Journal of the Operational Research Society

ISSN: 0160-5682 (Print) 1476-9360 (Online) Journal homepage: https://www.tandfonline.com/loi/tjor20

A multiple criteria information system for pedagogical evaluation and professional development of teachers

Marlene N M Filipe, Fernando A F Ferreira & Sérgio P Santos

Journal of the Operational Research Society (2015) 66, 1769–1782

www.palgrave-journals.com/jors/

A multiple criteria information system for pedagogical evaluation and professional development of teachers

Marlene NM Filipe[1], Fernando AF Ferreira[2,3]* and Sérgio P Santos[4,5]

[1]Polytechnic Institute of Santarém, Santarém, Portugal; [2]University Institute of Lisbon (ISCTE-IUL), Lisbon, Portugal; [3]University of Memphis, Memphis, TN, USA; [4]University of Algarve, Faro, Portugal; and [5]Center for Advanced Studies in Management and Economics of the University of Évora, Évora, Portugal

Information systems for the evaluation of pedagogical practices aim to improve teaching activity and encourage the pedagogical development of teachers. The implementation of these systems has been interpreted in different ways, depending on each school's administration board and/or the person responsible for the evaluation process. This ambiguity (and inherent subjectivity) is a barrier to the principles of transparency, equity and fairness in the pedagogical performance evaluation of teachers. From this premise, the main objective of this paper consists in proposing a multiple criteria information system to support the evaluation of pedagogical practices. By combining cognitive mapping and the measuring attractiveness by a categorical based evaluation technique (MACBETH), our framework allows for the conception of a more transparent evaluation system, bringing new insights to the pedagogical appraisal and professional development of teachers. The strengths and weaknesses of our framework are also discussed.

Journal of the Operational Research Society (2015) **66**(11), 1769–1782. doi:10.1057/jors.2014.129
Published online 4 February 2015

Keywords: decision making; cognitive mapping; multiple criteria decision analysis; management information systems; pedagogical practices; development of teachers

1. Introduction

Teacher training and development refers to the learning undertaken by teachers to upgrade their skills and stay informed and motivated. The process consists of several steps: (1) *initial training*; (2) *start of professional practice*; and (3) *professional development* (see Maskit, 2011; Richter *et al*, 2011; Stoklasa *et al*, 2011). In most cases, teacher training is a continuous process, since both trainees and qualified teachers must periodically evaluate and update their own pedagogical practices and/or methodologies (cf Altet, 2000; Huber, 2002; Timoštšuk and Ugaste, 2010; Inácio and Salema, 2011; Woods, 2011). Therefore, the process includes knowledge that is being accumulated during the teacher's career and, naturally, varies from teacher to teacher, depending on the academic environment in which each teacher operates.

Starting from this initial background, information systems for the evaluation of pedagogical practices emerged to improve the teaching activity and encourage the pedagogical development of teachers (see Ramli *et al*, 2010; Stoklasa *et al*, 2011). By pedagogical practices we mean all the teaching activity performed by teachers (within or outside of the classroom), which has the potential to lead to improvements in terms of the cognitive and effective development of students. However, the implementation of these systems has been interpreted in different ways, depending on each school's administration board and/or the person responsible for the evaluation process. As such, this ambiguity (and inherent subjectivity) is a serious obstacle to a transparent, equitable and fair evaluation of teachers' pedagogical performance, and there is an urgent need for restructuring and ensuring compliance with the principles of transparency, fairness and justice during the pedagogical evaluation of teachers.

In order to overcome this obstacle, this paper proposes a multiple criteria decision analysis (MCDA) information system that combines cognitive mapping and measuring attractiveness by a categorical based evaluation technique (MACBETH) to support the evaluation of the pedagogical practices of teachers. This information system aims to provide a better basis for the justification and explanation of decisions regarding the pedagogical evaluation and professional development of teachers, leading to a more transparent and fairer evaluation system. More specifically, cognitive maps have been characterized as a metacognitive negotiating tool that helps decision makers to understand and structure complex problems, thereby increasing the transparency of the evaluation process (Eden, 2004;

Correspondence: Fernando AF Ferreira, ISCTE Business School, BRU-IUL, University Institute of Lisbon, Avenida das Forças Armadas, Lisbon 1649-026, Portugal.
E-mails: fernando.alberto.ferreira@iscte.pt and fernando.ferreira@memphis.edu

Eden and Ackermann, 2004). MACBETH, in turn, has proved to be an extremely versatile and effective tool in supporting the calculation of trade-offs among evaluation criteria in a wide range of decision-making contexts (cf Bana e Costa *et al*, 1999; Belton and Stewart, 2002; Ferreira *et al*, 2011; Bana e Costa *et al*, 2012; Ferreira, 2013; Ferreira *et al*, 2014a). From these two premises, there seems to be considerable scope to explore the integrated use of these approaches (ie cognitive mapping and MACBETH) to support the pedagogical appraisal and professional development of teachers. It should be highlighted that this study is inspired on previous work by Bana e Costa and Oliveira (2012, p 425), who noted that the 'literature in the area does not provide comprehensive models for the evaluation of academic staff', and consequently proposed the use of MACBETH for faculty appraisal. It is worth stating, however, that our study is different from the one proposed by these authors in several ways. Firstly, our framework is aimed at assisting the assessment of teachers at the secondary education level, whereas the one proposed by Bana e Costa and Oliveira (2012) was targeted at higher education. Secondly, and as a consequence of the previous objective, our framework focuses on the pedagogical dimension of performance, which is the most relevant area of academic activity at this level of analysis. Because its emphasis was on the higher education context, Bana e Costa and Oliveira's (2012) framework focused on several different dimensions, including teaching, research, knowledge transfer and university management. Some of these dimensions are not relevant at the secondary education level. Finally, we use cognitive mapping to get a better understanding of the problem and to select the criteria to include in the assessment system. In Bana e Costa and Oliveira's (2012) work, the definition of the criteria was driven by legislation and the need for the system to reflect the main areas of academic activity normatively defined for higher education. We have found no prior documented evidence reporting the integrated use of these two approaches (ie cognitive mapping and MACBETH) in the context of academic staff pedagogical performance assessment. Therefore, in so doing, this research also contributes to the stream of Operational Research (OR) that argues for the relevance of using OR techniques, independently or in an integrated way, when developing performance assessment systems (cf Dyson, 2000; Santos *et al*, 2002; Smith and Goddard, 2002; Santos *et al*, 2008; Amado *et al*, 2012; Ferreira *et al*, 2014b).

The remainder of the paper is organized as follows. The next section discusses the importance of pedagogical performance appraisal of teachers and highlights some of the limitations of the current evaluation models/systems. The following section presents the methodological background of our study in order to show how the integrated use of cognitive maps and the MACBETH approach may contribute to overcome some of the limitations previously identified. The ensuing section explains the procedural steps of our framework, which follows a constructivist approach. The last section concludes the paper.

2. Background on pedagogical performance evaluation

Ramos and Moraes (2000, p 1) state that the performance evaluation of teachers '[…] contributes to improve the quality of education […]', namely when it pushes teachers to work towards achieving more effective and/or productive results at all levels of teaching. Starting from this premise, the Organization for Economic Cooperation and Development (OECD) (cf Santiago *et al*, 2009) considers teacher evaluation controversial but necessary, since its ultimate purpose is to improve the education system, enhancing the skills and performance of those who are evaluated (for further details, see Huber, 2002; Ramli *et al*, 2010; Inácio and Salema, 2011; Woods, 2011). Several authors support this statement by the OECD, defending that the controversy stems from the fact that, when negligently conducted, teacher evaluation may fail to differentiate teachers according to their merit. When this is the case, it may result in lack of motivation, increased anxiety and reduced self-esteem, jeopardizing students' learning (for a broader discussion, see Avalos and Assael, 2006; Elmore, 2008).

