Available online at www.sciencedirect.com
ScienceDirect
Omega 36 (2008) 754–765
omega
The International Journal of Management Science
www.elsevier.com/locate/omega

An AHP-based evaluation procedure for Innovative Educational Projects: A face-to-face vs. computer-mediated case study☆

Mónica García Melón*, Pablo Aragonés Beltran, M. Carmen González Cruz

Department of Projects Engineering, Politechnical University of Valencia (Spain), Camino de Vera s/n, 46020 Valencia, Spain

Received 16 June 2004; accepted 13 January 2006
Available online 26 May 2006

Abstract

In this paper a procedure to evaluate proposals for Educational Innovation Projects is proposed. This methodology should help the Institute of Educational Sciences of the Politechnical University of Valencia to choose the best *Educational Project*, the final aim being to provide the Administration with a stringent evaluation methodology, since the current evaluation methodology was found to be neither sufficiently objective nor systematic.

Since in the definition and evaluation of these Educational Projects diverse stakeholders are involved, the process has been approached as a MCDA carried out by a group of experts. Although a whole methodology is proposed, the paper has been focused on the weight assignment of the different criteria chosen by the experts.

The experts have been asked to act in two different ways: *in face-to-face meetings* in which a consensus or compromise had to be reached, and *meetings at distance* where the experts have given their individual judgements, which have been next combined using the geometric mean with the software EC 2000 [Expert Choice 2000 Team. Pittsburgh: Expert Choice, Inc.; 2001]. This procedure has allowed the authors to analyse the possible scenarios that the IES board team might come up against in the future. The main difference between the two ways of work is the dimension of physical space or the distance between the members of the evaluating team. This distance has a significant effect on the way team members relate to each other.

Keywords: AHP; Group decisions; Decision process; Education

1. Introduction

As a result of reforms in Study Plans, the Politechnical University of Valencia (henceforth University) has been developing an Educational Innovation Projects (hence forth EIP) for the last five years. The aim of these projects is to initiate a process of methodological and organisational innovation in teaching that should improve the academic results of the students.

Every year the University calls for a tender of EIP, to which all the lecturers can apply. The projects should include a proposal explaining the objectives, as well as the methodological innovations proposed, resources required and expected results. These projects have to be evaluated. Therefore, a follow-up commission headed by the Institute of Educational Sciences of the University (henceforth IES) was created in order to select those to be supported, since the economic resources are limited and the IES board has to justify their distribution.

The method currently used to evaluate these proposals is the face-to-face meeting of five or six experts designated previously by the IES board who allocate a score

☆ This manuscript was processed by Area Editor Prof. B. Lev.
* Corresponding author.
E-mail address: mgarciam@dpi.upv.es (M.G. Melón).

0305-0483/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.omega.2006.01.005

to proposals on the basis of pre-established criteria. It is a simple and very subjective process and therefore difficult to justify. For this reason the IES board contacted the authors of the present work, to design a methodology that would help them to rank order the projects, the final aim being to provide the University with a stringent evaluation methodology that should allow the traceability of the whole process. Given the amount of University lecturers and departments involved in the proposals and the number of projects on offer, this evaluation would be accepted by the collective as a whole as long as the evaluation process were to instil sufficient confidence.

2. Aims of the work

The aim of the present work is to create a procedure that would allow evaluation of proposals for EIP, bearing in mind the multiple criteria and the opinion of different experts who were to advise the decision-making body, the Institute of Educational Sciences. The whole process could be understood as a problem of ordering the different proposals (alternatives) on the basis of a series of criteria established with the aid of experts. The process was therefore approached as a discrete multi-criteria multi-expert decision analysis (henceforth MCDA).*Multicriteria* because, as mentioned above, the conflicting interests involved are multiple: academic, economic, strategic, etc. and *multi-expert* since it was considered that the participation of independent experts would add gravity and rigour to the process. In this case, the selection process for experts was essential.

Two different ways of working for the experts will be analysed: (i) one based on face-to-face meetings and (ii) another one based on computer-mediated meetings since according to the authors experience in this field both may be useful for the evaluation of the proposals. The advantages and disadvantages of each of them will be analysed by means of a case study. After that the results of both ways of working, (i) and (ii), will be compared with the results of the current one used until now by the IES board. This comparison will be based on how the MCDA processes give more information, take more aspects into account, integrate the opinions of more people and therefore, can be used as a means for justifying the decisions better.

3. Multicriteria multi-expert evaluation

The use of MCDA makes the participation of different experts possible at different stages of the evaluation process, while taking multiple criteria and viewpoints into account. Different authors have proposed the use of MCDA as support in the decision-making process in different areas related to higher education. Mustafa and Goh [1] made an analysis of the techniques most used in the bibliography and the fields of application in the higher education area. Politis and Siskos [2] proposed their use in evaluating an Engineering Department in Greece with a view to enhancing its educational quality and internal organisation. Caballero et al. [3] proposed applying Goal Programming in assigning financial resources efficiently within a university system. Davey et al. [4] used it to analyse the selection process for a Ph.D. course.

The selection of the mathematical model based on MCDA is not easy. Among all the published methods the best known are: (i) the ones based on Multiple Attribute Utility Theory [5], (ii) the Analytic Hierarchy Process [6] and (iii) the outranking methods such as ELECTRE [7] and PROMETHEE [8]. A review of MCDA concepts and approaches can be found in [9,10]. According to Bouyssou et al. [11] there are several models that can be used in a decision-making process. There is no best model. To date, it has been impossible to demonstrate the domination of one MCDA technique over the rest. They all have advantages and drawbacks. Their use depends on the context.

In this process, the use of the Analytic Hierarchy Process (AHP) is proposed with multiple experts and, as support to this method, the software EC 2000 [12]. The reason this method is proposed is because it allows the decision maker to structure the problem of establishing priorities by means of hierarchic breake down of the problem, taking into account the consistencies of the emitted judgements. It is easy to explain to the experts who are to assess the EIPs and allows them to propose and gather the information generated individually in a simple and systematic way. The support software also enables the calculations and presentation of the results to be done easily and quickly. This favours its application in complex problems where the time of the participants in the process is rare and very expensive. The EC 2000 software has been used in a large variety of decision types, some related to the academic aspects of higher education [2] and other related to the decision-making support system journals [13]. Likewise, the program enables easy working with a group of experts via its group decision module. The operations research literature contains many applied and theoretical papers that describe the use of AHP in group decision-making setting. Early observations and suggestions for using AHP in group decision making are given in [14,15]. In

particular, some of these papers have focused on how groups construct the hierarchy, compare elements in the hierarcy, and aggregate weights.

4. Group evaluation. Face-to-face vs. distance meetings

As mentioned previously, the evaluation was to be carried out by a group of experts, who will be responsible for establishing and assessing weights to criteria. There are three basic approaches that a group can use to assess weights: (i) consensus, (ii) vote or compromise and (iii) geometric or arithmetic mean of the individual judgements. In the first approach, the group of decision makers is required to reach a consensus on each judgement. If the group is unable to reach a consensus, then a vote or compromise is used in the second approach to set the judgements values. In the third approach, let a_{ij}^k denote the comparison of element i to element j for decision maker k $(k=1,2,\ldots,n)$ in pairwise comparison matrix A. The individual judgements of the n decision makers are combined using the geometric mean to produce the entry $a_{ij}=[a_{ij}^1\times a_{ij}^2\times\cdots\times a_{ij}^n]^{1/n}$. Aczel and Saaty [16] have shown that the geometric mean preserves the reciprocal property in the combined pairwise comparison matrix. Therefore, the geometric mean is the most common approach used by groups to set priorities and has been incorporated into EC 2000 software.

Each of the above-described approaches has problems in practice. As pointed out by Saaty [17], the consensus and vote or compromise approaches could require "...a considerable amount of discussion (and initial disagreement) among the participants...". In a large hierarchy with many pairwise comparison matrices, reaching a consensus or voting on every entry could be very time consuming.

Taking all this information into account, the authors of this paper proposed that the experts should act in two different ways: *in face-to-face meetings* in which a consensus or compromise has to be reached, and *meetings at distance* where the experts would give their individual judgements, which will be next aggregated by EC 2000. These two working ways would allow the authors to analyse the possible scenarios that the IES board team might come up against in the future. The main difference between them is the dimension of physical space or the distance between the members of the evaluating team. This has a significant effect on the way team members relate to each other and their way of reasoning since in meetings at distance there is no discussion and therefore, the experts cannot change their judgements due to the group effect.

Different studies found in the literature show that there is no general agreement about which of the two ways of work, FTF and CM, is better. On the first place, several studies carried out [18] indicate that geographically dispersed teams can often work just as efficiently as those working face-to-face (henceforth FTF). Virtual interaction via electronic media (henceforth CM "Computer Mediated"), especially using email or electronic chat in their interpersonal relations, seems to be effective in reducing different types of discrimination between group members. With the elimination of visual stimulation, the focus is centred more on the content and less on the person generating the content. Furthermore, Raman et al. [18] researched the effect of group decision support systems (henceforth GDSS) and the communications media by means of these variables: satisfaction with the decision process carried out, ability to resolve conflicts and confidence in the decision taken. They concluded that FTF groups showed better results in their study than the CM groups for all variables. Finally, Gallupe and McKeen [19] carried up a study comparing the behaviour of FTF and CM groups in a decision-making process whose aim was to prioritise a series of alternatives. The results of that study indicated that there was no difference between both working ways and the authors justified this affirming that the task was not sufficiently complex to require a GDSS.

