

Available online at www.sciencedirect.com**ScienceDirect**

Procedia Computer Science 55 (2015) 539 – 548

Procedia
Computer Science

Information Technology and Quantitative Management (ITQM 2015)

Professor Assessment Using Multi-Criteria Decision Analysis

Nelson Hein^{a*}, Adriana Kroenke^{b*}, Moacir Manoel Rodrigues Júnior^{c*}^{a,b,c} *Department of Mathematics - FURB, Antônio da Veiga, Blumenau 89012-900, Brazil*

Abstract

This paper presents results from the pilot project on teacher assessment of the Regional University of Blumenau. It contains discussions about the project was elaborated by the team of internal assessments of the university, discussing the dimensions, issues and quantitative analysis techniques and the construction of the FURB Professor's Quality Index (IFQD). The multi-criteria tools used were the consensus theory, information entropy and the TOPSIS multi-criteria decision analysis method. Two undergraduate courses were assessed totaling 56 professors were evaluated by a group of 1979 students.

© 2015 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of ITQM 2015

Keywords: Professor evaluation; Multi-criteria Decision Analysis; Consensus; TOPSIS; Entropy.

1. Introduction

The responsibility of universities, when facing their challenges and goals, requires self-evaluation regarding their daily work, regarding teaching. Regarding management research and all the science outreach that take place in it. This self-assessment requires the participation of the entire academic community, which formulates judgments about the value of institutional quality.

The faculty is an important entity when it comes to the quality of a university. The internal assessment of the faculty reflects the future institutional perspective, in which the professors' action is one of its elements. As highlighted by the Institutional Evaluation Program of Brazilian Universities (IEPBU), this evaluation should not be a punitive process, not even for restraint, but it must present indicators that aim at quality (BRAZIL, 1994; BRAZIL, 2004).

This project presents the results of a study that diagnoses and evaluates the management of education from the perspective of Chemical Engineering and Pedagogy students of the Regional University of Blumenau (Universidade Regional de Blumenau - FURB), located in Vale do Itajaí, in Blumenau (Santa Catarina). The project and its results are products of a pilot program, which is fully institutionalized in all of FURB's under graduation programs, summarizing a total of 40 programs, 842 professors and about 10200 students.

For the assessment, a number of issues that were previously questioned by the assessed professors were resolved: the significance of the sample and the consistency of the answers marked by the students - the evaluators. Regarding the statistical significance of the data, the size of the sample for each group of students

was calculated, divided into ranges, for each assessed discipline (allowing one teacher to be evaluated by more than one class). To solve the problem of consistency, the concept of average consistency per question was introduced, generating the information to the evaluation and the entropy, in order to generate the information weight.

All of those result in one indicator, called FURB Professor's Quality Index (IFQD), limited on a scale between 1-5, similar to the one adopted by the Brazilian institution called INEP (National Institute of Education and Research). Multi-criteria decision analysis techniques were used to obtain the aforementioned index, along with other quantitative techniques.

2. Theoretical Reference

The first university assessment projects developed by Brazilian universities emerged between the 80's and 90's, from the necessity of evaluating the university, due to criticism published in the media about these institutions (BOTH, 2005). From that moment on, the term "institutional assessment" was consecrated. Thus, the evaluation movement developed from the necessity to stimulate improvements in the university management.

The issue of institutional assessment led to the emergence, in 1993, of the Institutional Evaluation Program of Brazilian Universities (IEPBU), which is the instrument used by the government to define the project that finances other projects for institutional assessment. This instrument sets out the principles and objectives of the Brazilian institutional assessment, so that each university can perform activities in the pursuit of quality improvement (BRASIL, 1994).

In 1994, the IEPBU established seven guiding principles of the assessment process: globality, comparability, respect to institutional identity, no reward or punishment, voluntary membership, process legitimacy and process continuity. It is possible to realize that the characteristics of each institution must be observed, taking into account the forces acting on their specific environment, the institutional mission and its institutional design. That way, it can be understood that the institutional assessment is characterized as a systemic and continuous process that allows reflection upon the developed activities, allowing continuous improvement of quality in higher education.

Subsequent to IEPBU, Decree 2.036 of October 10, 1996 established the procedures for the assessment of programs in higher education institutions.

Thereafter, on April 14, 2004, Law 10,861 created the National Assessment System of Higher Education (SINAES) (BRAZIL, 2004) which is based on the necessity to promote the improvement of quality in the higher education system. In addition to that, it aims at providing guidance and expansion of its magnitude, at permanently increasing its social and academic effectiveness and, especially, at deepening its commitments and social responsibilities.

According to the INEP/SINAES (INEP: Instituto Nacional de Estudos e Pesquisas Educacionais; SINAES: Sistema Nacional de Avaliação da Educação Superior) document (BRAZIL, 2004) the assessment of higher education institutions have an educational character since it aims at improving the agents of the academic community and the institution as a whole.

The self-assessment at FURB began in 1995. Since then the institution is committed to the process of self-assessment, not just for a moment, but with a constant and renewable character. This fact allows an almost global view of the institution, which facilitates the resizing of policies and plans for the university.

The evaluation of professors started at the same time, and its model is modified from time to time, but without any major breakthrough in its analysis.

3. Theoretical Reference

The research is characterized as exploratory, with bibliographical, documentary and data collection procedures. The analysis and the data interpretation are quantitative, The pilot project involved a population comprised of 1979 students and 56 professors of the Regional University of Blumenau. The evaluation was performed by taking 58 groups of the programs offered in the 1st semester of 2013 from the programs of Chemical Engineering and Pedagogy.

The questionnaire consists of two dimensions: (a) **social and ethics** and (b) **teaching and learning**, distributed in 19 questions, which were organized as follows:

(a) Social and Ethics Dimension

Teacher attitude. Throughout the period did the professor:

- 1) The professor established a relationship of courtesy, respect and attention to students?
- 2) The professor encouraged the students to question fundamentals, theories, concepts, etc.?
- 3) The professor encouraged students to find and to participate effectively in discussions, exercises and activities?
- 4) The professor encouraged the students to integrate knowledge with other related disciplines?
- 5) The professor plans the classes, meeting the schedule set for the course plan?
- 6) The professor recorded the punctuality and attendance of the students?
- 7) The professor encouraged students to value the program, promoting a good image of the institution?
- 8) The professor demonstrated to be updated with the trends in the area of his/her discipline citing authors, events, experiences, and other professionals?

(b) Teaching and Learning Dimension

Regarding the organization of the professor's teaching/learning actions did:

- 1) The professor presented (discussed) the course plan?
- 2) The professor has made clear the goals of the course?
- 3) The professor taught the contents of the course according to the teaching/learning course plan?
- 4) The professor demonstrated clarity, preparation and objectivity in explaining the content of the course?
- 5) The professor pointed out the relevance and/or the application of the content that was studied?
- 6) The professor has indicated research sources (literature, websites, ...) that were appropriate to the ones proposed by the course?
- 7) The professor demanded the content developed in the learning assessments according to the objectives of the teaching/learning plan?

In the conduction of the teaching/learning actions, the professor did:

- 1) The professor elaborated and implement tools and assessment procedures according to the ones provided in the teaching/learning plan?
- 2) The professor used diversified and adequate didactic procedures to the goal(s) of the course?
- 3) The professor used diversified and adequate assessment instruments to the goal(s) of the course?
- 4) The professor analyzed and commented to the students the results of their tests, assessments and other activities undertaken by correcting errors and clarifying questions and doubts?

Only the groups of students that achieved samples with 95% of confidence were analyzed (BETHLEHEM, 2002). The scheme to legitimize the process will be given as following.

Table 01. Size of sample per group

# Students	# Sample
2<N≤10	N-1
10<N≤20	N-2
20<N≤30	N-3
30<N≤40	N-4
40<N≤50	N-6
50<N≤60	N-8
60<N≤70	N-10
70<N≤80	N-13

Source: Elaborated by the authors.

The scale used is the Likert ordinal of 5 points, scoring as follows: Strongly Disagree (1), Disagree (2), Neither agree nor disagree (3), Agree (4) and Strongly Agree (5). Even having a scale that features a number of technical limitations, the merit of being known by students and of having a neutral point cannot be taken away, in the case of a 5-point scale the value '3' is the neutral point, which can be understood as 0 (zero) on a scale of -2 to +2.

Besides the use of the scale, it was incorporated into the consensus the picture evaluated by the expression:

$$CONS(X) = 1 + \sum_{i=1}^n p_i \text{Log}_2 \left(1 + \frac{|X_i - \mu_X|}{d_X} \right)$$

Where X is an ordinal variable (question); p_i is the percentage associated to each X_i ; d_X is the scale range, which in this case is $d_X = 4$ and μ_X is the average. The consensus will underpin the average score obtained because it will serve as its variability criterion. According to Wierman and Tastle (2007) the consensus should be so interpreted.

Table 02 – Classification of the consensus degree

Range	Classification
$CONS(X) \geq 90\%$	Very Strong Consensus
$80\% \leq CONS(X) < 90\%$	Strong Consensus
$60\% \leq CONS(X) < 80\%$	Moderate Consensus
$40\% \leq CONS(X) < 60\%$	Balance
$20\% \leq CONS(X) < 40\%$	Moderate Dissensus
$10\% \leq CONS(X) < 20\%$	Strong Dissensus
$CONS(X) < 10\%$	Very Strong Dissensus

Source: Wierman e Tastle (2007).

The data in the simulation model below were collected in the survey conducted by the student body. However, they are part of the investigation and serve only to provide an algorithmic and numerical understanding regarding the design of the assessment of the professors.

Below, a simulation in which 10 professors are evaluated is presented, being each one of them assessed by 10 students in a questionnaire of 8 questions. To illustrate, a table with the simulated responses is shown in order to help understanding the technique.

Table 03 - Report of the answers to the questions given by students

Student	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
A1	3	2	1	1	1	5	4	1
A2	4	3	4	2	1	4	4	1
A3	5	4	5	3	2	3	5	1
A4	4	3	4	4	1	4	5	2
A5	5	2	5	5	1	3	5	1
A6	4	3	4	4	5	2	5	1
A7	5	4	5	3	3	2	5	1
A8	3	2	4	2	5	3	5	1
A9	2	1	3	1	5	4	5	1
A10	3	2	4	2	5	3	5	1
Average	3.8	2.6	3.9	2.7	2.9	3.3	4.9	1.1
Cons(X)	0.64	0.66	0.62	0.50	0.16	0.68	0.93	0.93

Source: Partial research data.

The columns of the questions Q7 and Q8 have both a 93% consensus. However, it should be understood that the "strong consensus" impacts on the average. since it confirms the high scores received in question Q7 and the low score received in question Q8. Note that in question Q4, with an average of 2.7. the consensus was of only 50%. and that in question Q5. with an average of 2.9. there was a consensus index of only 16%. The total score of each professor in every question will be given by the product of the average by their consensus: Value = Average × Consensus. In the case of this particular professor, his/her score will be determined as follows:

Table 04: Value associated to each question to the professor being assessed

Professor	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Value	2.43	1.72	2.42	1.34	0.46	2.24	4.56	1.02

Source: Research data

Assuming now that the table below shows the values obtained by consensus of the 10 professors assessed. Besides this analysis each question received an importance factor, calculated by entropy (ZELENY, 1982) present in each batch of responses, i.e., information theory tools were used, calculated by:

$$ENT(X) = -\alpha \sum_{k=1}^n \frac{x_i^k}{X_i} \ln \left(\frac{x_i^k}{X_i} \right)$$

Table 05 - Values of the questions of the assessed professors

Prof.*	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
PR1	2.43	1.72	2.42	1.34	0.16	2.24	4.56	1.02
PR2	3.46	2.33	3.42	3.45	1.23	4.78	4.32	2.34
PR3	3.78	2.67	3.78	4.02	1.56	4.56	3.46	4.31
P44	2.45	3.42	2.25	1.24	0.67	4.02	3.21	3.67
PR5	1.22	1.78	1.23	2.34	1.54	3.45	3.56	3.21
PR6	4.02	3.45	4.56	3.21	1.56	2.32	3.56	2.89
PR7	3.57	3.42	3.33	2.67	0.98	1.21	3.01	2.56
PR8	2.41	3.56	2.67	3.54	1.89	0.56	2.23	3.12
PR9	4.12	2.42	3.45	3.21	2.03	0.34	0.89	3.42
PR10	3.51	4.67	2.67	1.78	2.12	1.45	3.76	2.65
Max	4.12	4.67	4.56	4.02	2.12	4.78	4.56	4.31
Min	1.22	1.72	1.23	1.24	0.16	0.34	0.89	1.02
ENT*	0.980	0.981	0.980	0.972	0.950	0.907	0.975	0.980
Inf%	7.1	6.9	7.4	10.0	18.3	33.7	9.0	7.4

Source: Research data.

(*) Professor

(**) Entropy

The information value (**Inf%**), i.e., the weight of each point is given by the portion of the total entropy in the total of 8 issues:

$$\omega_i = \frac{1 - e(P_i)}{n - E}$$

and (P_i) is the entropy of each question, for $n = 8$ is the total possible entropy and $E = 7.725$, which was the entropy generated by the 8 questions.

Once the professor's scores matrix was generated, the problem became a multi-criteria analysis problem, which aims at ranking the group of professors by using as the criteria, the performance in each of the questions that were answered.

Over the past few decades, decision-making has gained popularity due to its frequent implications in managerial domains as it enables decision makers to come up with preeminent decisions. This explicit the importance of improved decision making processes given the competitive and dynamic business environment these days. Multi-criteria decision making (MCDM) - a well known decision making process is based on the progression of using methods and procedures of multiple conflicting criteria into management planning processes, whereas, Decision Support Systems (DSS) are considered powerful tools for decision-making. MCDM is widely used in conjunction with Decision support systems (DSS) by a large number of decision makers in variety of fields, such as financial analysis, flood risk management, housing evaluation, disaster management and Customer relationship management. Apart from several diversified advantages of using MCDM-DSS architecture, certain issues are also attached to this highly useful decision-making methodology (BRUHGA, 2004)

The multi-criteria analysis model to be used is the TOPSIS, which stands for **T**echnique for **O**rders **P**reference by **S**imilarity to **I**deal **S**olution (HWANG & YOON, 1981). The model starts with a decision matrix composed by alternative (professors) and criteria (questions).

$$A = \begin{bmatrix} v_{11} & \dots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \dots & v_{mn} \end{bmatrix}$$

The technique is composed by the following steps:

1st Step: Calculation of A⁺ positive ideal solutions (benefits) and A⁻ negative ideal solutions (costs), as follows:

$$A^+ = (p_1^+ \cdot p_2^+ \cdot \dots \cdot p_m^+) \text{ and } A^- = (p_1^- \cdot p_2^- \cdot \dots \cdot p_m^-).$$

where:

$$p_j^+ = \{Max_i p_{ij} \cdot j \in J_1; Min_i p_{ij} \cdot j \in J_2\}$$

$$p_j^- = \{Min_i p_{ij} \cdot j \in J_1; Max_i p_{ij} \cdot j \in J_2\}.$$

where: J₁ and J₂ represent the benefit and cost criteria respectively. In the subject in better the goal is the pursuit of benefits, i.e., the higher the performance in the Likert scale, the better.

2nd Step: The calculation of the Euclidean weighted distances between the benefits is conducted in such way:

$$d^+ = \sqrt{\sum_{j=1}^n w_j (p_j^+ - p_{ij})^2}, \text{ with } i=1 \dots m \text{ and}$$

$$d^- = \sqrt{\sum_{j=1}^n w_j (p_j^- - p_{ij})^2}, \text{ with } i=1 \dots m.$$

The values of w_i are the degree of importance of each question, obtained by the entropy of each one of them.

3rd Step: Calculus of the relative proximity: $\xi_i = \frac{d_i^-}{d_i^+ + d_i^-}$.

To exemplifying the situation is used a questionnaire for evaluate 10 professors. With the values are obtained the averaging and consensus.

Table 06 – Data of the imulation

Professor	V1	V2	V3	V4	V5	V6	V7	V8
PR1	2.43	1.72	2.42	1.34	0.16	2.24	4.56	1.02
PR2	3.46	2.33	3.42	3.45	1.23	4.78	4.32	2.34
PR3	3.78	2.67	3.78	4.02	1.56	4.56	3.46	4.31
PR4	2.45	3.42	2.25	1.24	0.67	4.02	3.21	3.67
PR5	1.22	1.78	1.23	2.34	1.54	3.45	3.56	3.21
PR6	4.02	3.45	4.56	3.21	1.56	2.32	3.56	2.89
PR7	3.57	3.42	3.33	2.67	0.98	1.21	3.01	2.56
PR8	2.41	3.56	2.67	3.54	1.89	0.56	2.23	3.12
PR9	4.12	2.42	3.45	3.21	2.03	0.34	0.89	3.42
PR10	3.51	4.67	2.67	1.78	2.12	1.45	3.76	2.65

Source: Research data.

For each question is marked the best (A⁺) and the worst (A⁻) result. These points are the references to calculate the distances for each professor to the best scenario (d⁺) and the worst scenario (d⁻). In this case:

$$A^+ = \{4.12, 4.67, 4.56, \dots, 4.31\} \text{ (Best)}$$

$$A^- = \{1.22, 1.72, 1.23, \dots, 1.02\} \text{ (Worst)}$$

Table 07 – Relative proximity and IFQD

Professor	d _i ⁺	d _i ⁻	ξ _i	IFQD
-----------	-----------------------------	-----------------------------	----------------	------

PR1	2.35	4.09	0.634	3.54
PR2	0.99	6.15	0.861	4.45
PR3	0.72	7.10	0.908	4.63
PR4	1.50	4.97	0.768	4.07
PR5	1.77	4.04	0.696	3.78
PR6	1.58	6.69	0.808	4.23
PR7	2.32	5.24	0.689	3.75
PR8	2.68	4.82	0.643	3.57
PR9	2.90	5.33	0.647	3.59
PR10	2.19	5.95	0.731	3.92

Source: Research data

It is important to add that the model generates a final rank on the scale adopted by INEP to be attributed to the Preliminary Program Concept (PPC) and the General Index of Programs (GIP) which is given by the range [1, 5] that in the pilot project received the name of IFQD (FURB Professor's Quality Index).

4. Results Analysis

The questionnaires were completed through an online form, forming a database in Access, once they are extracted, they generated reports, after that they went through all the techniques already presented.

The IFQD of each of the 58 professors, of the two analyzed programs were those that appear in Tables 8 and 9.

Table 08: IFQD of the faculty from the Chemistry Engineering program

Prof.	IFQD	Prof.	IFQD	Prof.	IFQD	Prof.	IFQD	Prof.	IFQD
PR1	2.61	PR11	2.80	PR21	3.51	PR31	3.17	PR41	3.40
PR2	3.28	PR12	2.98	PR22	3.03	PR32	2.89	PR42	3.66
PR3	3.20	PR13	3.91	PR23	3.15	PR33	3.00	-	-
PR4	3.27	PR14	2.57	PR24	4.33	PR34	3.80	-	-
PR5	3.09	PR15	1.82	PR25	2.90	PR35	1.59	-	-
PR6	4.42	PR16	3.33	PR26	3.61	PR36	1.91	-	-
PR7	2.94	PR17	2.65	PR27	3.92	PR37	1.53	-	-
PR8	2.80	PR18	1.97	PR28	3.03	PR38	3.80	-	-
PR9	4.48	PR19	3.97	PR29	2.13	PR39	3.02	-	-
PR10	2.45	PR20	2.75	PR30	3.62	PR40	2.86	-	-

Source: Research data.

