ORIGINAL ARTICLE

Quality in academic postgraduate studies: design and validation of an evaluation instrument

Calidad en el posgrado académico: diseño y validación de un instrumento de evaluación

Fridel Julio Ramos Azcuy[1] ✉, Rosa Mayelin Guerra Bretaña[1] ✉, María Beatriz Valencia Bonilla[2] ✉

[1]Universidad de La Habana, Cuba.

[2]Universidad Tecnológica de Pereira, Colombia.

RESUMEN

La autoevaluación de los programas de maestría impartidos total o parcialmente en entornos virtuales permite la mejora de estos y favorece la toma de decisiones basada en evidencias. La eficacia de la autoevaluación depende de los indicadores utilizados por lo que es indispensable determinar el valor de estos. El objetivo del trabajo fue implementar un instrumento de evaluación de los indicadores para la autoevaluación de la calidad de los programas virtuales de maestría. Se realizó la identificación y análisis de artículos científicos, guías de diseño y normas técnicas que abordan la evaluación de indicadores. Se empleó el motor de búsqueda de Google y Google Académico con las palabras clave: indicadores, sistemas de indicadores, evaluación de indicadores. Como resultado del análisis se identificaron seis criterios de calidad de los indicadores utilizados en la autoevaluación de los programas virtuales de maestría. Se elaboró un instrumento para medir estos seis criterios mediante una escala de Likert de cinco niveles. La prueba piloto del instrumento diseñado permitió su validación y mejorar los indicadores utilizados en la autoevaluación de los programas virtuales de maestría en la Cátedra de Calidad, Metrología y Normalización de la Universidad de La Habana.

Palabras clave: educación superior; instrucción basada en la web; mejora del programa; medición; autoevaluación.

ABSTRACT

Systematic self-assessment of master's programs taught totally or partially in virtual environments allows the improvement of these and favors evidence-based decision making. Self-assessment effectiveness is to a large extent due to indicators that are defined to carry it out. Value determination of these indicators for the self-assessment's objectives achievement is the first step to the realization of this. Given the above, the objective of this work was to implement an instrument for evaluating the self-assessment's indicators for quality of virtual programs master's. It was carried out identification and analysis of scientific articles, design guides and technical standards that address the evaluation of indicators. The Google and Google Scholar search engine was used with the keywords: indicators, indicator systems, indicator evaluation. As a result of the analysis, six indicators' quality criteria used in the master's virtual programs self-assessment were identified. An instrument was developed to measure these six criteria through a five-level Likert scale. Pilot test of designed instrument allowed its validation and improve the indicators used in master's virtual programs self-assessment in the Chair of Quality, Metrology and Standardization of the University of Havana.

Keywords: Higher education; web-based instruction; program improvement; measurement; self-evaluation.

INTRODUCTION

Today, educational organizations (EOs) are called upon to show their dedication to adopting effective management practices, have consistent self-assessment processes and tools to demonstrate and improve efficiency and effectiveness, and achieve more individualized learning and effective responsiveness for all

students. To this end, the implementation of a management system for educational organizations (SGOE) supported by internationally consensual practices is beneficial (Guerra Bretaña et al., 2020).

Improvement, as a principle of the SGOE, promotes an approach focused on the investigation and determination of the essential elements of the problem, followed by prevention and corrective actions (National Bureau of Standards, 2015; 2019). This principle induces SOs to maintain their performance levels, react to changes in the context, and take advantage of opportunities. When this approach is systematically applied, it can be referred to as continuous improvement and allows EOs to refine their management and optimize their results, including those related to academic postgraduate studies.

As a continuous improvement tool, self-assessment of academic postgraduate programs allows EOs to know the strengths and weaknesses, risks and opportunities for improvement, the level of maturity of these programs, and if repeated, their progress over time (National Bureau of Standards, 2018). The first step for self-assessment is to establish indicators relevant to the object of improvement.

An indicator is defined as an expression, either qualitative or quantitative, used to represent characteristics or behaviors over time by tracking a variable or establishing relationships between variables. By comparing this expression with previous periods or with a target, it facilitates the evaluation of performance and its evolution over time (Departamento Administrativo de la Función Pública [DAFP], 2018; Ministerio et al. y Política Económica [MIDEPLAN], 2018; Vinajera-Zamora et al., 2023).