In the case of the EIP evaluation the authors of this work were faced with:

- FTF presented the advantage a priori of scarcely modifying the decision-making structures currently used in the IES institution, as it allowed them to continue acting as they had done until now: FTF meetings of the four or five persons in charge, but with a more rigorous methodology.
- CM would enable the IES board team to work with experts from outside the organisation, which would probably contribute with greater independence and objectivity to the evaluation process. These experts would be able to work from their offices, which implies time and money saving.

Since no general agreement about the goodness of any of the two working ways has been found in the revised literature, the authors of the present work opted to study the differences between them, analysing their advantages and disadvantages by means of a practical case. To this end, two pilot experiments with two groups of doctorate students from the University took place, whose results were then analysed separately.

Table 1
Current evaluation procedure

Decision maker	IES board	
Definition and weighting of criteria	IES board by face-to-face meeting (5–6 people)	
Criteria	1. Clear and realistic approach in the application of the proposals. Resources indispensable for start-up.	16.66%
	2. Clarity of objectives and relevance of content and innovative quality.	16.66%
	3. Incidence upon improvement and increase in student activity, initiative and motivation.	16.66%
	4. Direct and preferably immediate repercussion in day-to-day educational practice.	16.66%
	5. Definition of clear evaluation criteria, both of the process and the desired outcome it is hoped to achieve through the project.	16.66%
	6. Participation and co-ordination of teams of lecturers.	16.66%
Evaluation of alternatives (EIPs)	IES board with a score type: "failed", "passed but with conditions", and "passed" for each of the criteria	

5. Methodology proposed

5.1. Current evaluation procedure

The procedure currently used to evaluate these proposals is synthesised in Table 1.

This present procedure is not considered satisfactory by the IES management because only little and subjective information is used, which means a lack of transparency. It is assumed that all the criteria have the same weight, since the decision maker has never thought about weighting before. If a proposal obtains a "pass" for most of the criteria, the project is adjudicated to the applicant, with advice on how to improve those aspects particularly related to the criteria with the poorest evaluation.

5.2. Proposed evaluation procedure

The new approach proposes to work with a group of external and in-house experts. They will be given the advice of a facilitating team (henceforth Facilitator) who have some knowledge of the problems inherent in group evaluation (Table 2).

Selection of group of experts: Since the decision makers are members of IES board and they hold a political post, they are not always experts in the issue on which is to be decided. Since it is public money what is being assigned and to avoid misunderstandings about the final decision a technical group is thus recommended to advise the political head of the institution.

This technical group or group of experts must select the evaluation criteria for the EIP proposals, consider them, assign scales and assess each proposal in line with the established criteria.

In selecting these experts it must be taken into account that when evaluating the quality of the proposals for EIPs the academic factors must be considered, both the innovative aspects and their content. Given the nature of the problem, there might be a trend towards participation of a large number of experts in different areas making up the field of Innovation and Education. However, the contribution of many experts could lead to an excess of information; this is hard to deal with and very costly. In any evaluation and decision-making process, it must be considered whether the cost of the process outweighs the cost of the possible consequences of an erroneous decision. At any rate, the aim of this evaluation task is to provide the final decision maker (IES board) with information to help ensure the success of the training project.

It is recommended that the group of experts should have from 6 to 12 participants [20,21].

Selection of criteria: In this stage, a list of criteria to evaluate the proposals should be proposed by the group of experts, with no restrictions of any kind.

Weighting of criteria: In this stage the criteria should be weighted by the experts using the AHP method. To this end, a questionnaire has to be designed (see Appendix A).

Evaluation of the EIP proposals: The correct evaluation of proposals against each criteria is crucial, and requires the establishment of evaluation scales. Making pairwise comparisons, as the AHP method suggests, in a model that has a large number of alternatives can be very time consuming, as you would have to perform a large number of comparisons for the alternatives (up to $n^*(n-1)/2$) with respect to each covering objective. Alternatives' priorities can also be established relative to each covering objective using ratio scaled rating

Table 2
Proposed evaluation procedure

Decision maker	IES board
Definition and weighting of criteria	Group of external experts properly selected. They may work either in FTF or CM meetings. No IES board member should be included.
Criteria	To be defined by the experts.
Evaluation of alternatives (EIPs)	Group of external experts assess values to each proposal. The Facilitator analyses the results and reports to the decision maker.

intensities [6]. The authors recommend the use of *direct* due to the high number of projects to be expected in response to each call.

Sensitivity analysis and report: The obtained results should not be taken to be an absolute truth since many of the judgements made during the decision process are subjective, ill defined, indeterminate or imprecise. Finally, the decision maker must know the degree of reliability of the results in order to be able to make the final decision. Therefore, a sensitivity analysis is recommended to be used once the global order of alternatives has been obtained. This consists on calculating again the rank order of the alternative but with a slight modification ($\pm 10\%$) in the weight of an individual criterion, keeping the remaining weights fixed [22]. This procedure should be carried out for each criterium. That will allow the decision maker to obtain several rankings for the alternatives, which will help him to analyse how results can be affected by small deviations in weights.

After having obtained these results it is necessary to elaborate a report to aid decision, in which the coherence of the obtained results should be studied and the circumstances and factors which have influenced the application of the methodology have to be borne in mind.

6. Pilot test development

6.1. Aim of the pilot test

The first aim of the pilot test was to verify that the two ways of work proposed were applicable and useful. The Facilitator drew up a questionnaire to measure the degree of satisfaction and trust of the experts who joined the test (see point 5.2). The test was also used to study the differences between indirect assessment of weights (AHP binary comparisons as described in Section 4), and direct assessment of weights (setting the weight of each criterion on the basis of a pre-assigned scale). This study was carried on with the purpose of having an easier methodology to propose to the IES board in case the one based on AHP was considered too complicated.

The development of the test is described in the following:

6.2. Definition of the two groups of experts

To carry out the pilot experiment a group of 14 doctorate students from the University was selected, following the recommendations proposed in Section 5.2. Two groups were made, each of them composed by three foreign university lecturers, two workers of national industrial companies and two workers of national research institutes.

- *Computer mediated* (*CM*) *group*: All sessions were conducted by electronic mail. This group was allowed no communication among themselves, communicating only by email with the decision Facilitator. (GDSS level 1 according to [14].)
- *Face-to-face* (*FTF*) *group*: All sessions were conducted by means of face-to-face meetings based on a consensus-seeking method.

6.3. Obtention of the hierarchised list of criteria

CM group: This group met in a room and each person was assigned a computer. In this way, the aim was to simulate a situation where the experts would be working on-line and at a distance. They were asked to draw up a list of criteria, which in their opinion should be used to evaluate the proposals, and send it per email. The lists of criteria arrived in the Facilitator's mailbox in an orderly fashion and after this first round the Facilitator synthesised all the information received. In the second round the experts were asked to vote on the list drawn up. After that, this new information was synthesised in table form. Beside each criterion appeared the number of votes received for and against. For the third round, the experts were asked to vote again in the light of the group results. Once the last vote had been received, the round and session were considered closed. The whole session lasted 2 h.

At this point the Facilitator began to calculate, in line with the last votes received, which criteria would be

selected and which rejected. To this end, a simple sum of votes for and against was made and it was decided whether the criterion would be chosen or not by simple majority.

FTF group: The experts met in the meeting room. The Facilitator gave them the documentation and explained the aim of the meeting. Comments were made on the objectives of the evaluation and several template proposals and the information they contained were analysed. The group members decided that it would be easier for them to draw up a list of criteria and then to try to group them. After 30 min, the group had listed 14 or 15 criteria for evaluation. Next, the experts proceeded to classify the criteria. This phase turned out to be more complex than expected and a heated debate took place, where the definitions of some of the previously established criteria were modified. Finally, consensus was reached regarding the hierarchy of criteria. The whole session lasted 1 h 30 min.

6.4. Weighting of the criteria and assessment of scales

In this stage both groups were to assess weights to the criteria both directly and indirectly by means of pairwise comparisons, so that the results obtained would allow the analysis of the differences between them.

In order to use the AHP method, a specific questionnaire was prepared (see Appendix A).

CM group: Each expert was requested to answer said questionnaire and return it by email to the Facilitator, who then introduced the data into the EC 2000 so that the program could find the synthesis judgement. This phase lasted 45 min.

For direct weights' assessment, another questionnaire was drawn up asking the experts to assign a weight on a scale of 1–5 to each of the criteria. This questionnaire was returned by email to the Facilitator, who then calculated the synthesis judgement by means of the geometric mean of all the weights obtained for each criterion and then those values were normalised. In this way, the value obtained was comparable to that calculated by the EC 2000 program. Thus, the aim was to verify whether the experts were consistent and if the method of weight assessment affected their judgements. This phase lasted 30 min.

Finally, the group had to define which scale each of the criteria was to be evaluated with. To this end, the Facilitator sent the group of experts the consensual list of criteria so that each of them could propose a scale to measure each one. They were also allowed to give any clarification they considered appropriate in order to understand the significance of the scales proposed. This phase lasted 1 h.

Once the scales had been stated and voted the Facilitator calculated which scale was rated best for each criterion, by means of a simple sum of all the scores obtained by each scale. In this way it was decided which scale was to be selected to evaluate each criterion.

FTF group: This session was carried out in two phases. The first, weighting of the criteria (indirect), and the second, setting out the intensities of each criterion (direct), they both took place in a meeting room. In the first phase, each expert was proposed to issue his individual judgement for each binary comparison based on a Saaty scale, and then a discussion for each comparison took place until the consensus judgement was found. This phase lasted 1 h.

For direct assessment of weights, the experts were asked to assign a weight on a scale of 1–5 to each of the criteria and then discussion to find consensus started. This second weight assessment method lasted only 20 min, because all the experts were at that time very familiarised with all the criteria. As it was said before, the aim was to verify whether the experts were consistent and if the method of weight assessment affected their judgements.