Table 09: IFQD of the faculty from the Pedagogy program

Prof.	IFQD	Prof.	IFQD	Prof.	IFQD	Prof.	IFQD
PR43	4.04	PR47	3.44	PR51	2.74	PR55	2.22
PR44	1.96	PR48	4.34	PR52	3.22	PR56	3.34
PR45	4.42	PR49	4.54	PR53	1.51	PR67	3.40
PR46	4.13	PR50	4.09	PR54	3.98	PR58	1.43

Source: Research data

The minimum IFQD rate among the Chemical Engineering professors is 1.53 (PR37) and the maximum is 4.48 (PR9). In the Pedagogy program the minimum was 1.43 (PR58) and the maximum was 4.54 (PR49).

5. Conclusion

The model that is in use in the Regional University of Blumenau is the result of many previous attempts to improve the system of self-evaluation. The presented model emerged from the professors' questionings, who are evaluated every six months, and the dissemination of the results from these evaluation, which caused embarrassment and discontent on the part of those involved.

The use of the arithmetic average by itself precluded further analysis. The use of the concepts of consensus, information weight and, specifically, the introduction of IFQD reduced complaints and currently serves as a way to hire professors for the university programs, serving as a decision aid by coordinators and department managers.

The TOPSIS method has numerous advantages. Is "an approach to identify an alternative which is closest to the ideal solution and farthest to the negative ideal solution in a multi-dimensional computing space" (QIN *et al.*, 2008, p. 2166). It has a simple process. It is easy to use and programmable. The number of steps remains the same regardless of the number of attributes (IC, 2012). A disadvantage is that its use of Euclidean Distance does not consider the correlation of attributes. It is difficult to weight attributes and keep consistency of judgment, especially with additional attributes. TOPSIS has been used in supply chain management and logistics, design, engineering and manufacturing systems, business and marketing management, environmental management, human resources management, and water resources management. Many of the uses seen in the literature review had TOPSIS confirm the answers proposed by other MCDM methods. The advantage of its simplicity and its ability to maintain the same amount of steps regardless of problem size has allowed it to be utilized quickly to review other methods or to stand on its own as a decision-making tool.

References

- [1] BETHLEHEM, J.G. Weighting nonresponse adjustments based on auxiliary information. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (Eds.), *Survey Nonresponse*. Wiley, New York, 2002, pp.275-287.
- [2] BOTH, Ivo. Da avaliação precursora ao comportamento da avaliação: PAIUB e SINAES construindo pontes. *Revista da Rede de Avaliação Institucional da Educação Superior*. n. 4. p. 61-73. 2005.
- [3] BRUGHA, C.M. Structure of Multi-Criteria Decision-Making, *The Journal of the Operational Research Society*, Vol. 55, No. 11, pp. 1156-1168, 2004.
- [4] IC, Y. An experimental design approach using TOPSIS method for the selection of computer-integrated manufacturing technologies. *Robotics and Computer-Integrated Manufacturing*, 28(2), pp. 245-256, 2012.
- [5] HWANG, C.L.; YOON, K.. *Multiple attribute decision making: methods and applications*. Springer, New York. 1981.

- [6] HIGHER EDUCATION SECRETARY. *Programa de avaliação das universidades brasileiras (PAIUB)*. Brasília: Ministério da Educação. Secretaria de Educação Superior. 1994.
- [7] HIGHER EDUCATION SECRETARY. *Sistema nacional de avaliação da educação superior (SINAES)*. Brasília: Ministério da Educação (MEC). Secretaria de Educação Superior. 2004.
- [8] Qin, X., Huang, G., Chakma, A., Nie, X., and Lin, Q. (2008). A MCDM-based expert system for climate-change impact assessment and adaptation planning – A case study for the Georgia Basin, Canada. *Expert Systems with Applications*, 34(3), 2008, pp: 2164-2179.
- [9] WIERMAN. M. J.; TASTLE. W. J.. Consensus and dissention – a measure of ordinal dispersion. *International Journal of Approximate Reasoning*. New York. pp. 531 – 545. 2007.
- [10] ZELENY. Milan. *Multiple Criteria Decision Making*. New York:Mc Graw-Hill. 1982.