Regardless of its importance, the pedagogical performance evaluation of teachers has been commonly reduced to the students' appraisal of their teachers through periodical questionnaires (see Mills and Hyle, 1999; Spooren, 2010). Several limitations can be pointed out to this practice mostly because students are not (completely) informed about all the aspects of the evaluation process and, as individuals, care only about a few aspects of the pedagogical dimension (for a broader discussion, see Sproule, 2002). In particular, as defended by Bana e Costa and Oliveira (2012, p 425), '[…] validation methods for students to assess the performance of their teachers […] are only able to capture a small part of the daily activities of the academic staff […]'. In this sense, the authors raise an interesting question: 'should the evaluation of teaching be performed only by students, or should peers also review the content of teaching?' (Bana e Costa and Oliveira, 2012, p 435). Based on the premise that peers should also be involved in the appraisal process, teachers' pedagogical practices have been positively influenced by the development of more complete evaluation systems. These (new) systems are beneficial because they allow for the enhancement of skills, which in turn leads to permanent updates and adaptation to a changing environment (cf Ramos and Moraes, 2000; Desselle *et al*, 2004; Politis and Siskos, 2004). In spite of the progress achieved in this area, there still is, however, considerable scope for further improvement. As observed by Bana e Costa and Oliveira (2012, p 424), '[…] there is a need to develop comprehensive evaluation systems, based on methodologically sound procedures that can adequately reflect the differences between the academic staff […]'. One should bear in mind, however, that the pedagogical evaluation of teachers should not be seen as an attack on teacher professionalism, but as a stimulus to their professional development (Avalos and Assael, 2006).

In practice, the evaluation of teacher training and development emerged to improve the teaching activity and encourage the pedagogical development of teachers (cf Mills and Hyle, 1999; Simões, 2000; Huber, 2002; Richter *et al*, 2011). Nevertheless, the processes (and/or the purposes) by which evaluation has been conducted have been criticized by several authors. For example, Liu and Teddlie (2003, p 244) state that 'teacher evaluation is used for three major purposes: accountability, promotion and staff development […] but rarely for teacher or school improvement'. This means that the main reasons for teacher evaluation have been associated with higher levels of responsibility, promotion and career progression, but (incorrectly) separated from the progress of the teacher and of the school itself. In fact, teacher evaluation has been more concerned with the assessment of teachers' performance than with the effectiveness of the process (see Elmore, 2008; Sahlberg, 2011). From this standpoint, adequate teaching evaluation systems should consider certain principles, which should be centred on the professional development of teachers, through the improvement of their skills and working practices, contributing to improve students' learning (Avalos and Assael, 2006; Elmore, 2008; Inácio and Salema, 2011; Woods, 2011). As such, teaching evaluation systems should be primarily seen as learning mechanisms, allowing for the establishment of performance objectives and providing improvement suggestions.

Despite the progress achieved over the last few decades in the evaluation process of teachers (eg Sparks and Loucks-Horsley, 1989; Wolansky, 2001; Desselle *et al*, 2004; Avalos and Assael, 2006; Murias *et al*, 2008; Uzoka, 2008; Ramli *et al*, 2010; Stoklasa *et al*, 2011), several authors consider that teaching practices cannot be fully measured given the subjective nature of the appraisal and lack of available indicators. In this sense, Bana e Costa and Oliveira (2012, p 425) defend that '[…] there has been little research on how to integrate objective and subjective approaches adequately […] "most of the evaluation methodologies used […] suffer major flaws in both substance and process" […]' (see also Keeney *et al*, 2006; Elmore, 2008). Specifically, as remarked by Bana e Costa and Oliveira (2012), most of the flaws are related to difficulties in: (1) differentiating performance indicators from evaluation criteria; (2) weighting criteria; (3) ignoring the notion of value trade-offs when applying additive aggregation models; and (4) summing up ordinal scores on the criteria, creating meaningless overall scores. In spite of the variety of criticisms and/or limitations that may be identified in the current evaluation systems, it seems clear, however, that these criticisms and limitations converge into two main categories, namely: (a) the way the evaluation criteria are selected and incorporated into the evaluation mechanisms; and (b) the way the trade-offs (ie weights or compensation rates) among evaluation criteria are calculated and/or made explicit. By combining cognitive mapping and the MACBETH approach, this study aims to propose a methodological framework able to address these two major categories of limitations, improving the transparency and fairness of the evaluation process. Simultaneously, we aim to illustrate the potential of and to provide meaningful insights into how cognitive mapping and the MACBETH approach can be applied to enhance and add value to the key stages of any performance measurement and management process. In particular, this paper discusses how the methodological framework we propose can assist the *design* of a performance measurement system, the *measurement* of key metrics, the *analysis* of metric results and the *improvement* of performance. From a research perspective, our proposal follows a process similar to the one used by Santos *et al* (2002) and should be viewed as part of a continuous learning cycle, as suggested by Kolb (1984) and illustrated in Figure 1.

Two of the authors have practical experience in the use of cognitive mapping and the MACBETH approach to improve understanding and to assist decision making across several organizational contexts. Observation and reflection on this experience, informed by a review of the relevant literature, led the authors to the proposal, initially as an abstract concept, that there is synergy to be obtained from the integrated use of cognitive mapping and the MACBETH approach in the context of performance measurement and management. The development of this concept has led us to the framework proposed in this paper, which has been successfully tested in the banking context (eg Ferreira *et al*, 2011, 2012, 2014a, b, c), but which requires further applications to ensure generalization. The testing of this framework in a new situation, involving the pedagogical evaluation of teachers is, therefore, an important stage of this research. It is this stage that we discuss in detail in the following sections.

3. Brief methodological background

3.1. Cognitive maps

Cognitive mapping is an approach that assists decision makers in structuring their own ideas in real time, facilitating the reorganization of different lines of thinking and boosting collaborative debate between participants of the decision-making process (Eden and Ackermann, 2001a, b; Eden, 2004; Gul and Boman, 2006).

Belton and Hodgkin (1999), Tegarden and Sheetz (2003) and Eden and Ackermann (2004) present cognitive maps as tools

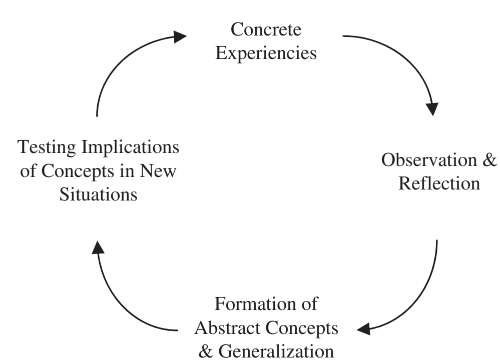

Figure 1 Kolb's experiential learning cycle.

for structuring complex problems. In spite of their subjective nature, the main advantage of these maps stems from the fact that they promote learning through a deeper understanding of the cause-and-effect relations among criteria, while may also be used to reduce the omission rate of important criteria (cf Ferreira *et al*, 2012). Therefore, cognitive mapping brings together a set of specific advantages, including the ability to: (1) deal simultaneously with quantitative and qualitative factors; (2) structure difficult decision situations; (3) underpin group work; and (4) be useful in the development and implementation of strategic directions. Consequently, cognitive maps hold great potential in the structuring phase of a performance evaluation process, namely because of their constructivist and recursive nature (for details, see Eden and Ackermann, 2004).

3.2. Basics of the MACBETH technique

The MACBETH approach was created by Carlos Bana e Costa and Jean Claude Vansnick in the early 1990s (cf Bana e Costa *et al*, 1999; Bana e Costa *et al*, 2012), and is based on numerical representations of semi-orders for multiple thresholds. Taking $X = \{a, b, \ldots, n\}$ to be a finite set of n alternatives (also known as actions), the technique consists in the association of each alternative to a value x (which results from a value function $v(.)$: $X \rightarrow R$) such that differences $v(a)$–$v(b)$ (with a strictly more attractive than b (ie $a\ P\ b$)) are as compatible as possible with the decision maker's (or group of decision makers') value preferences. Seven pre-defined categories of semantic differences of attractiveness (ie C_0—*null,* C_1—*very weak,* C_2—*weak,* C_3—*moderate,* C_4—*strong,* C_5—*very strong* and C_6—*extreme*) are used to support the application of the technique. This means that for all pairs (a, b) allocated to a certain category C, the differences $v(a)$–$v(b)$ will belong to the same interval. For example, if a is more attractive than b and the difference between both alternatives is considered *extreme*, then $(a, b) \in C_6$. Bana e Costa *et al* (2008) clarify that whereas two contiguous ranges correspond to two consecutive categories, the technical procedure consists in associating asymmetric partitions of the ray of positive real numbers to partition classes of ordered pairs (a, b) (with $a\ P\ b$), and intervals between consecutive categories are obtained based on a value function v and function thresholds s_k as defined in formulation (1), where $P^{(k)}$ represents a value preference that is stronger the greater the k.

$$a\, P^{(k)}\, b: \quad s_k < v(a) - v(b) < s_{k+1} \tag{1}$$

Because the thresholds s_k are positive real constants, Bana e Costa *et al* (2005, p 412) further explain that 'the basic idea underlying the initial development of MACBETH was that limits of these intervals should not be arbitrarily fixed a priori, but determined simultaneously with numerical value scores for the elements of X'. Hence, grounded on the projection of value judgments, MACBETH consists in allocating the difference of attractiveness between each pair $(a, b) \in X$ to one of the semantic categories C_k. The value judgments' consistency is then analysed following formulations (2) and (3) (cf Junior, 2008; Ferreira *et al*, 2014a).

$$\forall a, b \in X: v(a) > v(b) \Leftrightarrow aPb \tag{2}$$

$$\forall k, k^* \in \{1, 2, 3, 4, 5, 6\}, \forall a, b, c, d \in X \text{ with } (a, b) \in C_k$$
$$\text{and } (c, d) \in C_{k^*}: k \geqslant k^* + 1 \Rightarrow v(a) - v(b) \geqslant v(c) - v(d) \tag{3}$$

Specifically, formulation (2) shows that if $a\ P\ b$, then the value of a should be greater than the value of b. If no cognitive difference exists (ie $a\ I\ b$), then $v(a) = v(b)$, and $(a, b) \in C_0$. As for formulation (3), Bana e Costa *et al* (2008, p 28) state 'that all of the differences allocated to one semantic preference difference category are strictly larger than those allocated to a lower category'. With consistent value judgments, linear programming is then applied according to formulation (4) (cf Junior, 2008), which allows $v(n)$ to be minimized and an initial scale for discussion to be obtained.

$$\begin{aligned}
&\text{Min } v(n)\\
&\text{S.T.}: \forall a, b \in X: aPb \Rightarrow v(a) \geqslant v(b) + 1\\
&\qquad \forall a, b \in X: aIb \Rightarrow v(a) = v(b)\\
&\qquad \forall (a, b), (c, d) \in X, \text{ if the difference of attractiveness between}\\
&\qquad\qquad a \text{ and } b \text{ is bigger than between } c \text{ and } d\text{, then:}\\
&\qquad\qquad v(a) - v(b) \geqslant v(c) - v(d) + 1 + \delta(a, b, c, d)\\
&v(a^-) = 0
\end{aligned}$$

where:

n is an element of X so that $\forall a, b, c, \ldots \in X: n\, (P \cup I)\, a, b, c, \ldots$

a^- is an element of X so that $\forall a, b, c, \ldots \in X: a, b, c, \ldots\, (P \cup I)\, a^-$

$\delta(a, b, c, d)$ is the minimal number of categories of difference of attractiveness between the difference of attractiveness between a and b and the difference of attractiveness between c and d. (4)

It is worth clarifying that n is the most attractive element of X (or at least as attractive as the others) (ie $n\ (P \cup I)\ a, b, c, \ldots$), and its value minimization guarantees the minimal length of the scale. In turn, a^- is the least attractive element of X (or at least as attractive as the others) (ie $a, b, c, \ldots\ (P \cup I)\ a^-$), and its value is considered the 'zero' of the scale (cf Bana e Costa *et al*, 2008). Next, an overall score for each alternative is calculated based on the additive model presented in formulation (5), where $V(a)$ is the overall value of alternative a; w_i stands for the weight of criterion i and v_i

represents the performance level of a in criterion i.

$$V(a) = \sum_{i=1}^{n} w_i v_i(a) \text{ with } \sum_{i=1}^{n} w_i = 1 \text{ and } w_i > 0$$

$$\text{and } \begin{cases} v_i(Good_i) = 100 \\ v_i(Neutral_i) = 0 \end{cases} \quad (5)$$

In addition, it should be explained that $v_i(Good_i)$ and $v_i(Neutral_i)$ represent the partial scores of two specific performance levels (ie *Good* and *Neutral*, respectively), which are usually included in the process to facilitate cognitive comparisons (cf Ferreira *et al*, 2014b). Other MCDA techniques could have been applied in the context of our study (for details, see Belton and Stewart, 2002; Zavadskas and Turskis, 2011). However, as pointed out by Weber and Borcherding (1993) and Ananda and Herath (2009), there is no superior elicitation technique as each method has strengths and limitations. In light of this reasoning, the fact that MACBETH is based on simple question-answer procedures, which ask decision makers to pairwise compare different alternatives and provide semantic judgments about the difference in attractiveness between them, is easy to understand, able to consider qualitative and quantitative evaluation criteria and is solidly grounded on mathematics, supported our methodological option. Furthermore, it is worth noting that MACBETH allows the consistency of the answers to be tested, offering suggestions to bypass inconsistent situations and sequentially assisting decision makers to enter into the domain of cardinal measurement (Bana e Costa *et al*, 2005). Next, we explain how the integrated use of cognitive maps and MACBETH can support the design of information systems to assess the pedagogical performance of teachers.

4. A 'new' system for pedagogical performance evaluation

This study was organized in three main phases: (1) *the structuring phase*, where cognitive mapping was applied to identify areas of concern and evaluation criteria for pedagogical performance evaluation; (2) *the evaluation phase*, where MACBETH was used to calculate trade-offs among impact levels and among criteria; and (3) *the recommendations phase*, which explores the value added by the methodological approach proposed to the existing practices regarding the pedagogical performance evaluation of teachers. Advantages and shortcomings of our framework are also presented and discussed.

4.1. The structuring phase

In this study, the decision-making process involves the conception of an evaluation system, through the identification, prioritization and coordination of the evaluation criteria (or Fundamental Points of View (FPVs) as reported in the literature) that best characterize the pedagogical activity of teachers. In order to do so, it is fundamental to identify the decision makers that will ultimately use the system. Considering that the literature suggests that decision groups should have between 6 and 10 key individuals (cf Eden and Ackermann, 2004, p 618), our study had the collaboration of six decision makers (ie teachers with teaching evaluation responsibilities in their schools). Two facilitators (ie researchers), responsible for conducting the negotiation process and recording the results, also participated in the group sessions (ie three intensive group meetings with an average duration of 4 hours).

The structuring phase of the problem started with the following trigger question: 'From a teacher pedagogical performance perspective, and based on your own values and experience, how do you define the best teacher?'. We then applied the post-its technique which, according to Bana e Costa *et al* (2002, p 229), 'helps to identify clusters of linked aspects'. In practice, the technique consists of writing relevant criteria, from the decision maker's point of view, on post-its (ie one criterion per post-it) and producing evaluation clusters (ie areas of concern). The decision makers are then asked to focus their attention on the internal analysis of each cluster, in order to discuss the relationships among criteria. According to Ferreira *et al* (2011), this type of analysis serves to support the definition of hierarchies among the criteria considered in each cluster. After completing this procedural step, and using the *Decision Explorer software* (http://www.banxia.com), it was possible to obtain a final cognitive map, which served as a metacognitive tool to stimulate discussion among decision makers about how the problem had been structured.

The final version of the map contained about 150 interrelated criteria, and for this reason it could not be displayed in this paper. However, its construction proved very valuable to structure and improve understanding regarding the pedagogical performance evaluation of teachers. Additionally, as highlighted by one of the group members, some of the criteria presented in the map are rarely taken into consideration in the current evaluation systems (eg capacity to stay calm, to be open minded, to deal with sporadic situations), but 'the process followed allowed for their identification' (quoting that teacher's words). These new criteria were very important in defining the descriptors for the FPVs that formed the assessment system. Furthermore, the iterative nature of the process allowed the decision makers to share their thoughts and/or to explore new ones, allowing for the conclusion that the use of cognitive mapping 'contributes significantly to reduce the rate of omitted criteria' (in the panel members' own statement).

After reviewing the collective map, which was approved by the decision makers in terms of form and content, and reflected the most relevant aspects of the pedagogical evaluation of teachers, the next step consisted in the creation of a tree of FPVs. Following Keeney's (1992) methodological guidelines, the value tree was constructed using the M-MACBETH software (www.m-macbeth.com), and was presented to the decision makers for further discussion. It is important to note that several tests were carried out to ensure mutual preferential independence among the criteria, 'giving rise to a non redundant set of criteria' (Bana e Costa and Oliveira, 2012, p 429). Figure 2 shows the value tree, which represents the agreement reached by the decision makers.

As can be observed, the tree is composed of six FPVs (marked in bold in Figure 2), namely: FPV_1—*Acquired Skills*, which represents those skills that teachers develop over time (ie along their daily lives and/or careers); FPV_2—*Innate Skills*, which stands for the innate characteristics of an individual; FPV_3—*Practices Outside of Classroom*, understood as the set of practices undertaken outside of the classroom, but essential from a pedagogical point of view; FPV_4—*Practices Within the Classroom*, perceived as the set of pedagogical attitudes undertaken in the classroom; FPV_5—*Curriculum Updating*, which reports those formal practices that teachers go through to upgrade their skills and stay informed; and FPV_6—*Constraints*, understood as the set of variables controlled by the evaluator (ie school administration or person responsible for the evaluation process) that are external to teachers but able to influence their evaluation. Again, several tests were carried out to ensure mutual preferential independence among the FPVs. At the end, the tree of criteria was re-checked and considered concise and complete by the participating decision makers.

In a second group session, the decision makers focused their attention on the tree of FPVs to carefully define a *descriptor* (and respective *impact levels*) for each FPV. As an example, Table 1 presents the descriptor created for FPV_5—*Curriculum Updating*, which resulted from the negotiation process established with and among the decision makers.

Based on the information presented in Table 1, it should be pointed out that a *Good* level and a *Neutral* level were included in all descriptors to facilitate cognitive comparisons (cf Bana e Costa *et al*, 2001, 2006). With the definition of a descriptor for each of the FPVs included in the valuation model, the structuring phase was considered completed. It is important to note, however, that due to its time-consuming nature, the process was demanding. Even so, it was possible to identify the following advantages: interactivity; dynamism; learning; versatility and simplicity in the application. Moreover, and despite the inherent subjectivity, the application of cognitive mapping techniques allowed for the identification of criteria and relevant information that, according to the decision makers, are often omitted in existing evaluation models. The next step was the evaluation phase.

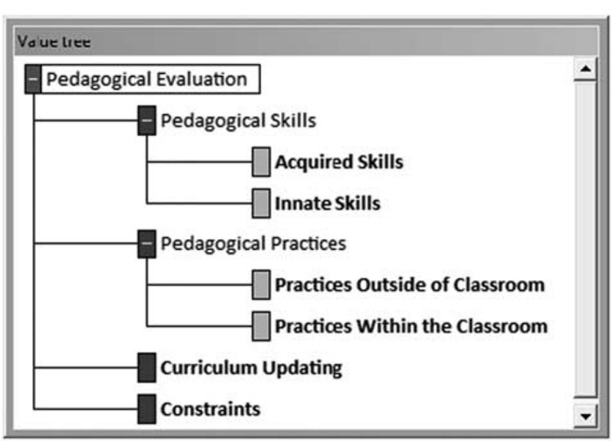

Figure 2 Tree of FPVs.

Table 1 Descriptor and levels of impact of FPV_5

Level	*Descriptor*
L1	Recent Conclusion of a Post Graduate Course; Participation in many Training Initiatives; Coordination of Several Educational Projects; High Number of Pedagogical Publications.
L2	Conclusion [Less Recent] of a Post Graduate Course; Participation in many Training Initiatives; Coordination of Several Educational Projects; Significant Number of Pedagogical Publications.
Good	Conclusion [Recent or Not] of a Post Graduate Course; Participation in Training Initiatives; Significant Participation in Educational Projects; Some Pedagogical Publications.
L4	Frequency of Post Graduate Course; Participation in some Training Initiatives; Participation in Educational Projects; Some Pedagogical Publications.
Neutral	No Post Graduate Course; Participation in some Training Initiatives; Some Involvement in Educational Projects; Low Number of Pedagogical Publications.
L6	No Post Graduate Course; Participation in few Training Initiatives; Low Involvement in Educational Projects; No Pedagogical Publications.
L7	No Post Graduate Course; Absence of Participation in Training Initiatives; No Involvement in Educational Projects; No Pedagogical Publications.

4.2. The evaluation phase

The evaluation phase consisted of filling matrices of value judgments for each of the descriptors previously defined, with the aim of creating local preference scales. For example, Figure 3 shows the value judgments regarding the descriptor of the FPV_1—*Acquired Skills*. After applying MACBETH, L_1 obtained a score of 180 points, while the lowest level (ie L_7) obtained a negative score of 440 points. The analysis of the results shows that the *Good* level was associated to 100, while the *Neutral* level was associated to 0. In terms of interpretation of the scale, this means that performance levels above *Good* were associated to scores over 100, while performance levels below *Neutral* were associated to negative scores. In addition, it should be noted that all data inputs were automatically verified by the M-MACBETH software. Whenever incompatibility between semantic judgments was noticed, the software suggested alternatives to overcome the inconsistencies, and the scale was afterwards subjected to discussion and validated by the participating decision makers. Figure 3 shows the process followed by the group members and the value scale obtained for FPV_1.

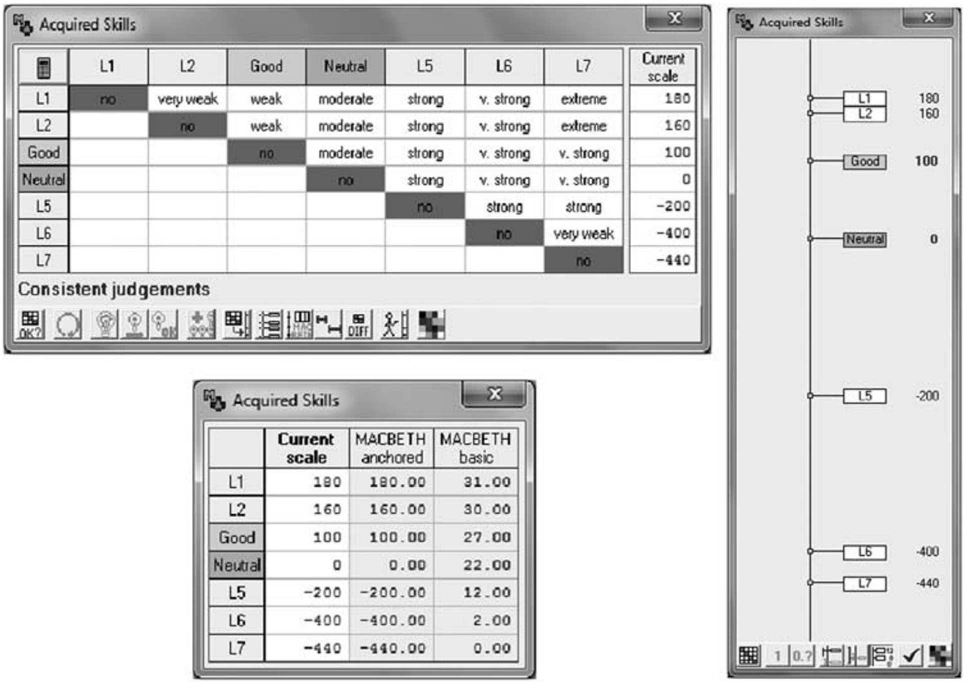

Figure 3 Value judgments and value scale proposed for FPV_1.

Table 2 Ranking of FPVs by overall attractiveness

		FPV_1	FPV_2	FPV_3	FPV_4	FPV_5	FPV_6	*Total*	*R*
Acquired skills	FPV_1		1	0	0	1	1	3	3
Innate skills	FPV_2	0		0	0	0	1	1	5
Practices outside the classroom	FPV_3	1	1		0	1	1	4	2
Practices within the classroom	FPV_4	1	1	1		1	1	5	1
Curriculum updating	FPV_5	0	1	0	0		1	2	4
Constraints	FPV_6	0	0	0	0	0		0	6

This process was repeated six times in order to obtain a local performance scale for each of the six FPVs identified in the model. With the six scales discussed and validated by the participating decision makers, the next step consisted in obtaining the trade-offs (ie weights or substitution rates) among FPVs. For that purpose, the FPVs were firstly ranked by decreasing order of preference. This procedural step was based on a process that consisted in filling in a matrix of pairwise comparisons, where the decision makers attributed the value ‘1’ whenever a FPV was considered more attractive than another (and ‘0’ otherwise) (Table 2).

Once the FPVs were ordered by decreasing attractiveness, a second matrix was filled in based on the semantic categories previously identified. Figure 4 shows this matrix of value judgments, which allowed us to obtain the trade-offs among evaluation criteria.

Following previous steps, it should be highlighted that the weights obtained were provided to the decision makers in order to be discussed and validated. The trade-offs shown in Figure 5 represent the agreement reached by the decision makers.

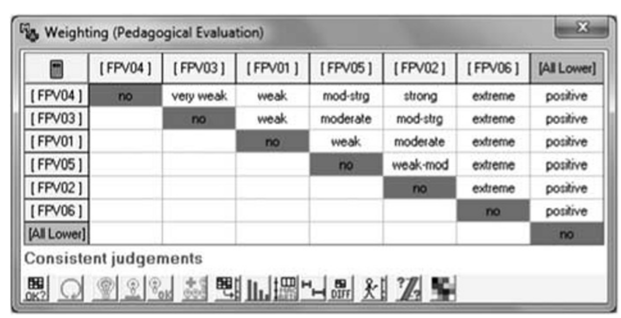

Figure 4 Matrix of value judgments for obtaining the trade-offs.

As illustrated in Figure 5, the FPV that stands out with the greatest weight is *FPV_4—Practices Within the Classroom* (24.28%). On the opposite side, the *FPV_6—Constraints—* presents the lowest weight according to the decision makers’ preferences (1.43%). At this stage, it should be remembered that the information gathered is based on the value judgments of the decision makers, who revealed hesitations in certain periods of

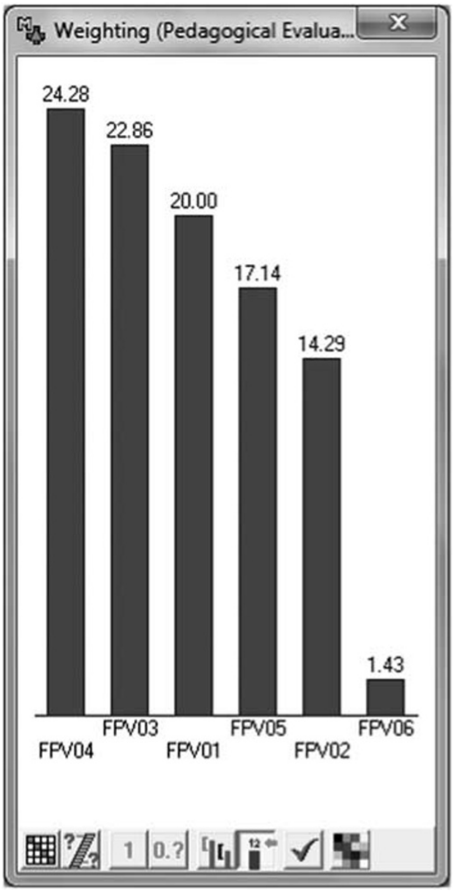

Figure 5 Normalized weights.

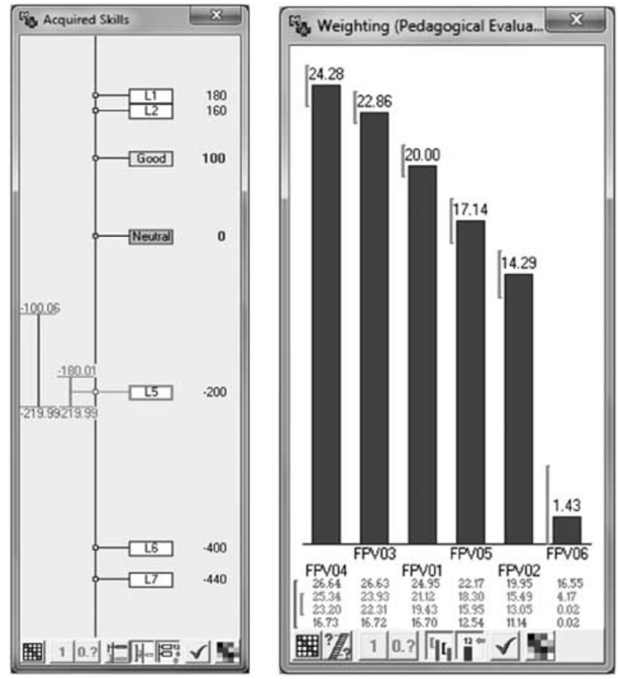

Figure 6 Acceptable variations in the proposed trade-offs.

Table 3 Partial performance revealed by the alphas

Options	*FPV*01	*FPV*02	*FPV*03	*FPV*04	*FPR*05	*FPV*06
Alpha1	L2	L4	L2	L2	Good	L6
Alpha2	Good	L4	L2	L3	L4	Neutral
Alpha3	L2	Good	L2	L1	Good	Neutral
Alpha4	Good	L4	Neutral	L4	L6	Neutral
Alpha5	Good	Neutral	Good	L2	Good	L6
Alpha6	Good	L4	Good	L3	Good	Good

the evaluation phase. However, it should be noted that the weights defined are endowed with sufficient flexibility such that in case of variations within certain limits (ie intervals), the consistency of the judgments delivered is not lost. Figure 6 presents those limits.

Once the trade-offs among the FPVs were obtained, it was possible to test the new system for pedagogical performance evaluation of teachers. To this end, we used information regarding the pedagogical practices of the six participating decision makers (referred to as Alphas from now on), in order to validate the results and, above all, increase the discussion among the group members. The lessons learned at this stage are the basis for the recommendations phase.

4.3. Application of the (new) system and recommendations phase

As mentioned in the previous subsection, before starting this phase it was necessary to gather information regarding pedagogical practices. In this sense, the decision makers were asked to use information related to their own careers and professional experience, in order to test the system that had been developed. To this end, we asked the participating decision makers to fill in a spreadsheet, indicating their respective level of performance in each of the six FPVs incorporated in the model. Table 3 presents the levels of partial performance revealed by the Alphas.

Once this phase was completed, the next step consisted in aggregating the partial performances in order to obtain an overall score for each Alpha. This was done using the additive aggregation model presented in formulation (5) (see subsection 3.2). As already discussed, this additive model allows for the aggregation of the partial scores $v_i(a)$ and the calculation of an overall score $V(a)$ for each Alpha, which represents a holistic measure of pedagogical performance for each teacher evaluated. Table 4 shows the partial and overall scores revealed by the Alphas during the application of the decision model developed. Two fictitious Alphas (ie *Good* and *Neutral*) were also included in the model to facilitate cognitive comparisons (for discussion, see Ferreira *et al*, 2014b).

Once the aggregation of partial values was completed and the overall scores for the different Alphas were calculated, it was

Table 4 Partial and overall scores revealed by the alphas

Options	*Overall*	*FPV*01	*FPV*02	*FPV*03	*FPV*04	*FPR*05	*FPV*06
Alpha1	130.49	160.00	83.33	150.00	163.64	100.00	–320.00
Alpha2	114.05	100.00	83.33	150.00	150.00	66.67	0.00
Alpha3	139.66	160.00	100.00	150.00	172.73	100.00	0.00
Alpha4	51.08	100.00	83.33	0.00	118.18	–55.56	0.00
Alpha5	95.16	100.00	0.00	100.00	163.64	100.00	–320.00
Alpha6	109.76	100.00	83.33	100.00	150.00	100.00	100.00
Good	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Neutral	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Weights:		0.2000	0.1429	0.2286	0.2428	0.1714	0.0143

possible to formulate a few recommendations. In this sense, it should be noted that the framework developed in this study assumes idiosyncratic characteristics and, therefore, cannot be extrapolated to other contexts without caution. In fact, the greatest contribution of this research is, perhaps, to illustrate the potential learning that can occur when cognitive mapping and MACBETH are combined to design pedagogical performance evaluation systems. By following a constructivist approach, the use of these techniques can result in more effective evaluation systems. These systems allow us to identify not only the best and worst performers, but also those criteria where improvement is deemed fundamental. For example, Figure 7 shows the partial scores of Alpha 1 and Alpha 4, making it possible to identify which criteria require priority of intervention.

As seen in Figure 7, Alpha 1 and Alpha 4 present very different profiles, clarifying the results shown in Table 4. Alpha 1 presents a good performance on all but one FPV. Despite the poor performance revealed in FPV_6, because the relative importance of this criteria is low, this teacher still obtains a good score overall (130.49). The fact that Alpha 4 presents unsatisfactory performance levels on three out of the six FPVs considered in the analysis, explains his/her relatively low overall performance score (51.08). In particular, it is fundamental that this teacher develops efforts to improve performance on FPV_5. This type of analysis clearly reflects the partial performance evaluations obtained for each Alpha, and points out which aspects should and/or could be improved in order to achieve a higher performance level.

After careful analysis of the evaluation outputs, several sensitivity and robustness analyses on different parameters were performed and discussed with the participating decision makers. In this sense, the ability of the framework proposed to assess the merit of each teacher, to identify the best and worst performers, to identify the areas in which improvement can take place and to explore the sensitivity/robustness of these results to changes in priorities and values, is very encouraging. This is also corroborated by the fact that at the end of the third (and last) session of our intervention, the model was considered 'more complete and discerning than other practices of pedagogical performance evaluation of teachers' (in the decision makers' own words).

Although it can be argued that the intervention only reflects the priorities and values of a limited number of decision makers, it is important to emphasize that the framework is flexible enough to accommodate the inclusion of different criteria weights, allowing the impact of these on the overall performance of teachers to be immediately assessed. Additionally, whenever there is uncertainty related to the criteria weights, the assessment process can be complemented by sensitivity and robustness analyses allowing decision makers to test how sensitive and/or robust the results are to changes in priorities and values (for details about this type of analyses, see Ferreira *et al*, 2012).

4.4. Discussion and implications to theory and practice

The system discussed in the previous sections was under trial and opened to discussion during six months at a public (State) secondary school in the region of Santarém, Portugal. Unfortunately, due to the stringent austerity measures implemented by the Portuguese government to address the debt crisis faced by the country, most public sector institutions experienced severe budget restrictions preventing them from implementing new management systems. The secondary school where the system was tested was no exception, and the school administration decided that due to the economic crisis it was not the right time to embrace a new project such as the one discussed here. It was felt, however, that the system proposed had considerable potential. In an attempt to gain further insights into the extent to which the proposed system was preferred to the one in place, we carried out a short survey. This survey aimed to assess the perception of six key decision-makers regarding the potential of the proposed system to assist them in the implementation of the four stages of the performance measurement process (ie *design, measurement, analysis* and *improvement*). In order to collect the decision-makers' perceptions, we used a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Table 5 presents the results obtained.

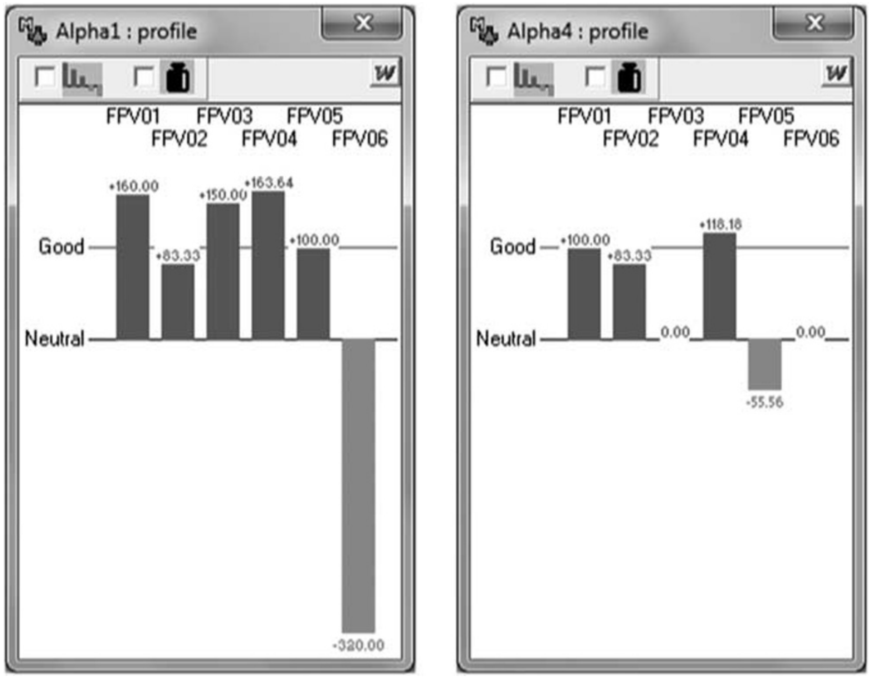

Figure 7 Impact profiles of alpha 1 and alpha 4.

As indicated in Table 5, the decision makers showed strong agreement regarding the value of the framework proposed in this paper to assist them in successfully carrying out the four stages of the performance measurement process.

With respect to the design stage, it was felt that the use of cognitive mapping and the MACBETH approach provide an important focus for discussion, which can be helpful in identifying appropriate performance indicators, in grouping them in a manner that makes them easier to be interpreted and understood and in making explicit what the trade-offs between the different performance indicators are. In addition, it was also agreed that the modelling process adopted is valuable to explore and reconcile important differences in the value systems held by the different stakeholders involved in the process, increasing consensus, ownership and commitment among them. Overall, it was agreed that the approaches used and the modelling process adopted can be valuable to design performance measurement systems reflecting the strategic priorities of schools.

In what regards the measurement stage, the decision makers agreed that the MACBETH approach can be very helpful in setting targets for each of the performance indicators and in assessing, in a quick and justifiable way, how each teacher performs both against each individual indicator and overall. There was strong consensus among the decision makers that the performance measurement system proposed can assist teachers in gaining a greater understanding of their own and others' pedagogical performances.

The value of cognitive mapping and the MACBETH approach to help decision makers analyse the performance results, and gain insights about the reasons why a particular level of performance was observed, was also acknowledged by the respondents of the questionnaire.

Finally, when asked about the value of the framework we proposed to aid the design of improvement solutions, the decision makers agreed that it can assist teachers to identify those areas where actions can lead to more significant performance improvements and to design corrective actions consistent with the objectives of the school. Overall, it was agreed that the fact that the performance assessment framework assists teachers to test the impact of alternative courses of action on their performances before they become operational is very valuable in deciding which corrective actions to implement.

These findings are important not only because they corroborate previous research (eg Santos *et al*, 2002; Ferreira *et al*, 2011) but also because they demonstrate the practical value of the framework. All in all, the results of a direct comparison between systems indicate that in the particular context under analysis, the evaluation system developed in this study surpasses the existing one in terms of *transparency, comprehension and functionally, time spent and practical utility*, and *fairness of the results*. Indeed, the decision makers considered the new system the 'overall best' option.

5. Conclusion

Teacher training is a continuous process, since both trainees and qualified teachers must periodically evaluate and update their own pedagogical practices. In this sense, teacher training and development is considered of extreme importance in the

Table 5 Results of the survey

		Minimum	*Maximum*	*Average*	*Mode*	*Standard Deviation*
Group I—Design						
The use of cognitive mapping and the MACBETH approach provides an important focus for discussion between the different stakeholders involved in the design of a performance measurement system, facilitating communication and learning regarding the pedagogical performance of teachers.		3	5	4.3	5	0.81650
The use of cognitive mapping and the MACBETH approach is very helpful in identifying appropriate performance indicators and grouping them in a manner that makes them easier to be interpreted and understood by the users of the system.		4	5	4.5	5	0.54772
The MACBETH approach provides valuable assistance in making explicit what the trade-offs between the different performance indicators are (ie in explicitly setting their relative importance).		3	5	4.2	5	0.98319
The modelling process combining cognitive maps and the MACBETH approach is helpful in exploring and reconciling important differences in the value systems held by the different stakeholders involved in the design and implementation of the performance measurement system.		4	5	4.7	5	0.51640
The use of cognitive mapping and the MACBETH approach is an important means to increase consensus, ownership and commitment among stakeholders involved in performance assessment.		4	5	4.7	5	0.51640
Overall, the approaches used and the modelling process adopted can be valuable to design performance measurement systems reflecting the strategic priorities of the school.		4	5	4.5	5	0.54772
Group II—Measurement						
The fact that the MACBETH approach allows levels of acceptable and unacceptable performance to be defined is helpful to set targets for each of the performance indicators.		4	5	4.5	5	0.54772
Through the MACBETH approach it is easy to understand how each teacher performs against each individual indicator and what the overall performance of each teacher is.		3	5	4.3	5	0.81650
The MACBETH approach adds transparency to the process of performance assessment making it easier to justify and explain to others why a particular level of performance is observed.		4	5	4.8	5	0.40825
The performance evaluation system developed based on cognitive mapping and MACBETH is quick to implement.		4	5	4.3	4	0.51640
Overall, the performance measurement system developed through the use of cognitive mapping and the MACBETH approach assists teachers in gaining a greater understanding about their own and others' pedagogical performances.		4	5	4.8	5	0.40825
Group III—Analysis						
The performance assessment framework developed through the use of cognitive mapping and the MACBETH approach is very valuable in assisting teachers to gain a greater understanding about the reasons why a particular level of performance is observed.		4	5	4.8	5	0.40825
The fact that the MACBETH approach allows to test how robust the performance scores obtained by each teacher are to changes on the weights of the different criteria is helpful to increase the acceptance of the results.		3	5	4.3	5	0.81650
Group IV—Improvement						
The fact that the MACBETH approach allows to assess how successfully each teacher is achieving his/her objectives is important to assist teachers identify corrective actions consistent with these objectives, whenever improvement is envisaged.		4	5	4.7	5	0.51640
The performance assessment framework developed through the integrated use of cognitive mapping and the MACBETH approach allows the decision makers to identify those areas where actions can lead to more significant performance improvements.		4	5	4.5	5	0.54772
Overall, the fact that the performance assessment framework assists teachers to test the impact of alternative courses of action on their performances before they become operational is very valuable to decide which corrective actions to implement.		3	5	4.5	5	0.83666
Group V—Overall evaluation						
Transparency, Comprehension and Functionally	Current System	1	3	2.0	2	0.89443
	Cognitive Mapping—MACBETH System	3	5	4.3	5	0.81650
Time Spent and Practical Utility	Current System	1	3	2.2	3	0.98319
	Cognitive Mapping—MACBETH System	3	5	4.0	5	0.89443
Fairness of the Results	Current System	1	3	2.0	2	0.89443
	Cognitive Mapping—MACBETH System	4	5	4.5	5	0.54772
Overall Evaluation	Current System	1	3	1.8	2	0.75277
	Cognitive Mapping—MACBETH System	4	5	4.5	5	0.54772

professional life of teachers because it helps them to acquire and/or improve their teaching skills. Starting from this premise, this paper has highlighted that information systems for the appraisal of pedagogical practices have emerged to improve the teaching activity and encourage pedagogical development of teachers. Still, the implementation of these systems has been interpreted in different ways, depending on each school's administration board and/or the person responsible for the evaluation process. As pointed out, this ambiguity is a serious obstacle to a transparent, equitable and fair appraisal of teacher pedagogical performance, and there is an urgent need for restructuring and ensuring compliance with the principles of transparency, fairness and justice during the evaluation of teachers.

Inspired in previous work by Bana e Costa and Oliveira (2012), who proposed the use of MACBETH for faculty appraisal, our study explored the possibility of creating pedagogical performance evaluation systems using cognitive maps and MACBETH. We know of no prior work reporting the integrated use of these two approaches in the context of academic staff performance evaluation.

By combining cognitive mapping and MACBETH, our study allowed for the conception of a comprehensive and defensible evaluation system. In particular, the information system proposed in this study allowed us to provide the participating decision makers with a more informed, transparent and accurate pedagogical performance evaluation system. The use of cognitive maps during the structuring phase of the decision-making process allowed us to: (1) promote discussion among the participating group members; (2) 'reduce the omission rate of important criteria' (in the participants' own words); and (3) learn about the evaluation of pedagogical practices, based on a deeper understanding of the cause-and-effect relations among criteria. In its turn, by generating cardinal value functions capable of representing semantic preferences, MACBETH proved to be valuable in dealing with compensations among evaluation criteria, while adding simplicity and transparency into the decision-making process. Unlike the current evaluation practices, which are often difficult to interpret and are sometimes applied in an ambiguous manner, the evaluation system developed in this study allows each teacher to think about her/his pedagogical skills and, from a constructivist perspective, analyse her/his performance through the identification of strengths and weaknesses. Following this, it seems evident that the integrated use of cognitive maps and the MACBETH technique can inform and support the development of more effective performance evaluation systems by bringing new insights to the pedagogical performance context.

Notwithstanding the potential application demonstrated by the performance appraisal system proposed here, its idiosyncratic nature recommends caution in extrapolating the results to other contexts. As previously outlined, factors such as the duration of the sessions, facilitator skills, participants involved, circumstances undertaken, and so on, may impact the final result achieved. This means, for example, that the final version of the collective map could have been different, had the context or the participants involved been different or had the group sessions lasted longer (cf Ferreira *et al*, 2014b). In this regard, it seems important to note that the proposed evaluation system and the process that gave rise to it are important tools for learning, promoting reflection and identifying opportunities for performance improvement. As already observed by Zavadskas and Turskis (2011, p 398), 'most importantly perhaps was the finding that decision analysis can be useful to help multiple stakeholders understand what they agree and disagree about, focus on the things that they disagree about and explore options that are better for everyone involved'.

In looking ahead to future research, it seems important to: (1) conduct a panel study with a different set of decision makers (eg school administrators) to determine the robustness of our findings; (2) conduct a panel study within a different region or country to increase generalizability; and (3) compare the use of different MCDA techniques in this context (for a variety of MCDA techniques, see Belton and Stewart, 2002; Zopounidis and Doumpos, 2002; Ananda and Herath, 2009). Improvement or updates will hopefully strengthen the approach proposed herein.

Acknowledgements—A previous and unpublished version of this paper was presented at *8th International Forum on Knowledge Assets and Dynamics*, held in Zagreb, Croatia. The authors wish to thank the superb contribution and infinite willingness of the panel members: Anabela Gonzaga, Carla Jesus, Celina Justino, Liliana Gens, Vanda Martins and Vera Vieira. The authors gratefully acknowledge use of the services and facilities of the School of Management and Technology of Santarém, Portugal. The third author also acknowledges financial support from Fundação para a Ciência e a Tecnologia and FEDER/COMPETE (grant PEst-C/EGE/UI4007/2013).

References

Altet M (2000). *Análise das Práticas dos Professores e das Situações Pedagógicas*. Porto Editora: Porto.

Amado C, Santos S and Marques P (2012). Integrating the data envelopment analysis and the balanced scorecard approaches for enhanced performance assessment. *Omega: The International Journal of Management Sciences* 40(3): 390–403.

Ananda J and Herath G (2009). A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological Economics* **68**(10): 2535–2548.

Avalos B and Assael J (2006). Moving from resistance to agreement: The case of the chilean teacher performance evaluation. *International Journal of Educational Research* **45**(4/5): 254–266.

Bana e Costa C and Oliveira M (2012). A multicriteria decision analysis model for faculty evaluation. *Omega: The International Journal of Management Sciences* 40(4): 424–436.

Bana e Costa C, Corrêa E, Corte J and Vansnick J (2002). Facilitating bid evaluation in public call for tenders: A socio-technical approach. *Omega: The International Journal of Management Sciences* 30(3): 227–242.

Bana e Costa C, Corte J and Vansnick J (2012). MACBETH. *International Journal of Information Technology & Decision Making* **11**(2): 359–387.

Bana e Costa C, De Corte J and Vansnick J (2005). On the mathematical foundations of MACBETH. In: Figueira J, Greco S and Ehrgott M (eds). *Multiple Criteria Decision Analysis: The State of the Art Surveys*. Springer: New York, pp 409–442.

Bana e Costa C, Ensslin L, Corrêa E and Vansnick J (1999). Decision support systems in action: integrated application in a multicriteria decision aid process. *European Journal of Operational Research* **113**(2): 315–335.

Bana e Costa C, Fernandes T and Correia P (2006). Prioritisation of public investments in social infrastructures using multicriteria value analysis and decision conferencing: A case study. *International Transactions in Operational Research* **36**(3): 279–297.

Bana e Costa C, Oliveira C and Vieira V (2008). Prioritization of bridges and tunnels in earthquake risk mitigation using multicriteria decision analysis: Application to Lisbon. *Omega: The International Journal of Management Sciences* **36**(3): 442–450.

Bana e Costa C, Silva F and Vansnick J (2001). Conflict dissolution in the public sector: A case-study. *European Journal of Operational Research* **130**(2): 388–401.

Belton V and Hodgkin J (1999). Facilitators, decision makers, DIY users: Is intelligent multicriteria decision support for all feasible or desirable? *European Journal of Operational Research* **113**(2): 247–260.

Belton V and Stewart T (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers: Dordrecht, MA.

Desselle S, Mattei T and Vanderveen R (2004). Identifying and weighting teaching and scholarship activities among faculty members. *American Journal of Pharmaceutical Education* **68**(4): 1–11.

Dyson R (2000). Strategy, performance and operational research. *Journal of the Operational Research Society* **51**(1): 5–11.

Eden C (2004). Analyzing cognitive maps to help structure issues or problems. *European Journal of Operational Research* **159**(3): 673–686.

Eden C and Ackermann F (2001a). Group decision and negotiation in strategy making. *Group Decision and Negotiation* **10**(2): 119–140.

Eden C and Ackermann F (2001b). SODA—The principles. In: Rosenhead J and Mingers J (eds). *Rational Analysis for a Problematic World Revisited: Problem Structuring Methods for Complexity, Uncertainty and Conflict*. John Wiley & Sons: Chichester, pp 21–41.

Eden C and Ackermann F (2004). Cognitive mapping expert views for policy analysis in the public sector. *European Journal of Operational Research* **152**(3): 615–630.

Elmore H (2008). Toward objectivity in faculty evaluation. *Academe* **94**(3): 38–40.

Ferreira F (2013). Measuring trade-offs among criteria in a balanced scorecard framework: Possible contributions from the multiple criteria decision analysis research field. *Journal of Business Economics and Management* **14**(3): 433–447.

Ferreira F, Santos S and Rodrigues P (2011). Adding value to branch performance evaluation using cognitive maps and MCDA: A case study. *Journal of the Operational Research Society* **62**(7): 1320–1333.

Ferreira F, Santos S, Marques C and Ferreira J (2014a). Assessing credit risk of mortgage lending using MACBETH: A methodological framework. *Management Decision* **52**(2): 182–206.

Ferreira F, Santos S, Rodrigues P and Spahr R (2014b). Evaluating retail banking service quality and convenience with MCDA techniques: A case study at the bank branch level. *Journal of Business Economics and Management* **15**(1): 1–21.

Ferreira F, Santos S, Rodrigues P and Spahr R (2014c). How to create indices for bank branch financial performance measurement using MCDA techniques: An illustrative example. *Journal of Business Economics and Management* **15**(4): 708–728.

Ferreira F, Spahr R, Santos S and Rodrigues P (2012). A multiple criteria framework to evaluate bank branch potential attractiveness. *International Journal of Strategic Property Management* **16**(3): 254–276.

Gul R and Boman J (2006). Concept mapping: A strategy for teaching and evaluation in nursing education. *Nurse Education in Practice* **6**(4): 199–206.

Huber M (2002). Faculty evaluation and the development of academic careers. *New Directions for Institutional Research* **2002**(114): 73–83.

Inácio M and Salema M (2011). The competence of learning to learn in vocational training: Promoting reflective thinking of the trainer of trainers. *Procedia: Social and Behavioral Sciences* **29**(1): 611–619.

Junior H (2008). Multicriteria approach to data envelopment analysis. *Pesquisa Operacional* **28**(2): 231–242.

Keeney R (1992). *Value-Focused Thinking: A Path to Creative Decisionmaking*. University Press Harvard: Harvard, MA.

Keeney R, See K and von Winterfeldt D (2006). Evaluating academic programs: With applications to U.S. graduate decision science programs. *Operations Research* **54**(5): 813–828.

Kolb DA (1984). *Experiential Learning*. Prentice-Hall: Englewood Cliffs, NJ.

Liu S and Teddlie C (2003). The ongoing development of teacher evaluation and curriculum reform in the People's Republic of China. *Journal of Personnel Evaluation in Education* **17**(3): 243–261.

Maskit D (2011). Teachers' attitudes toward pedagogical changes during various stages of professional development. *Teaching and Teacher Education* **27**(5): 851–860.

Mills M and Hyle A (1999). Faculty evaluation: A prickly pair. *Higher Education* **38**(3): 351–371.

Murias P, Miguel J and Rodríguez D (2008). A composite indicator for university quality assessment: The case of Spanish higher education system. *Social Indicators Research* **89**(1): 129–146.

Politis Y and Siskos Y (2004). Multicriteria methodology for the evaluation of a Greek engineering department. *European Journal of Operational Research* **156**(1): 223–240.

Ramli N, Mohamad D and Sulaiman N (2010). Evaluation of teaching performance with outliers data using fuzzy approach. *Procedia—Social and Behavioral Sciences* **8**(1): 190–197.

Ramos G and Moraes R (2000). Avaliação do desempenho de professores numa perspectiva qualitativa: Contribuições para o desenvolvimento profissional de professores universitários. *Revista Iberoamericana de Educación* **108**: 1–18.

Richter D, Kunter M, Klusmann U, Lüdtke O and Baumert J (2011). Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. *Teaching and Teacher Education* **27**(1): 116–126.

Sahlberg P (2011). Developing effective teachers and school leaders: The case of Finland. In: Darling-Hammond L and Rothman R (eds). *Teacher and Leader Effectiveness in High-Performing Education Systems*. Alliance for Excellent Education and Stanford Center for Opportunity Policy in Education, pp 13–22.

Santiago P, Roseveare D, Amelsvoort G, Manzi J and Matthews P (2009). *Teacher evaluation in Portugal - OECD Review*. OECD Review: Paris.

Santos S, Belton V and Howick S (2002). Adding value to performance measurement by using systems dynamics and multicriteria analysis. *International Journal of Operations & Production Management* **22**(11): 1246–1272.

Santos S, Belton V and Howick S (2008). Enhanced performance measuring using OR: A case study. *Journal of the Operational Research Society* **59**(6): 762–775.

Simões G (2000). *A Avaliação do Desempenho Docente: Contributos para uma Análise Crítica*. Texto Editora: Lisbon.

Smith P and Goddard M (2002). Performance management and operational research: A marriage made in heaven? *Journal of the Operational Research Society* **53**(3): 247–255.

Sparks D and Loucks-Horsley S (1989). Five models of staff development for teachers. *Journal of Staff Development* **10**(4): 40–57.

Spooren P (2010). On the credibility of the judge: A cross-classified multilevel analysis on students' evaluation of teaching. *Studies in Educational Evaluation* **36**(4): 121–131.

Sproule R (2002). The underdetermination of instructor performance by data from the student evaluation of teaching. *Economics of Education Review* **21**(3): 287–294.

Stoklasa J, Talaŝová J and Holeček P (2011). Academic staff performance evaluation: Variants of models. *Acta Polytechnica Hungaric* **8**(3): 91–111.

Tegarden D and Sheetz S (2003). Group cognitive mapping: A methodology and system for capturing and evaluating managerial and organizational cognition. *Omega: The International Journal of Management Sciences* 31(2): 113–125.

Timoštšuk I and Ugaste A (2010). Student teachers' professional identity. *Teaching and Teacher Education* **26**(8): 1563–1570.

Uzoka F (2008). A fuzzy-enhanced multicriteria decision analysis model for evaluating university academics' research output. *Information Knowledge Systems Management* **7**(3): 273–299.

Weber M and Borcherding K (1993). Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research* **67**(1): 1–12.

Wolansky W (2001). A multiple approach to faculty evaluation. *Education* **97**(1): 81–96.

Woods N (2011). Evaluation matters: Lessons learned on the evaluation of surgical teaching. *The Surgeon* **9**(1): S43–S44.

Zavadskas E and Turskis Z (2011). Multiple criteria decision making (MCDM) methods in economics: An overview. *Technological and Economic Development of Economy* **17**(2): 397–427.

Zopounidis C and Doumpos M (2002). Multicriteria classification and sorting methods: A literature review. *European Journal of Operational Research* **138**(2): 229–248.

Received 4 March 2013;
accepted 3 December 2014 after two revision