Finally, each expert was asked to assign each criterion a direct weight on a scale of 1–5. This phase lasted almost 2 h.

6.5. Evaluation of the proposals

For this stage, the Facilitator drew up an evaluation questionnaire for the proposals, which was likewise sent by email to the CM group experts and delivered personally by the Facilitator to the FTF group experts. The experts were asked to assign a score to each proposal according to the scale defined in the previous stage.

It was not the aim of the pilot test to measure the time needed for this last task of the experts, evaluation of the proposals after the proposed criteria. At this last stage of the methodology, the aim was to make sure that the methodology was correct and executable and that nothing was misunderstood. Therefore, the results obtained for each of the proposals are not analysed in this paper.

7. Obtained results

7.1. Criteria hierarchical structure obtained for both procedures

The following results were obtained after the first work session: having analysed the results of this first

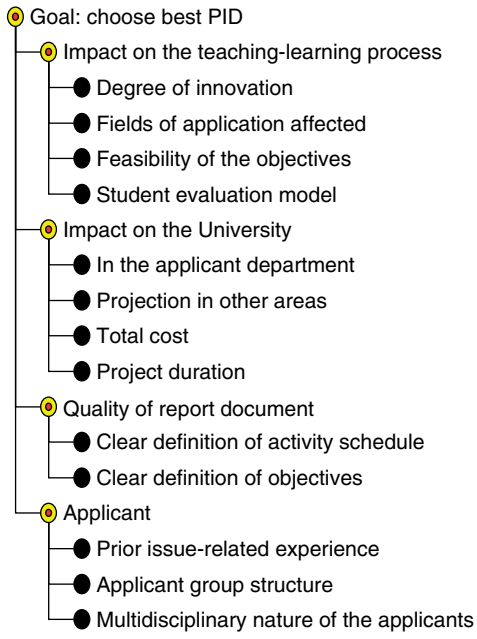

Fig. 1. Criteria hierarchy obtained by the FTF group of experts.

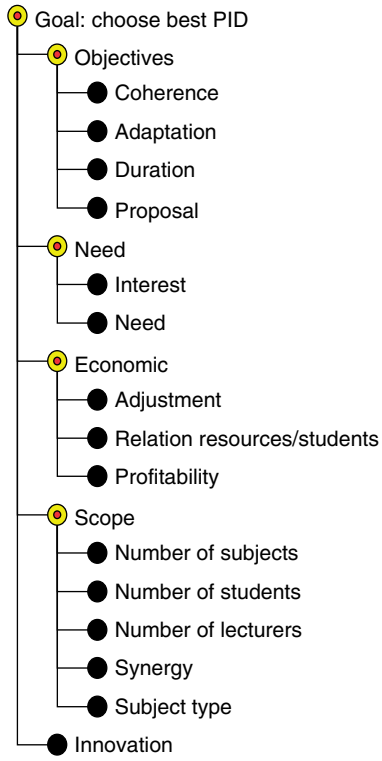

Fig. 2. Criteria hierarchy obtained by the CM group of experts.

session, one aspect that should be underlined is that there is a remarkable difference in the points of view adopted by each of the groups to draw up the list of criteria: each of the groups came up with completely different criteria hierarchies. That means that the criteria clearly depend on the experts and for that reason the authors of the present work insisted upon an extremely careful selection of them.

This may also be due to the significant role of the Facilitator in the CM group (Fig. 1). His influence when ranking or classifying the criteria into a hierarchy was very important, given that no communication was permitted between experts. The experts were simply asked to propose and score the list of criteria and the Facilitator ranked them conceptually and generated the hierarchy on two levels. However, in the FTF group, the hierarchy was established by consensus of all the experts. It may therefore be affirmed that the first hierarchical list of criteria comes under the clear influence of the Facilitator whereas the second does not (Fig. 2).

7.2. Criteria weights obtained for both procedures

The lists of criteria hierarchised and weighted by the two methods are presented below: the assessment of global weights corresponds to the aggregation of the judgements of the group members by the calculation of the geometric mean of the values in the individual matrices.

In Tables 3 and 4 it may be seen that in the case of direct assessment of weights there is a higher concentration of values. It is notable that if the mean of the weights obtained by the criteria is equal for both types of assessments, the standard deviation for the CM group is: 3.16 times greater for assessment by AHP (0.038) than direct assessment (0.012) and for the FTF group: 3.53 times greater for assessment by AHP (0.060) than direct assessment (0.017). Therefore, we can conclude that with the AHP method weights are much more dispersed than with the direct assessment method.

These differences in weight assessment, due to the different methods of weighting used, are in agreement with the works of several authors who showed that weights elicited through holistic procedures tend to be more dispersed than those elicited directly [23].

It must be pointed out that some of the decision makers were inconsistent in issuing their judgements, i.e. obtained a ratio of inconsistency of 15% calculated by EC 2000. There have been few studies on the

Table 3
List of hierarchised and weighted criteria obtained by the CM group

Criteria	Indirect assessment (AHP)	Direct assessment (scale 1–5)
1. Impact on the teaching–learning process		
1.1. Degree of innovation	0.14	0.102
1.2. Fields of application affected	0.067	0.079
1.3. Feasibility of the objectives	0.148	0.084
1.4. Student evaluation model	0.042	0.071
2. Impact in the University		
2.1. In the applicant department	0.041	0.068
2.2. Projection in other areas	0.045	0.069
2.3. Total cost	0.081	0.091
2.4. Project duration	0.047	0.069
3. Quality of proposal document		
3.1. Clear definition of activity programme	0.035	0.069
3.2. Clearly defined objectives	0.105	0.094
4. Applicant		
4.1. Previous subject related experience	0.108	0.074
4.2. Applicant group structure	0.054	0.059
4.3. Multidisciplinary nature of applicant	0.087	0.069
Mean of the values	0.077	0.077
Standard deviation of values	0.038	0.012

Table 4
List of hierarchised and weighted criteria obtained by the FTF group

Criteria	Indirect assessment (AHP)	Direct assessment (1–5)
1. Objectives		
1.1. Coherence	0.125	0.081
1.2. Adaptation	0.056	0.074
1.3. Duration	0.035	0.037
1.4. Proposal	0.065	0.072
2. Need		
2.1. Interest	0.184	0.065
2.2. Need	0.153	0.091
3. Economic		
3.1. Adjustment	0.037	0.069
3.2. Relation resources/students	0.032	0.066
3.3. Profitability	0.017	0.057
4. Scope		
4.1. Num. subjects	0.034	0.066
4.2. Num. students	0.025	0.062
4.3. Num. lecturers	0.013	0.037
4.4. Synergy	0.032	0.074
4.5. Subject type	0.021	0.066
5. Innovation	0.17	0.081
Mean of the values	0.067	0.067
Standard deviation of the values	0.060	0.017

inconsistency of the judgements of assessors. In some of them, inconsistencies are seen as mistakes, not intentionally made by the decision maker. They are fed back to enable the decision maker to learn about the procedure, and they are either ignored or reconciled by asking the person for a final judgement. If no feedback were possible or the decision makers did not want to change their judgements, weights would be derived mathematically through averaging or estimating the best fitting weights [6]. However, in our opinion, as the most important objective is to get the group to make better decisions and consistency is not so important, it was decided to accept a degree of consistency slightly higher than 10% as suggested by other authors [15].

7.3. Degree of satisfaction with the consecution of each stage of the procedure

The differences between both types of methodology will be analysed from the viewpoint of the Facilitators and the experts. To find out the opinion of the experts, they were asked to respond to the questionnaire presented below, the results of which are synthesised in Tables 5 and 6.

Here, it may be seen that the scores obtained for each of the aspects evaluated in the form of average of all the experts.

1. How would you describe the experience of doing the assigned task?								
Very pleasant	1	2	3	4	5	6	7	Very frustrating
2. In your opinion, the task was:								
Very easy	1	2	3	4	5	6	7	Very difficult
3. In your opinion, the process was:								
Efficient	1	2	3	4	5	6	7	Inefficient
4. In your opinion, the time needed to find solutions was:								
Reasonable	1	2	3	4	5	6	7	Not reasonable
5. How strongly do you feel that the group solution is correct?								
Very sure	1	2	3	4	5	6	7	Very unsure
6. How committed are you to the group solution?								
Very sure	1	2	3	4	5	6	7	Very unsure
7. How do you feel regarding the quality of the group solution?								
Satisfied	1	2	3	4	5	6	7	Not satisfied

Through the analysis of the values contained in Tables 5 and 6 it can be concluded: firstly, the scores awarded by the experts to each of the methodology stages of the pilot experiment were all situated around 3.5–4. However, according to the degree of satisfaction of the experts, FTF is slightly higher ranked than CM. In our opinion this may be due to the fact that the time required for the complete evaluation of the proposals was greater for the distance group. The fact that there is no communication between the group members means that any stage of the methodology requires many back and forth unilateral explanations, which causes delays.

Table 5
Evaluation by the CM group experts of each phase of the procedure

	Task to be done (questions 1, 2 and 3)	Solution obtained (questions 5, 6 and 7)	Time (question 4)
Phase 1	3.62	3.19	3.43
Phase 2	3.48	3.24	3.43
Phase 3	3.19	3.14	2.92
Mean of the values	3.43	3.19	3.26

Table 6
Evaluation by the FTF group experts of each phase of the procedure

	Task to be done (questions 1, 2 and 3)	Solution obtained (questions 5, 6 and 7)	Time (question 4)
Phase 1	3.27	3.27	4.60
Phase 2	3.47	3.60	3.20
Phase 3	3.27	2.67	3.20
Mean of the values	3.34	3.18	3.66

8. Conclusions

A detailed methodology for evaluation of proposals on Projects of Educational Innovation is presented. Its worth has been proven after realisation of the pilot test, given the low number of problems encountered.

The proposed methodology has the advantage, in comparison with the current one, that it brings more information to the decision process and improves the chance to explain the decisions made. Objectivity also improves because more point of views can be obtained following the group technique for criteria hierarchysation [5,6]. Moreover, the way the criteria are weighted and its later sensitivity analysis improve, in our opinion, what is being done now: consider a few criteria and give them all the same weight. In addition to that, it has to be highlighted that the procedure allows discussion both in the presence of the experts (FTF case) or through a Facilitator (CM case). All in all, transparency to the process is added, which is always recommended when we are dealing with distribution of state economic resources.

Both CM and FTF ways of work are evaluated by the experts in a similar way, as they obtained very close scores in each of the phases, although at all times the degree of satisfaction of the experts was higher for the FTF case. The results obtained by the authors in this case study are no different from the results obtained by other authors: studies which examined the effects of computer-mediated interaction on group processes and decisions vs. face-to-face interaction, found that computer-mediated interaction suppressed information exchange and led to poor group decisions [24]. Furthermore, face-to-face communication vs. computer-mediated communication has been found to affect evaluation performance by changing

the accuracy of the evaluation judgement. Research results showed that computer-mediated groups lowered judgement accuracy, obtained poorer outcomes and distributed resources less equally than face-to-face groups [25]. The rationale was that during e-negotiation, the parties could make their decisions and judgements on the basis of text only. The absence of non-verbal or verbal signals, which could help the parties attribute the "true meaning" to the message, may result in misunderstanding, misjudgement and undesired outcomes. The authors maintain that these affirmations agree with the results that they observed.

This may also be due to the fact that there is no communication between the members of the group; this means that any stage of the methodology may require several unilateral explanations back and forth, giving rise to delays. It is to be expected that in a real situation the person who actually "wastes time" is the Facilitator, as it is foreseen that all the experts will not be on line at the same time, rather choosing to do so when they have time to spare for the work.

Thus, it is estimated that the integral evaluation of the proposals might take from 4 to 6 weeks, but in contrast, the degree of satisfaction of the experts would be greater. When using FTF our opinion is that working with experts with no direct links with the Facilitator (in the pilot trial the Facilitator was the lecturer and the experts the students), the degree of freedom and spontaneity in expressing opinions will be greater, and the satisfaction with the work carried out will also improve.

To sum up, the authors of the present work recommend:

- the use of FTF methodology if the IES management works with people linked with the University as experts;
- the use of CM methodology whenever there is interest in the IES in working with experts from outside the University.

Finally, one aspect that should also be pointed out is that there is a notable difference concerning the point of view adopted by each of the groups in order to draw up the list of criteria. This means that the criteria clearly depend on the experts and therefore the authors of the present work insisted upon a very scrupulous selection of the experts.

As for the method of assessment of weights recommended, the authors recommend indirect assessment of weights, by means of the AHP method, given that a great concentration of values of the weights is considered to be not very realistic neither by the experts not by the authors of this paper.

9. Future works

The experts were asked to assign a score to each proposal according to the scale defined by themselves. They were allowed to bring the EIP proposals home and to take as much time as they needed to evaluate the proposals. Since it was not the aim of the pilot test to measure the time needed for this last task of the experts, and to analyse the evaluation of the proposals after the proposed criteria, it is planned to evaluate and study carefully the results obtained for each of the proposals. These proposals will be rank ordered using different MCDA techniques, in order to analyse the advantages and disadvantages of each of them and give further recommendations to the IES board.

Glossary

List of abbreviations and terminology used

- *EIP*: Educational Innovation Project
- *Proposals of Educational Innovation Projects*: alternatives that have to be evaluated and rank order in the present Study
- *UPV*, *University*: Polytechnic University of Valencia
- *IES*, *IES management*: Institute of Educational Sciences, Institution that orders this Study
- *AHP*: Analytic Hierarchy Process, procedure used to assess weights to the criteria
- *Procedure*, *methodology*: Set of phases proposed to evaluate the EIPs
- *FTF*, *Face to Face*: way of working for the group of experts of this Study
- *CM*, *Computer Mediated*, *Meetings at distance*: way of working for the group of experts of this Study

List of roles used

- *Decision Maker*, *IES board*: people responsible for the results of the evaluation of the Education Innovation Projects. People who order this Study.
- *Group of experts*: people responsible for the definition of the criteria and their weights used for the evaluation procedure of the Study
- *Facilitator*, *facilitating team*: People giving advice related to MCDA to the experts
- *Authors of the paper*, *authors of the work*: people in charge of carrying out the Study

Acknowledgements

The authors of this paper would like to thank the Foreign Language Co-ordination Office at the Polytechnic University of Valencia for their help in translating this paper.

