The Evaluation of Adult Education Staff

EDUEVAL Handbook

Extended Version

edited by
EDUEVAL Consortium

3.
The evaluation of adult education staff
by Loredana Perla and Viviana Vinci[10]
University of Bari Aldo Moro, Italy

In order to contextualize the EDUEVAL model of evaluation of adult education staff, the plurality of theoretical approaches underpinning the *evaluation object*, conceptualized as a powerful regulator of the functioning of systems, should be understood first of all. It is based on the collection and scientific interpretation of data and oriented at improving the processes and products of a system.

To be extremely concise, the international debate on evaluation will be referred to, starting from a tripartite pattern (Stame, 2001) which groups together evaluation studies in three main approaches, describing, for each approach, both the main meanings and models of evaluation that emerge and how the evaluation of adult staff is (or is not) considered. The intention, taking this tripartite model as reference, is to understand and refer to those models of evaluation, the characteristics of which appear more coherent with the requirements of the evaluation of AE staff.

---

10 Loredana Perla is the author of sub-sections 3.1 and 3.1.1; Viviana Vinci is the author of sub-sections 3.1.2, 3.1.3 and 3.1.4.

3.1 The theoretical framework : theories and models of evaluation

The meanings and functions of the act of evaluation are multiple and fundamentally vary between two poles, *measurement* and *evaluation*, from which different approaches stem with different considerations of evaluation, as "measurement", "estimate", "appreciation", "comprehension", and which refer, with a different importance, to criteria such as determining the results obtained and the efficacy, efficiency and the performance of the object being evaluated. The three main approaches of evaluation, focused differently on one or more dimensions of those described, are summarized below in graphic form (in a table and a figure).

Table The approaches to evaluation (Stame, 2001)

	Positivist-experimental	Pragmatist-quality	Constructivist
Benchmark	The objectives	The standards	What the *stakeholders* define "success"
Authors	Hyman, Suchman, Campbell, Rossi and Freeman, Chen	Scriven, Wholey, Donabedian, NPM (New Public Management) tradition	Stake, Stufflebeam, Guba and Lincoln, Cronbach, Patton, Fetterman, Hirshman, Tendler
Questions	Do the results correspond to the objectives?	Do the results correspond to the criterion of quality?	What happened? Is what happened good?
Direction of the investigation	*Top down*	*Top down*	*Bottom up*
Attitude towards values	Relativism: the values are those of the programme	The evaluator judges with respect to the values (his own or of the existing concept of quality)	The values are those of the *stakeholders*: at times they agree, at other times they are conflicting
Theory	With good planning all the effects can be foreseen	There is a concept of quality to aspire to in every situation	Reality is richer than can be foreseen; the importance of unexpected events

Main method of investigation	Experiments and quasi-experiments	Scriven's "logic of evaluation"; multicriteria analysis	Comparative analysis; exploration; participated analysis
Techniques	Survey	Analysis of user satisfaction; opinions of the experts	Case studies; interviews, focus groups, observations
When and where it is normally applied	In programmes; in European Structural Funds; wherever there are objectives with respect to which it is possible to identify means and results (social and work policies etc.)	In training and education institutions for adults; in cultural and literacy centres; in services (health, education etc.); in university evaluation; in charters of services (standards of quality); in programmes of public sector reform	In innovative situations; in pilot projects etc.
Area of use	Instrumental for political decision	Instrumental for the management and functioning of the administration	Fact-finding; *empowerment*
Theoretical problems	The black box: why should there be this result?	What is quality? How are values formed?	Where to start?
Problems of empirical research	The objectives are not clear: there is no data	How are standards of quality fixed?	Where to look?
Answers- Solutions	Analysis of evaluability; taxonomic evaluation, conceptual maps, evaluation based on theory: Weiss, Toulemonde	If there are no standards of quality, use those from a comparison with other situations or with one's own past. Involve the users in defining quality	One thing leads to another; the reflective practice of the evaluator
Advantages	It helps to plan better	It helps for good management	There is something to learn for all the stakeholders

3.1.1 The *positivist-experimental* approach

In the *positivist-experimental* approach, evaluation is under–stood as the analysis and verification of the attainment of pre-established objectives. Alongside methodological rigour and therefore validity and reliability, the coherence, pertinence and neutrality of the evaluator are important in the models with this approach. Particular emphasis is given to *measurement*, the *quantitative* dimension[11]. The conditions necessary

---

11 In practice, this approach coincides with the logic of examinations and tests.

for an evaluation understood as "measurement" are very careful planning of the objectives – including classified taxonomically in terms of observable behaviour – and reliable tools to analyse the expected results. The resulting evaluation model is of a *rationalist* type (Galliani, 2014, p. 28), in which evaluation is associated with the ability to foresee – owing to clear planning of objectives – not only the outcomes of the training process but also of the possible changes/improvements. This approach is affected by a certain methodological rigidity and is not always able to reconcile *grey area* variables.

This approach includes *measurement* models and *goal-oriented* models, which have been applied almost exclusively in scholastic contexts. Some procedures and tools (questionnaire) have also been borrowed from the pragmatist-quality approach and then applied to the evaluation of educational actions.

"Measurement" models (authors: Galton, Wundt, Binet & Simon): according to these models, evaluation and measurement are synonymous. The evaluator takes on the role of a technician who has to suitably record, with procedures that are mostly quantitative, all the variables necessary for the evaluation. These are models which are can be situated at the dawn of docimology (dating back to the 1950s and 60s) but still have extensive influence, in the form of procedures which use psychometric techniques of measuring performances, objective tests of profit and in general all the tests that aim at reducing the subjectivity of the evaluator. This traditional concept of evaluation (evaluate the product, measure and select) was not to be challenged until the 1970s, but it is nevertheless still used, especially in schools[12]. The theoretical frameworks, which form a background to these models, are those of studies on psychometrics and experimental pedagogy.

12 For example the international PISA objective tests.

Goal-oriented models (author: R. Tyler): This focuses on the comparison between goals and results (with the adoption of procedures of checking in progress). Resumed by B.S. Bloom (1956), who formulated a taxonomy of the cognitive behaviour expected from the learning subject, this model understands evaluation as *measurement through tests of difference* (between expected goals and behaviour). Initiated in the United States with the precise aim of revising school curricula and identifying ways of evaluation that could play down spontaneous behaviour, which had emerged in the innovative school experiments of the New Schools movement, it has been applied in the training context only with reference to the procedures of instructional evaluation of formal learning skills.

3.1.2 The *pragmatist of quality* approach

The *pragmatist of quality* approach, on the other hand, stresses the dimension of the comparison and definition of standards and criteria, conceiving of educational evaluation "as *management of the organizational procedures* to guarantee attaining the training standards defined inside or outside the system" (Galliani, 2014, p. 31). The implied evaluation model is of a *functionalist* type, in which evaluation takes on a supporting role for the *decision-makers* and meets the external requests of the *stakeholders*.

In the models that can be ascribed to this approach, particular significance is given to the opinion (and therefore to the "voice" of the different players involved in the evaluation process). The risk of these models is that of self-referentialism and being anchored to indicators established only in the system where the evaluation takes place.

This approach includes various models, such as *Company-Wide Quality Control*, *Goal-free evaluation*, the CAF model and the Servqual model.

Company-Wide Quality Control Model: this came into being in the United States in the first half of the last century and was perfected in Japan in the corporate context. It has the aim of identifying a managerial strategy and a culture that can sustain competition in quality and through quality. *Company-Wide Quality Control*, by shifting the axis of evaluation from the out-comes to the processes (of transforming incoming resources into "products"), is characterized by the central role assigned to some factors deemed essential to achieve quality: the context the company belongs to (with its system of relations and peo-ple); the definition and the clear attribution of roles and tasks to the different players in the system; agreement on the form-ative aims by all the subjects involved in the process; institu-tionalized innovation; a system rewarding personal commitment; constant attention to satisfying the user, i.e. *cus-tomer satisfaction* and the production of data for both internal and external control. By retrieving the whole contribution of the comparative and measuring tradition of evaluation (Tyler, Provus, Fisher), for the first time this model focuses attention on the "black-box" of the training path, i.e. on the black box of actions between the objectives and the outcomes of the path. Quality is divided into *expected quality*, including all the needs expressed by the client; *designed quality*, i.e. all the processes and outcomes expected in relation to the expecta-tions of the client and the organization; *produced quality,* i.e. all the characteristics of the product and of the service delivered; *perceived quality*, i.e. all the user's representations about the serv-ice or the product delivered. In its radical form, that expressed by T.Q.C. (*Total quality control*) or T.Q. (Total Quality) control, the model defines "negative quality" as a *non-conformity* or, more in general, any negative shift of the performance of the product or of the service with respect to the expectations of the users, whilst any positive shift with respect to these expec-tations is defined "positive quality". "Competition" is thus played out around the minimization of "negative" quality" until

it is reduced to "zero" and the maximization of the "positive quality", especially that perceived by the user. The *process quality* is considered as something more than the *product quality*: the former is defined the "means" to obtain the result, the latter is only "one of the results" (efficacy), as the other results are *improved efficiency* (minimum costs and time of production) and the *increased flexibility* (ability to adapt rapidly to change). Alongside these models, there is one which was given great emphasis, particularly in the early 1990s: that of "Quality defined by standards"[13], and which was by some considered irreconcilable with the previous ones. The standard was developed to protect the interest of the "client" and – the sole point of contact of this model with the others – the culture and the practice of quality are seen from the user's perspective. Unlike the other models, however, which theorize an idea of quality inspired by *proactivity*, i.e. aiming for continuous improvement, the model of "Quality defined by standards" theorizes an idea of quality linked to the need for *control* and *certification of conformity with predefined standards.*

Evaluation in these models is therefore of two types:

1. *Evaluation for the purpose of improvement (or self-evaluation)*
2. *Evaluation for the purpose of control*

Through *self-evaluation,* which represents the central element of the evaluative procedure, it is the company (not a

---

13 The family of "standards" for quality came into being in the military field with the aim of guaranteeing the quality of the supplies and was subsequently extended to the space, nuclear and energy fields and in general to the vast area of purchase transactions regulated by a contract. In the extensive history of standards, three different generations corresponding to three different approaches to the subject of the quality of organizational systems can be identified (Perla, 2004).

client or an external body) that intends to evaluate itself. The term self-evaluation, however, must not give rise to misinterpretations: the company must not express opinions of value on itself, but make users and players speak and, above all, *be able to listen*.

Evaluation for the purpose of control, on the other hand, is of the "external" type and, whilst having as its objective the improvement of quality, acts through *accreditation* and *certification*, promoting comparison with a system of standards and competition between different subjects through *ranking* (i.e. the different positioning in a classification of merit). Usually, the two types of evaluation reach *certification* and *accreditation*[14]. In this way, evaluation provides an explicit and public recognition of the quality levels ascertained (which can consist of, for example, granting funding or admission to a public tender). The evaluation model, in this case, has the task of predetermining the requisites unanimously deemed necessary (*agreed standards*) to guarantee the desired levels of quality, which companies must necessarily possess in order to be accredited.

All the *standards* taken as a whole represent the "minimum threshold" of accreditation, which the companies can supersede but never disregard, on pain of losing their accreditation.

There are three essential aspects of a system of certification and accreditation: the presence of standards of quality/quantity (*agreed standards*); *cyclical evaluation* and the *published statement*.

Wishing to summarize, whilst improvement evaluation is

14 A definition of accreditation that is exemplary in its clarity is taken from the CRE-Lisbon document: *Accreditation is a formal, published statement regarding the quality of an institution or a programme, following a cyclical evaluation based on agreed standards.* Source Ministry of Education and Universities, Final report by the "Accreditation of courses of study" working group, June 2001, p. 1.

characterized as a dynamic process, as it tends to analyse/point out the weak and strong points of a process, in order to identify their causes and provide indications in the direction of change, certification and accreditation are "static" procedures as their objective is to state the existence of a conformity with respect to a standard. This model has been very successful and is widely used in the evaluation of organizational contexts of public and private companies, schools and universities and social education.

As an evaluation framework for AE staff, the model presents both strengths and weaknesses.

The *strong points* include: an interpretation of quality as a dynamic construct and not as a universal parameter (the culture of quality gives priority to the design/evaluation more than to the control; to the process–efforts relationship rather than to that of objectives–results; it is built up in time and evolves according to the personal and material investments that an organization can offer); the attention to the users of the service/product; the importance assigned to the processes and to working out a method which makes improvement faster; making the most of the "human capital" and personal commitment and qualities; incentives on results; investment in innovation; the organization as a strategic variable of improvement (educational management).

The *weak points* include: the concrete risk of a mechanistic interpretation of educational work: the quality system was created for companies and has to be adapted with critical intelligence to the context of educational relations, processes and products; an economistic vision of quality, which leads to excessive dependence on the "third pole", i.e. the *market* and which is based on the principle of competition and on the continuous improvement of the service/product in order to "dominate" the latter; the interpretation of quality as a "competitive strategic variable" and not as a value in itself; "complexity": the efficacy of evaluation lies in succeeding in

involving all the members of the AE staff but also, contextually, the other variables of the context, including the stakeholders.

Goal-free evaluation model (author M. Scriven 1967, 1991, 2007): the syntagm "goal-free evaluation" is an extreme position where it is better for the external evaluator not to know the goals of the project and the objectives of the planning, in order not to narrow his vision and to better investigate the effective outcomes of the process, rather than those planned. The investigation concentrates on the activities carried out, on the effects produced and on the overall impact of the programme, evaluated not in relation to the objectives of the project but according to their relevance and significance from a social point of view. For Scriven, evaluation is understood as attributing value to a programme, project, product, performance: it is the process that can determine "merit" (the quality of the object evaluated in accordance with the specific standards of the sector concerned), "worth" (i.e. the value connected with the benefits of the sector concerned) and "value" (i.e. the coherence between the intervention carried out and the needs that originated it) (Torre, 2010, p. 14). Scriven states the importance of global evaluation focused on the multi-dimensional nature of the evaluee.

There are several aspects worthy of note in his model and we will refer to them by points.

– Having identified two of the most important functions of evaluation, initially coined by Scriven in relation to the "social" evaluation of a programme of development and then successfully "imported" them into the field of educational/ training action: the *formative* function and the *summative* function of evaluation. In Scriven's original interpretation, the former, i.e. the *formative* function, has to be understood as support and progressive (and improving)

construction of a project of development that is increasingly suitable for the need of the user. The *summative* function, on the other hand, is exactly the sum of the effects that the programme has produced in the direct and indirect consumers. It is an operation that certifies the impact of the programme on the user.

− Having identified as a central construct of the evaluation not so much the decision but the *judgment of value*, at the basis of which there has to be the analysis of the needs and of the values of the users (whom he defines "consumers" of the educational service).

− Having assigned greater importance to the addressee of the educational actions and to the *effects* of the evaluation on the latter. The evaluator is not so interested in observing the points of view of the subjects involved in the process, or the intentions of the project, but the impact of the latter on the addressees. The focus on the evaluation system is on the *user* of the educational–didactic offer, on the *consequences* that the implementation of the didactic project produces and on the *processes,* implemented to meet the demand of the "consumers" of education.

− Having released evaluation from the objectives of the planning. The evaluator, for Scriven, should not even know what the objectives of a project are, as his evaluation should be oriented towards investigating the effective outcomes of the process, not the planned ones: knowing the latter, on the contrary, would produce only a dangerous conditioning on the evaluator. This is why Scriven's proposal has been defined *goal-free evaluation*.

− Having increased the objectivity of the evaluation investigation: the evaluator is not included in the field of observation and should keep his distances from the evaluee, estranging himself from the context of observation, reducing to a minimum contacts with those in charge of the

project, in order to reduce that involvement in the process which would end up by nullifying his "impartiality".

– Having been amongst the first to draw up a system of *check-lists* to evaluate the impact of the product on the consumer. Scriven recommends the use of check-lists to succeed in reaching a judgement that is concretely analytical and comprehensive of all the dimensions evaluated. An adapted example is shown as follows:

Analysis of the needs: in order to show that the educational product *being evaluated* will contribute to improving the system
Analysis of the resources available to evaluate the product
Analysis of the consumer: who is he and what does he expect from the educational offer?
Analysis of the product: is it suitable for meeting the consumer's need?
Analysis of the process: checking the educational process complies with the standards established in response to the demand and referred to the product to be evaluated
Compared analysis of the performances of competitive products
Analysis of the effects: goal-free evaluation of the *outcomes*
Analysis of the cause-effect relations: to show that the effects derive directly from the use of the product
Analysis of statistical significance
Analysis of the educational significance: which improvements does the product make to the educational system?
Analysis of costs: estimate of the economic and non-economic costs and comparison with similar educational products
Analysis of the maintenance supports of the product: how to maintain the need for education? How to control the potential of disappointment of the user?
Analysis of the forms of report: relative to the communication of the results of the evaluation

The *Key Evaluation Checklist by M. Scriven* (2007) indicates different criteria including: background and context, detailed description of the programme, effects of the programme on the direct and indirect users, financial and social resources,

values analysed, positive and negative results, costs, comparisons with other models, generalizability, needs of the stakeholders. Scriven speaks of *meta-evaluation* and its criteria: validity, utility, credibility, efficacy linked to costs, legality, attention to ethical rules, protection of the human rights of the subjects.

Scriven's model supports an idea of evaluation understood as certification of product quality that should meet precise specifications, more than as an "internal" process of constructing the evaluation judgement. Where the pragmatist models of quality focus the evaluator's attention on the process and on the decision, here the evaluation object is only the product and its ability to correspond to the needs of a hypothetical user. The criterion for the choice of the evaluation standards, i.e. the rules for determining the "success" or "failure" of a pathway or of a programme, is not given in this case by the objectives, but by the needs and the expectations of the users. As a consequence, the evaluation structure significantly highlights the time of checking and observing the judgements of value of the stakeholders, capable of contributing to triggering off improvement dynamics of the quality of the process/product offered. Operations such as the detailed and specific descriptive analysis of all the components of the product, the estimate of the effects of impact and the comparison with similar products, therefore take on great importance. The person of the evaluator, the only one who can manage the evaluation procedure and inform on the outcomes of the operations, takes on great power.

Scriven would seem to follow an epistemological model of the positivist type, but this is not the case. This atypical position is probably the cause of all the methodological ambiguities that have been attributed to him. Scriven refuses experimental approaches but also phenomenological ones; he is not interested in knowing the intentions of the project or the opinions of those who are involved in it. Only the evaluator has the credit

– and the power – to succeed in observing reality and developing a valid and reliable judgement of evaluation. Is this possible? According to many critics of Scriven, it is not. If the reliability of evaluation must not be conditioned by the objectives, it can, however, be influenced by the subjectivity of the evaluator. "The evaluator possesses in any case his own ideas and opinions on which he constructs hypotheses and conjectures; he has personal ways of reading and interpreting reality and it is on this basis that he organizes information. In other words, if it is not the objectives of a project that condition the fidelity of an evaluation, then it can be the preconceived ideas of the evaluator" (Tessaro, 2000, p. 74). The problem exists and Scriven does not offer any answers for clarification. He is not able to go beyond a perspective of an evaluation approach understood as a mere "service" to the user (and conversely to the "producer" who has every interest in "selling" a product). It is curious that the scholar of evaluation who has gone down in history as the father of *formative evaluation* actually conceives of evaluation in a sense that is not fully "formative", at least according to the meaning that today tends to be given to the term; i.e. constant verification of the educational pathway, in order to accompany the student in a personalized way and orient their subsequent developments. If Scriven's formative evaluation has as a reference the product and is carried out in the course of doing a programme, in order to acquire information useful for improving the programme whilst it is still under way, today formative evaluation coincides with the same didactic process that assumes as the privileged reference the person being education in order to support, promote and guide them. If in Scriven, formative evaluation "lives outside" the educational process, today it tends to be included in the process, of which it becomes the self-regulatory dimension par excellence.

It is precisely the *self-regulatory dimension,* in the evaluation processes of AE staff that is the most important legacy today that can be borrowed from Scriven's model.

CAF model - Common Assessment Framework: the CAF is the result of the collaboration between national experts of the EU and the European Network of Public Administration (EUPAN), when they met at the 1st European Conference on Quality in Lisbon, in May 2000. It represents a real common European tool for quality management for the public sector and was developed by the public sector itself. Over the last 10 years, almost 2000 organizations of the public sector throughout Europe have adopted this model and the number of CAF users is still growing. In the first years, the model was mainly used to introduce the principles of *Total Quality Management* (TQM) into the public sector and it was gradually adapted and personalized in various sectors, such as justice, local administration and education. As a tool of Total Quality Management, the CAF is inspired by the EFQM mode of excellence of the *European Foundation for Quality Management* (EFQM) and the *Speyer* model of the German University of Administrative Sciences. The CAF is based on the principle that excellent results relative to organizational performance, citizens/clients, personnel and the company are obtained through a leadership that guides policies and strategies, personnel management, partnerships, resources and processes. The CAF considers the organization from various points of view simultaneously, according to the holistic approach of analysis of organizational performances. A pilot version was presented in May 2000 and a first revised version was launched in 2002. On the decision of the DGs, a CAF Resource Centre (CAF RC) was established in Maastricht, at the EIPA (European Institute of Public Administration). With a strategic perspective, the EIPA indicated its intended role and objectives as CAF RC. The CAF has been designed to be used in any sector of the public administration, at all levels: national, regional and local. It can be used, depending on the circumstances, both as part of a systematic programme of reforms and as a base to direct the improvements actions in individual public organizations.

The CAF has four main aims: to introduce the public administrations to the principles of TQM and guide them progressively, through using and understanding the process of self-evaluation, from the current Plan–Do sequence to the fully integrated Plan–Do–Check–Act sequence; to facilitate the self-evaluation of a public organization in order to obtain a diagnosis and undertake actions of improvement; to act as a bridge between the various models in use for quality management; to facilitate *bench-learning* between the organizations of the public sector.

Various elements have been studied in depth in support of these aims; the structure with 9 criteria, the 28 sub-criteria with the examples, the diagrams for the evaluation of the qualifying factors and the results, the guidelines for self-evaluation, the improvement actions and the projects of *bench-learning* and a glossary. The structure with nine criteria describes the main aspects that have to be taken into consideration in the analysis of any organization: the qualifying factors of an organization (what the organization does and the approach used to achieve the pre-established results) and the results obtained in relation to the citizens/clients, the personnel, the company and the key performances, through measures of perception and indicators of functioning. Each criterion is divided into sub-criteria, which identify the main dimensions that have to be considered when an organization is evaluated. They are illustrated with examples, which explain their contents in detail and suggest the possible areas to take into consideration, to explore how the organization meets the requisites expressed in the subcriteria. As the CAF is a tool that is suitable for all the areas of the public administration, its personalization for the different sectors is encouraged, as long as its constitutive elements are respected: the 9 criteria, the 28 sub-criteria and the score system. The examples and the process of self-evaluation, as described in the guidelines, are flexible, but the key passages of the guidelines should be

taken into consideration in order to keep an important function of the model unchanged: to promote a common culture among the European public organizations, acting as a bridge between the various TQM models and fostering *bench-learning*. The CAF model is widely used in the evaluation of organizational contexts, including educational.

Servqual model (authors Parasuraman, Zeithaml, Berry, 1985): also known as the model of "gaps", measures in a standardized way the opinion of clients and the expectations of the users in relation to the quality of the services. Structured in 22 questions structured into two repeated groups, concerning respectively the expectations of the users on the service and the opinion on the various aspects of the service (or distributed in compact form in a unique series of questions), the Servqual model allows measuring the perceived quality and the expectations separately, narrowed down to 5 dimensions deemed indispensable to judge the quality of the service:

1. Tangible elements (aspect of the physical structures, equipment and personnel)
2. Reliability (ability to dispense the service promised reliably and accurately)
3. Ability of response (willingness to help clients and promptly provide the service)
4. Ability of reassurance (competence and courtesy of employees and ability to inspire trust and security)
5. Empathy (caring and personalized assistance which is reserved for clients and users)

These dimensions include a set of characteristics such as communication, security, competence, courtesy, ability to understand the needs of the client, possibility of access to the service focused mainly on the relationship, the characteristic component of every service relationship.

Regalia and Bruno (2000, p. 18) describe the differences, which represent important obstacles in offering a service of quality, that can be observed in an evaluation according to the Servqual model:

– *Difference between expectation of the consumer and perception of the management (*difference 1): the managers of the service companies and the teachers do not always identify in advance the characteristics that connote high quality in the eyes of the client. Those offering services do not always succeed in understanding what the consumer expects from a service.
– *Difference between perception by the managers and specific qualities of the service* (difference 2): the difference between perception, by the management, of what the consumer's expectations are and the specific qualities of the company's service ends up by having repercussions on the quality of the service in the eyes of the consumer
– *Difference between specific qualities and effective supply of the service* (difference 3): this is the crucial phase of the "front office", in which the personnel in their different components – management, administrative and operative – come into contact with the client.
– *Difference between the quality of the service and the external communications* (difference 4): the advertising and the other communications issued by a company can influence the expectations of the consumer. In addition, these communications can also influence the perception of the service provided
– *Difference between the service expected and the service perceived* (difference 5): the key to guarantee a good quality of services lies in satisfying or exceeding what the consumers expects from the service (also in terms of illustration of the meaning of the service provided). The quality observed in a service depends on the entity and the direction (lesser

or greater) of the difference existing between the service expected and the service received.

The quality of the service perceived by the client depends on the entity and direction (positive or negative) of difference 5, where $S\% = f (S1, S2, S3, S4)$. The quality of a service can be situated on a continuous line that goes from ideal quality to totally unacceptable quality. Satisfactory quality will be found at some point on this scale (Regalia & Bruno, 2000, p. 18).

3.1.3 The *constructivist* approach

The *constructivist* approach values the subjectivity of the players involved in the evaluation process and aims at interpreting and understanding, by *hermeneutic evaluation* (Perla, 2004), more than *measuring* the phenomena and the actions, which are the object of evaluation. At the centre of the models included in this approach, there is attention to the qualitative dimension of evaluation, the pluralism of values held by the various stakeholders, which requires a process of negotiation, phenomenological understanding of the meanings, languages and cultures emerging from the community in which they are inserted (Galliani, 2014). The model of educational evaluation underlying the constructivist approach is of a procedural type. It sees evaluation almost as an *act of communication*, which can be interpreted and negotiated, characterized by continuity, recursivity, creativity, unforeseeability, progressiveness, collaboration, cognitive and metacognitive regulation of the quality of individual learning and organizational systems (*ibid*, p. 35; Guba & Lincoln, 1989).

This is an approach that is closer to the possibility of understanding the implicit elements of processes that are not grasped by the methodologies of traditional evaluation. However, it is not always possible to guarantee generalization and

the use of the knowledge and results obtained. In this case too, there are multiple models: the CIPP model (Struffle-beam, 1967, 1983, 2003); the Responsive Evaluation model (Stake, 1975, 1988); the Multi method model (Patton, 1990, 1997); the Model of reflection in the course of action (Schön, 1983, 1987); Model of formative evaluation (Calonghi, 1992; Hadji, 1995).

CIPP model (author D. Stufflebeam, 1967, 1983, 2003): this was developed by Stufflebeam in 1967 on the grounds of three basic convictions: evaluation must be inclusive of all the components of a context of education (and therefore focus on the processes as well as on the products); all the comparative approaches, i.e. focused on the objectives-result relationship[15], are insufficient to evaluate the quality of "formative objects/subjects"; a good evaluation must in the first place help to make "good decisions". By putting evaluation at the service of the decision, the author distinguishes four different types of evaluation, correlated to taking specific decisions.

1. *Context Evaluation*
2. *Input Evaluation*
3. *Process Evaluation*
4. *Product Evaluation*

---

15 Cf. Provus's discrepancy model (1973), Tyler's "*rational*" model, the experimental model of Campbell and Stanley (1985).

TYPE OF EVALUATION	OBJECTIVE	TYPE OF CORRELATED DECISION
Context Evaluation	Knowledge of the context in which the programme and analysis/definition of the needs is intended to be applied	*PLANNING DECISION* Identification of the objectives and strategies appropriate for the context
Input Evaluation	Determination of the resources and the potential available to achieve the project	*STRUCTURING DECISION* Choice of the strategies and finalization of the planning
Process Evaluation	Collection of information on the trend of the project; identification of the strong and weak points, monitoring of the processes; comparisons	*IMPLEMENTATION DECISION* Decisions relative to rebalancing interventions of the pathway
Product Evaluation	Collection of the judgements on the formative outcomes; comparison with the outcomes of previous programmes.	*RECYCLING DECISION* Decisions on the opportunity of continuing, modifying or terminating the programme; whether to redefine the structure of the project and how

This is one of the first and most successful formalizations of a comprehensive and genuinely formative evaluation. Stufflebeam invites us to pay attention no longer only to the outcomes of an educational pathway but also to the context, the processes, the changes that can be induced "on the way", on the basis of a nature decision-making reflectivity. On the other hand, the evident *formative* function of evaluation is emphasized here: the evaluating actions make *sense* if they provide the deciders with useful information, not only for a summative purpose but also to improve and guide the pathway.

The model also has a high degree of flexibility and dynamism, which seems to have greatly anticipated some acquisitions relative to theories of a programming-curricular type, which are much later. The author brings into focus the global design that an evaluation system ought to have, whilst he is less concerned about providing indications on the procedures and instrumentation to have to translate it into practice. The reference to the structure and an eclectic use of methods more than to the procedural criteria seems almost implicit.

Stufflebeam does not seem to place much importance on the epistemological option (whether for "quantitative" or "qualitative") but the possibility of reaching a sufficiently clear, comprehensive and procedural observation/interpretation of the object *being evaluated* which must not be "enclosed" in rigid and pre-defined frameworks, but observe and monitored in its evolution in order to mature decisions. "Decision-making" is the objective of evaluation for Stufflebeam. It appears in line with the "philosophy" of a phenomenological and hermeneutic approach to the subject of evaluating the quality of educational work.

In the most updated version, the model proposes a *Cipp Evaluation Model Checklist* (2007), for a summative evaluation of the intrinsic and extrinsic value, of moral integrity and honesty, of the importance and significance of the programme as a whole. It is focused on: evaluation of context, needs, resources and problems (collection of information, interviews with those in charge of the programme and the stakeholders: those who have an interest at stake), evaluation of the input, i.e. of the resources, of the strategies, of the work plan, of the budget, evaluation of the process, documentation and monitoring of the planned activities, evaluation of the product, divided into evaluation of the impact (effects of the programme); of the efficacy (quality and importance of the results); of the sustainability (degree in which the effects produced by the programme are constant in time); of the transferability to a different context. At the end of every phase, a report is scheduled to be given to the client, In these phases, "meta-evaluation" is important; the evaluation of the evaluation focused on the documentation of the whole process.

The CIPP model reflects an evaluation of an educational/formative type that has aims of improvement and guidance and considers the evaluees in their contest, in the process of change, in order to make decisions. Applied to the evaluation of AE staff, it allows seizing elusive dimensions of

educational work and bringing into focus functions which, in educational contexts, do not appear to be very formalized. The essential characteristics of the proposals are expressed well in Stufflebeam's famous phrase "not to prove but to improve" (Stufflebeam, Shinkfield, 1985), i.e. that evaluation must not "test" but above all "improve" the processes under way and, in this direction, it would accentuate the hermeneutic dimension of the evaluators themselves.

Responsive evaluation model (author R. Stake 1975, 1988): i.e. which is developed from the needs and questions asked by the stakeholders, by the context, by the players involved, according to a *bottom-up* logic. This is an idiographic model focused on the individual concrete activities of the programme and on the judgements and personal interpretations of the programme triggered off by those who operate in it, which recovers the diversity of the perspectives. R. Stake's responsive evaluation takes a further step in the direction of a phenomenological, hermeneutic and reflective evaluation marked out by Stufflebeam. Stake, by starting from an epistemological position with a clear subjectivistic-phenomenological matrix, maintains that the *value* of a programme or of a performance cannot be expressed by a score, but by the best "description" and "interpretation" of the programme by the direct beneficiaries; he therefore invites giving up the precision of measuring to the advantage of the *meanings* attributed to the evaluating actions by the people involved. The essence of the evaluation lies in acquiring information that is truly useful for understanding the complexity of the educational-didactic undertaking as it is performed and this information can be brought out in the first place through the analysis of the points of view and the opinions of the "players" involved in the didactic situation.

Stake's position evolves in time and matures through two phases of theoretical elaboration: the first is formalized in

countenance evaluation; the second arrives at *responsive evaluation*. In *countenance evaluation,* the objectives are still the important parameter of the evaluating actions Stake, however, identifies others (beyond the traditional objectives relative to the programme and the results): objectives relative to the context, to the organization of the teaching actions, to the didactic "methods" etc. The aim is to reach *countenance evaluation*, i.e. an evaluation which expresses as far as possible the didactic reality and not only some of its parts. Later, Stake integrated the structure of the *countenance evaluation* into *responsive evaluation* which is a fully-fledged *Evaluation in situation*. Anticipating the coordinates of later theories (constructivism and evaluations focused on organizational development), *responsive evaluation* focuses the responsibility of the main players of the didactic in building up the evaluation structure.

The main assumptions can be summarized as follows:

1. The importance of the context in which the educational actions take place;
2. The inadequacy of a comparative evaluation (objectives-results) to be able to understand (and evaluate) the *quality* of the educational actions;
3. The opportunity that the evaluation is oriented more directly towards investigating the actions and meanings that a programme takes on for the people involved rather than the "intentions" of those who drafted the programme;
4. The need that the evaluation is developed from the nodes of conflict, the needs and the questions deemed significant by the participants and by the observer, i.e. from the issues;
5. A good evaluation must aim at bringing out the different systems of value that guide the actions of the people involved;
6. The use of "triangulation" as an elective tool of validation in the observations and opinions collected.

Stake's *responsive* evaluation challenges the rational frameworks typical of the previous models which tend to reduce the evaluating process in the educational context to a mere measuring technique and to subject it to purposes of an exclusively pragmatic nature. The greatest merit of Stake lies in having highlighted the *singularity* of the evaluating action and the holistic, systemic and hermeneutic character of every evaluation process, especially if started in an educational context. Evaluator and *evaluee*, educator and trainee, are placed in the same field of relations and observations and both take part in a process which makes them grow *together* and which should guide them in the direction of a permanent reflectivity/self-reflectivity.

One important aspect which emerges in Stake's model is the need to represent *all the points of view* in the opinion, Evaluation can be said to be such, for Stake, only on condition of recovering the diversity of the perspectives and promoting the qualitative growth of those who, even if located in reciprocally distant positions, are involved. This is the only way to give meaning to evaluation and to foster innovation. Evaluation, therefore, is conceived of by R. Stake as a "formative" and "qualitative" way for innovation. To these aspects of positivity we cannot help adding a comment. The only one, but an important one: the complexity of the actions of building up criteria of judgement that are really comprehensive of all the subjectivities involved in the evaluation process; the difficulty of negotiating meanings, of reconciling the multiplicity of points of view of the different *stakeholders*, of the translation of the many voices into a harmonious "melody". Evaluating is not only understanding and interpreting, but also succeeding in making a judgement which, unfortunately, seldom expresses unity of positions. Stake, while not offering "pragmatic" answers to the problem, invites us to consider the problem from another point of view: not that of the need to reach a *single* opinion, but the need to represent *all the points of view* in the opinion.

People (and their opinion) are taken as the main references of the process: the evaluating action is built up thanks to the interpretations of each stakeholder and evolves according to the argumentative and critical competence that each offers. The evaluator is the director of this dynamic: he does not have "positivistic" interests, i.e. he is not in search of generalization. His interest is of an almost prevalently *hermeneutic* type: to bring out what has value for the various *stakeholders*, to stimulate awareness, to collect and document knowledge that is the result of listening to everyone and the expression of the experiences and expectations, revelation of the meanings that each person attributes to the situation to be evaluated. According to Stake, therefore, any educational project should come into being on the basis of the issues of those directly concerned and the involvement of everyone, including the designers and the evaluators, in the same field of actions. In this sense, we could say that *responsive* evaluation gives great credit to the person and to "thinking": it is entrusted to the responsibility of those who are involved in the evaluation path and their critical capacity, their authentic interests in changing for the better.

Multi-method model (author M. Q. Patton, 1990, 1997): evaluation, according to Patton, consists of the systematic collection of information on activities, characters and results of a project that can form opinions on it, to improve its efficacy and direct future decisions. According to Patton, evaluators must be able to use and choose a variety of tools flexibly, adapting the search to the different questions of evaluation and the needs of the context. In the *Qualitative Evaluation Checklist* (Patton, 2003), evaluation has to be planned with particular attention to its practical consequences and its real use by the subjects that belong to the context evaluated, without being structured from a theory or a pre-established model, but identifying together with the subjects involved

the most suitable methodology for an evaluation process suitable for the context.

A famous book by M. Q. Patton has an emblematic title: *Utilization-focused evaluation* (1998); in the evaluation of AE staff, the methodology according to this model teaches to "think like an evaluator" and therefor to orient the educational actions already in the direction of how they can be evaluated. Patton thought a great deal about the utility and the usability of evaluation for the professionalism involved in evaluating actions: the techniques and the procedures should not be confined to a hermeneutic specialization, known to a small "caste" of evaluators, but have to be mastered by all the players involved; and the term *face validity* indicates precisely this: reaching a validity evident for the players. Evaluation has to be a process made to measure for the players, using their languages and where they can find a trace of their values.

Reflection in the course of action model[16]*:* in the reflective conceptual picture, it is also possible to understand a series of

---

16 According to this theory, in the reality of professional practices – and therefore of evaluation as well – we have to face two realities: the first is the singularity of each practical situation and the second is that the problems in practice never appear as already given. They " have to be set" in situations dominated per se by uncertainty and conflict of values. "When we set the problem," writes D. Schön, one of the most representative authors of this theory, "we select those that we will treat as 'objects' of the situation, defining the borders of our attention, and we impose on them a coherence that allows us to say that is wrong and in which directions the situation has to be modified" (1993, p. 68). "The situations that are face in professional practice are characterized by unique events… and the case that is presented as unique requires an art of practice that could be taught if it were constant and known, but it is not constant" (pp. 44-45). For Schön all the models that can be ascribed to an approach of technical rationality are unable

models of evaluation for which quality is not an external parameter to which processes and products are to be referred, but a dynamic construct which comes into being and is structured from inside an organizational context through the negotiated sharing of actions and objectives. We speak of *negotiated sharing* to refer to the fact that evaluation, in the theory maintained by these models, puts on the table all the aspects which deserve reflection, operating a mediation between the various players present in a context in order to gradually obtain the best result possible. In this last aspect, reflective evaluation presents analogies with the model of *Company-Wide Quality Control* but, unlike the latter which admits the existence of objective criteria with respect to which qual-

to "understand" or to solve the problems of the "practice" as they are based on the presupposition that these can be solved by applying theories and techniques with a scientific basis. More effective, however, is the recourse to an *epistemology of practice*, implicit in artistic processes and more suitable for situations of uncertainty, uniqueness and conflict of values. What are the nodal points of this epistemology? The first: knowing is in our action (*Knowing in action*). An expert action often cannot be "said" bit reveals a much wider cognitive activity than can be expressed. The second: reflecting in the course of action (*reflection in action*). The professional expert succeeds in thinking about what he is doing as he does it. The third: reflecting in practice. Professionals reflect on their knowledge in practice, they do not depend on categories consolidated by theory and technique, but each time build up a new theory of the unique case. "The dilemma between rigour and pertinence can be removed if we are able to develop an epistemology of the practice which places the technical solution of the problems in a wider context of reflective investigation, which shows that reflection in action can be rigorous for its own merits, and that relates the art of exercising practice in conditions of uncertainty and uniqueness to the art of research which is proper to the scientist. It is therefore possible to increase the legitimacy of reflection in action and encourage a wider, deeper and more rigorous use of it" (p. 95).

ity can be compared, reflective evaluation takes quality as a dynamic and problematic construct, to be achieved through processes of the *autopoietic* type.

Evaluating *reflectively* means activating an internal process of monitoring the formative path in which the evolution, initially not very visible, is made the systematic object of reflection: "doing" is intercalated with the observation/analysis of what is being done and the logic is not of opposition between knowledge, action and reflection, but of dialectic and reciprocal correlation between the three moments. This is why we prefer to speak of reflective evaluation rather than of *self-analysis* or *self-evaluation*, to underline the recursivity existing between theory and practice and the perfect identity between the roles of protagonist of the formative actions and subject appointed for the evaluation expressed by the process (even when re-course is made to evaluators "external" to the context).

The "educational work" "object" is thus taken as the "place" of structuring of an all-round reflective conversation with the experience of research and education, with its subjects and its "problems", whilst the focus on the dimension of the *formativity* of evaluation is accentuated. Self-evaluating oneself (*evaluating oneself reflectively)* means growing in awareness and meta-cognitivity, looking with greater depth at the motivations with which the main players enter the experience and develop their contents, considering the elements of decision-making which intervene in the relationship and which involve all the protagonists, bringing out the meanings that each one attributes to the formative situation.

Formative evaluation model (authors L. Calonghi and C. Hadji): Calonghi and Hadji, in Italy and France respectively in the middle of the 1970s, aimed to emancipate evaluation from any form of pragmatism or functionalism and tried to establish it as an "educational dimension" for the promotion of man and piloting educational actions (Perla, 2004, p. 87).

The ultimate goal of evaluation ought to be promoting self-regulating competence and the passage from evaluation to self-evaluation. The reflection by the French author starts from an assumption: every act of evaluation in the educational field is *over-determined and multi-dimensional*. To realize this, writes Hadji, you only have to ask the question "who does what?" to observe the pervasiveness and specularity of every action of evaluation.

Let's now bring into focus the main features of the articulated proposals of Calonghi and Hadji.

The first: the "epistemological" effort of clarification of what evaluation in the educational and training context is. If for Calonghi, evaluation has to tell us to what extent we are close to the ultimate goal and when the intermediate stages are reached, for Hadji there is evaluation every time someone makes an effort to observe a reality to say its value. Evaluating means making an oriented interpretation of reality. As to why we evaluate, the question leads to consider in both authors the intentions that emerge through the relationship that unites the evaluator and the evaluee. Taking up again some principles of C. Rogers, A.H. Maslow and Don Bosco, Calonghi shows how evaluation has to always express the confidence that the educator has in the possibilities of the evaluee and has to be implemented in such a way as to infuse confidence and optimism in the life project that the evaluee is making specific to himself and those around him.

According to Hadji, the evaluator's intentions are of three types and can be translated into: weighing up or measuring the object; appreciating it; understanding it. In the first case, following J. Ardoino and G. Berger, Hadji speaks of *estimative* evaluation which bases the judgement on measuring the performance with respect to certain intentions or certain pre-defined objectives. The resulting judgement is the result of a comparison not to a standard (referred to an absolute standard) but based on criteria, i.e. resulting from the comparison

of the performance to criteria, intentions and objectives defined before the work of evaluation. In the last case the evaluation is carried out to *understand*, to make reality more intelligible and to try to interpret it: in this case it is not so much the objectives and performances that are placed at the centre of the evaluating actions but the reflection on why this is the point that has been reached. This evaluation tends to involve all the people involved in the evaluation process.

As far as the answer to the question "*for what do we evaluate*?", if Calonghi makes explicit reference to the purpose of human maturity intended not only as education and cultural training, but also as ethical maturity, as an exercise habitually and morally supported by freedom, Hadji introduces the concept of "pedagogical utility" of the evaluation. Overlooking the evaluations of an implicit and spontaneous type (meaning those which, although without a foundation of reflectivity in the "strong" sense, nevertheless allow most of us to decide how to act), Hadji thematizes the "instituted" evaluation (moreover, already thematized by J. M. Barbier, 1985, p. 34), or that explicit one which terminates with a certifying judgement. It has a precise social function which is defined by the use that can be made of the judgement of value. This use depends on the type of decision that can be taken depending on the results of the evaluation. The different functions of evaluation can also be derived from the different way of using the result of the evaluation. L. Calonghi, resuming and extending a subject of Acriven, distinguishes between *continuous*, *formative* and *overall* evaluation. The first accompanies the didactic process and is carried out at the end of every learning sequence. Overall evaluation is when stock is taken of the activities for various purposes; such as an examination at the end of a course, to attribute eligibility, to certify admission to the next course. Formative evaluation is, on the other hand, the premise to projects of educational intervention and follows them step by step: evaluation is done to better educate,

to improve the interventions, the structures and the people implicated. According to this model, there are three stages and six tasks of the evaluator:

1. SPECIFICATION OF THE EVALUATION PROJECT
   a. To specify one's intentions as an evaluator
   b. To specify what is "expected" from the object evaluated
2. MAKING EXPLICIT THE PROJECT OF ACTION
   a. To specify the system of purposes and values which give meaning to the project fo action
   b. To indicate the "spaces" of evaluation
3. ACT OF EVALUATION
   a. To make a forecast of the expected effects
   b. To look for indicators that show the success or the failure of the action carried out.

With respect to the evaluation of AE staff, tis model presents some interesting peculiarities: having highlighted that formative evaluation is indissoluble from the educational process and therefore, the close relationship existing between *planning* and evaluation, between educational action and evaluating action.

Evaluation cannot be separated from the dynamic in which it is placed, i.e. the educational action, nor can it be reduced only to the conclusive moment of a pathway. It is, on the other hand, at the service of educational actions and accompanies them from the beginning to the end; having rooted evaluation in a paradigm of an anthropological type; any evaluating action must not lose sight of the purpose for which it is carried out (which model of man do I want to evaluate and therefore educate?) and the functions which it must fulfil (of facilitation, accompaniment, regulation piloting of formative actions); having clearly defined what has to be meant by *formative evaluation*. Thanks to Calonghi and Hadji

the meaning of formative evaluation is made explicit in the terms of encouraging guide/accompaniment in the educational pathway; the reference, especially in Calonghi's proposal, to the importance of the assessment for evaluation; to the need for the strictness of a "critical" docimology that is not limited to a descriptive contribution but which ensures continuous control and valid, reliable and coherent documentation for the evaluation; insistence on self-evaluation: the directional movement of evaluation ought to go from the outside towards the inside: from formative, writes Hadji, evaluation ought to become *forming* and promote the self-regulatory value of what has been evaluated.

3.1.4 *Certification and evaluation of competences in adult education*

After having identified the three main approaches of evaluation, reference should be made to the one which, although it does not represent a real scientific "model" of evaluation, includes a set of procedures – many of which are being experimented in the field – which are verifying the possibility of validating and certifying the informal and non-formal competences of European workers in AE. As is generally known, the goal of the Europe 2020 strategy is to reach intelligent, sustainable and inclusive growth. It is in this direction that the Proposal for a Recommendation of the European Council on the validation of non-formal and informal learning (Brussels 05.09.2012) should be read, which repeated the invitation to all the Member-States to establish as soon as possible a homogeneous system of certification and evaluation of competences, to allow recognizing competences matured during adults' professional lives. The White Paper of the Bauer Committee already pointed out, as early as 1997, the need to establish a better system of recognizing and defining non-formal competences (cf. Cedefop Glossary). At European level,

the Recommendation of the Council of the European Union on the validation of non-formal and informal learning was published on 20/12/2012, with which the Member-States were urged to set up national systems for the validation of non-formal and informal learning by 2018. The urgency was felt at the same time to promote the development of methodologies for evaluating competences acquired outside the standard contexts of education and training, i.e. in non-formal and informal learning contexts. The recognition of these categories of competences would make mobility and re-employment of workers on the job market easier. This is also a necessary action in view of the growing need for new professional profiles in some sectors such as services to the person (known as *white jobs*).

3.2 Aims of the evaluation: why the evaluation?
by Kleio Koutra, George Kritsotakis, Lina Pelekidou,
Nikoleta Ratsika, School of Health and Social Welfare, Greece

Why evaluate? What does evaluation in contexts of adult education aim at? Answering these questions is not simple and entails referring to a multiplicity of interconnected elements. The aims of evaluation in adult education can vary enormously, as it is a process linked to numerous variables, including the characteristics of the context of reference, the heterogeneity of the subjects involved in the contexts where the evaluation takes place, the complexity of the activities and the aims of the organization, the dynamics of power inside and outside the context, the relations with the local area and the interests of the stakeholders.

Some of the main functions of evaluation are:

Managerial function of the evaluation as control. One function of evaluation is managerial and organizational control, "i.e.