Through the use of indicators, important information can be obtained on the level of achievement of previously established objectives and the evolution of the critical success factors of the process or organization being analyzed. The information obtained allows a better understanding of the processes, which contributes to their improvement (Asociación et al., 2003; ISOTools Excellence, 2023).

Guerra Bretaña & Meizoso Valdéz (2019) state that for indicators to be effective, they must provide relevant, unequivocal, and objective information. To achieve this, Rossi and Illescas (2022) explain that it is essential that indicators are measurable, interpretable, accurate, verifiable, reliable, comparable, and accessible. Furthermore, the indicators used must be clear, relevant, pertinent, technically feasible, and economical (Consejo Nacional de Evaluación de la Política de Desarrollo Social, 2021).

Therefore, the purpose of an instrument for evaluating indicators is to record data that allow the assessment of the degree to which the inherent characteristics of the indicator meet the requirements. The functions of such an instrument are to evidence the suitability of the indicator and to identify opportunities for improvement (Spanish Association for Standardization, 2003).

The effectiveness of a measuring instrument is determined by its ability to represent the object to be measured. To achieve this, the instrument must meet three basic requirements (Hernández Sampieri et al., 2014; García-Valcárcel et al., 2020):

- Reliability: The measuring instrument is capable of arriving, under given circumstances, at similar results when repeatedly measuring the same object.
- Validity: The instrument actually measures the object it intends to measure. It can be evidenced through content, criterion, and construct validity.
- Objectivity: "the degree to which the instrument is or is not permeable to the influence of the biases and tendencies of the researchers who administer, score, and interpret it." (p. 206)

The indicators designed for EOs to self-evaluate their master's degree programs, designed to be taught in the virtual modality, should be adequate to measure both the general and intrinsic characteristics of these programs, considering their planning, execution, and control.

The above constitutes the foundation of the problem that motivated this research that was presented during the design of a tool for the systematic, comprehensive, and detailed self-evaluation of the quality of master's programs in the virtual modality (Ramos et al., 2023). Since it was determined that the criteria for the evaluation of the indicators were not sufficiently established. Therefore, the objective of this work is to implement an instrument for the evaluation of indicators for the self-evaluation of master's degree programs in the virtual modality.

METHODS

This research used a quantitative approach; its scope was descriptive and explanatory, and its design was quasi-experimental (Hernández Sampieri et al., 2014). A literature review was conducted to identify scientific articles, design guides, and technical standards that address the definition, development, and evaluation of indicators. The Google and Google Scholar search engines, the Scopus and Scielo databases, and the online browsing platforms of the International Organization for Standardization and the Spanish Association for Standardization were used with the keywords indicators, indicator systems, and indicator evaluation.

Subsequently, a preliminary study of the identified bibliographic sources was carried out with the purpose of selecting the documents that met the established inclusion criteria. The relevance of the content of the

literature identified in relation to the evaluation and selection of performance indicators was observed, giving priority to those bibliographic sources belonging to the space of time from 2018 to the present in English and Spanish languages. In the case of technical standards and other normative documents, their validity was verified through the information provided by the organizations in charge of their elaboration and edition.

Based on the analysis of the selected documents, the information obtained was systematized to identify the quality criteria that, in the authors' opinion, are theoretically justified and allow the evaluation of the indicators for self-assessment.

On the basis of the identified criteria, the instrument that allowed the evaluation of the indicators included in the tool for self-evaluation of master's degree programs in the virtual modality was constructed. These criteria are measured for each indicator using a five-level Likert scale, where each value corresponds to a label that assigns an ascending degree of agreement. Thus, the values and their labels are 1 (Strongly disagree), 2 (Disagree), 3 (Neither agree nor disagree), 4 (Agree), and 5 (Strongly agree).

Validation of the instrument

The instrument was designed for a pilot test to evaluate the self-evaluation indicators of the two master's degree programs managed by the Chair of Quality, Metrology, and Standardization of the University of Havana. The evaluation of the indicators was carried out using the designed instrument, and consultation with experts using the Delphi method was conducted to structure group communication (Linstone & Turoff, 2002).

