A 360-Degree Evaluation Framework for Doctoral Programs

María del Carmen Calatrava Moreno
Electronic Commerce Group
Vienna University of Technology
Vienna, Austria
mc.calatrava.moreno@ec.tuwien.ac.at

***Abstract*—Due to increasing international competitiveness, higher education institutions face the need to assess and enhance the quality of their activities. Usually, industrial quality management models are employed, which successfully assess administrative functions but fail to address educational processes.**

In this paper we present an assessment framework that overcomes the shortcomings of existing quality evaluation models. Inspired by the 360-degree feedback, we collect the perceptions of the individuals related to an educational program. Furthermore, with the introduction of a dual-scale assessment, we enable stakeholders to judge not only the fulfillment of each criterion but also its relevance. Together with the self-definition of the program, this allows the validation of the program's objectives, as well as quality assessments with a focus on both the program's and the stakeholders' priorities. We provide a formal definition, a set of quality criteria and a categorization of the relevant stakeholders for the assessment of doctoral programs in Computer Science. Furthermore, we present the analytical methods for the weighting and aggregation of information.

Keywords*—*Quality evaluation, quality assessment, 360-degree feedback, higher education, doctoral education.

I. Introduction

Despite the increasing efforts to implement quality measures in Higher Education Institutions (HEIs), there is still no consensus on how to best manage quality. Sophisticated industrial quality models are now being used within a large number of HEIs but have only been applied with partial success. The managerial and economic rationalism behind them make them suitable to assess administrative and service functions, as well as to account resources and quantitative results. They fail, however, to address the complex nature of educational processes, such as teaching, learning, research or knowledge transformation [1].

In this paper, we introduce a framework that overcomes the main shortcomings of existing industrial quality frameworks applied in HEIs. It is based on the 360-degree feedback methodology, a widely recognized method used in human resource management. The main contribution of this paper is the adaptation of this framework as an internal audit of educational programs in the context of HEIs. This includes the explanation of the analytical methods for the weighting and aggregation of information, as well as the description of its implementation for the assessment of doctoral programs in Computer Science, based on our ongoing implementation efforts.

II. Related Work

In this section we briefly review the methodologies most extensively used for internal assessment at HEIs. A complete survey of various models is beyond the scope of this paper and we refer to the works of Grant et al. [2], Srikanthan et al. [3].

The most basic approach to education assessment is the course experience questionnaire, which is frequently used to evaluate teaching and learning performance. The increasing importance of quality issues motivated the use of more sophisticated approaches that rely on industrial quality management models [1]. Among them, *Total Quality Management* (TQM) is one of the most frequently employed [4]. TQM is a generic umbrella for quality frameworks for the management of overall quality measures, organization activities and members within the organization.

In the context of HEIs, TQM frameworks such as EFQM [5], QFD [6], or SERVQUAL [7] and its adaptations have proved to address service and administrative functions successfully [8], [9]. They fail, however, to assess educational processes. Their strong focus on customer satisfaction is often inappropriate to measure educational matters [10]. In the educational context, TQM frameworks equate students with customers and ignore the opinions of other crucial stakeholders. Consequently, there is a growing concern that the increasing use of industrial assessment approaches might affect the quality of educational processes [3], [11], [12].

III. 360-degree Feedback

In this section we review the methodology that serves as a basis for our framework: the 360-degree feedback, also known as 360-degree performance appraisal.

In its original context, the 360-degree approach aims to provide a better understanding of the strengths, weaknesses and results of an individual, the appraisee, during a period of time. Information about the practice and performance of the appraisee is collected from different points of view. This involves the gathering of feedback from various *stakeholders*. These are both people within the organization such as staff, colleagues, subordinates and supervisors, as well as people from outside the organization who are professionally related to the appraisee. The evaluation is performed according to a set of criteria, and in some cases a self-evaluation is incorporated. The basic idea is that the feedback is forthcoming from

all individuals "around" the appraisee, hence the term "360-degree". Some of the main advantages of 360-degree feedback over single-source or top-down feedback are [13]:

- **Multiple-perspective judgment:** Often particular aspects under assessment are not observable by all stakeholders. In order to obtain a full picture of the appraisee, inputs from all perspectives are needed.
- **Assessment validity increase:** The involvement of numerous stakeholders with different expertise reduces the possible bias by providing complementary opinions. The aggregation of all points of view contributes to deliver a trustful rounded portrait of personal performance.
- **Performance and strategy alignment:** The result of such a comprehensive feedback can be used to better align individual's competencies with the overall mission and vision of an organization.

