A Multicriteria Decision Analysis Model for Faculty Evaluation

Carlos A. Bana e Costa, Paulo A. F. Martins, Mónica D. Oliveira, Amílcar Sernadas, Carlos A. Mota Soares

Instituto Superior Técnico, Technical University of Lisbon, Portugal

June 21st, 2010

AGENDA

- Context
- Model building process
- Faculty evaluation model
- Application of the model to IST
- Discussion
Universities: a changing environment

Global trends
- ↑ contribution of universities to society
- ↑ transparency in decisions related to academic staff performance

Portugal
- All universities presently defining evaluation systems to apply to all academic staff
- Opportunity for strategically aligning the activity of the academic staff with the university goals

Opportunity

Review
- Few studies attempting to evaluate the overall activity of the academic staff (Elmore, 2008)
- Different opinions:
  - Possible to measure with some precision faculty performance and performance measurements can be used in university management (Arreola, 2007)
  - Scientific activities cannot be fully measured given current knowledge and available indicators, and use of measurement tools might affect researchers’ autonomy and lead to undesirable effects (Adler & Harzing, 2009)
- Methodological difficulties in faculty performance evaluation
  - “existing metrics do not capture the full range of activities that support and transmit scientific ideas” (Lane, 2010)
- Major flaws in both substance and process in previous studies
Key challenges in faculty evaluation

Develop comprehensive evaluation systems
...based on methodologically sound procedures
...capable of reflecting differences between academic staff
...taking into account the university mission
...applicable to all faculty members and scientific areas while respecting their specificities.

...following a request: model to be used by Instituto Superior Técnico (IST), an engineering school with 778 faculty members working in a wide variety of scientific domains

Key questions not previously addressed in a comprehensive and systematic manner within faculty evaluation literature

1. When STRUCTURING the faculty evaluation model:
   - How to design a model reflecting the strategic objectives of the school and useful for human resources management?
   - How to define a coherent set of evaluation criteria projecting, in the various areas of academic activity (pedagogical, scientific, etc), stakeholders’ values and concerns about academic careers and institutional policies?
   - How to describe, as objectively and unambiguously as possible, the performance on each one of the criteria, taking into account and adequately integrating its quantitative and qualitative dimensions?
   - How to care for specificities of each one of the scientific domains of the school?
2. When modelling the MEASUREMENT of academics’ VALUE:
   – How to convert individual performance into perceived added value to the school?
   – How to assign relative weights to the criteria adequately reflecting value trade-off judgements between criteria?
   – How to appropriately aggregate added value on multiple criteria, within and across areas of activity, respecting the autonomy of each faculty member to choose to invest more in some activities rather than in others, while not allowing extreme performance compensation phenomena inconsistent with achieving an adequate balance among objectives?

3. How to set BOUNDARIES for the RATING CATEGORIES imposed by law so that the classification of each faculty member may reflect her or his intrinsic value to the school?
Methodological and contextual issues

Multicriteria approach for faculty evaluation

Structuring tasks
- Areas of activity and respective criteria
- Descriptors of quantitative & qualitative performance
- Targets and ceilings

Value measurement tasks
- Value functions
- Criteria weights and areas’ interval weights

Optimization procedure for overall scoring

Hierarchical additive aggregation procedure

Assignment to faculty rating categories

Make explicit what should be considered a “good performance” on each criterion, for a given scientific domain (and in a given evaluation period)

To prevent that a very high performance level in a single criterion may play an excessive undesirable role in compensating very weak performance in all remaining criteria

Converting performance into value

Should reflect the relative importance of achieving the targets (in a given evaluation period, from the perspective of the school)
The overall score resulting for each faculty member reflects the value of her or his specific profile.

Additive value procedure, firstly to aggregate value scores on criteria within each area of activity, and then to aggregate values across areas.

Optimization procedure for overall scoring

Hierarchical additive aggregation procedure

Assignment to faculty rating categories

Model requisiteness

Assignment to associate each faculty member with one rating category, with the several categories separated by thresholds of increasing overall value, combined or not with other assignment rules.

...when its form and content are sufficient to provide satisfactorily uncontroversial answers to the questions that motivated its development (as defined by (Phillips 1984)).

Methods to build multicriteria value measurement models

- Several theoretically sound methods (von Winterfeldt and Edwards 1986) (Kirkwood 1997) (Belton and Stewart 2001)

- We propose the use of MACBETH, the Measuring Attractiveness by a Categorical Based Evaluation Technique:
  - Asks only for qualitative pairwise comparison judgements of difference in value between stimuli (Bana e Costa, De Corte et al. Forthcoming)
  - Theoretical foundations and has been extensively applied in various evaluation contexts (Bana e Costa, De Corte et al. 2005) (Bana e Costa, Lourenço et al. 2008)
  - Interactive application with the M-MACBETH software (Bana e Costa, De Corte et al. 2003)
PROPOSED FACULTY EVALUATION MODEL

- Faculty member
- Evaluation criteria i from area of activity j
- Performance
- Quantitative performance
- Qualitative performance

\[
\begin{align*}
    P_{ij} &= Q_{ij}^t \times Q_{ij}^l \\
    V_{ij}^d &= V_{ij} \left( p_{ij}^d \right)
\end{align*}
\]
Building value functions

S-shaped value function with:
• Two branches defined by exponential functions;
• Exponential function respecting the delta property (or the constant trade-off attitude condition (Kirkwood, 1997));
• Desirable properties for faculty evaluation...

Hierarchical additive procedure

Optimization procedure for overall scoring

With:

Subject to:

\[ V^d_j = \sum_{i,j=1}^{N_j} V^d_{ij} \times w_{ij} \]

\[ V^d = \max \sum_{j=1}^{M} V^d_j \times w_j \]

\[ \sum_{i,j=1}^{N_j} w_{ij} = 1 \]

\[ w_{ij} > 0, \forall i, j \]

\[ V^d_j = 100 \]

\[ V^d_j = 0 \]

\[ w_j \leq \overline{w_j} \leq \underline{w_j} \]

\[ \sum_{j=1}^{M} w_j = 1 \]

\[ w_j \geq 0 \]
APPLICATION OF THE MODEL TO IST

Model applied for retrospective evaluation:


Two level hierarchical structure

Area of activity

Evaluation criteria
### Value judgements for building quantitative descriptors of performance

<table>
<thead>
<tr>
<th>Type of publication</th>
<th>Equivalent units</th>
</tr>
</thead>
<tbody>
<tr>
<td>International book</td>
<td>5.5</td>
</tr>
<tr>
<td>Article published in a type A journal</td>
<td>3</td>
</tr>
<tr>
<td>Article published in a type B journal</td>
<td>1.75</td>
</tr>
<tr>
<td>Chapter in an international book or Edition of international book</td>
<td>1.0</td>
</tr>
<tr>
<td>Article published in a type C journal</td>
<td>0.3</td>
</tr>
<tr>
<td>Article published in conference proceedings</td>
<td>0.2</td>
</tr>
</tbody>
</table>

### Qualitative descriptor of performance

<table>
<thead>
<tr>
<th>Levels of performance</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly positive</td>
<td>There is at least one ‘strong’ determinant point and no ‘weak’ determinant points</td>
<td>1.5</td>
</tr>
<tr>
<td>Positive</td>
<td>There are no ‘strong’ nor ‘weak’ determinant points, and ‘strong’ points more than compensate ‘weak’ points</td>
<td>1.25</td>
</tr>
<tr>
<td>Neutral</td>
<td>There was no identification of neither ‘strong’ nor ‘weak’ determinant points, or the ‘strong’ points are balanced with ‘weak’ points</td>
<td>1.0</td>
</tr>
<tr>
<td>Negative</td>
<td>There are no ‘strong’ nor ‘weak’ determinant points, and ‘weak’ points more than compensate ‘strong’ points</td>
<td>0.75</td>
</tr>
<tr>
<td>Highly negative</td>
<td>There is at least one ‘weak’ determinant point and no ‘strong’ determinant points</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Examples of targets and ceilings

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>Target</th>
<th>Examples of targets (to be interpreted with quality = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedagogical related publications</td>
<td>Ecp 1.5</td>
<td>1 book chapter and 1 pedagogical text</td>
</tr>
<tr>
<td>Students supervision</td>
<td>Eao 6</td>
<td>2 supervision of MSc thesis per year</td>
</tr>
<tr>
<td>Courses teaching</td>
<td>Euc 9</td>
<td>9h of teaching courses per week with normal evaluation by students</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>Pedagogical related publications</th>
<th>Students supervision</th>
<th>Courses teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ecp</td>
<td>Eao</td>
<td>Euc</td>
</tr>
<tr>
<td>$V(\text{ceiling})$</td>
<td>500</td>
<td>300</td>
<td>300</td>
</tr>
</tbody>
</table>

Value functions
### Building weights

<table>
<thead>
<tr>
<th>Area of activity</th>
<th>Pedagogical</th>
<th>Research</th>
<th>Knowledge transfer</th>
<th>University management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight by area of activity</td>
<td>20%-40%</td>
<td>40%-60%</td>
<td>0%-20%</td>
<td>0%-20%</td>
</tr>
<tr>
<td>Evaluation criteria</td>
<td>Ecp</td>
<td>Eao</td>
<td>Euc</td>
<td>Ipj</td>
</tr>
<tr>
<td>Weight by evaluation criteria</td>
<td>1/6</td>
<td>2/6</td>
<td>3/6</td>
<td>6/8</td>
</tr>
<tr>
<td>Weight by evaluation criteria (normalized)</td>
<td>6.7%</td>
<td>13.3%</td>
<td>20%</td>
<td>45%</td>
</tr>
</tbody>
</table>

### Model outputs

#### 1. STAFF MEMBER PROFILE

- **V(target)** = 100
- **V(ceiling)** = 500

#### 2. STAFF SCORE

- 600 (Excellent)
- 300 (Very good)
- 500 (Good)
- 300 (Inadequate)

#### 3. FACULTY RATING CATEGORY

- Excellent
- Very good
- Good
- Inadequate
DISCUSSION

Effectiveness of the model yet to be confirmed by large scale application!

1. Only the implementation of the model will show whether it is effectively dealing with differences across scientific domains, and whether adjustments are required.

2. Concerns with the calibration of the model when different targets and other values across scientific domains are used.

3. The incentives motivated by the model adoption have not been studied in detail.

4. It is not clear which is the level of acceptability of the model within the school.
Improvements to the model (I)

- Being informed by literature in specific areas, e.g.,
  - Should evaluation of teaching be done only by students, or peers should also review the content of teaching? If peer review of the context of teaching is required, which methods for evaluation are available and have been validated?
  - Which is the best indicator for the impact of publications on the community?

- Using detailed information on the performance of IST academic staff:
  - The use of high-quality data and of scientific metrics might contribute to build a sounder model and to a higher level of acceptation.

Improvements to the model (II)

- Using participatory mechanisms.

- Developing multiple criteria interactive analysis tools for the collection and analysis of model inputs and outputs.

- Developing multicriteria methods, for example:
  - Which procedures should be used for validating the chosen descriptors of performance?
  - Should thresholds be linked with targets and ceilings?

- Testing the model within schools other than engineering.
Some questions for discussion…

- Performance appraisal: which viewpoint?
  - Rules for dividing the score for publications by the number of authors: individual vs. university viewpoints

- Value functions: avoiding complexity?

- Which incentives are being created? Simultaneous to faculty evaluation, which other tools might be used?
  - Regulating levels of teaching activity
  - Rewards to teaching

References