The tool designed by Dobrov & Smirnov (1972) was used to select the experts, with the modifications introduced by Cruz Ramírez and Martínez Cepena (2012), to determine their expertise. Once the experts had been selected, they were sent the instrument by e-mail, implemented in an Excel workbook, accompanied by each of the adequately formalized indicators.

With the data obtained in the pilot test, the internal consistency analysis was carried out by calculating Cronbach's Alpha statistic. The concordance between experts was also analyzed based on the results of Kendall's test, Friedman's two-way test, and the ANOCHI coefficient (García Pulido et al., 2023).

RESULTS AND DISCUSSION

Based on the analysis of the bibliographic sources chosen, seven tools were identified, as shown in Table 1, which establish requirements for the selection, design, and evaluation of indicators.

Of the 67 requirements proposed in these tools, six were eliminated, four from the global assessment stage established by the National Council for the Evaluation of Social Development Policy (CONEVAL, 2021), and two from ILPES (2003) because they were focused on the use of the indicators and not on their quality. From the comparative analysis of the 59 requirements selected, it was observed that 18 of them were present in several tools, as shown in Table 2. However, their denomination is sometimes different.

The requirements "Reliability," "Economy," "Functionality," "Interpretability," "Relevance," and "Relevance" are included in five or more tools, which implies that they are more relevant for the evaluation of indicators. On the other hand, the requirements "Accessibility," "Applicability," "Coherence," and "Technical feasibility" are only described in one tool. However, in the analysis performed, it was determined that the contents of these requirements are diluted in the contents of the rest of the requirements, with the exception of the "Accessibility" requirement, which is a novelty introduced by CONEVAL (2021).

The "Reliability" requirement requires that the information sources chosen be reliable and that this information possesses statistical quality attributes. It also requires that the data to be measured be sufficient, demonstrable, and auditable so that the results are always the same (ILPES, 2003; ISOTools Excellence, 2023; Ministry of National Planning and MIDEPLAN, 2018).

For its part, the "Economy" requirement requires that the cost associated with obtaining information to develop the indicator be reasonable (Departamento et al. [DANE], 2012). In other words, "an indicator is economic if the benefit of generating the information is greater with respect to the economic or human cost necessary to calculate it" (CONEVAL, 2021, p. 16).

The "Functionality" requirement allows checking that the indicator is "measurable, operable, and sensitive to changes registered in the initial situation" (DAFP, 2018, p. 55). The above relates to the indicator's ability to be monitored and is based on the fact that the information from verification sources must be accurate and unambiguous (CONEVAL, 2021).

Interpretability" is related to the ease with which the indicator can be understood. In this sense, MIDEPLAN (2018) states that the indicator "should be as direct and unambiguous as possible" (p. 14). For its part, the "Relevance" requirement demands that the indicator refers to the organization's core processes, products, and services (DAFP, 2018; ILPES, 2003; UNE, 2003). Finally, the "Relevance" requirement requires that the indicator focus on the most important aspects related to its objective; that is, "it must be defined on some important aspect with practical sense" (CONEVAL, 2021, p. 12).

Table 1. Tools for the selection, design and evaluation of indicators.

