School external evaluation: theory and models

José Augusto Pacheco

Universidade do Minho

In his search related to the making up of a theory of the field of evaluation, Sobrinho\(^2\) refers to multi-faceted references and complexity as central elements, not only because evaluation is a “field whose domain is disputed by various disciplines and social practices of distinct academic, political and social institutions” but also evaluation is a “social phenomenon,” which can be seen to include “actions, attitudes and values of individuals from diverse backgrounds.”

If these words of a renowned author in the field of evaluation serve to characterize evaluation as a complex social activity, on the other hand, they also validate this theoretical incursion of mine into institutional evaluation, because evaluation of a school is an activity that requires specialist knowledge of Curricular Theory and Development. Above all, if this epistemological field would be considered as an interdisciplinary study of educational experience,\(^3\) where the common thinking and in an update of Schwab,\(^4\) are the context (political, social, cultural, economic and ideological), the actors (students, teachers, parents and educational guardians, local authorities, publishing houses, trade unions, etc.) and knowledge.

One is dealing with a markedly curricular vision that I intend to include in this contribution not forgetting my situation as a teacher of Theory and Models of Evaluation in the University of Minho but also my experience acquired as an external evaluator of the model, which is at the centre of this article.

My inclusion in the process of the external evaluation of schools or groups of schools, later designated by schools in the role of external specialist, does not compromise me in terms of a critical epistemological analysis, in so far as I assume a

---

\(^1\) Conference presented to Nacional Seminar on external Evaluation, Braga, Universidade do Minho, 2010.
\(^3\) Cf. William F. Pinar, O Que é a Teoria do Currículo?, 2007, p. 18
position of distance in this academic world, even that my exhaustive knowledge of the object under analysis (the school) enables me to reconcile judgements that are based on experience and with value judgements.

1.

In the bibliography consulted for the writing of this text, there is a question/doubt that I would like to share, which is raised by Figari\(^5\) in the following way: “unlike that which happens in Portugal, the debate about questions of school evaluation seems to have stopped, which is regrettable and a surprise.”

The designation of the school as an object of evaluation cannot have been the cause of much academic output and discussion in the last few years, but also it is certain that institutional evaluation has undergone an intense approach, mainly through transnational and supra-national organizations, which have been defining the political agenda and imposing logic and regulation of models. There are the cases of the OCDE,\(^6\) whose contribution to the evaluation of schools is a current reference in political agendas and the European Union,\(^7\) with reports that link evaluation of schools to autonomy and the responsibility of teachers. One cannot, therefore, maintain that institutional evaluation has not been at the centre of the political agenda and even though it has not produced a large amount of academic output, which has often not been widely disseminated and is little known due to political actors.

The first question that I would like to analyze in this process of evaluating schools - initiated in Portugal in 2005 - is its poor representativeness in terms of writing up of studies and discussions in academic and social forums. If the model adopted has as its origin – How good is our school? – we should compare what was produced in the respective national areas, so as to discover how far away we are in Portugal from the Scottish reality in terms of both theoretical and methodological output as well as social discussion.

\(^5\) Cf. Gérard Figari, A avaliação de escola: questões, tendências e modelos, p. 41.
Another question, which appears to be endemic in all educational systems, is the distancing that exists between the production of the law (Law 31/2002 of 20 December) and the implementation of a model for external evaluation (fulfilled in a pilot study in 2007 and generally implemented in school year 2007/08\(^8\)). This political distancing signifies the incorporation of a building of ideas in the law that originates from international and supra-national agendas, which takes time to appropriate nationally, but without a significant academic contribution. Here we are dealing more with the technical appropriation of a timing with the help of politicized technical decisions, that is to say, the creation of working groups of a technical nature to whom is entrusted the task of creating the mechanisms of response imposed by law.

In this case, one hypothesis that I would like to raise in my analysis of the model for the external evaluation of schools concerns an insufficient theoretical and methodological discussion of the model, because academia is poorly represented in respect of the contribution that the educational sciences and other related areas could have given for this purpose.

Without intending to confirm or invalidate this hypothesis, since it is simply a statement in a space for debate such as this one, I propose to analyse the actual model using ideas about the theory, paradigms and models of evaluation.

I could indicate that the model lacks a theory of evaluation. In the analysis of the document entitled “External evaluation of schools: References and Instruments for Evaluation” one cannot find any reference to the notion of evaluation, neither to paradigms. The document - a synthesis of other documents - is essentially a technical selection, which explains objectives and its internal structure.

Evaluation being “the systematic judgement of the value or merit of an object (Stufflebeam and Shinkfield, 1989, p. 19),” as well as “a process for taking decisions (...) that require a political and moral judgement (Stake, 2006, p. 38)” and for Scriven (1999), the logical foundation of evaluation is based on factual premises and premises of value, so a non-theoretical model of evaluation can become a technical reference.

The act of evaluation signifies the formulation of a value judgement that, in its turn, implies the attribution of a meaning between an object of evaluation (product) and

---

\(^8\) It is certain that between 1999 and 2002 the General Inspectorate for Education proceeded to use Integrated Evaluation based on the methodology of the *European Foundation for Quality Management* from 1999 to 2002.
a reference (criteria) for the purpose of taking a decision, which in this case would be a classification. Now, the model implemented in these three years is not sufficiently discriminatory for an understanding of what is to be considered - in the language of evaluation – namely, a criterion, a standard and an indicator. For Stake (2006) the criteria is a descriptive variable, the standard is the measure of the attribute necessary for the description of a criteria and the indicator is a specific descriptive item that is always exemplifying.

By reading the objectives of the model, it is possible to catch a glimpse of the elements of a formative theory (Domingues, 2007) – the articulation of the external evaluation with the internal evaluation - although the normative theory calls attention to too many objectives. This discussion about the theory of evaluation combines with the discussion about paradigms, even so, nowadays, when evaluation becomes a central instrument of educational reforms (Sobrinho, 2003) - having as its underpinning budgetary rationality and accountability (Lima & Afonso, 2002) – it has turned into a powerful instrument of competitiveness. Many references could be summoned up to support this statement but I judge that there are sufficient that can be found in the works of Almerindo Afonso (2009) and of Leonor Torres and José A. Palhares (2009).

The externality of evaluation, which requires a degree of comparability, can be defined through an understanding of a paradigm to be implemented knowing that the evaluator looks for knowledge about the value of something and cannot even dissociate his/her “personal experience” (Stake, 2006, p. 61) from this same object about which he/she has to give a value pronouncement. Again as Stake holds (Ibid., p. 61), “evaluation consists always in the determination of the merits and defects. Sometimes it is much more but its essential function is the establishment of the merit of something. This is its prime purpose” and, consequently, demands from the evaluator the taking of a position, almost always - at least at the level of external evaluation - measured with an objectivistic rationality (Rodrigues, 2002), even when the external evaluation of schools becomes a markedly state strategy. There is, like this, a certain parallelism between the normative function of evaluation linked to regulation and “parameterization” than can be evaluated in the school and the objectivity of the results - converted on a quantitative scale – that are supported by descriptions.

The playing out of these two theories and two paradigms in evaluation links to the usefulness that external evaluation has for the various participants. It depends - over and
above other motives - in the way that the results of the evaluation are read and manipulated by the educational authorities, by the school itself and by the community. Not being disposed on the side of this hypothesis: there are schools that assign a more formative and comprehensive value to evaluation reports than other schools. In the latter cases the emphasis of interest is only to register the number of goods and very goods than to a critical analysis of their functioning.

In the words of Stake (2006, p.144) the comprehensive paradigm “is a way of searching for and documenting the quality of a program. In as much to use the measurement based on criteria, as far as the interpretation is concerned. Its essential aspect is the intelligibility (receptivity, sensibility) of questions or key problems, especially those that are experienced by people in their position or program.”

Nevertheless, evaluation in theory and in practice, lives through a contradiction existing between, on the one side, the search for objectivity (and comparability in the case of external evaluation) and, on the other side, to proceed with a sense of comprehensibility. For that reason, as Sobrinho (2003, p. 149) recognizes, “evaluation exists in the conflict between two worlds and two registers of words,” that is to say between the objective and the subjective, the quantitative and the qualitative.

2.

In a revision of models of evaluation of schools one can observe that evaluation is presented as a resolution of the problems that the schools have, thereby making it an activity that helps them to have better results and to know better the way they function (Stufflebeam, 2003). Thus, “evaluation of the school can be defined as the systematic investigation of the quality of the school and the way it can best serve the needs of the community” (Sanders & Davidson, 2003, p. 807).

As there are many models and differences in the practices of evaluation, the implementation of a model reflects their own political options and makes it as if it is not possible to have one total vision about what the school is and the way in which it functions. Therefore, for Sanders and Davidson (Ibidem, p. 808,) the practice of evaluation of schools involves two basic activities: collecting information and data that can be available and accessible; using criteria to judge the level of adequacy of the data, which is desirable.
In spite of the various ways of collecting data, it is common to the models that the schools have an active part in the provision of data both by the submission of reports and through active participation in the production of empirical data (by participating in evaluation panels, replying to questionnaires, presenting data, etc.).

In the model proposed by Sanders and Davidson (2003), the Inspection has active participation, while in the model of Stufflebeam (2003, p. 783), the evaluation is as much an individual responsibility as a collective one of all of the people entrusted with the work of the school for the benefit of the students and the community.”

If for Stufflebeam (Ibidem, p. 782), the purposes of the school evaluation are the improvement, the accounts rendered, the understanding and the dissemination (adoption of good practices), the system of evaluation must involve as well as students, teachers, parents/educational guardians and elements of the community, since the school has social aims and should regularly furnish feedback about the achieved results, according to its agenda for decision making.

The responsibility of the school in external evaluation tends to grow, as self-evaluation becomes an institutional practice and serves to support appropriate external evaluation, and not, as happens in Portuguese reality, to an evaluative practice of conformity. However, the understanding about what is an institutional evaluation on the part of the teachers and head teachers is a question that requires to be widened, in so far as the schools cannot do of this evaluation the prolongation of the collection and analysis of the trimestral results of learning.

One of the most complex questions of the model certainly is the selection of what constitutes the object of evaluation and what are its facets or evaluative dimensions. For Stufflebeam (2003, p. 776) all important aspects of the school should be evaluated on a constant basis to procure its individual and collective improvement.” And among these aspects the author enumerates three main areas: the students, the personnel and the programs (including the curriculum, as well as its many diverse services). In this respect Sanders and Davidson (2003, p. 814) declare that “school evaluation covers more than the elements referred to by Stufflebeam and should include, moreover, the

---

9 The authors, p. 811, mention that in an international review of external evaluation of schools, there is a growing consensus between European countries so that the Inspection should have an active role in this process.
“performance” of the students, the facilities, the finances, the school atmosphere, the school policies and the records of the school.”

Besides the clarification about the object of evaluation, Stufflebeam (2003. p. 777) raises the question about which level the evaluation should be carried out, asserting that “it is crucial to evaluate all of the school levels,” considering the school as a whole, including the classrooms, but this whole should not be considered as a sum of individual parts. For this reason it is fundamental that the evaluation proceeds in a collaborative way among the various participants, who should each know the data which is to be collected and the criteria to be applied, since only the use of “criteria helps to determine if something is good, adequate or poor” (Sanders & Davidson, 2003, p. 817).

The process of an evaluation of quality is completed with the questions to be raised, which Stufflebeam invariably identifies at the level the variables of the context, of the input, of the process and of the output. These variables, thus, constitute the generic reference categories of the model of evaluation for the particular audience and include a definition of the criteria, the choice of the methodology for collecting the data (both qualitative and quantitative), a definition of the standards (ethical adequacy; usefulness; feasibility) and rigor or exactness (Rodrigues, 2009). Of all the elements in the process of evaluation the “definition of the criteria is not easy and can become controversial” (Sanders & Davidson, 2003, p. 817), because it is fundamental that they should be defined and shared in an open way by all of the evaluators without leaving doubts over much of their interpretation. Without the definition of criteria it is also not possible to have a school evaluation of quality and, therefore, as mentioned by Stufflebeam (2003, p. 804), the evaluation of the school is a “monumental task” that “constantly needs to improved.”

Besides a more systemic model that is oriented to the taking of decisions, such as that of Stufflebeam (the CIPP model), there are others that can be referred to as is the case of the models responding to Stake (2006) and Sanders and Davidson (2003). They state that the practices for the evaluation of schools are multiple and because of this they cannot be combined with the application of a model that calls for the consideration of variables more or less in the long run. As far as the model centered on the Inspection is concerned, they considered an evaluation based on a visit that focuses on three more models: evaluation centered on the indicators, that is, an evaluation of conformity,
which has tended to become the norm for international studies centered on the results of the students; evaluation centered on local initiatives; an ad-hoc evaluation of the school held on the initiative of a group external to the school but relating to the community.

With these possibilities, the first cycle of external school evaluations in Portugal (2007-2011) has followed a uniform model at the national level, in spite of the existence of schools that begged for a more flexible model.

3.

After a more general questioning, I propose to follow an approach centered more on the content of the model by discussing what can be considered a school evaluation and which parameters can be proposed.

When the school constitutes an object of evaluation, it is necessary to ask what kind of perspective projects itself more at the moment of defining parameters. Being limited by pressure for the production of norms in Portugal, the perspective is more sociological and organizational (Canário, 2005; Barroso, 1996) than curricular (Pacheco, 2008) - even though the didactical perspective has been more present in the last few years, external evaluation is inscribed in the final cycle of reform of the educational system after the publishing of the 1986 Base Law of the Education System (BLES).

As it is a difficult object to define (Figari, 2008), evaluation of schools in Portugal tends to overvalue the more visible aspects of their functioning with a tendency for an emphasis on organizational questions. By force of the political agenda for “accountability,” school evaluation tends towards focussing first and foremost on the results and on what is more positive and less positive, such as information detected by a SWOT analysis. School evaluation is a legitimate legal activity in accordance with Article 49 of the BLES fulfilled by a governmental agency (the Inspectorate-General for Education) and completed by the participation of external experts. This makes it a less desirable activity for the school.

Notwithstanding his/her neutrality and impartiality, the evaluator always tries to save some thing, as Stake (2006) states, as it is not possible for an external evaluator to refrain from sharing common aspects of the model that he/she finds adequate, ethical,
feasible and rigorous, in the outcome of the standards delineated by the Joint Committee on Standards for Educational Evaluation.

With respect to some of these standards, I propose only to make an analysis of the overlapping of the “dimensions” and the “factors,” as well as the “references to take account of in the evaluation of the factor.”

When the primary element of the model is made by the results, which does not happen with the model inspiring How good is our school?, the tendency is to adopt the productivist models and the rendering of accounts, since the results of the students are a reflection of the great amount of work carried out in the school and the community. This element contains “factors” that include academic success, participation and civic development and the valuing and impact of learning. In spite of there not being an application of a weighting factor in the model, it can be easily assumed that there are items that have more weight than others, by which it would be advantageous to distinguish what are the results of the learning processes, including the management of processes of quality in the monitoring of the students in the classroom, the school and the community.

The “dimension” of the educational services rendered includes, in a superficial way, curricular questions and which should be explored in the model in terms of vertical and horizontal articulation, so that the curriculum should embrace projects and justice/equity (present in the “dimension” Organization and School Management), as well as the valuing and impact of learning (from the “dimension” Results) and an opening for pedagogical innovation (present in the “dimension” Leadership).

There is, so to speak, an overlapping of the elements of evaluation, not having, sometimes, an understanding of what is more appropriate to include in one or other dimension, being that the organizational questions and the curricular questions could have, in this model, a strong element of discrimination, registering the overlapping of various “factors.”

One other question, where the model can be improved, would be at the level of clarifying the evaluation criteria in a clear and objective statement for the evaluators. For example, what can be considered as “open to innovation” and as an “innovative solution?”

The definition of many of these criteria should have to correspond directly with legal procedures and curricular instructions, even knowing and accepting that the
schools do not always follow the official directives in terms of curricular and organizational practices.

Yet in my experience, as an external evaluator, I can record that the model was not contested and serves as a counterbalance to the evaluation of the teacher, which is a revelation of the authority, which the schools have in time have been making some conquests.

Nevertheless and so far as evaluation takes effect, it is necessary to know what is the impact of this evaluation in schools and in the community and to know in which way the quality of the school can be constantly improved.

If each one of us involved in the school and/or the community would give his/her contribution to a critical analysis, directly or indirectly, it is most likely that Portuguese schools will have an even higher referential of quality.

References:


