

Um sistema fuzzy de suporte à decisão para meta-avaliação uma nova abordagem e um estudo de caso desenvolvidos no Brasil*

Resumo

Este artigo apresenta uma nova metodologia para meta-avaliação que utiliza conjuntos fuzzy e lógica fuzzy. Esta metodologia é composta por um instrumento de coleta de dados e por um sistema hierárquico de inferência fuzzy. As vantagens da metodologia proposta são: (i) o instrumento de coleta de dados, que permite respostas intermediárias; (ii) a capacidade do sistema de inferência de se adaptar a necessidades específicas; e (iii) transparência, através da utilização de regras lingüísticas que facilitam tanto o entendimento como a discussão de todo o processo. As regras foram construídas com base nas diretrizes estabelecidas pelo Joint Committee on Standards for Educational Evaluation (1994) e por informações fornecidas por especialistas. O sistema pode auxiliar avaliadores que ainda não têm experiência em meta-avaliação. Um estudo de caso é apresentado como validação da metodologia proposta.

Palavras-chave: Meta-avaliação. Lógica fuzzy. Avaliação de programas educacionais. Padrões de avaliação.

Abstract

A fuzzy decision support system for meta-evaluation a new approach and a case study performed in Brazil

This paper presents a new methodology for meta-evaluation that makes use of fuzzy sets and fuzzy logic. It is composed of a data collection instrument and of a hierarchical fuzzy inference system. The advantages of the proposed methodology are: (i) the instrument, which allows intermediate answers; (ii) the inference process ability to adapt to specific needs; and (iii) transparency, through the use of linguistic rules that facilitate both the understanding and the discussion of the whole process. The rules are

based on guidelines established by the Joint Committee on Standards for Educational Evaluation (1994) and also

Ana Carolina Letichevsky
Dra. em Engenharia Elétrica,
PUC-Rio
Prof^a. do Departamento de
Informática, PUC-Rio
Estatística da Fundação Cesgranrio
estatistica@cesgranrio.org.br

Marley Maria Bernardes
Rebuzzi Vellasco
Ph.D., University of London,
Grã-Bretanha
Prof^a do Departamento de
Engenharia Elétrica, PUC-Rio
marley@ele.puc-rio.br

Ricardo Tanscheit
Ph.D., University of London,
Grã-Bretanha.
Prof. do Departamento de
Engenharia Elétrica, PUC-Rio.
ricardo@ele.puc-rio.br

*A revised and improved version of the paper presented at the 2006 American Evaluation Association Conference "The Consequences of Evaluation", Portland, Oregon, United States

represent the view of experts. The system can provide support to evaluators that may lack experience in meta-evaluation. A case study is presented as a validation of the proposed methodology.

Keywords: Meta-evaluation. Fuzzy logic. Evaluation of educational programs. Evaluation standards.

Resumen

Un sistema de soporte a la decisión para meta-evaluación y sus consecuencias: nueva abordaje y estudio de caso realizado en Brasil

Este artículo presenta una nueva metodología para meta-evaluación que utiliza conjuntos fuzzy e lógica fuzzy. Esta metodología es compuesta por un instrumento de coleta de dados y por uno sistema de inferencia fuzzy. Las ventajas de la metodología propuesta son: (i) el instrumento de coleta de dados, que permite respuestas intermediarias; (ii) la capacidad del sistema de inferencia de adaptarse a necesidades específicas; (iii) transparencia, a través de la utilización de reglas lingüísticas que facilitan tanto el entendimiento como la discusión de todo el proceso. Las reglas fueron construidas con base en las directrices establecidas por el Joint Committee on Standards for Educational Evaluation (1994) y por informaciones fornecidas por especialistas. El sistema puede auxiliar evaluadores que aún no tienen experiencia en meta-

evaluación. Un estudio de caso es presentado como validación de la Metodología propuesta.

Palabras clave: Meta-evaluación. Lógica fuzzy. Evaluación de programas educacionales. Padrones de evaluación.

Introduction

Assuring the quality of an evaluation is a great challenge to evaluators. Meta-evaluation (SCRIVEN, 1991) is the mechanism used nowadays to face this challenge. The main focus of the discussions about meta-evaluation is the excellence criterion for an evaluation. This is only the starting point to obtain a quality meta-evaluation; it is necessary to go beyond and find new methodologies that confer more flexibility to the execution of evaluative processes and that supply a precise and timely answer.

Meta-evaluation can be carried out in different ways, but frequently checklists (STUFFLEBEAM, 2001) are used as an instrument to collect data. Checklists are instruments with items or assertions about a specific focus with options of closed answers. In case of a meta-evaluation process, those assertions must investigate the presence of each one of the standards of a true evaluation in the target evaluation process. Traditionally, data are collected and treated based on classic logic, where the frontier between a point of the scale and another one is always clear. However, it is not easy for a human being to precisely define this frontier, since it may be of a fuzzy nature. In Fuzzy Set Theory, one given element can belong to more than one set with different grades of membership (TANSCHHEIT, 2004).

