehungswissenschaft* (Weinheim, Germany: Beltz), 207–224.
- HOPMANN, S. (in press) *Didaktikkens didaktikk* (Copenhagen: Unge pædagoger).
- HOPMANN, S. and RIQUARTS, K. (1995) *Didaktik and/or Curriculum* (Kiel, Germany: Institut für die Pädagogik der Naturwissenschaften).
- JACKSON, P. W. (ed.) (1996) *Handbook of Research on Curriculum: A Project of the American Educational Research Association* (New York: Macmillan).
- KAUFMANN, F.-X., MAJONE, G. and OSTROM, V. (eds) (1986) *Guidance, Control, and Evaluation in the Public Sector: The Bielefeld Interdisciplinary Project* (Berlin and New York: W. de Gruyter).
- KÜNZLI, R. and HOPMANN, S. (eds) (1998) *Lehrplanarbeit in der Schweiz und der Bundesrepublik Deutschland* [Curriculum making in Switzerland and the Federal Republic of Germany] (Aarau and Basle, Switzerland: Sauer).
- LESCHINSKY, A. (ed.) (1996) *Die Institutionalisierung von Lehren und Lernen: Beiträge zur einer Theorie der Schule 34*. Beiheft der Zeitschrift für Pädagogik (Weinheim, Germany: Beltz).
- LUHMANN, N. (1964) *Funktionen und Folgen formaler Organisation* (Berlin: Duncker and Humboldt).
- LUHMANN, N. (2000) *Organisation und Entscheidung* (Wiesbaden, Germany: Westdeutscher Verlag).
- MADDAUS, G. F. and STUFFLEBEAM, D. L. (2000) Program evaluation: a historical overview. In D. L. Stufflebeam, G. F. Madaus and T. Kellaghan (eds), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (Boston, MA: Kluwer), 3–18.
- MCLAUGHLIN, M. W. and PHILLIPS, D. C. (eds) (1991) *Evaluation and Education: At Quarter Century*, 90th Yearbook, Part 2, of the National Society for the Study of Education (Chicago, IL: National Society for the Study of Education).
- MEHRENS, W. A. (1998) Consequences of assessment: what is the evidence. *Education Policy Analysis Archives*, 6 (13). <http://olam.ed.asu.edu/epaa/v6n13.html>. (Visited 11 November 2002.)
- NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION (1983) *A Nation at Risk: The Imperative for Educational Reform: A Report to the Nation and the Secretary of Education, United States Department of Education* (Washington, DC: US Government Printing Office).

- NESJE, K. and HOPMANN, S. (eds) (2002) *En Lærende Skole: L97 i Skolepraksis* (Oslo: J. W. Cappelens Forlag).
- PRINCE, J. T. (1892) *Methods of Instruction and Organization of the Schools of Germany for the Use of American Teachers and Normal Schools* (Boston, MA: Lee and Shepard).
- SCHMALZ-BRUHNS, R. (1989) *Ansätze und Perspektiven der Institutionentheorie. Eine bibliographische und konzeptionelle Einführung* (Wiesbaden, Germany: Universitätsverlag).
- SCHÜLEIN, A. (1987) *Theorie der Institution. Eine dogmengeschichtliche und konzeptionelle* (Opladen, Germany: Westdeutscher Verlag).
- STUFFLEBEAM D. L., MADAUS, G. F. and KELLAGHAN, T. (eds) (2000) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (Boston, MA: Kluwer).
- TEDDLIE, C. and REYNOLDS, D. (eds) (2000) *The International Handbook of School Effectiveness Research* (London and New York: Falmer).
- VEDUNG, E. (1997) *Public Policy and Program Evaluation* (New Brunswick, NJ: Transaction Publishers).
- WESTBURY, I., HOPMANN, S. and RIQUARTS, K. (eds) (2000) *Teaching as Reflective Practice: The German Didaktik Tradition* (Mahwah, NJ: Lawrence Erlbaum Associates).