The results of a 360-degree evaluation are not only used by management teams to design administrative or human decision strategies. Appraisees may also use the results to develop specific plans for their professional development.

IV. 360-DEGREE DOCTORAL ASSESSMENT

Conceptual and methodological alterations of the 360-degree methodology are required in order to fit it to the purpose of assessing education at HEIs. In our approach, we will not examine individuals but higher education programs - in particular, doctoral programs.

A. Operationalization

Our framework has been designed in the context of the Faculty of Informatics of an European university where different doctoral programs in Computer Science (CS) co-exist. The different doctoral programs run parallel, following different doctoral models. They not only have different administrative structure but also different curricula. The application of our framework in such environment would be especially useful for mutual comparison and improvement.

B. Criteria and their Items

The following criteria aim to assess the academic aspects that influence the educational process of doctoral students becoming independent researchers. The selection of these criteria stems from several sources. Among them is a qualitative study of the experiences of doctoral students in the field of CS based on interviews and a focus group, as well as a review of quality evaluations of doctoral programs [14], [15], [16], [17], [18], and a series of discussions with management teams of doctoral programs in CS in Europe.

- **Resources and infrastructure** refer to information about the existence and use of university-provided resources that are available to the doctoral program and facilitate the doctoral activity.
- **Formal education** assesses the teaching activities of the doctoral program.
- **Mobility** evaluates the importance of mobility, the opportunities given, and the mobility undertaken.
- **Sense of community** assesses the involvement and integration of the program and its students into the scientific community and their research environment.
- **Program's competence** evaluates the general quality of the program, including a reflection on the its competence compared to other doctoral programs.
- **Career path** includes the professional background of doctoral candidates, their current working activities and their future career aspirations.
- **Academic work** covers the participation of doctoral students in academic activities such as research, teaching, undergraduate supervision or organizational tasks.
- **Student's progress** evaluates the development of the students on aspects such as knowledge acquisition, publications, dissertation writing, etc.
- **Working condition** refers to aspects such as duration of the program, employment situation, funding, social insurance, cultural and gender integration, etc.
- **Doctoral training and supervision** assess the elements of students' training (e.g., theory, methods, transferable skills), and the role that supervision plays in the doctoral training.

These criteria are evaluated by the stakeholders in terms of their perceptions through a series of questions in a survey. Due to the different perspectives of the collectives of stakeholders, the formulation of *items* (i.e., questions) differ for each of them. Each item is a close-ended question that depending of the aspect under evaluation, is rated using different domains (e.g., numerical, interval) and scales.

C. Stakeholders

For the implementation of our framework we have identified the following collectives of stakeholders:

- **Students and graduates:** Given that students and graduates are involved in the whole educational process, they assess the complete set of criteria.
- **Faculty:** This collective includes all faculty members in contact with the doctoral program, such as supervisors or lecturers of students in the program.
- **External scholars:** They are faculty members of other HEIs that are in contact with the doctoral program or its students.
- **External / industry partners:** These are external organizations that cooperate with the doctoral program or its students.

Given that not all stakeholders have the necessary knowledge to assess all criteria, different subsets of criteria need to be assigned to each collective.

The separation of the possible multiple roles of a single appraiser is necessary for obtaining an unbiased measurement. Measures should be taken to ensure that the same person does not undertake different roles.

D. Doctoral Program's Self-Definition

In our model the doctoral program also takes part in the assessment. It defines itself by assigning a weight to each item of each criterion in order to state its nature and objectives. Weighting criteria items could be seen as equivalent to stating the importance of an issue for the doctoral program. The doctoral program, however, does not provide any score to any criteria item. Only stakeholders provide scores to criteria items. Therefore, the program does not evaluate its own performance and is limited to state only the importance of each item. This approach differs from the traditional 360-degree feedback, where the appraisee also judges his/her own performance.

The program's self-definition, in combination with the stakeholders' judgments, allows the evaluation of the relationship between results and established objectives or priorities of the doctoral program, in a similar way as in the EFQM model. This is important because educational institutions might want to focus on quality criteria or indicators that are regarded as more important.

E. Weighting

In our model, stakeholders do not only judge the performance of the doctoral program according to the criteria, but also the importance of each criterion. This is effectively done through a dual-scale with which our framework operates. We evaluate both a *score* that assesses fulfillment and a *weight* that assesses relevance. This dual-scale compared to the single one used in SERVQUAL provides additional information by comparing the weighting (the scale of values) of the stakeholders and the doctoral program. This can be used to determine if the objectives and mission statement of the program align with the interest of the stakeholders.

Weighted evaluations are not part of the traditional 360-degree feedback methodology. Therefore, we outline the mathematical foundations on how to integrate the dual-scale evaluation scheme into our framework. Given the assessment by all stakeholders of their assigned criteria, we employ a multi-stage approach to compute the final results. After a normalization stage, we aggregate the data to coarser categories until the desired level of detail is reached. Following the notations given in Table I, our framework can be stated formally as follows: Each stakeholder h_k^r belongs to a collective $H^r \subset H$ of all stakeholders with role $r \in R$. Each criterion c out of the global set of criteria C is assessed via one or more items I^c, e.g., different questions related to it. A given item i_j^c of category c is assessed by a stakeholder h_k^r with role r by assigning both a score s_{jk}^{cr} and weight w_{jk}^{cr}. As introduced in the beginning of Section IV, the score assesses how well this item is fulfilled by the doctoral program while the weight yields a judgment of its relevance or importance. Note that not all combinations of j and k have a score s_{jk}^{cr} (resp. weight w_{jk}^{cr}) associated with them as not all stakeholders assess all items (see Section IV-C).

In the normalization stage, all assessments are converted to a uniform scale. This is necessary as they can be expressed in different domains (e.g., numerical, as an interval) and at different scales. We use a normalization to the unit interval $[0, 1]$ for convenience. Note that the weights w_{jk}^{cr} that are assessed by each stakeholder have to be further normalized to sum to one, i.e. $\sum_c \sum_j w_{jk}^{cr} = 1$.

TABLE I. NOTATION USED IN THE FORMAL SPECIFICATION.

R	Set of all roles
H	Set of all stakeholders
H^r	Collective of all stakeholders with role r
h_k^r	k-th stakeholder of the collective H^r
C	Set of all criteria
I	Set of all items
I^c	Set of all items of criterion c
i_j^c	j-th item of the set I^c
s_{jk}^{cr}	Score given by stakeholder h_k^r to item i_j^c
w_{jk}^{cr}	Weight given by stakeholder h_k^r to item i_j^c
s_j^{cr}	Score given by stakeholders of H_k to item i_j^c
w_j^{cr}	Weight given by stakeholders of H_k to item i_j^c
s_j^c	Score given by all stakeholders to item i_j^c
w_j^c	Weight given by all stakeholders to item i_j^c
s^c	Score given by all stakeholders to category c
w^c	Weight given by all stakeholders to category c
s	Score given by all stakeholders to the program

F. Aggregation Process

The subsequent aggregation stage is another contribution of this paper. It allows the derivation of scores for whole criteria or for whole collective of stakeholders. Through the aggregation of scores and weight, we obtain intermediate results that provide detailed information of the scores and weights of items and criteria per collective of stakeholder.

Given a set of values $V = \{v_1, \ldots, v_M\}$ and their respective aggregation weights $\Omega = \{\omega_1, \ldots, \omega_M\}$ an aggregation procedure agg yields a representative value v^* for the whole set by $v^* = \mathrm{agg}(V, \Omega)$.

This problem can be reformulated as a multi-objective optimization method and in our framework we employ Goal Programming [19] to obtain a solution. The involved distance functions $\|\cdot\|$ that measure the disagreement, i.e. $\sum_i^M \omega_i \|v_i - v^*\|$, can be chosen to be a general L^p norm [20] or convex sums of several such norms [21]. Similar to the performance appraisal method of de Andrés et al. [22] we perform an iterative aggregation of the raw scores and weights to global values. For the aggregation of scores $s_{\cdots}^{\cdots}$ we use themselves and their respective weights $w_{\cdots}^{\cdots}$ as input to the aggregation procedure, i.e. $\mathrm{agg}(\{s_{\cdots}^{\cdots}\}, \{w_{\cdots}^{\cdots}\})$.

To aggregate the weights $w_{\cdots}^{\cdots}$ we use themselves as input to the procedure and use uniform weighting, i.e. $\mathrm{agg}(\{w_{\cdots}^{\cdots}\}, \mathbf{1})$ with $\mathbf{1} = \{1, \ldots, 1\}$.

We start by aggregating both item scores s_{jk}^{cr} and weights w_{jk}^{cr} given by the stakeholders of a single role. This yields a representative score s_j^{cr} (resp. weight w_j^{cr}) for each collective H^r and corresponds to

$$s_j^{cr} = \mathrm{agg}(\{s_{jk}^{cr} : h_k^r \in H^r\}, \{w_{jk}^{cr} : h_k^r \in H^r\}) \quad \text{and}$$
$$w_j^{cr} = \mathrm{agg}(\{w_{jk}^{cr} : h_k^r \in H^r\}, \mathbf{1}).$$

Note that the weights have to be renormalized after each aggregation step to sum to one again.

The further aggregation steps yield

$$s_j^c = \mathrm{agg}(\{s_j^{cr} : r \in R\}, \{w_j^{cr} : r \in R\})$$
$$s^c = \mathrm{agg}(\{s_j^c : j \in I^c\}, \{w_j^c : j \in I^c\})$$

with corresponding steps for the weights to obtain w_j^c and w^c. The aggregated values s_j^c give the score for the j-th item i_j^c of category c when taking all relevant stakeholders into account. The values s^c denote the corresponding scores per criterion. As described in the beginning of Section IV, the doctoral program gives a self-definition by assigning a weight to each criterion to state its nature and objectives. A direct comparison of these values with the weights w^c obtained by the aggregation shows if the stakeholders' perception of the importance of the various criteria matches with the program definition.

The last aggregation step combines all scores to a single global value by performing $s = \text{agg}(\{s^c : c \in C\}, \{w^c : c \in C\})$. Note that the last operation can be performed with two different sets of weights. The use of the aggregated weights w^c gives an assessment weighted according to the relevance assigned by the stakeholders. Using the weights given by the self-definition of the program instead of w^c, we obtain an evaluation that focus on the priorities set by the program.

Using this multi-stage aggregation process we obtain intermediate results that can be used for a thorough evaluation of the program. While the later stages give a global quality assessment of the program, the first stages allow a detailed investigation of single criteria.

V. Discussion

The complexity of the concept of quality in education is a barrier for the implementation of a holistic management model [10], [23]. The combination of different reliable instruments is, therefore, necessary. The results provided by our framework could be enriched with other instruments such as interviews to analyze reasons behind the results, or statistical data that reflect the performance of the program.

This framework could be applied to other levels and fields of education defining different criteria and stakeholders. In this paper we focus on doctoral level in CS, as it is in the process of being implemented in a small selection of doctoral programs in CS. We plan to validate the consistency of the construct and its usefulness with a user study relying on the feedback of the management teams of such doctoral programs. Furthermore, we also plan to analyze the results to provide a better understanding of the different programs' models of doctoral education in CS.

VI. Conclusion

We proposed a quality assessment framework for doctoral programs designed as an internal audit to facilitate quality evaluations. Our framework focuses on the educational processes that take place in doctoral programs and assesses them from the perspective of its stakeholders. The program itself provides a self-definition that allows a coherent final evaluation according to its nature and objectives. The presented framework not only obtains a global assessment for a doctoral program, but also intermediate collective evaluations of each criterion and group of stakeholders.

Apart from the general definition of our framework, a main contribution of this paper is the introduction of analytical methods for the weighting and aggregation of information.