The same difficulty that exists in the stage of data collection is also faced in the treatment of information, since, in order to make an evaluation, excellence criteria must be established. These serve to elaborate a value judgment and can constitute rule bases, generally supplied by experts, that are used to verify whether or not the result meets a certain criterion. This study presents a methodology developed in Brazil for meta-evaluation that makes use of the concepts of fuzzy sets and fuzzy logic. This allows for the use of intermediate answers in the process of data collection. In other words, instead of dealing with crisp answers ("accomplished" or "not accomplished", for example), it is possible to indicate that an excellence criterion was partially accomplished in different levels. The answers of this instrument are treated through the use of a Mamdani-type inference system (MAMDANI; ASSILIAN, 1975), so that the result of the meta-evaluation is eventually obtained.

Meta-evaluation

The Concept

When an evaluative process is designed, several aspects, such as evaluative questions, methods and techniques of data collection, identification of the respondents, should, from the beginning, be negotiated with whomever is in charge of the evaluation and with the representatives of those being evaluated. Evaluators and clients must be aware of the bias in the evaluative process, and seek to minimize it whenever possible; when this is not possible, they must report it. After all, meta-evaluations are carried out so that the bias is minimized and the quality of an evaluative process in all its

stages is ensured. This includes decisions concerning the execution of the evaluation, the definition of its purpose, design, information collection and analysis, elaboration of budget and contract, management, setting up of the team, among others.

In the same fashion that it is recommended that evaluations be carried out in the formative and summative perspectives (SCRIVEN, 1967), meta-evaluations should also be carried out having in view those two perspectives, which up to a certain point complement one another. The formative meta-evaluation is conducted along the evaluative process to improve on the evaluation. Ideally, its starting point should coincide with that of the evaluation. The main objective is to provide the team responsible for carrying out the evaluative process with useful information, in order to improve the process while it is still in progress. The summative meta-evaluation is carried out at the end of the evaluation, in search of conclusive answers about its merit and relevance to those who ordered it, as well as to users and others interested in the process. The aim is to give credibility to the evaluation and to the final results generated by it. In other words, while the role of the formative meta-evaluation is to improve the evaluative process throughout its development, the role of the summative evaluation is to give an account to those involved and to the community at large, and to contribute to the improvement of future processes.

The concern with the creation of standards (principles obtained through

consensus among the people involved in the practice of evaluation, which, if achieved, will guarantee the quality of an evaluation) that the evaluative process should follow is an old one and, perhaps, as old as the concern with evaluation itself. This is a hard task, not only on account of the technical difficulty inherent to it, but also, above all, as a result of the difficulty to sensitize, mobilize, and reach consensus among different pertinent people, so as to produce a technically good work, accepted by those who carry out the evaluation, those who are evaluated, and those who make use of it (LETICHEVSKY et al., 2007).

To discuss procedures of meta-evaluation is to discuss the quality of the evaluative process. Therefore, it is fundamental to consider also the standards that an evaluator should follow, in the light of pre-established criteria of an evaluation of quality.

Practical aspects

It is necessary to ensure that instruments for data collection are adequate for obtaining the data one really intends to collect. The validation of the instruments for data collection can be made in different ways. In the case of instruments that collect qualitative information, such validation can be made with the assistance of specialists in the area or through a comparison among different evaluators, techniques, and instruments. In the case of quantitative information, the use of a Confirmatory Factor Analysis is recommended (BOLLEN, 1998). This is a technique of reduction of data dimension just like Exploratory Factor

Analysis, which is better known and more frequently used. The fundamental difference is that Confirmatory Factor Analysis is carried out from the application of a model of structural equation and, therefore, a theoretical model is assumed beforehand relating latent variables (not observable) to the observable variables. In the Exploratory Factor Analysis, on the other hand, each and every latent variable may have an influence on the observable variables, since the number and nature of the latent factors before processing the analysis is unknown. It is precisely this difference that makes the first one confirmatory and the second one eminently exploratory. Another important difference is that in the Confirmatory Factor Analysis the errors are also modelled and may (or may not) be correlated, whereas in Factor Exploratory Analysis it is assumed that the errors may not be correlated (which is not always true). In the Confirmatory Factor Analysis, the model previously established is adjusted for the purpose of minimizing calculated residues through the difference between the variance and covariance matrixes observed and calculated. The Factor Exploratory Analysis is vastly employed, without any type of test for checking whether the errors are not, in fact, correlated. Ideally, different types of instruments for data collection should be used, considering both quantitative and qualitative information.

As to the quality of information, there are two aspects that must be observed: (i) the quality and adequacy of the sources of information and (ii) the adequate treatment of databases. When choosing

the sources of information, it is important to ensure (whenever applicable) that all different groups of possible informants about the evaluative focus are considered. On the other hand, after collecting data and before calculating the indicators, it is fundamental to remove from the databases information that does not reflect the latent trace one wants to measure. Thus, in the case of instruments that collect quantitative data, those that present responses with always the same pattern, no responses, or indication of objection, must be excluded, as well as any other instruments that when filled in do not reflect a useful information regarding to what one wants to measure (BRYK; RAUDENBUSH, 1992). When the information is a qualitative one, this problem may be avoided through data triangulation (FRANSES; GELUK; HOMELEN, 1999).

When choosing the most adequate technique for modelling and data analysis, it is important to be clear about which evaluative question or questions one intends to answer, since an adequate technique to search for the answer to a given question may not be adequate for another one (LETICHEVSKY, 2004). For example, if an evaluation of student performance is carried out with the aim of determining in which schools the best students are, it is possible to work directly with the students' scores (results). On the other hand, if one's intention is to identify which are the most efficient schools, it is necessary to consider that students present some differences and bring diversified life experiences, both with regard to formal education and to general knowledge (GOLDSTEIN, 1995).