Appendix A. Questionnaire AHP criteria weighting

For each pair of criteria please indicate **highlighting in black** which of the two you consider to be most important and to what extent. Remember that these are criteria to be used in the evaluation of training proposals.
The criteria must be compared two by two, asking to what degree criterion C_i is better compared with criterion C_j, using the following scale:

$C_{ij} = 1$: considered equally important criterion and criterion j
$C_{ij} = 3$: criterion i is considered slightly more important than criterion j
$C_{ij} = 5$: criterion i is considered considerably more important than criterion j
$C_{ij} = 7$: criterion i is considered much more important (or demonstrably more important) than criterion j
$C_{ij} = 9$: criterion i is considered absolutely more important than criterion j

---

C1: Impact on teaching-learning process
C2: Impact in the *University*

Which criterion do you consider most important?	C1	C2			
To what extent?	1	3	5	7	9

C1: Impact on teaching-learning process
C3 Quality of proposal document

Which criterion do you consider most important?	C1	C3			
To what extent?	1	3	5	7	9

C1: Impact on teaching-learning process
C4: Applicant

Which criterion do you consider most important?	C1	C5			
To what extent?	1	3	5	7	9

References

[1] Mustafa A, Goh M. Multi-criterion models for higher education administration. Omega, International Journal of Management Science 1996;24:167–78.
[2] Politis Y, Siskos Y. Multicriteria methodology for the evaluation of a Greek engineering department. European Journal of Operational Research 2004;156:223–40.
[3] Caballero R, Galache T, Gómez T, Molina J, Torrico A. Efficient assignment of financial resources within a university system. Study of the University of Malaga. European Journal of Operational Research 2001;133:298–309.
[4] Davey A, Olson D, Wallenius J. The process of multiattribute decision making: a case study of selecting applicants for a Ph.D. program. European Journal of Operational Research 1994;72:469–84.
[5] Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley; 1976.
[6] Saaty TL. Fundamentals of decision making and priority theory. Pittsburgh: RWS Publications; 1980.
[7] Roy B. The outranking approach and the foundations of Electre methods. Theory and Decision 1991;31:49–73.
[8] Brans JP, Vincke Ph, Mareschal B. How to select and how to rank projects: the Promethee method. European Journal of Operational Research 1986;24:228–38.
[9] Belton V, Stewart TJ. Multiple criteria decision analysis, an integrated approach. Dordrecht: Kluwer Academic Publishers; 2003.
[10] Figueira J, Greco S, Ehrgott M, editors, Multiple criteria Decisión Análisis, state of the art surveys. Springer Science; 2005.
[11] Bouyssou D, Marchout Th, Pirlot M, Peruy P, Troukias A, Vincke Ph Evaluation and decision models. A critical perspective. Dordrecht: Kluwer Academic Publishers; 2000.
[12] Expert Choice 2000 Team. Pittsburgh: Expert Choice, Inc.; 2001.
[13] Forgionne GA, Kohli R, Jennings D. An AHP analysis of quality in AI and DSS journals. Omega 2002;30:171–83.
[14] DeSanctis G, Gallupe. A foundation for the study of group decision support systems. Management Science 1987;22: 589–609.
[15] Dyer RF, Forman EH. Group decision support with the analytic hierarchy process. Decision Support Systems 1992;8: 99–124.

[16] Aczel J, Saaty T. Procedures for synthesizing ratio judgements. Journal of Mathematical Psychology 1983;27:93–102.
[17] Saaty T. Group decision making and the AHP. In: Golden BL, Wasil EA, Harker PT, editors. The analytic hierarchy process: applications and studies. Heidelberg: Springer; 1989. p. 59–67.
[18] Raman KS, Tan BC, Wei KK. An empirical study of task type and communication medium in GDSS. Proceedings of the 26th annual Hawaii international conference on systems sciences 1993. Silver Spring, MD: IEEE Computer Society.
[19] Gallupe RB, McKeen JD. Enhancing computer-mediated communication: an experimental investigation into the use of a group decision support system for face-to-face vs. remote meetings. Information and Management 1990;18:1–13.
[20] Aragonés P, Gómez-Senent E, Pastor J. Ordering the alternatives of a strategic plan of Valencia (Spain). Journal of Multicriteria Decision Analysis 2001;10:153–71.
[21] Greenbaum TL. The handbook for focus group research. New York: Lexington; 1993.
[22] Mareschal B. Weight stability intervals in multicriteria decision aid. European Journal of Operational Research 1988;33:54–64.
[23] Bordeching K, Eppel T, von Winterfeldt D. Comparison of weighting judgments in multiattribute utility measurement. Management Science 1991;37:1603–19.
[24] Holsaple C. Decision support in multiparticipant decision makers. Journal of Computer Information Systems 1991;18: 37–45.
[25] Baltes BB, Dickson MW, Sherman MP, Bauer CC, LaGanke JS. Computer mediated communication and group decision making: a meta-analysis. Organisational Behaviour and Human Decision Processes 2002;87:156–79.