References

[1] N. Becket and M. Brookes, "Quality management practice in higher education - what quality are we actually enhancing?" *JoHLSTE*, vol. 7, no. 1, pp. 40–54, Nov. 2008.

[2] D. Grant and E. Mergen and S. Widrick, "A comparative analysis of quality management in US and international universities," *Total Quality Management*, vol. 15, no. 4, pp. 423–438, 2004.

[3] G. Srikanthan and J. Dalrymple, "A synthesis of a quality management model for education in universities," *Int. Journal of Educational Management*, vol. 18, no. 4, pp. 266–279, 2004.

[4] E. Sallis, *Total Quality Management in Education. Third Edition*. London, UK: Kogan Page Ltd, 2002.

[5] EFQM, "European Framework for Quality Management," http://www.efqm.org/en/, July 2013.

[6] Y. Akao, *Quality Function Deployment - Integrating Customer Requirements into Product Design*. New York, USA: Productivity Press, 2004.

[7] V.A. Zeithaml, A. Parasuraman and L.L. Berry, *Delivering Quality Service: Balancing Customer Perceptions and Expectations*. New York, USA: The Free Press, 1990.

[8] G. McElwee and G. Holmes, "Total quality management in the new universities: How to approach human resource management," *Higher education in Europe*, vol. 21, no. 2-3, pp. 182–192, 1996.

[9] E. Coate, "The introduction of total quality management at Oregon State University," *Higher Education*, vol. 25, pp. 303–320, 1993.

[10] G. Srikanthan and J. Dalrymple, "Developing a holistic model for quality in higher education," *Quality in Higher Education*, vol. 8, no. 3, pp. 216–224, 2002.

[11] B. Dollery, "Knaves or knights, pawns or queens?: An evaluation of Australian higher education reform policy," *Journal of Educational Administration*, vol. 44, no. 1, pp. 86–97, 2006.

[12] L. Harvey, "A history and critique of quality evaluation in the UK," *Quality Assurance in Education*, vol. 13, no. 4, pp. 263–276, 2005.

[13] J.W. Fleenor and J.M. Prince, *Using 360-Degree Feedback in Organizations. An Annotated Bibliography*. North Carolina, USA: Center for Creative Leadership, 1997.

[14] European University Association, "Rankings in institutional strategies and processes (RISP)," EUA, Brussels, Belgium, Tech. Rep., 2013.

[15] J.P. Ostriker and C.V. Kuh, *Assessing research-doctorate programs: a methodology study*. Washington: The National Academies Press, 2003.

[16] J.P. Ostriker, C.V. Kuh and J.A. Voytuk, "A data-based assessment of research-doctorate programs in the United States," National Research Council, Washington, Tech. Rep., 2010.

[17] J.P. Ostriker, P.W. Holland, C.V. Kuh and J.A. Voytuk, "A revised guide to the methodology of the data-based assessment of research-doctorate programs in the United States (2010)," National Research Council, Washington, Tech. Rep., 2010.

[18] The European Council of Doctoral Candidates and Junior Researchers, "The first Eurodoc survey on doctoral candidates in twelve European countries," Eurodoc, Tech. Rep., 2011.

[19] A. Charnes and W.W. Cooper and R. Ferguson, "Optimal estimation of executive compensation by linear programming," *Manage. Sci.*, vol. 1, pp. 138–151, 1955.

[20] P. Yu, "A class of solutions for group decision problems," *Manage. Sci.*, vol. 14, no. 1, pp. 21–34, 2010.

[21] C. Romero, "Extended lexicographic goal programming: A unifying approach," *OMEGA, The Int. Journal of Management Science*, vol. 29, pp. 63–71, 2010.

[22] R. de Andrés and J.L. Garcí-Lapresta and L. Martínez, "A multigranular linguistic model for management decision-making in performance appraisal," *Soft Computing*, vol. 14, no. 1, pp. 21–34, 2010.

[23] A. N. Pears, "Does quality assurance enhance the quality of computing education?" in *Proc. of the 12th Australasian Conf. on Computing Education*. ACS, 2010.