Socioeconomic levels of families vary, and so does the previous knowledge of students (FLETCHER, 1997). Thus, students with a higher socioeconomic level, or students with a broader scope of pre-existing knowledge, would tend to display a better performance. Interactions between the student and the school environment also interfere with his performance and should also be incorporated into the models (YANG et al., 1999a). It is in this context that there appears the need to isolate the effects that do not depend on school, that is, those that are not, and will never be, under the control of administrators, teachers, pedagogical and support team. In this sense, one generally intends to isolate in particular the effects of the socioeconomic level of students and the schools which, somehow, have impact on their performance. Thus, what one wants to measure is the value added by the school (YANG et al., 1999b). Traditional methods of study of cause and effect relationship involve models of regression made at a single level, where one dependent variable is explained by a set of independent variables plus one error. In this specific case, the dependent variable is the student proficiency (estimated by his performance in content tests). However, is it possible to accept the validity of these models ignoring the relations between different hierarchical levels and the way these relations impact on the results of the study? The intuitive answer is no, and statistical studies confirm this. When one analyses multilevel questions through models of a single level, errors may possibly be committed, and, therefore, multilevel models must be used (RASBASH, 1999).

Similarly to what happens with schools, the same care must be taken in the evaluation of the performance of sections or boards of directors within a business company or the impact achieved by the beneficiaries of social programs with similar objectives (LETICHEVSKY, 2004).

A meta-evaluation can be carried out in several ways, through the use of different instruments, but the use of checklists has been the procedure most adopted by many evaluators and evaluation centers, generating satisfactory results (PENNA FIRME; LETICHEVSKY, 2002). Checklists are "lists of things to be checked or done". In practice, when a checklist is transformed into an instrument for data collection, a set of instruments is created where each item is a statement and the respondent must simply indicate whether the statement is true or not. In the specific case of meta-evaluation, such statement tries to investigate the presence of the patterns adopted in the evaluative process that is the focus of meta-evaluation. Checklists represent an efficient instrument, in a friendly format, for sharing lessons learned in practice (STUFFLEBEAM, 1994).

Fuzzy Logic: basic concepts

Concepts of Fuzzy Set Theory and of Fuzzy Logic can be used to translate, in mathematical terms, the imprecise information expressed by a set of linguistic IF-THEN rules. If a human being is capable of articulating its reasoning as a set of IF-THEN rules, it is possible to create an inference system, the algorithm which may be implemented through a computer

program, where Fuzzy Set Theory and Fuzzy Logic provide the mathematical tools for dealing with such linguistic rules (TANSCHHEIT, 2004).

Fuzzy Logic studies the formal principles of approximate reasoning and is based on Fuzzy Set Theory. Fuzzy Logic deals with intrinsic imprecision, associated with the description of the properties of a phenomenon, and not with the imprecision associated with the measurement of the phenomenon itself.

In ordinary sets theory, the concept of membership of an element to a set is very precise: either the element belongs or does not belong to a given set. Given a set A in a universe X , membership of an element x of X to the set A is expressed by the characteristic function f_A :

$$f_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

L.A. Zadeh generalized the characteristic function so that it could assume an infinite number of values in the interval $[0, 1]$. Given a fuzzy set A , in a universe X , membership is expressed by the function $\mu_A(x): X \rightarrow [0,1]$. A is now represented by a set of ordered pairs, $A = \{\mu_A(x)/x \mid x \in X\}$, where $\mu_A(x)$ indicates to what extent x is compatible with the set A .

The *support set* of A is the set of elements in the universe X for which $\mu_A(x) > 0$. Thus, a *fuzzy set* may be seen as the mapping of the support set in the interval $[0, 1]$.

A linguistic variable is a variable whose values are names of fuzzy sets. For example, the answer to an item of a certain instrument of data collection

may be a linguistic variable assuming the values "poor", "good", and "excellent". These values are described by means of fuzzy sets, defined by membership functions.

Consider the linguistic variable scale (in an instrument for data collection), with values *poor*, *good* and *excellent*, defined by the membership functions shown in Figure 1. Answers up to 2.5 have a membership grade equal to 1 in the *poor* set; the membership grade in this set decreases as the response increases. An answer of 5 is considered "totally compatible" with the *good* set, whereas responses above 5 present a membership grade different from zero in *excellent*.

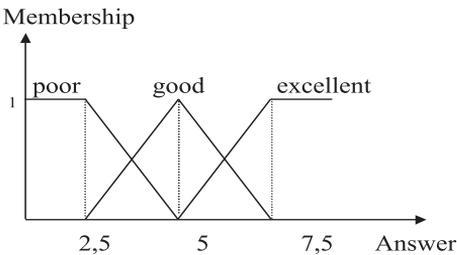


Figure 1 - Example of membership functions.

Membership functions may have different shapes, depending on the concept that one wishes to represent and the context in which they will be used. Context is extremely important, since the concepts of *poor*, *good* and *excellent*, for example, are extremely subjective. Membership functions may be defined by the user, but they are usually of a standard form (triangular or gaussian, for example). In practice shapes can be adjusted, in accordance with the results, in a trial-and-error procedure.

In fuzzy logic, a conditional statement (IF x is A THEN y is B) is expressed mathematically by a membership function, which denotes the *degree of truth* of the *implication* $A \rightarrow B$.

As to inference, the *modus ponens* of propositional logic (premise 1: x is A ; premise 2: IF x is A THEN y is B ; consequence: y is B) is extended to the *generalized modus ponens*, described as:

Premise 1: x is A^* .

Premise 2: IF x is A THEN y is B .

Consequence: y is B^*

While in classical logic the rule generates a consequence only if premise 1 is the exact antecedent of the rule (and the result is exactly the consequent of that rule), in fuzzy logic a rule is activated if there is a degree of similarity different from zero between premise 1 and the antecedent of the rule. The result will be a consequent with a degree of similarity to the consequent of the rule.

A Fuzzy Inference System, shown in Figure 2, does the mapping from precise (crisp) inputs to a crisp output. The crisp inputs may be measurement or observation data, which is the case of the large majority of practical applications. These inputs are *fuzzified* (mapped to fuzzy sets), which can be viewed as the activation of relevant rules for a given situation. Once the output fuzzy set is computed through the process of *inference*, a *defuzzification* is performed, since in practical applications crisp outputs are generally required. The *rules* are linguistic IF-THEN statements sentences and constitute a key aspect in the performance of a fuzzy inference system.

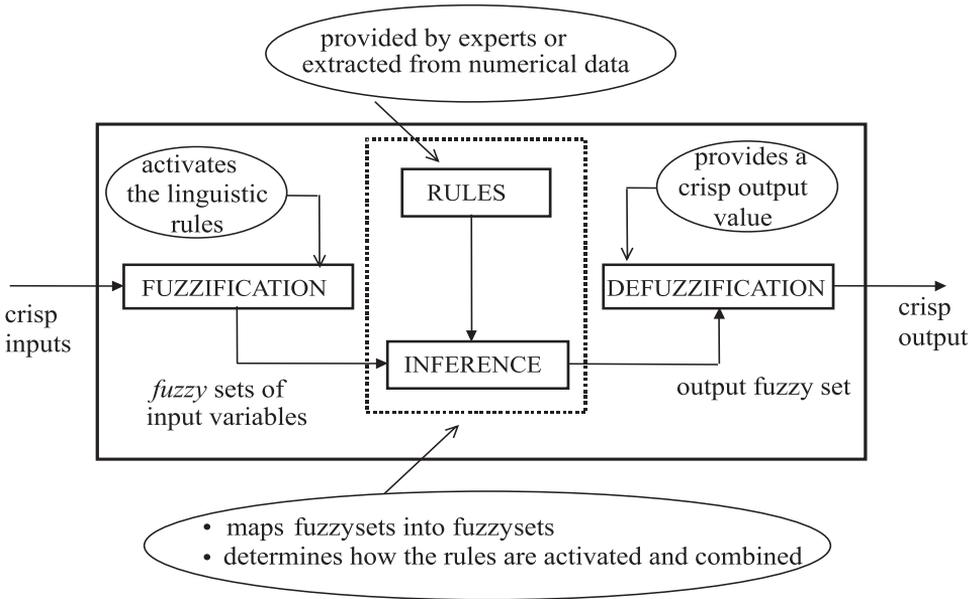


Figure 2 – Fuzzy Inference System

The methodology

The methodology adopted in this work (LETICHEVSKY et al., 2007) is composed of an instrument for data collection (Checklist for the Meta-evaluation of Programs/Projects) and of a fuzzy inference system to treat data related to

the meta-evaluation of projects and programs (Figure 3). The instrument for data collection was constructed from the adaptation of the checklist for meta-evaluation developed by the Evaluation Center of the Western Michigan University, whereas the fuzzy inference system was based on the thirty standards developed by the Joint Committee on Standards for Educational Evaluation.

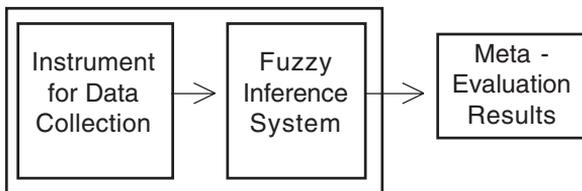


Figure 3 –Methodology for Meta-Evaluation.

Due to the complexity of the problem, the fuzzy inference system was subdivided into thirty-six rule bases, organized into a hierarchical structure composed of three levels (Figure 4):

- Level 1: Standards rule bases

- Level 2: Categories rule bases
 - Level 3: Meta-evaluation rule bases
- The hierarchical inference system, with the proposed three levels, is shown in Figure 4. The whole system was implemented by using the *MatLab*® *Fuzzy Toolbox*.

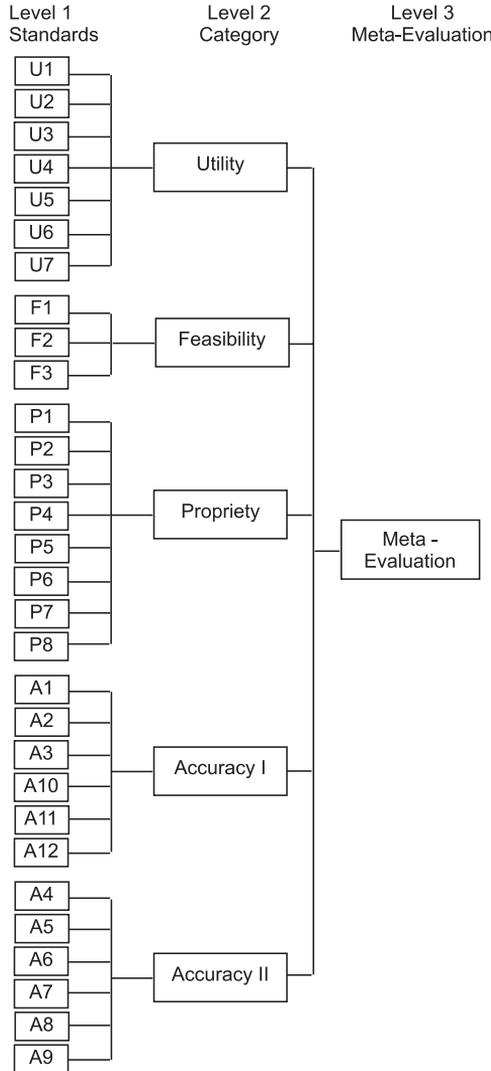


Figure 4 – Inference System Rule bases.

First, the standards rule bases (level 1) were built. Since each standard is evaluated on the basis of six criteria, the rules at this level have at most six antecedents. Each criterion has three linguistic values (*insufficient*, *satisfactory* and *excellent*) associated with it. The membership functions of the 3 fuzzy sets are shown in Figure 5.

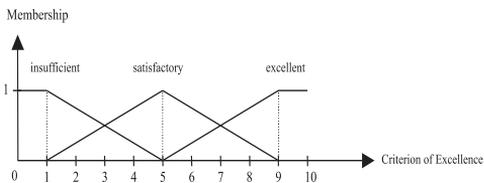


Figure 5 – Linguistic values for level 1 input variables.

To the output variables of this level the values *insufficient*, *satisfactory* and *excellent* are also associated, as shown in Figure 6. Thirty standards rules were developed.

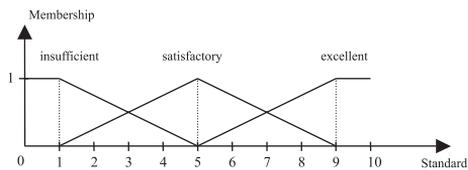


Figure 6 – Linguistic values for level 1 output variables.

As an example of linguistic rules of this level, consider the U2 rule base, which refers to the credibility of the evaluator and aims to measure to what extent people conducting the evaluation are both trustworthy and competent to perform it. The following linguistic variables were considered in the rules antecedents:

- U21: competent evaluators
- U22: trustworthy evaluators
- U23: evaluators address stakeholders' concerns;

- U24: evaluators are responsive to issues of differences;
 - U25: evaluators help stakeholders understand and assess the evaluation plan and process;
 - U26: evaluators pay attention to stakeholders' criticisms and suggestions.
- Some of the linguistic rules generated for this rule base are:

If U22 is *insufficient* then U2 is *insufficient*
 If U21 is *satisfactory* and U22 is *satisfactory* and U23 is *excellent* and U24 is *insufficient* and U25 is *satisfactory* and U26 is *insufficient* then U2 is *insufficient*.

In the rule bases of level 2, the number of antecedents varies in accordance with the number of standards present in each category, that is: category Utility has seven; category Feasibility has three; category Propriety has eight; and category Accuracy has twelve. In the case of the category Accuracy, the large number of input variables would jeopardize the development and understanding of linguistic rules. Therefore, the solution was to create two rules bases for accuracy: one with the standards that are directly related to information and its quality (Accuracy I rule bases) and another with standards that refer to the analysis and disclosure of information (Accuracy II rule bases).

The inputs to the inference systems of level 2 are the outputs from level 1 (linguistic variables with three values each). In the rule bases of level 2, the outputs are the linguistic variables that represent the category and have five associated values (*insufficient*, *regular*, *satisfactory*, *good* and *excellent*), specified by fuzzy sets defined by the

membership functions shown in Figure 7. The category *Feasibility*, for example, considers the following standards:

F1: practical procedures

F2: political viability

F3: cost effectiveness.

Examples of rules are:

If F1 is *insufficient* and F2 is *insufficient*

and F3 is *insufficient* then F is *insufficient*

If F1 is *excellent* and F2 is *sufficient* and

F3 is *excellent* then F is *good*

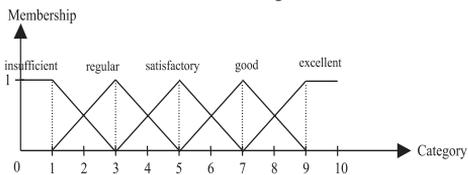


Figure 7 - Linguistic values for level 2 output variables.

Results by category are obtained at level 2 and not only facilitate the elaboration of recommendations and adjustments but also enable some categories to go through the process of meta-evaluation at different moments, when the instrument is used in a formative character.

The meta-evaluation rule base is responsible for the generation of the final result. Rules at this level have five antecedents (outputs from level 2). The consequent is a linguistic variable with five values, as shown in Figure 8. Examples of rules at this level are:

If utility is *excellent* and feasibility is *good* and propriety is *excellent* and accuracy I is *excellent* and accuracy II is *excellent* then meta-evaluation is *excellent*

If utility is *insufficient* then meta-evaluation is *insufficient*

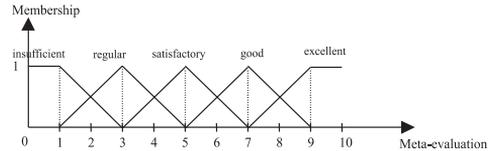


Figure 8 – Linguistic values for level 3 output variables.

When the instrument is totally filled out, the inference system computes thirty six results. Thus, besides calculating a result for each standard, the system also calculates one for utility, one for feasibility, one for propriety, two for accuracy, and the general result of evaluation. Each one of the standards results may be *insufficient*, *satisfactory* or *excellent* and the others: *insufficient*, *regular*, *satisfactory*, *good* or *excellent*.

A case study

A case study was developed, based on the evaluation of the initial stage of an educational program aimed at a low-income population. The length of the program was of about six months, and it was applied in ten towns in the North, Northeast, and Southeast of Brazil, all with low HDI (Human Development Index). Steps taken for data collection and validation of the proposed methodology are shown in Figure 9.

The first step for data collection was the selection of respondents (*meta-evaluators*) (1). Five types of respondents were selected:

EPP: Evaluator who Participated in the Process (1a), integrated by those evaluators who took part in the evaluative process focus of this case study.

ENPP Evaluator who did Not Participate in the Process (1b), consisting of

evaluators in general who did not know the evaluative process focus of the meta-evaluation and whose first contact with it was through the document 'Description of the Evaluative Process - Focus of the Case Study'.

PIIP: Professionals who Participated in the Implementation of the Evaluated Programs (1c), including those who are interested in and are users of the evaluative process focus.

META: Meta-evaluator who carried out the meta-evaluation external to the evaluation focus of the case study (1d).

STU: STUdent of evaluation who did not participate in the process (1e), integrated by students who did not know the evaluative process focus of the meta-evaluation and who had their first contact with it through the same document handed to the ENPP evaluators.

The next step was to *contact selected respondents (2)* that could ask for *additional explanation (3)* on the evaluative process focus. After the *data collection instrument was given back (4)*, the *fuzzy inference system (5)* was fed so that the *initial processing (6)* of data could be carried out. Based on the fuzzy inference system outputs and on the answers of the other instruments, *preliminary results were discussed (7)*, the results being considered in accordance with the category of the respondent. After that, adjustments were done to *fuzzy inference system (8)*, especially in correction and enlargement of rule bases. Afterwards, a *new processing (9)* was carried out, and with the new results it was possible to conclude a validation of the proposed *methodology (10)*. Eventually, *final results were discussed (11)*.

The analysis of results was based on the comparison between the results generated by the fuzzy inference system (starting with inputs from different evaluators) and the grades given.

Table 1 presents a summary of the results of the inference system and the grades given to the meta-evaluation and to utility, feasibility, propriety, and accuracy, according to the type of respondent.

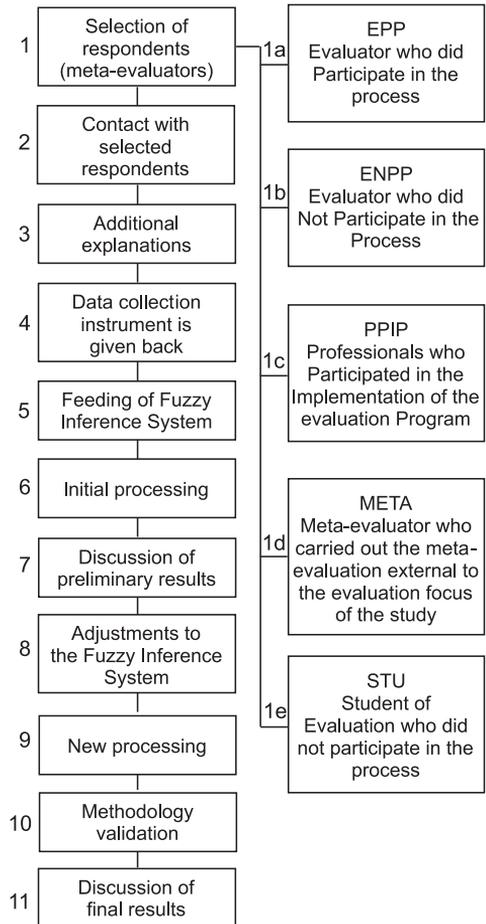


Figure 9 – Procedure to validate the proposed methodology.

Results provided by the inference system and the grades given by the EPP, ENPP and the meta-evaluator are very coherent. This coherence is justified by the fact that the three groups are composed, in general, by evaluators who have a wide experience, and, in many cases, with specific education in the area. All human beings are able to evaluate, and generally they do it many times a day, but the actual evaluator is one who is capable of judging the value on the basis of excellence criteria and previously established values. This is a competence achieved through the study of evaluation and the development of an evaluative culture.

In the case of PPIPs, the results are also coherent. The exception is the discrepancy, in propriety, between the fuzzy inference system's result of 5.34 and the grade good (G).

In the STU group, results were incoherent when compared to the grades given. Students probably do not have much experience in the area, and are still acquiring theoretical knowledge about the evaluation field. It is natural for them to have some difficulties to perform the evaluation and to separate their personal opinions, based only on their own standards and values, from a value judgment done in the light of previously expressed excellence criteria.

Table 1- Inference System Results and Grades* according to the type of respondent.

Type of respondent	META		Utility		Feasibility		Propriety		Accuracy		
	Grade	Result	Grade	Result	Grade	Result	Grade	Result	Grade	Result	
										PI	PII
EPP	G	7.00	G	7.98	G	7.79	G	6.68	G	7.28	7.28
	G	7.00	E	8.19	E	8.19	G	7.27	S	6.17	7.84
	G	7.00	G	7.00	G	8.19	S	5.00	G	6.73	7.79
	G	7.24	G	8.19	E	8.19	G	7.20	G	7.33	7.84
	G	7.82	G	7.79	G	8.19	G	7.60	E	8.19	8.19
ENPP	S	5.00	S	5.53	S	6.72	S	6.05	G	7.62	7.98
	S	5.00	G	7.00	S	6.72	S	4.70	G	7.62	7.62
	G	7.00	G	7.00	G	7.62	G	6.74	S	6.27	7.62
	G	7.00	G	7.31	E	8.19	G	6.65	G	7.62	7.79
	G	7.00	G	7.18	G	7.16	G	7.20	G	6.46	7.62
	G	7.00	G	6.87	G	7.28	G	6.65	G	7.68	7.86
	G	7.00	E	7.18	G	7.62	G	6.65	G	7.62	7.68
	G	7.01	G	7.00	G	7.62	G	7.05	G	7.62	7.79
	G	7.01	G	7.00	E	7.62	G	7.27	G	7.62	7.98
	G	7.01	G	7.00	E	7.62	E	7.27	G	7.98	7.98
	G	7.34	E	7.79	E	8.19	G	6.70	G	7.31	7.98
	G	7.44	E	7.79	E	8.19	G	7.27	E	8.19	8.19
	G	7.98	G	6.56	G	7.62	G	6.74	G	7.62	7.79
	G	7.98	G	6.70	G	7.28	G	6.65	G	7.33	7.92
	PPIP	G	7.00	G	7.02	G	7.79	G	5.34	G	7.86
G		7.01	G	7.00	G	7.79	G	7.60	G	8.19	7.68
G		7.28	G	7.02	G	7.31	G	7.10	E	8.19	8.19
G		7.60	G	8.19	G	8.19	G	7.60	E	8.19	8.19
G		7.71	G	7.98	G	8.19	G	7.60	G	7.98	7.98
META	G	8.28	S	4.65	G	8.19	G	7.60	G	8.19	8.19
	G	7.68	E	8.19	E	8.19	G	7.60	G	7.98	8.06

(Continuação)

STU	G	5.00	G	7.28	G	5.27	G	5.71	G	7.98	8.19
	G	5.00	G	7.79	G	6.88	S	5.86	G	7.86	8.19
	G	5.00	S	8.19	G	6.73	G	5.73	G	7.98	7.86
	S	7.00	E	8.19	G	7.79	S	5.48	S	6.23	8.19
	G	7.00	E	8.19	G	7.79	S	5.71	G	7.98	7.86
	G	7.00	E	8.19	G	7.79	G	6.68	G	7.98	8.19
	G	7.00	G	7.02	G	7.79	G	3.92	G	7.86	7.68
	G	7.00	G	7.62	E	8.19	G	5.94	G	7.30	7.02
	G	7.00	G	7.79	G	7.31	S	5.34	G	7.98	7.28
	E	7.00	S	7.02	E	8.19	G	6.70	E	8.19	7.98
	S	7.00	G	7.02	G	7.62	S	6.68	E	8.19	8.19
	S	7.00	G	7.62	G	7.79	S	5.00	G	7.98	7.79
	S	7.00	G	7.18	E	8.19	S	6.48	G	7.98	8.19
	G	7.00	G	7.31	G	7.28	G	6.68	G	7.86	7.62
	G	7.00	G	7.35	G	6.43	E	7.20	G	7.98	7.62
	S	7.00	E	8.19	E	8.19	S	6.70	S	7.98	7.62
	E	7.00	E	8.19	G	8.19	G	6.65	G	8.19	7.98
	E	7.00	E	8.19	G	7.79	G	6.37	G	8.06	8.06
	G	7.00	E	8.19	G	8.19	G	5.88	G	8.06	7.39
	S	7.11	S	7.02	G	7.62	G	7.27	E	8.06	8.19
	E	7.35	E	8.19	E	8.19	G	7.27	G	7.98	7.86
	E	7.36	E	7.62	G	7.16	E	7.60	G	7.98	7.28
	G	7.39	G	7.68	G	7.79	E	7.20	G	7.86	7.30
	G	7.54	E	8.19	G	7.62	G	7.27	G	8.19	7.00
	G	7.61	G	7.79	G	7.31	G	7.60	E	8.19	8.19
	G	7.72	G	7.31	G	7.16	G	7.10	G	7.98	7.62
	E	7.77	E	7.62	G	7.28	E	7.19	E	8.19	7.98
	G	7.81	E	7.62	G	7.28	G	7.20	G	8.19	8.19
	G	7.89	G	7.62	G	7.16	E	7.19	G	7.98	8.19
	E	7.98	E	6.37	E	8.19	G	7.20	G	7.86	8.19

*I - insufficient; A - average; S - satisfactory; G - good; E – excellent.

The harmony between the fuzzy inference system's results and the grades given by EPP, ENPP and meta-evaluators validate the methodology proposed here. It is important to emphasize that the coherence between results was confirmed by the feedback given by EPP, ENPP and META. Of 20 respondents, 16 have given feedback. In general, these evaluators presented their feedback according to the four categories of the Joint Committee on Standards for Educational Evaluation.

Conclusion

This work presented a methodology for meta-evaluation based on fuzzy sets concepts. This new methodology makes use of a fuzzy inference system and

consists of 36 rule bases organized in three levels: Standards (level 1), Category (level 2), and Meta-evaluation (level 3). The hierarchical structure and the rule bases were built in accordance with the standards of a true evaluation proposed by the Joint Committee on Standards for Educational Evaluation.

In this methodology the inference system employs linguistic rules provided by experts; this favors understanding and the update of rules. It may incorporate contradictory rules, which is not possible when traditional logic is used, and can deal with intrinsic imprecision that exists in complex problems, as is the case of meta-evaluation. The system was built on

the basis of evaluation standards of the Joint Committee on Standards for Educational Evaluation; thus, it is able to reach a broad range of users. It is expected that this work may help evaluators, those who order evaluations, and those who make use of results.

As for future work, the methodology shall be applied to performance, business and

institutional evaluations. It may also be adapted to other patterns, as, for example, those suggested by the European Society and by the Joint Committee on Standards for Educational Evaluation. The use of a model of Structural Equations and Factorial Analysis may be used to aggregate different meta-evaluators results.

References

- BOLLEN, K. A. *Structural equations with latent variable*. New York: John Wiley & Sons, 1998.
- BRYK, A.; RAUDENBUSH, S. *Hierarchical Linear Models*. Newbury Park, CA: Sage Publications, 1992.
- FLETCHER, P. *À procura do ensino eficaz: relatório de pesquisa*. Brasília, DF: PNUD/MEC/SAEB, 1997.
- FRANSES, P. H.; GELUK, I.; HOMELEN, V. P. Modeling item nonresponse in questionnaires. *Quality & Quantity*, Netherlands, n. 33, p. 203-213, 1999.
- GOLDSTEIN, H. *Multilevel statistical models*. London: Edward Arnold, 1995.
- JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION. *The Program Evaluation Standards*. 2nd ed. Newbury Park, CA: Sage Publications, 1994.
- LETICHEVSKY, A. C. La categoría precisión en la evaluación y en la meta evaluación: aspectos prácticos y teóricos. In: CONFERENCIA DE RELAC, 1., 2004, Lima. *Trabajo presentado...* Peru, Lima, 2004.
- LETICHEVSKY, A. C. et al. A new approach to meta-evaluation Using Fuzzy Logic. In: NEDJAH, N. et al. (Ed.). *Intelligent educational machines: methodologies and experiences: systems engineering*. [S.l.: s. n.]. 2007. Books in Series. In press.
- MAMDANI, E. H.; ASSILIAN, S. An experiment in linguistic synthesis with a Fuzzy Logic Controller. *International Journal of Man-Machine Studies*, v. 7, n. 1, p. 1-13, 1975.
- MENDEL, J. M. Fuzzy Logic Systems for Engineering: a tutorial. *Proceedings of the IEEE*, Raleigh, NC, v. 83, n. 3, p. 345-377, 1995.
- MORTIMORE, P. The nature and findings of school effectiveness research in primary sector. In: RIDDELL, S.; BROWN, S. (Ed.). *School effectiveness research: its messages for school improvement*. London: HMSO, 1991.

PENNA FIRME, T.; LETICHEVSKY, A. C. O desenvolvimento da capacidade de avaliação no século XXI: enfrentando o desafio através da meta-avaliação. *Ensaio: avaliação e políticas públicas em educação*, Rio de Janeiro, v. 10, n. 36, p. 289-300, jul./set. 2002.

PFEFFERMANN, D. et al. Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society*, London, Serie B, n. 1, p. 23-40, 1998.

RASBASH, J. et al. *MlwiN Beta version: Multilevel Models Project*. London: Institute of Education, University of London, 1999.

RAUDENBUSH, S.; Bryk, A. *Hierarchical Linear Models: applications and Data Analysis*. 2nd ed. Newbury Park, CA: Sage, 2002.

SCRIVEN, M. The methodology of evaluation. In: AMERICAN EDUCATIONAL RESEARCH ASSOCIATION. *Perspectives of curriculum evaluation*. Chicago, IL: Rand McNally, 1967. (Monograph Series on Curriculum Evaluation; No. 1).

_____. *Evaluation Thesaurus*. 4th ed. Newbury Park, CA: Sage, 1991.

SHADISH, W. R. et al. *Guiding principles for evaluators*. San Francisco: Jossey-Bass, 1995. (New Directions for Program Evaluation; No. 66).

STUFFLEBEAM, D. Empowerment evaluation, objectivist evaluation, and evaluation standards: where the future of evaluation should not go and where it needs to go. *Evaluation Practice*, Beverly Hills, CA, v. 15, n. 3, p. 321-38, 1994 .

_____. The methodology of metaevaluation as reflected in metaevaluations by the western Michigan University. *Journal of Personal Evaluation*, v. 14, n. 1, p. 95-125, 2001.

TANSCHKEIT, R. *Sistemas Fuzzy*. Rio de Janeiro: Departamento de Engenharia Elétrica, PUC- Rio, 2004.

YANG, M. et al. *MlwiN macros for advanced multilevel modelling: version 2.0: Multilevel Models Project*. London: Institute of Education, University of London, 1999a.

_____. The use of assessment data for school improvement purposes. *Oxford Review of Education*, Oxford, n. 25, p. 469-483, 1999b.

Recebido em: 30/07/2007

Aceito para publicação em: 09/08/2007