Author	Description of the proposed tool
National Council for the Evaluation of Social Development Policy [CONEVAL] (2021)	Terms of Reference for the Evaluation of Indicators. Objective: to develop the evaluation of indicators in state organizations in Mexico. It establishes a procedure composed of three stages (integral evaluation, specific evaluation, and global evaluation) and 12 requirements for the quality of the indicators.
Departamento Administrativo de la Función Pública [DAFP] (2018)	Guide for the construction and analysis of management indicators. Objective: to adequately guide Colombian government institutions in the development and analysis of indicators. It establishes five quality criteria to validate indicators.
National Administrative Department of Statistics [DANE] (2012).	Introduction to the design, construction, and interpretation of indicators. Objective: To improve statistical data management in order to facilitate decision-making in Colombian government institutions. Methodological elements for the design, construction, and interpretation of indicators are addressed, including 15 criteria for the selection of indicators.
Latin American and Caribbean Institute for Economic and Social Planning [ILPES] (2003)	Performance evaluation indicators: a tool for results-based management in Latin America. Objective: To analyze the indicators used to evaluate performance as a tool used by state organizations in Latin America for results-oriented management. It provides ten criteria for the construction of indicators.
Ministry of National Planning and Economic Policy [MIDEPLAN] (2018).	Indicator guide. Basic guidelines for their elaboration. Objective: To provide guidance on concepts, terminology, and basic elements for the development of indicators in Costa Rica. Provides 11 attributes that account for the quality in the definition of indicators.
Ministry of Finance and Public Credit [SHCP] (2016).	Guide for the design of strategic indicators. Objective: to present in a simplified, schematic, and uniform way the essential methodological foundations for the development of indicators in Mexico. It provides six criteria for the choice of indicators and a template for the application of these criteria.
Spanish Association for Standardization [UNE] (2003)	UNE 66175: 2003 Quality management systems. Guide for the implementation of indicator systems. Technical standards were developed by the Spanish Association for Standardization, whose objective is to establish guidelines for the formulation and development of indicators, with the purpose of making them effective and efficient in decision-making to promote the improvement of organizations. It provides eight characteristics of the indicators.

Based on the study of the requirements identified, in particular, those most frequently used in the tools analyzed, six criteria were established for the evaluation of indicators:

• Clarity of wording: The indicator is characterized by precision and clarity, meaning that there is no ambiguity as to its measurement purpose. It allows for the evaluation that the indicator is precisely formalized and does not contain ambiguities in the formulation of its constituent elements. This criterion pursues a similar objective to the "Interpretability" requirement and measures that the indicator's design facilitates its understanding.

• Relevance: The indicator is appropriate and evaluates significant aspects of program quality. It requires that the indicator be linked to a critical activity in the management of the virtual master's program. This criterion combines the objectives of the "Relevance" and "Relevance" requirements since it measures that the indicator is relevant and appropriate in terms of the objectives of the self-evaluation.

• Objectivity: The indicator is not affected by the evaluator's thinking, perspective, or emotions. This criterion assesses that external factors do not condition the results of the indicator and requires that the method for obtaining the information is precisely defined and unbiased so that the results of the indicator are always the same.

• Direct correspondence: There is a connection between what the indicator is intended to measure and the criterion with which it is associated. This criterion requires that the indicator be capable of detecting and reacting to small differences. It also requires that the indicator be measurable, operable, and auditable.

• Adequacy of the rating scale: The method used to evaluate the indicator is appropriate. This criterion requires that the scale and unit of measurement of the indicator be consistent with the expected results of the indicator.

• Relevance of evidence: The required evidence has a meaningful relationship to the indicator.

The instrument consisted of two sections. The first section contained information related to the objective of the instrument, the six criteria to be evaluated, the scale used, and its mode of use, which consisted of assigning a degree of agreement for each indicator evaluated. The second section consisted of a table with columns representing the six evaluation criteria and rows representing the indicators to be evaluated, as well as a row for the evaluator to write down suggestions, observations, proposals for improvement, and any other information considered relevant.

Table 2. Quality requirements of the indicators proposed by the tools analyzed.

Requirements	Tools						
	CONEVAL (2021)	DAFP (2018)	DANE (2012)	ILPES (2003)	MIDEPLAN (2018)	SHCP (2016)	UNE (2003)
Accessibility			X				
Applicability			X				
Consistency			X				
Comparability			X				X
Reliability		X	X	X	X		X
Credibility			X				X
Availability		X	X		X		
Economy	X			X	X	X	X
Technical feasibility	X						
Functionality	X	X	X		X	X	X
Homogeneity				X	X		
Independence				X	X		
Interpretability	X		X	X	X	X	X
Non-redundancy	X		X	X		X	
Timeliness			X	X	X		
Pertinence	X	X	X	X	X		X
Relevance	X		X		X	X	X
Utility	X	X	X		X		

In the validation of the instrument, Cronbach's Alpha reliability statistic was calculated for the measurement scale of each of the criteria with the SPSS version 21 statistical software. The results obtained for this statistic exceeded the value of 0.7 in all cases, which evidences the reliability of the scale used to measure each of the criteria.

On the other hand, the data generated by Kendall's test, Friedman's two-way analysis, and the ANOCHI coefficient, shown in Table 3, evidenced that the agreement between experts with respect to the value given to each of the evaluation criteria of the indicators is good.

After a first round of consultations with the experts, in which a consensus of 52.5% was obtained, the data provided by them was analyzed. As a result, six tasks were identified to optimize the indicators designed and improve their effectiveness. These were:

• Five indicators were eliminated.
• Three pairs of indicators were merged to create three new indicators.
• The objectives of the two indicators were clarified.
• The wording of the description of the seven indicators was optimized to achieve greater clarity and consistency.
• The measurement scale of two indicators was improved.
• The specificity and accuracy of the evidence requested were increased for ten indicators.

Table 3. Results of statistical tests for the estimation of agreement between experts.

Criteria	W Kendall[a]	Friedman[a]	ANOCHI
Clarity of wording	0,57	516,6	0,82
Relevance	0,57	524,0	0,82
Objectivity	0,60	559,5	0,82
Direct correspondence	0,60	556,9	0,83
Adequacy of the rating scale	0,55	520,6	0,76
Relevance of the evidence	0,59	549,3	0,83

Note: [a] p-value = 0,00 for all cases.

In the next round of consultations with the experts, they were provided with the instrument, together with the indicators, which were revised and updated. On this occasion, a consensus of over 70% was obtained, and no suggestions, observations, or proposals for improvement were made. In this way, sufficient evidence was obtained of the adequate quality of the indicators designed for the self-evaluation of master's degree programs in the virtual modality.

The study determined the six quality criteria that are most suitable for the evaluation of management indicators, particularly for the self-evaluation of virtual master's degree programs. These criteria are established in more than 71% of the bibliographic sources analyzed and provide both qualitative and quantitative information that is valuable for the choice or development of indicators that allow knowing, in a comprehensive manner, the state of the management of master's programs taught in the virtual modality.

The evaluation of the six established criteria helps to ensure that the indicators designed are specific, measurable, achievable, realistic, and of limited duration. It also favors the robustness and balance of these indicators since it allows the determination of possible biases. The above allows indicators to exhibit high performance in acquiring meaningful data for the purpose of supporting decision-making (Boada & Alzate, 2020).

Although the statistical tests performed to check the concordance between experts provided sufficient evidence to ensure the content validity of the instrument implemented, it was found that in the case of the criterion "Adequacy of the rating scale," the result of these tests is lower than the result obtained by the other criteria. For this reason, a more precise description of the analyzed criterion is recommended in order to strengthen and improve the instrument's performance.

CONCLUSIONS

Self-evaluation is a way of verifying whether the objectives and activities necessary to guarantee the quality of virtual master's programs are being met. Based on the state-of-the-art analysis of the evaluation of indicators, the six most relevant quality requirements for the design, selection, and evaluation of management and performance indicators were identified. The analysis of the identified requirements and their contextualization to the environment in which the virtual master's degree programs are taught made it possible to establish six criteria for the evaluation of the indicators.

The results of the statistical techniques applied to the data obtained in the pilot test of the designed instrument showed its validity in evaluating the indicators used in the self-evaluation of virtual master's degree programs.

The implemented instrument is effective for the evaluation of self-evaluation indicators of virtual master's degree programs, and its use can be generalized to other forms of academic postgraduate programs, such as specialties and doctorates in any modality of studies.

REFERENCES

Asociación Española de Normalización. (2003). Sistemas de gestión de la calidad. Guía para la implantación de sistemas de indicadores (UNE 66175). https://bit.ly/3tR9dT0

Boada, A., & Alzate, I. (2020). Debilidad simétrica en los indicadores de gestión: consecuencias e impacto estructural. Revista de Métodos Cuantitativos para la Economía y la Empresa, 29, 319-336. https://doi.org/10.46661/revmetodoscuanteconempresa.3720

Consejo Nacional de Evaluación de la Política de Desarrollo Social. [CONEVAL[(2021). Términos de Referencia de la Evaluación de Indicadores. http://bit.ly/3I5e9Zo

