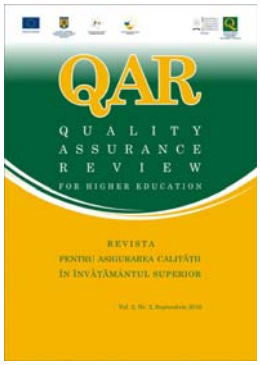

Quality Assurance Review for Higher Education

Defining Quality in Higher Education – Practical Implications

Mirela Cerkez

Quality Assurance Review, Vol. 2, Nr. 2, Septembrie 2010, p. 109 – 119

Publicat de: Consiliul Agenției Române de Asigurare a Calității în Învățământul Superior - ARACIS
Locul publicării: București, România
Tipul publicației: tipărit, online

Quality Assurance Review este editată în cadrul proiectului „Asigurarea calității în învățământul superior din România în context european. Dezvoltarea managementului calității academice la nivel de sistem și instituțional", Contract POSDRU/2/1.2/S/1, cod proiect 3933.
Pentru a reproduce un text din revistă este necesar acordul scris al colegiului de redacție al revistei Quality Assurance Review.

Răspunderea pentru textele scrise aparține autorilor.
Conținutul acestui material nu reprezintă în mod obligatoriu poziția oficială a Uniunii Europene sau a Guvernului României.

FONDUL SOCIAL EUROPEAN

Investește în OAMENI

Defining Quality in Higher Education – Practical Implications

Mirela Cerkez *National School of Political Studies and Public Administration*
mirela_state@yahoo.com

Abstract: *This article explores the practical implications of the various understandings that we attach to the idea of quality in the context of educational program evaluation. In order to highlight the major theoretical and practical implications that the various understandings of quality have, I will follow the major lines of development in the history of program evaluation. The presentation starts with early periods of development, when there was no delineation between the assessment of the students and the quality of the program, and ends with today's open quality definition. The last paragraph is a rather short analysis on the idea of quality that lies behind the ARACIS methodology of evaluation and ends with some suggestions for improvement. The main line of argument draws around the difference between accreditation and quality assurance.*

Key words: *accreditation, evaluation, quality assurance, educational program.*

Introduction

As the data presented below will show, the establishment of the Romanian Agency for Quality Assurance in Higher Education (ARACIS) and the Romanian Agency for Quality Assurance in Pre-university Education (ARACIP) in 2005 is a solid and natural step in creating and developing evaluation theory and practice in Romania. In the United States, evaluation's country of origin, it was given birth related to the educational programs and up until the present the most important theoretical developments, that are then imported by other fields, are presented in academic reviews dedicated to the educational field amongst which I name: *Review of Educational Research*, *Educational Evaluation and Policy Analysis*, *Studies in Educational Evaluation.* Also, the most experienced evaluators in academia are not strangers to evaluating educational programs and institutions and the most prestigious publishers in the USA and UK have many titles in educational evaluation.

Hereby, I explore the practical implications of various understandings of an educational program's quality and, in order to do that, I will follow the stages of development that educational evaluation went through, as mentioned in the literature. References to developments from the 1900s are not to be neglected as the creation of more sophisticated approaches does not mean that the old ones should be abandoned. The new approaches are able to deal with complex situations but that does not mean that, depending on how we define our evaluation interest, we could not undertake an older approach.

Specifically, my intention is to argue that, depending on how we define an educational program's quality, the evaluation process tends to highlight certain aspects of the program and shadow others. Given that, the main suggestion in this article is that, each time we start an evaluation process with the purpose of judging its quality, we should state the understanding that we have on quality, otherwise the evaluation process will not necessary deliver the information that we expect it to. More so, the way we understand quality has direct effects on the indicators we use in the evaluation process and that influences the selection of the evaluation model and the specific procedures that we use to produce information on the program's quality.

The argument in this article draws its significance from the fact that, usually, evaluation processes work with implicit assumptions on the definition of quality; the evaluator and the client do

not necessarily establish a common understanding on that. Because we are in the situation of having an established quality evaluation procedure in higher education that consists of the methodology developed by ARACIS, I will dedicate the last paragraph of the article to a short analysis meant to underline the understanding of quality that dominates the Romanian higher education system and place it on the 'map' developed in the first paragraph of the article. This analysis may contribute to potential future ARACIS strategies of refining educational program's quality.

Quality – Views from the Field of Theory and Practice

There are several levels (from abstract to concrete) of understanding an educational program's quality, but this is only an analytical separation as in practice they manifest themselves almost simultaneously. At the most abstract level we have the general vision over the educational process followed by more concrete versions of an educational program that suits this general vision, the program that has quality. This ideal program image is usually broken into indicators that we apply as a matrix to the program under evaluation. The term matrix suggests that the view of an ideal program shapes reality as real programs try to fit into those standards.

Coming back to the most abstract level, we deal with what Madaus and Kellaghan (2002, 20 – 25) named "metaphors". These metaphors are ideals that dominate society at a certain moment and they are not about a specific field, but influence all policies. These metaphors need our attention as their influence over specific fields of practice tends to go by unnoticed and unquestioned. Precisely, the two authors argue that, as long as society was driven by the ideas of industrialization, standardization, uniformity, precision, rationality and quantification, the dominant metaphor in education was that of the school as a "factory" meaning that what we usually expect from a factory was applied to the process of education. This was the era of Fredrick Taylor's scientific management who's rationalization and standardization of work studies are, according to Madaus and Stufflebeam (2002, 7), considered the first steps in personnel evaluation. Thus, given the adoption of instrumental rationality that characterised the society that I have just described, the educational process was dominated by a set of consistent ideas: the evolution of the educational process can be predicted, meaning that we can anticipate the level of performance that it will reach; we can control and manipulate the educational process, meaning that it is possible to correct problems through special techniques and a better knowledge of the causal relationships; teacher's activity can be evaluated in quantitative terms, etc.

Madaus and Kellaghan state that, as modern values were replaced by the post-modern ones, the metaphor of the school as a factory was replaced by the one of the school as a "journey". This metaphor is attached to a less rigid and 'scientific' vision of the educational process. It was considered to be unpredictable as students are not machines which can provide the same output for the same input. By consequence, the quality of the educational process as such is more important than the results that it produces. The educational process should be an enriching journey and the teacher a talented and experienced companion. According to today's developments in the field of evaluation, we should not make a choice between the metaphors and we should try to design evaluations so as to apply criteria that follow from both of them.

Next, I will present several understandings of the idea of quality that are implied in these metaphors of the educational process in order to explore the implications that they have on the way we evaluate the quality of the educational programs.

As stated in the introduction, I will follow the stages of development in evaluation theory and practice: the age of reform – before 1900; the age of efficiency and testing – 1900 – 1930; the Tylerian age – 1930 – 1945; the age of innocence – 1946 – 1957; the age of development – 1958 – 1972; the age of professionalism – 1973 – 1983; the age of expansion and integration[1] (Guba and Lincoln, 1989; Madaus and Stufflebeam, 2002; Stufflebeam and Shinkfield, 2007).

[1] This stage, which continues up to present, will not be discussed in this article because the focus now is manly on refining theory and practice than on theorising new fundaments.

Madaus and Stufflebeam (2002) state that the history of program evaluation starts in 1792 when William Farish invented the quantitative mark to score examinations. Thanks to this invention, the ranking of pupils or aggregating scores was now possible. This is a time when student's performance levels and the quality of the educational program were perceived as one single issue, and teacher's payment was related to these values.

The first formal attempt to evaluate schools' performance was in 1845 in Boston, according to Madaus and Stufflebeam (2002, 5 – 6). In order to be able to compare them, grammar schools in Boston had to introduce written examinations. Students' performance was presented as proof of the schools' quality and used to sustain the annual appointments of headmasters. This is one of the first moments in the history of evaluation when there was an awareness of the fact that political decisions get involved in the process of evaluation as the above mentioned procedures were in fact used to eliminate headmasters who opposed the abolition of corporal punishment. Fortunately, this was a case of good intentions. Besides having hidden political purposes, the Boston evaluation was criticised because the written examination consisted of an essay and that meant lack of standardization which lead to less reliable comparisons.

Between 1887 and 1898, Joseph Rice conducted the first formal educational program evaluation in the United States, thus becoming the father of comparative research design to study student achievement. Rice compared student's performance in spelling according to the time per week dedicated to the study of spelling and noticed that there was no difference between the ones that spend 200 minutes per week studying spelling and the ones studying only 10 minutes.

Another point of evolution in this period was the establishment of the first accrediting institution – North Central Association of Colleges and Secondary Schools, but the accreditation movement did not spread until the 1930s with the establishment of six more accrediting regional associations in the USA (Madaus and Stufflebeam, 2002, 6; Stufflebeam, 2002, 67).

It is interesting to point out that in this stage of development there was not a clear demarcation between a program's quality, students' performance levels and accreditation. As shown below, in later stages of development, accreditation became a distinct form of evaluation, among other forms of establishing the quality of a program and students' achievements is only one of the elements taken into consideration when we evaluate the quality of a program or we undertake a process of accreditation.[2]

Stufflebeam and Shinkfield (2007, 32 – 33) named the period between 1900 and 1930 the pre-Tylerian stage and state that the main characteristic is the standardization of existing practices. As the main practice consisted of measurement, Guba and Lincoln (1989, 22 – 26) stated that evaluation was nothing more than measurement at that time and that the right name for that period is the age of measurement.

During this stage, the idea of quality starts to develop on several other variables such as the level of spending or the rates of school drop out. But the main point of reference in establishing quality continued to be student's performance levels. This was due to the fact that the quantitative values which resulted from the measurement of the mentioned variables were not put together into an extensive analysis. By consequence, the data had no profound meaning attached to them; they explained nothing and that made them easy to use in political battles (Madaus and Stufflebeam, 2002, 7). Data were considered *a priori* unbiased, the mirror of the state of being under investigation, the right assessment

[2] The difference between accreditation and the evaluation of a program's quality is more difficult to perceive in the Romanian context where evaluations are undertaken only when there's an obligation in that respect. Indeed, accreditation is about establishing the quality of a program, but in the United States program quality is evaluated in several other contexts than that of accreditation. As accreditation is a sensitive issue, it is done by standard procedures that are agreed upon long before the accreditation of one program or another. But when programs are evaluated for other purposes, evaluators use several other methods, different from accreditation. This is due to the fact that the ones responsible with running the program really want to know what's good and what's bad in their programs so that they could better them. Their interest is way beyond accreditation. In this context, evaluations target certain elements of the program by methods rarely used in accreditation. That doesn't mean that accreditation is bad. On the contrary, it is useful in establishing minimum quality standards in a system, but is bad when it is the only type of evaluation that one can see in that system because it is not that profound that an evaluation can be if done properly. For a critique of accreditation practice, see Scriven (2002,"*Evaluation Ideologies*").

of the quality of a program. The many critics that led to the development of qualitative methods in social research and program evaluation were not highly developed at that time.

Ralph W. Tyler is recognised as the father of educational evaluation. According to him (2002), the main concern between 1930 and 1945 was the development of educational curricula as a set of experiences meant to help students reach certain behavioral outcomes that were specified in advance. This is the first time when educational programs are attached specific objectives and the quality of a program was determined by the comparison of intended and obtained results. This type of evaluation laid the foundations of today's outcome evaluation.

According to Madaus and Stufflebeam (2002, p. 9), Tyler's evaluations, unlike Rice's, were less costly because there was no need of data concerning other programs. Also, the fact that Tyler's method did not need to compare two or more programs to make a statement concerning one of them presented the advantage of overcoming the difficulty to compare student's achievements that are obtained in completely different systems. More so, Tyler's method diminished the possibility of disagreement over the results of evaluation as the program was evaluated according to its own standards and not according to a set of criteria imposed from the outside.

But there were several problems with regard to objective-based evaluation. One of the most important is the fact that unintended outcome could not be identified because the elements of the evaluation consisted only of the variables that resulted from the objectives. Unintended outcomes are very important in an evaluation, especially if they are bad and weight more than the intended ones. Also, we need to be aware of them because it might be the case that a program does not reach its objectives, but has valuable unintended outcomes. A second shortcoming in this type of evaluation is that the evaluator is not critical as to the desirability and value of objectives. S/he does not question the objectives, thus legitimising potential undesirable outcomes. Also, this type of evaluation can be applied only after the program has ended, meaning that it can not produce information to improve the program. Another issue is that if a program reached its objectives, this should not stop us from further investigations as we might find out that, with some few improvements, the results might be even better. Another reason is that comparing results with objectives tells us nothing about the implementation process, about the reasons why the objectives were reached; and that is a problem if we want to transfer the program. Another problem with not knowing the inside of a program is that, when it does not reach the objectives, we tend to throw it all out to the garbage, when, maybe, there's only one or two components of the program that are not working (Stufflebeam and Shinkfield, 2007, 8; Vedung, 2005, 43, 45 – 47; Chen, 1990, 18; Chen, 2005, 9). More so, the results of such evaluations can not be compared (Madaus and Stufflebeam, 2002, 9) and this makes ranking educational programs difficult.

Thus, understanding quality as the fulfillment of objectives proved to be rather problematic and, as shown below, the idea of quality expanded quite a lot since the 1960s.

The stage between 1946 and 1957 was characterised by expansion meaning that large sums were spent on education, there was a lot of innovation, but there were no analysis of the effects, costs and role of those innovations. Data were collected, achievement tests were improved, but there was no concern for the identification of the needs of educational programs' beneficiaries and for the extent that the programs responded to those needs (Madaus and Stufflebeam, 2002, 10).

It was during this stage that Lindquist (1953) settled the basis for experimental designs in educational evaluation, but the success of this type of design came in the 1960s.

The next stage, that between 1958 and 1972, does not abandon the developments of earlier periods, instead it refines them and puts them together in more complex approaches that are able to produce information about how to improve the quality of programs. The concerns that characterise this period were mainly the result of the social and political context at that time. The evaluation of social programs expanded a lot during the 1960s in the USA due to extensive social programs initiated under Kennedy and continued by Johnson and Nixon. As public expenditure on social programs grew, so was the concern (mainly of the Congress) for the control over these sums and over the results that their spending would produce. Pressure for evaluation also grew due to some corruption and fraud accusations and political factors (Vedung, 2005, 6; Bovens et al., 2006, 324).

Coming back to education, according to Madaus and Stufflebeam (2002, 12 – 14), there were a series of problems which, I think, claimed the redefinition of the idea of quality in education and, by consequence, of the way to evaluate it. Thus, given the War on Poverty context, the idea of Good Society that drove the above mentioned reforms, Robert Kennedy formed a group of initiative that managed to amend the Elementary and Secondary Education Act so as to compel schools to annually evaluate the improvement in disadvantaged children's achievements. Because the achievement tests were standard, designed for normal children, all schools reported bad result. From a series of studies that followed that period, it was concluded that educational programs were context-dependent when it came to producing a certain performance level (which continued to be largely understood as student's scores in achievement tests). If students were smart by nature, they would rank high, if not the program would not help them much. If students were rich, they would get easy access to education[3], if not it would be difficult for them to learn. There was a big problem for the field of evaluation during that time because it could not provide answers to such questions as: Did the challenged students have any progress in learning? How come that school programs do not overcome the natural and social condition of students?

The above mentioned shortcomings of evaluation at that time fostered a strong attention to developing new approaches to evaluation. Because the biggest need was to understand *why*, the new developments in evaluation theory and practice focused mainly on determining causal relationships and this meant a change of focus from the results to the internal components of a program.

The first major change that characterised this period is that evaluation undertook the role of indicating or selecting the best program (Wittrock and Lindström *apud* Vedung, 2005, 26). The idea of 'the best program' has also changed. Evaluations focus their attention on program beneficiaries' needs and that is why Scriven (2002) calls his approach "the consumerist ideology". But the direct beneficiaries were not the only ones taken into consideration; programs had to contribute to the good of the society (Shadish et al., 1991, 472 – 473). In this context, the official objectives of the program were not the equivalent of a good quality program and the evaluator had to investigate if the objectives of the program are in accordance with the needs of the direct beneficiaries and the society in general (Stufflebeam and Shinkfield, 2007, 367 – 370). So, the evaluator presented her/himself as an objective stance that can identify the needs of the target group and those of the society and uses those needs to judge the objectives of the program. Also, the evaluator would apply an experimental design in order to find out what in a program makes it good or bad. This would be a scientific method of finding out the good programs and so evaluation would be able to replace political judgement (see Vedung, 2005).

I will not insist here on the methodological critiques that were formulated in relation to this stage of development in evaluation practice, but I will mention that they were mainly about the shortcomings of the experimental design (that it can not determine causal relationships; it can not prove the validity of a theory; it is immoral, etc. – see Stufflebeam and Shinkfield, 2007, 175, 389 – 390; Shadish et al., 1991, 145) and about using predefined designs in the evaluation process (the evaluator can not know in advance the informational needs that the evaluation has to satisfy and that is why evaluation designs have to be established after the collection of some data on the evaluation, and it also has to be flexible and change according to unanticipated needs – Stake, 2002).

These methodological critiques were not the only ones that this type of evaluation had to deal with. The biggest problem was that the new developments in humanities, such as Kuhn's *The Structure of Scientific Revolutions* or Popper's earlier *The Logic of Scientific Discovery*, argued that there's no absolute objective observation or confirmation of a theory. By consequence, the evaluators could not possibly identify the needs of the beneficiaries or what is good for society and, even if they could, there was no way that we could be absolutely sure that a certain set of actions will lead to the desired state of affairs. This was the more difficult if we take into consideration that the experiment, so said the critiques, could not identify causal relationships (Vedung, 2005, 263; Shadish et al., 1991, 463 – 465).

[3] By access to education I do not mean only the process of enrolment, but the conditions to do their homework, to study and all sorts of other elements that, for instance in Romania, are left for the family to provide.

Briefly, the general conclusion was that a quality program can not be some entity described by the evaluator after s/he collected data on needs and desires of those affected by the program. Despite this conclusion, this generation's merits are not to be diminished. The approaches developed during this stage went a big step forward by highlighting the fact that there is no direct connection between students' performance levels and the quality of the educational program, that this relation is mediated by several factors and, by consequence, the quality of a program has to be redefined on more dimensions than learning outcomes. This is the period when the metaphor of the school as a factory started to be replaced by that of a journey and the educational program is defined in a more complex manner than as a simple act of teaching.

This new vision was materialised in 1974 when the *American Psychological Association* modified the *Standards for Educational and Psychological Tests* recognising the need to create separate standards for evaluating programs and students, thus coining the difference between students' performance and the quality of the program (Madaus and Stufflebeam, 2002, 10 – 11). This had the effect of recognising the fact that evaluation developed into a distinct science and there is a need to train specialists in evaluation. Educational evaluators were the vanguard of newly established professional associations, academic reviews, and universities such as University of Illinois, UCLA, Western Michigan University, Stanford University, University of Minnesota or Boston College developed graduate programs in evaluation.

So far, we can resume that evaluation theory and practice reached two big conclusions: quality is not necessarily obtained by the achievement of program's objectives, and quality cannot be defined by the evaluator alone and s/he cannot point to a specific program and state that it represented the very idea of quality or not.

Given those general conclusions, from the 1970s onward, evaluation embarked on a rather modest enterprise in the sense that, if it is not the source of the absolute right answers, it should try to create conditions to help programs perform better than before, and this can only be done with the help of those involved in the program, even if they are the beneficiaries, the ones that designed or implemented the program, or the ones that financed it. That means that evaluation generally ceased to start from predefined ideas of quality and tried to reveal this idea through the evaluation process. Simplifying dramatically decades of theory and practice, discovering the idea of quality through the evaluation process was generally done in three ways: the utilisation-focused evaluation way, the theory-driven evaluation way and the constructivist evaluation way.

The main purpose of utilisation-focused evaluation was to produce as much knowledge as possible about a program's results (intended and unintended), its components, or the effects that it has over the target group and society at large so that those involved in designing programs could use this knowledge to make them better (Weiss, 1999). Patton's (1997, 2002) version of utilisation focused evaluation focuses on revealing hidden values so that those involved in designing, implementing and financing the program could interpret the results of the evaluation according to their interest, to the values they promote through the program.

Chen's (1990, 2005) theory-driven evaluation determines the quality of a program by comparing it to the theory that describes what the ideal version of that program should look like. This ideal version of the program is established by the evaluator and the clients of the evaluation in the evaluation process. According to Chen, there are several options: either we have a scientific theory at hand, which is rarely the case, and the evaluator and the client agree that this theory should be the term of comparison, or this ideal theory is already formulated by the client and the evaluator checks on it by appealing to substantive knowledge, or the client does not have an explicit theory on how the program should work and, guided by the evaluator, they build it together in the evaluation process.

Constructivist evaluations (Guba and Lincoln, 1989; Lincoln and Guba, 1985; Lincoln, 2005; Stake, 1970, 2002, 2006, 2010) developed around the idea of conflict of interests, needs, values, interpretations, meanings, etc. and, coupled with the idea that none of them is better or worthier than the other, found methods to establish consent. As quality is not an object waiting to be discovered, it is what we agree it is, but that does not mean that all the agreements should be valued equally, only

those that are reached through a constructivist evaluation process. Basically, that means identifying those directly and indirectly affected by the program, collecting data on their views of the evaluation through qualitative methods (mainly interviews, narratives, etc.), writing extensive reports an these views, disseminating the reports and trying to establish consent through debates and negotiations.

If we were to shortly describe the general stand on the idea of quality that developed since the 1970s onward, we could state that there's a general agreement that there are several legitimate ways to describe quality and there are several legitimate actors to do that and that is why evaluation should identify and take into consideration as many versions and actors as possible, or a selection of them, depending on the approach. As regard to how we should do that, where do we look for several descriptions of quality; what do those descriptions consist of (numbers, targets, stories, ideals, needs, values, etc.) … for that we need the space of a book, and a really big one.[4]

In conclusion, quality is not a simple word; different people attach different meanings to it. By consequence, the different ideas of quality have distinct effects on the way we evaluate it in the sense that it leads us to investigate different aspects of the program and in different styles.

Now that we have established a more profound understanding of the idea of an educational program's quality, I will dedicate the next paragraph to a very short overview of the idea of quality that is implicit in ARACIS's methodology of program evaluation. As mentioned in the introduction, placing this idea on the 'map' that I have just drawn in this paragraph may be enriching in the sense that opportunities for future refinements of the methodology become easier to notice.

Before I go on and do this analysis, I have to add that, to my understanding, no idea of quality is *a priori* better than the other and it depends on the context. So the statements I might make about the ARACIS methodology are not judgemental; they are not about this methodology being right or wrong. My intention is to show if there is some type of understanding of quality that this methodology hasn't covered yet so that future line of development could be identified.

Quality – Views from ARACIS

My understanding on the manner that ARACIS establishes quality is based on the documentation that the Agency has posted on its website, the little involvement that I had in the process of institutional accreditation at my university and the article wrote by Florian and Hâncean (2009).

The activity of ARACIS is more than welcomed in the Romanian higher education system, but unfortunately it is the only body that deals with evaluating quality in a more complex and sophisticated manner.[5] The problem with ARACIS being the only one dealing with issues of quality is that it is an accreditation institution and, as mentioned before, accreditation and quality assurance are not at all the same thing (see also Florian and Hâncean, 2009). Accreditation compels you to deal with indicators that are general enough to fit all universities and, for reasons of transparency and equal treatment, one cannot adapt this methodology according to the situation at hand. That means that, even if the evaluator feels that there is something wrong about a program or something extremely good that needs further investigation, s/he cannot alter the standards of evaluation to deepen the investigation. More so, ARACIS needs to be very synthetic and work mainly with quantitative data because it needs to able to function with a lot of information. To be brief, the context that ARACIS works in gives it no opportunity for deep analysis, for keeping a detailed track of all the specificities in an educational program.

The problem with this type of evaluations is that it can be easily misleading. Working with really precise indicators means that universities will try their best to score high on those indicators

[4] See Henry, 2002; Fischer, 1980; Lindblom, 1959; Guba and Lincoln, 1989; Stronbach and MacLure, 1997; House and Howe, 1999; Stake, 1970, 2002; Scriven, 2002, 1994, 1998; Shadish et al., 1991; Stufflebeam and Shinkfield, 2007; Vedung, 2005; Chelimsky, 1999; Stake and Migotsky, 1997, to name just a few.

[5] I do not forget CNFIS or CNCSIS, but I will not take them into consideration as those methodologies are rather clear-cut, based exclusively on quantitative measures of outputs. In this case the analysis would be rather simple, with all the shortcomings just shown in the first paragraph of the article when I presented the age of measurement in evaluation.

and ignore other aspects. ARACIS had a good effect by imposing minimum standards, but the problem is that universities do not try their best to do better and I think that mixing the idea of quality assurance with that of accreditation is part of the problem.

Otherwise, ARACIS has very modern procedures: it tries to identify several types of actors (students, teachers, alumni, employers, evaluators form outside the country, etc.); it has several procedures for collecting qualitative data (as meetings and discussions with students and employers); the university presents it's point of view through auto-evaluative documents; it tries to see benefits for society that the program might bring (for instance, the university presents the number of students that got hired), the university's relation with the environment. But if we look at the way indicators are formulated and proportioned, we could easily see that these practices are secondary, with rules, regulations, financial management, complement to the laws, numbers, etc. being the most important. For instance, one element in the evaluation methodology is university's operational and strategic plans. But there is no analysis of those documents: do they have desirable objectives; are they realistic; are they fulfilled? What matters in the process is if the documents exist or not.

Once again, ARACIS has a very good accrediting methodology, but is just that. It is not a quality assurance methodology. An educational program can score high within this methodology and still be a bad program, especially as regards to student's interests. A quality assurance methodology would deeply analyse the objectives that a university or an educational program has; would try to estimate the impact of those objectives on society; would try to identify potential groups to be affected by the fulfilment of those objectives and register their opinion on that; would compare them to the objectives of other programs. Also, a quality assurance methodology would try to find out if the official objectives of the program can be reached in better ways, if the process could have more efficacity, etc.

What can ARACIS do? Certainly the answer is not that of modifying its existing accrediting methodology. This would only lead to lack of precise standards, to subjectivism, lack of transparency, etc. The need is to develop a methodology that is more client-responsive, that can be used for quality assurance purposes, not for accrediting purposes. This quality assurance methodology would help educational programs identify opportunities for development, for raising standards. This would be a methodology that the ones responsible for the management of an educational program would appeal to with no pressure of being judged in terms of right or wrong (as is the case with accrediting procedures). This would be a methodology flexible enough as to investigate what the ones involved in the program need to investigate, not what ARACIS needs to investigate, as is the case with accreditation. It would be a methodology that would make ARACIS a companion in the journey of development, not a harsh teacher that lets you do the homework and then just checks on you.

Florian and Hâncean (2009) present the example of other countries, especially the UK, where accrediting and quality assurance institutions are separate. To my understanding, in the Romanian context this is not possible. Our little experience with decentralization in the administrative/ governmental structures showed that it tends to lead to corruption and the creation of pressures for lowering standards. There is no reason to think that education would make an exception. I think that if we create a different institution, dedicated to quality assurance, it will be used by universities to put pressure on ARACIS to lower its accrediting standards. Theoretically speaking, there is no reason to do that because the quality assurance institution would produce information for improving the educational program, not overall judgements about the value of the program. The results of those two types of evaluation (quality assurance and accreditation) are, theoretically, not comparable. But politically … I am afraid that they will be comparable.

In conclusion, my suggestion is that ARACIS should think of establishing a distinct department within its structure that deals exclusively with quality assurance. This department should work more like a consultancy body that universities and educational programs particularly could come to seek advice for development. For instance, new programs or universities that are not accredited could buy the services of this consultancy body in order to prepare themselves for the accreditation process.

Also, accredited universities could use this quality assurance consultancy body in order to improve their activities and score high on CNFIS indicators. Some other possibilities of using this body could be: improve the attractiveness of its students in the labour market; the internationalisation of its products (articles, books, research results, students, etc.); assistance to revising existing development strategies; management of resources, etc.

References

Bovens, Mark et al. (2006), The Politics of Policy Evaluation, in Moran, Michael; Rein, Martin şi Goodin, Robert E. (eds.), *The Oxford Handbook of Public Policy*, Oxford University Press, New York, 319 – 335;

Chelimsky, Eleanor (1998), The role of experience in formulating theories of evaluation practice, in *American Journal of Evaluation*, Vol. 19(1);

Chen, Huey-Tsyh (1990), Theory-Driven Evaluations, Sage, Newbury Park;

Chen, Huey-Tsyh (2005), *Practical Program Evaluation: Assessing and Improving Planning, Implementation and Effectiveness*, SAGE, Thousand Oaks;

Fischer, Frank (1980): *Politics, Values, and Public Policy. The Problem of Methodology*, Westview Press, Boulder;

Florian, Bogdan şi Hâncean, Marian-Gabriel (2009), Metode şi instrumente de evaluare a calitcii în învcmântul superior, în *Evaluarea programelor şi politicilor publice. Teorii, metode şi practici*, Cerkez, Mirela (coord.), Polirom, Iaşi, 83 – 114;

Guba, Egon G. and Lincoln, Yvonna S. (1989), *Fourth Generation Evaluation*, Sage, Newbury Park;

Henry, Gary T. (2002), Choosing Criteria to Judge Program Success. A values Inquiry, in *Evaluation*, Vol. 8(2), 182-204;

Hirschon Weiss, Carol (1999), The Interface between Evaluation and Public Policy, in *Evaluation*, Vol. 5(4), 468-486;

House, Ernest R. and Howe, Kenneth R. (1999), *Values in evaluation and social research*, Sage Publications, Thousand Oaks;

Kuhn, Thomas S. (1976), *Structura revoluţiilor ştiinţifice*, Editura Ştiinţifică şi Enciclopedică, Bucureşti;

Lincoln, Yvonna S. (2005), Fourth-generation evaluation, in Mathison, S. (ed.), *Enciclopedia of evaluation*, Sage, Thousand Oaks, 161 – 164;

Lincoln, Yvonna S. and Guba, Egon G. (1985), *Naturalistic Inquiry*, Sage, Beverly Hills;

Lindblom, Charles E. (1959), The Science of Muddling Through, in *Public Administration Review*, Vol. 19(2), 79 – 88;

Lindquist, E. F. (1953), *Design and analysis of experiments in psychology and education*, Houghton-Mifflin, Boston;

Madaus, George F. and Kellaghan, Thomas (2002), Models, Methaphors, and definitions in evaluation, in Stufflebeam, Daniel L. , Madaus, George and Kellaghan, Thomas (eds.), Evaluation models. Viewpoints on educational and human services evaluation. Second Edition, Kluwer Academic Publishers, New York, 19 – 31;

Madaus, George F. and Stufflebeam, Daniel L. (2002), Program evaluation: a historical overview, in Stufflebeam, Daniel L. , Madaus, George and Kellaghan, Thomas (eds.),

Evaluation models. Viewpoints on educational and human services evaluation. Second Edition, Kluwer Academic Publishers, New York, 3 – 18;

Patton, Michael Quinn (1997), *Utilization-Focused Evaluation: The New Century Text.* Third edition, Sage Publications, Thousand Oaks;

Patton, Michael Quinn (2002), Utilization-Focused Evaluation, in Stufflebeam, Daniel L. , Madaus, George and Kellaghan, Thomas (eds.), *Evaluation models. Viewpoints on educational and human services evaluation.* Second Edition, Kluwer Academic Publishers, New York, 422-438;

Popper, Karl R. (1981), *Logica cercetrii*, Editura Științifică și Enciclopedică, București;

Scriven, Michael (1994), The Final Synthesis, in *Evaluation Practice*, Vol. 15, 367-382;

Scriven, Michael (1998), Minimalist theory: The least theory that practice requires, in *American Journal of Evaluation*, Vol. 19(1);

Scriven, Michael (2002), Evaluation Ideologies, in Stufflebeam, Daniel L. , Madaus, George and Kellaghan, Thomas (eds.), *Evaluation models. Viewpoints on educational and human services evaluation.* Second Edition, Kluwer Academic Publishers, New York, 249 – 278;

Shadish, William R. jr.; Cook, Thomas D. and Leviton, Laura C. (1991), *Foundations of Program Evaluation. Theories of Practice*, Sage, Newbury Park;

Stake, Robert E. (1970), Objectives, Priorities and other Judgment Data, in *Review of Educational Research*, Vol. 40(2), 181 – 212;

Stake, Robert E. (2002), Program evaluation, particularly responsive evaluation, in Stufflebeam, Daniel L. , Madaus, George and Kellaghan, Thomas (eds.), *Evaluation models. Viewpoints on educational and human services evaluation.* Second Edition, Kluwer Academic Publishers, New York, 338 – 362;

Stake, Robert E. (2006), *Multiple Case Study Analysis*, The Guilford Press, New York;

Stake, Robert E. (2010), *Qualitative Research: Studying How Things Work*, The Guilford Press, New York;

Stake, Robert E. and Migotsky, Christopher (1997), The evolving syntheses of program value, in *Evaluation Practice*, Vol. 18(2);

Stronach, Ian and MacLure, Maggie (1997), *Educational research undone: The postmodern embrace*, Open University Press, Philadelphia;

Stufflebeam, Daniel L. (2002), Foundational models for 21st century program evaluation, in Stufflebeam, Daniel L. , Madaus, George și Kellaghan, Thomas (eds.), *Evaluation models. Viewpoints on educational and human services evaluation.* Second Edition, Kluwer Academic Publishers, New York, 33-86;

Stufflebeam, Daniel L. and Shinkfield, Anthony J. (2007), *Evaluation Theory, Models, and Applications*, Jossey-Bass, San Francisco;

Tyler, Ralph W. (2002), A Rationale for Program Evaluation, in Stufflebeam, Daniel L. , Madaus, George and Kellaghan, Thomas (eds.), *Evaluation models. Viewpoints on educational and human services evaluation*. Second Edition, Kluwer Academic Publishers, New York, 87 – 96;

Vedung, Evert (2005), *Public Policy and Program Evaluation*, Transaction Publishers, New Brunswick.

Online resources:

www.aracis.ro

www.cnfis.ro

www.cncsis.ro