Cruz Ramírez, M., & Martínez Cepena, M. C. (2012). Perfeccionamiento de un instrumento para la selección de expertos en las investigaciones educativas. Revista Electrónica de Investigación Educativa, 14(2), 167-179. https://bit.ly/cruzramirez2012

Departamento Administrativo de la Función Pública. [DAFP]. (2018). Guía para la construcción y análisis de indicadores de gestión. https://bit.ly/3G9QV2Y

Departamento Administrativo Nacional de Estadística. [DAFP]. (2012). Introducción al diseño, construcción e interpretación de indicadores. http://bit.ly/3Vn3wEg

Dobrov, G. M., & Smirnov, L. P. (1972). Forecasting as a means for scientific and technological policy control. Technological Forecasting and Social Change, 4(1), 5–18. https://doi.org/bxc7jp

García Pulido, Y. A., Frías Jiménez, R. A., & Medina León, A. A. (2023). Validación de procedimientos para la gestión empresarial. Retos de la Dirección, 15(2), 152-178. https://bit.ly/3SxepG1

García-Valcárcel Muñoz-Repiso , A., Casillas Martín, S., & Gómez-Pablos, V. (2020). Validation of an Indicator Model (INCODIES) for Assessing Student Digital Competence in Basic Education. Journal of New Approaches in Educational Research, 9(1), 110-125. https://doi.org/10.7821/naer.2020.1.459

Guerra Bretaña, M., & Meizoso Valdéz, M. C. (2019). Gestión de la calidad. Conceptos, principios, modelos y herramientas.

Guerra Bretaña, R. M., Roque González, R., & Ramos Azcuy, F. (2020). Aplicación de la norma ISO 21001:2018 a la calidad de los programas de posgrado académico. Educación Médica Superior, 34(1), e2050. https://bit.ly/3PRxi4O

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. P. (2014). Metodología de la investigación (6 ed.). McGRAW-HILL.

Instituto Latinoamericano y del Caribe de Planificación Económica y Social. (2003). Los indicadores de evaluación del desempeño: una herramienta para la gestión por resultados en América Latina. Repositorio digital de la Comisión Económica para América Latina y el Caribe: http://hdl.handle.net/11362/9939

ISOTools Excellence. (2023). Criterios para la selección de indicadores de calidad en la salud. http://bit.ly/3C0JRmZ

Linstone, H. A., & Turoff, M. (2002). The Delphi Method. Techniques and Applications. Addison-WesleyA. Linstone. https://bit.ly/3L3erOJ

Ministerio de Planificación Nacional y Política Económica. [MIDEPLAN]. (2018). Guía de indicadores. Orientaciones básicas para su elaboración. http://bit.ly/3G8rXkB

Oficina Nacional de Normalización. (2018). Gestión de la calidad. Calidad de una organización. Orientación para lograr el éxito sostenido (NC-ISO 9004:2018). https://bit.ly/3MjF9G0

Ramos Azcuy, F. J., & Guerra Bretaña, R. M. (2023). Instrumento para la autoevaluación de programas virtuales de maestrías. RECIE. Revista Caribeña de Investigación Educativa, 7(1), 7-29. https://doi.org/10.32541/recie.2023.v7i1.pp7-29

Rossi, S., & Illescas, G. (2022). Una propuesta de diseño de indicadores de I+D en espacios de educación superior. Ingeniería Industrial, (42), 139-158. https://doi.org/10.26439/ing.ind2022.n42.5863

Secretaría de Hacienda y Crédito Público. [SHCP]. (2016). Guía para el diseño de indicadores estratégicos. http://bit.ly/3ve6wZ0

Vinajera-Zamora, A., Gaus, N., & Rodríguez-Martínez, Y. (2023). Framework and Analysis of Key Performance Indicators in Cuban Higher Education. Journal of Hispanic Higher Education, 22(2), 205-218. https://doi.org/10.1177/15381927221074505

Conflicts of Interest:

The authors declare that they have no conflicts of interest.

Author Contributions:

Fridel Julio Ramos Azcuy, Rosa Mayelin Guerra Bretaña, María Beatriz Valencia Bonilla: Conceptualization, data curation, formal analysis, investigation, methodology, supervision, validation, visualization, drafting the original manuscript and writing, review, and editing.

Disclaimer/Editor's Note: