ABSTRACT

The objective of this research was to identify the development of literature on performance evaluation. It is an exploratory and descriptive research study, with a quali-quantitative approach through bibliometric analysis, using the ProKnow-C intervention instrument. The analyses showed that the main featured article was by Neely, Gregory and Platts (1995). The most influential article in the portfolio was Bourne et al. (2000), with the highest level of betweeness. There are four large clusters of prominent authors Neely, Bourne, Platts, Kennerley, Franco-Santos, Martinez and Bititci. The field of operations covers 50%; management and strategy cover 42% of the works while accounting covers only 8% of empirical studies. There is great emphasis on performance measurement (55%). The Balanced Scorecard is the tool that predominates in 23% of studies; in 30% of the works, was developed by proposing models based on the literature while in 24% of them it was based on a proposal developed by the authors. The study contributed to broaden the knowledge about the subject in order to allow an overview of the literature, when knowing the main authors working on the theme, periodicals, articles, themes, tools and areas of practice. Emerging views on the subject are also presented.

Keywords: performance evaluation; measurement; management; review of the literature; systematic analysis.
DEVELOPMENT OF PERFORMANCE EVALUATION THEME: A SYSTEMATIC ANALYSIS OF THE LITERATURE

RESUMO
Esta pesquisa objetivou identificar o desenvolvimento da literatura sobre avaliação de desempenho. Trata-se de pesquisa exploratória e descritiva, com abordagem quali-quantitativa por meio de análise bibliométrica, com a utilização do instrumento de intervenção ProKnow-C. Com base nas análises realizadas foi possível constatar que o artigo de destaque é Neely, Gregory e Platts (1995). O artigo de maior influência dentro do portfólio é de Bourne, Mills, Wilcox, Neely e Platts (2000), citado por 38 trabalhos do portfólio. Evidenciaram-se quatro grandes clusters de autores, conectados entre si, Andy Neely, Mike Bourne, Ken Platts, Mike Kennerley, Monica Franco-Santos, Veronica Martínez e Umit Bititci. A área de operações abrange 50% dos estudos; administração e estratégia concentram 42% e área contábil apenas 8% dos trabalhos empíricos. Há ênfase na mensuração do desempenho (55%). O Balanced Scorecard é a ferramenta que mais predomina nos estudos (23%); 30% desenvolveram suas pesquisas propondo modelos baseados na literatura e 24% basearam-se em proposta desenvolvidos pelos autores. O estudo contribuiu para ampliar o conhecimento sobre o tema de modo a permitir uma visão geral da literatura, ao se conhecer os principais autores atuantes no tema, periódicos, artigos, temas, ferramentas e áreas de aplicação. Ainda, oferece visões emergentes sobre o tema.

Palavras-chave: avaliação de desempenho; mensuração; gestão; revisão de literatura; análise sistemática.

RESUMEN
Esta investigación objetivó identificar el desarrollo de la literatura sobre evaluación de rendimiento. Se trata de investigación exploratoria y descriptiva, con abordaje cuali-cuantitativo por medio de análisis bibliométrico, con la utilización del instrumento de intervención ProKnow-C. Con base en los análisis realizados fue posible constatar que el artículo destacado es Neely, Gregory y Platts (1995). El artículo de mayor influencia dentro del portafolio es Bourne, Mills, Wilcox, Neely y Platts (2000). Si demuestran los cuatro principales grupos de autores, Andy Neely, Mike Bourne, Ken Platts, Mike Kennerley, Mónica Franco-Santos, Verónica Martínez y Umit Bititci. Operaciones abarca el 50%; administración y estrategia concentra el 42% y el área contable solo el 8% de los trabajos empíricos. Hay un énfasis en la medición del desempeño (55%). El Cuadro de Mando es la herramienta que predomina en los estudios (23%); El 30% desarrolló sus investigaciones proponiendo modelos basados en la literatura y el 24% se basó en propuesta desarrollados por los autores. La investigación contribuyó a ampliar el conocimiento sobre el tema para permitir una visión general de la literatura, al conocerse los principales autores actuantes en el tema, periódicos, artículos, temas, herramientas y áreas de aplicación. Aún se presentan visiones emergentes sobre el tema.

Palabras clave: evaluación de rendimiento; medición; gestión; revisión de literatura; análisis sistemático.

1 INTRODUCTION

Performance evaluation is fundamental to the management of any organization (CHOONG, 2014; MELNYK et al., 2014). Organizations can use such an evaluation to direct efforts to control and correct strategies, thus establishing goals and the level of desired performance, as well as compare the latter with the level actually achieved. They can also use it for communicating their strategic intention and highlight, for everyone in the organization, the importance of what has been measured, and how important it is in order to achieve the strategic objectives of the organization (MELNYK et al., 2014).

The use of performance measurement and management systems is often recommended to facilitate the implementation of strategies and improve organizational performance (LEBAS, 1995; MELNYK et al., 2014; CUCCURULLO; ARIA; SARTO, 2016). In addition, previous studies have shown that performance evaluation influences people’s behaviour, organizational capabilities and organizational performance (FRANCO-SANTOS; LUCIANETTI; BOURNE, 2012).
Performance evaluation is a topic that has received considerable interest from researchers, in view of the large number of professional and academic conferences, and the high number of articles published on the topic, which has been growing exponentially as of the second half of the 1990s (NEELY; GREGORY; PLATTS, 1995; NEELY; 1999; BOURNE et al., 2000).

According to Bititci et al. (2012), performance measurement began with double-entry bookkeeping, which enabled not only registration of transactions but also monitoring of wealth evolution. It was improved over time, and other ways to monitor performance were added by managerial accounting, always with a focus on financial measurement. Later, after the industrial revolution, the focus of accounting data was moved to operational aspects, such as cost monitoring, productivity, time spent, etc. However, focus was still placed on aspects which were essentially financial. Later, the focus of performance measurement was moved to more strategic aspects, involving product quality, production flexibility, and satisfaction of customers and stakeholders, thus moving toward a more strategic type of control, covering the financial and non-financial dimensions, and resulting in the emergence of several other criteria and indicators.

In this sense, it can be seen that the literature has been developing towards the resolution of practical problems, whose emphasis is to measure the performance of a particular aspect and submit the result of this measurement, without a concern for an effective use of such information for managerial purposes (NEELY; GREGORY; PLATTS, 1995; NEELY, 1999; NUDURUPATI et al., 2011; MICHELI; MARI, 2014; VALMORBIDA; ENSSLIN, 2016). In addition, as Performance Evaluation evolved, it began to be recognized as a tool for information about measurement for an effective use in organizational management (OTLEY, 1999; BERRY et al., 2009). Instead of emphasizing the control of organizational performance, the focus has been shifted to understanding what such performance means and how it can be improved (BITITCI et al., 2012). This shift of emphasis poses challenges to the practice of performance evaluation when one seeks to understand what specific conditions can lead to an improvement in performance. However, this shift opens up opportunities for research.

Thus, there is a need to rethink research on performance evaluation by recognizing the challenges faced by managers as well as offering scientific contributions for the purpose of resolving practical problems experienced in the organizational context (BITITCI et al., 2012).
For this reason, the literature on the topic has to be mapped in order to offer insights on advances and identify opportunities for future research.

Therefore, the aim of the present study is to identify the production of relevant literature on performance evaluation, in order to describe authors, journals, relevant articles, the development of performance measurement and management, tools in use and fields of development of the research.

It should be noted that this article is aimed at highlighting the literature about the theme in order to provide an overview of the literature with a view to promoting the development of new research studies and, hence, align performance evaluation with organizational needs.

2 METHODOLOGICAL PROCEDURES

The present research, in terms of the nature of its objective, is characterized as an exploratory and descriptive study. First, a selection was made of a representative fragment of the literature produced on performance evaluation, seeking to build a robust portfolio which consisted of theoretical and empirical studies. In this portfolio, the authors seek to describe the characteristics of publications with information about authorship, journals and outstanding articles. They also present research networks on the topic, network of citations and co-citations (GRAY, 2013).

To approach the research problem, data were collected from secondary sources in international databases and then analyzed under a qualitative perspective. Although the research is based on bibliometric analysis, an in-depth analysis was made of the results, hence they differ from the simple count of occurrences (CRESWELL, 2009).

The instrument of intervention Knowledge Development Process - Constructivist (ProKnow-C) was used in this research to undertake an analysis of the characteristics of the publications because it enables the selection of a representative portfolio on the topic, thus reducing the bias inherent in this activity.

The ProKnow-C was developed from 1994 by the Laboratory of Constructivist Decision Aid Methodologies (LabMCDA-C), of the Federal University of Santa Catarina (STAEDELE et al., 2019). This instrument has been used to support research about the performance evaluation theme, in both theoretical and empirical terms. The ProKnow-C was
originally conceived to assist a researcher who has no consolidated knowledge about a particular subject and faces the question of where and how to select relevant studies on the subject (TASCA et al., 2010).

Due to the constructivist aspect of the process, the knowledge gives fundamentals that allow the researcher to justify the scientific choices, supported by relevant theoretical contributions (VALMORBIDA; ENSSLIN, 2016; TASCA et al., 2010; LYRIO et al., 2007). The scientific recognition of the instrument was established through dozens of international publications (STAEDELE et al., 2019; ENSSLIN et al., 2015). Several researchers conducted research related to the performance evaluation theme using the intervention instrument ProKnow-C: Staedele et al. (2019), Martins et al. (2018), Thiel et al. (2017), Nuernberg et al. (2017), Valmorbida e Ensslin (2016), Dutra et al. (2015), Tasca et al. (2010), among others.

The next sections will describe (i) data collection procedures; and (ii) data analysis procedures.

2.1 Data collection procedures

The development of this step is motivated by the interest of researchers in select a representative fragment of the literature relative to the topic “Performance Evaluation”, addressed both in theoretical and empirical research. Thus, for conducting this research, the instrument Knowledge Development Process - Constructivist (ProKnow-C) was used because it is a structured process for selection and analysis of literature by researchers, for the purpose of construction of knowledge on a particular subject, under the interests and boundaries of the researchers who put it into practice, according to a constructivist view, which allows for a critical analysis of the bibliographic portfolio (BP) built from the delimited fragment of the literature (ENSSLIN et al., 2015; LOOS; MERINO; RODRIGUEZ, 2016; VALMORBIDA; ENSSLIN, 2017; MARTINS; ENSSLIN; DUTRA, 2018).

To achieve the objective of this research, steps 1 and 2 of ProKnow-C were followed: (i) selection of the bibliographic portfolio; and (ii) bibliometric analysis. The first step, selection of the bibliographic portfolio, was performed as shown in Figure 1.
For the operationalization of ProKnow-C, groups of keywords are defined to represent the theme to be researched in the databases. The keywords used in the research were identified from seminal studies of the area, written by relevant theoretical authors, and from the previous experience of the researchers, published in other studies. The representativeness of these keywords was confirmed by the ProKnow-C adherence test. In this test, the articles of raw database were randomly selected, and their keywords were tested to be incorporated into the selection process. When new keywords are found, the process is restarted until the keywords representing the theme finished. The purpose of this constructivist procedure is to include all possible articles about the subject.

When the adherence of keywords is confirmed, the raw database of articles is defined, composed, in this research, by 608.291 gross references. Then, a fragment of the literature about the topic being addressed is selected after alignment analysis of title, abstract and full-text articles, recent articles and articles from the test of representativeness (analysing the references of the aligned articles). Through this structured process, 116 articles were considered aligned and representative in relation to the studied topic, composing the bibliographic portfolio of the present research. The articles that composed the bibliographic portfolio are presented and codified in Appendix A. The codes are used to refer the articles in the results.
2.2 Data analysis procedures

After selection of the 116 articles which composed the Bibliographic Portfolio, they were analyzed in seven points.

(I) Identification of most representative articles

The first point involves the identification of the most representative articles in the fragment of the selected literature. To identify the most representative set of articles, some methods were adopted. First, the Google Scholar website (scholar.google.com/) was consulted to identify the number of citations received by the works in general.

The co-citation network allows the identification of the degree of importance and representativeness of the work in relation to peers of the same BP. Thus, it is recognized which of these articles are references common to other studies in the subject of performance evaluation. The Ucinet software (BORGATTI; EVERETT FREEMAN, 2002) is applied to the construction of the network of co-citation among the articles of the BP.

In this aspect, through the Ucinet software, were performed the determination of: (i) density, corresponding to the number of relationships divided by the maximum number of possible relations; (ii) centrality, corresponding to the number of immediate connections that a node possesses, subdividing into InDegree – Interaction of the node with the others – and OutDegree – interaction of the other nodes with the node analyzed; and (iii) network centralization, expressed as a percentage, reveals how the network may be more or less centralized around particular nodes or sets of nodes (PARK; YOON; LEYDES DORFF, 2016).

(II) The most productive authors

To analyze the participation of the different authors in the literature, the identification of the co-authorship networks and the distinction between authors of theoretical and empirical articles were carried out. The number of articles of each author in the BP was counted, without distinction between authorship and co-authorship. Afterwards, they were grouped by frequency.

The developing of the authorship map is performed with the support of the VOSViewer software (VAN ECK; WALTMAN, 2010). The software allows the visualization
of clusters representing different research groups devoted to studying the theme. Relationships between different authors and how these groups can be integrated in some moments in the development of specific themes are presented.

The distinction analysis allows the recognition of the authors devoted to developing concepts and fundamentals on performance evaluation, from theoretical articles, and the authors devoted to the application and development of systems, from empirical articles. It is also possible to identify authors who permeate the two types of research, where they discuss the concepts and apply them to seek evidence, aiming the development of the area.

(III) Journals most devoted to the topic

For the analysis of the most representative journals on performance evaluation, the JCR (Journal Citation Reports), the SJR (Scientific Journal Rankings) and the H-index of each journal present in the BP were searched, making a distinction between theoretical and empirical articles.

The integrated analysis of the indicators allows the identification of the most devoted journals to the theme and those with prominent impact at the literature, knowing the journals that have more citations and more prestige in the field of knowledge (MOTKE; RAVANELLO; RODRIGUES, 2016).

(IV) Keywords most frequently used in the articles

The analysis of the keywords is performed by counting the number of occurrences. The VOSViewer software (VAN ECK; WALTMAN, 2010) is used to construct the distribution map of these keywords, presenting the relationship between the different keywords, to demonstrate and understand the terms more aligned to the theme.

(V) Field of development of the study

In general, when analysing the areas that use performance evaluation, three main research currents can be identified, originating from a number of disciplines: i) Accounting perspective; ii) the perspective of Production and Operations Management; and iii) perspective of Strategic Control and Business (GHALAYINI; NOBLE, 1996; PAVLOV; BOURNE, 2011; BITITCI et al., 2012; FRANCO-SANTOS et al., 2012). The content
analysis approached in the articles allows the identification of the areas and classification in one of the three possibilities. After, the number of articles in each discipline is counted. From this information, the degree of maturity of the studies can be assessed, considering the distinction between performance measurement and management.

(VI) Emphasis on performance measurement and management

The classification in 'performance measurement' comprises the processes of goal setting, collection, analysis and interpretation of performance data (MELNYK et al., 2014). The classification in 'performance management' encompasses the processes of assessing differences between real and desired results, identifying and signalling the differences that are critical (ensuring management intervention), understanding the causes of deficiencies that have occurred, and introducing actions to fill significant performance gaps (MELNYK et al., 2014). This analysis allows the visualization, among the areas of knowledge, of those that stand out, originating possible opportunities for future research.

(VII) Tools used by empirical studies

The analysis was performed through the identification of the tools used in the articles classified as empirical (point VI). After the tools were identified, the frequency of presentation was counted.

3 RESULTS

The first variable of analysis refers to articles with greater scientific recognition by peers, which compose the Bibliographic Portfolio. Together, 116 articles contained a total of 34,238 citations. Out of this total, the 10 major theoretical and empirical studies (five of each type) accounted for 13,818 citations.

As shown in Table 1, the main featured article is Neely, Gregory and Platts (1995), “Performance measurement system design: A literature review and research agenda”, published in 1995 and republished, upon invitation of the editor in 2005, to celebrate 25 years of the International Journal of Operations and Production Management, because of its relevance and timeliness, even 10 years after its publication. In this study, the authors sought
to highlight the main problems about performance measurement and the proposition of a research agenda. Although the authors come from the field of engineering, in this research they presented concepts of fields such as production, administration and accounting.

In addition, it can be seen that Andy Neely stands out not only because he published theoretical and empirical research studies which received the greatest scientific recognition, but also because he authored most studies among the major ones (5 articles). A diversity of fields could also be noted; Production, Accounting and Administration are the main fields of research for performance evaluation; moreover, there was an exchange of knowledge between fields, as in Bhagwat and Sharma (2007) and Kloom and Martin (2000): these studies used a methodology from administration and accounting (Balanced Scorecard) and applied it in production supply chain and in the public sector.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Theoretical studies</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Citation</th>
<th>Empirical studies</th>
</tr>
</thead>
</table>

Source: Developed by the authors.

Despite the major studies of the field are acknowledged by the scientific community, analyses should be made of the sources that they use to build their theoretical basis. While the activities of publication and innovation produce great amounts and various types of research
data (PARK; YOON; LEYDESDORFF, 2016), the analysis of co-citation of authors is an important method to discover the intellectual structure of a given scientific field (ZHAO; CHEN, 2014; MA et al., 2009), because a quality indicator for the analysis of authorship can play a guiding role by informing the research community (PARK; YOON; LEYDESDORFF, 2016).

Thus, to determine the density of the co-citation network of the BP, the number of relations divided by the maximum number of possible relations is calculated (PARK; YOON; LEYDESDORFF, 2016). The density of a network is simply the average value of binary inputs and, therefore, density and average value are identical. In this way, density found for the network of co-citation of this study (0.036) indicates that 3.6% of all possible collaborations occurred, which is considered to be a low percentage. After density was identified, the centrality of the network was measured.

The network of co-citation (Appendix B) allows to identify the most influential articles on the network. The determination of degree centrality is calculated by the number of articles with which a given article is directly connected. In-degree centrality (InDegree) corresponds to the sum of interactions that this particular article has with others (being cited), while the out-degree centrality (OutDegree) corresponds to the sum of the interactions that other articles present with that one (citing other works). Table 2 shows the main in-degree and out-degree centrality values of the articles in the BP.

The most influential article in the BP is [T11] - Bourne, Mills, Wilcox, Neely and Platts (2000), entitled “Designing, implementing and updating performance measurement systems”. Its influence is highlighted with 38 citations among the 115 (116-1) possible citations in the BP, because in this study, the authors address and discuss the phases of the life cycle of performance measurement systems (PMS): design, implementation, use and continuous update of the PMS. In this way, by segmenting the life cycle of PMS, each part of this cycle could be analyzed in more detail for the following research studies.

The article [T03] has out-degree centrality of 32. Although this research of Neely, Gregory and Platts (1995) does not stand out in terms of degree centrality, it is s a reference for scientific discovery, as previously mentioned.

are seven main reasons why business performance measurement has become so up-to-date: the changing nature of work; increased competition; initiatives for specific improvements; national and international awards for quality; changes in organizational roles; changes in external demands; and the power of information technology. In addition, the author describes the historical evolution of the theme of performance evaluation.

Even if there is a large volume of co-citations, it is interesting to emphasize that this factor does not necessarily represent an advance of the propositions performed by these studies, especially in relation to theoretical studies. The article by Bourne et al. (2000) - the greater centrality in the network of co-citation - contributes to the literature with the insertion of the concept of life cycle of the performance evaluation system, with the phases of design, implementation, use and review of the systems. In the article, Bourne et al. (2000) suggest the necessity of conducting longitudinal studies related to the complete cycle of the system, however among the articles of BP there are no studies that have effectively promoted the complete cycle proposed. In this case, only articles covering separate phases of the system life cycle, such as design and implementation, are not concerned with the evolution and feedback of the system, making it a potential opportunity for future research.

Table 2 – InDegree and OutDegree centrality of the articles of the BP and descriptive statistics

<table>
<thead>
<tr>
<th>Code</th>
<th>InD.</th>
<th>OutD.</th>
<th>Code</th>
<th>OutD.</th>
<th>InD.</th>
<th>Mean</th>
<th>InDeg</th>
<th>OutDeg</th>
<th>NrOutDeg</th>
<th>NrmInDeg</th>
</tr>
</thead>
<tbody>
<tr>
<td>T11</td>
<td>38.000</td>
<td>4.000</td>
<td>T46</td>
<td>26.000</td>
<td>0.000</td>
<td>4.190</td>
<td>4.190</td>
<td>3.643</td>
<td>3.643</td>
<td></td>
</tr>
<tr>
<td>T03</td>
<td>32.000</td>
<td>0.000</td>
<td>T35</td>
<td>18.000</td>
<td>7.000</td>
<td>5.029</td>
<td>6.872</td>
<td>4.373</td>
<td>5.976</td>
<td></td>
</tr>
<tr>
<td>T07</td>
<td>26.000</td>
<td>1.000</td>
<td>T33</td>
<td>18.000</td>
<td>6.000</td>
<td>486.000</td>
<td>486.000</td>
<td>422.609</td>
<td>422.609</td>
<td></td>
</tr>
<tr>
<td>T10</td>
<td>23.000</td>
<td>2.000</td>
<td>T44</td>
<td>17.000</td>
<td>0.000</td>
<td>25.292</td>
<td>47.223</td>
<td>19.124</td>
<td>35.707</td>
<td></td>
</tr>
<tr>
<td>T14</td>
<td>21.000</td>
<td>7.000</td>
<td>T20</td>
<td>15.000</td>
<td>8.000</td>
<td>4,970.000</td>
<td>7,514.000</td>
<td>3,758.034</td>
<td>5,681.664</td>
<td></td>
</tr>
<tr>
<td>E10</td>
<td>21.000</td>
<td>2.000</td>
<td>T38</td>
<td>15.000</td>
<td>1.000</td>
<td>2,933.828</td>
<td>5,477.828</td>
<td>2,218.395</td>
<td>4,142.025</td>
<td></td>
</tr>
<tr>
<td>T04</td>
<td>19.000</td>
<td>1.000</td>
<td>T34</td>
<td>15.000</td>
<td>1.000</td>
<td>70.498</td>
<td>86.683</td>
<td>61.303</td>
<td>75.377</td>
<td></td>
</tr>
<tr>
<td>E21</td>
<td>18.000</td>
<td>8.000</td>
<td>T47</td>
<td>14.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>T08</td>
<td>18.000</td>
<td>0.000</td>
<td>T39</td>
<td>14.000</td>
<td>1.000</td>
<td>26.000</td>
<td>38.000</td>
<td>22.609</td>
<td>33.043</td>
<td></td>
</tr>
<tr>
<td>T23</td>
<td>16.000</td>
<td>3.000</td>
<td>E48</td>
<td>13.000</td>
<td>0.000</td>
<td>116.000</td>
<td>116.000</td>
<td>116.000</td>
<td>116.000</td>
<td></td>
</tr>
<tr>
<td>T25</td>
<td>15.000</td>
<td>11.000</td>
<td>E25</td>
<td>13.000</td>
<td>5.000</td>
<td>Network Centralization (Outdegree)</td>
<td>19.130%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T05</td>
<td>13.000</td>
<td>2.000</td>
<td>T40</td>
<td>12.000</td>
<td>3.000</td>
<td>Network Centralization (Indegree)</td>
<td>29.656%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T02</td>
<td>12.000</td>
<td>0.000</td>
<td>T25</td>
<td>11.000</td>
<td>15.000</td>
<td>116.000</td>
<td>116.000</td>
<td>116.000</td>
<td>116.000</td>
<td></td>
</tr>
<tr>
<td>T09</td>
<td>10.000</td>
<td>3.000</td>
<td>T42</td>
<td>11.000</td>
<td>1.000</td>
<td>4.0100000</td>
<td>4.0100000</td>
<td>4.0100000</td>
<td>4.0100000</td>
<td></td>
</tr>
<tr>
<td>E06</td>
<td>10.000</td>
<td>0.000</td>
<td>T22</td>
<td>10.000</td>
<td>7.000</td>
<td>4.0100000</td>
<td>4.0100000</td>
<td>4.0100000</td>
<td>4.0100000</td>
<td></td>
</tr>
</tbody>
</table>

Source: Developed by the authors.

In the research “Dynamics of performance measurement systems”, Bititci, Turner and Begemann (2000) explored the use of IT-based management tools in order to ensure that the
performance measurement system of an organization continues to be integrated, efficient and effective at all times. The article shows that the levels of understanding at the time, together with the methods, tools and techniques available, were sufficient to develop truly dynamic performance measurement systems.

In Kennerley and Neely (2002), the authors seek to present a picture of the factors that affect the evolution of performance measurement systems, with data describing the forces that shape the evolution of measurement systems used by different organizations.

Among the first 15 studies that highlight degree centrality, only three are empirical. Articles [E10], of Neely, Platts, Richards, Gregory, Bourne and Kennerley (2000), with 21 citations; [E21] of Kennerley and Neely (2003) with 18 citations and [E06] of Flapper, Fortuin and Stoop (1996) with 10 citations. Although these studies have predominantly empirical characteristics, some of the authors are the most prominent in the field, with a history of academic research on the theme. In terms of importance for the performance evaluation literature, these studies have become fundamental for the advancement of the literature at the time, being precursors of new concepts until then theoretically discussed, as processes for the construction of indicators based on the organizational strategy and consequent design of the performance evaluation system, as described by Neely et al. (1995), Ghalayini and Noble (1996), Neely et al. (1997), Otley (1999) and other theoretical studies subsequently published. These studies have become references for applications and design of new models and systems.

Network centralization, expressed in percentage, reveals particular properties of the network structure as a whole and refers to general cohesion or to the integration of the network (PARK; YOON; LEYDESDORFF, 2016). Networks, for example, can be more or less centered around nodes or sets of specific nodes. In this research, centralization indexes were InDegree (19.30%) and OutDegree (29.656%).

The analysis of degree of betweenness corresponds to the possibility of a node (article) to mediate the communication between the pairs of nodes (other articles). The intersection is, therefore, a measure of the number of times that a vertex occurs in a geodesic. Normalized interaction in centrality is when interaction is divided by maximum possible interdependence expressed in percentage. The Table 3 shows this analysis.
Table 3 – Major degrees of betweenness of the co-citation network of articles in the BP and descriptive statistics

<table>
<thead>
<tr>
<th>Degrees of betweenness</th>
<th>Descriptive statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Betweenness nBetweenness</td>
</tr>
<tr>
<td>[T11]</td>
<td>101.767 0.776</td>
</tr>
<tr>
<td>[T20]</td>
<td>100.340 0.765</td>
</tr>
<tr>
<td>[T25]</td>
<td>99.182 0.757</td>
</tr>
<tr>
<td>[T35]</td>
<td>89.963 0.686</td>
</tr>
<tr>
<td>[T14]</td>
<td>85.129 0.649</td>
</tr>
<tr>
<td>[T33]</td>
<td>74.280 0.567</td>
</tr>
<tr>
<td>[E21]</td>
<td>63.042 0.481</td>
</tr>
<tr>
<td>[E25]</td>
<td>49.387 0.377</td>
</tr>
<tr>
<td>[T04]</td>
<td>47.285 0.361</td>
</tr>
<tr>
<td>[T05]</td>
<td>46.194 0.352</td>
</tr>
<tr>
<td>[T32]</td>
<td>43.647 0.333</td>
</tr>
<tr>
<td>[T36]</td>
<td>42.806 0.327</td>
</tr>
<tr>
<td>[T21]</td>
<td>39.840 0.304</td>
</tr>
<tr>
<td>[T16]</td>
<td>39.525 0.301</td>
</tr>
<tr>
<td>[T22]</td>
<td>38.149 0.291</td>
</tr>
</tbody>
</table>

Source: Developed by the authors.

Again, the article [T11], of Bourne et al. (2000), stands out with the highest number of betweenness. Because this research has a higher degree of betweenness, it is an article with a privileged position to the extent that readers "fall" into the geodesic paths between other pairs of this network. The first value (101,767) represents the total numbers of pairs of nodes that the article is able to connect. The second value (0.7760) corresponds to the normalized degree of betweenness, in percentage.

The authors of the articles in the BP are presented on an authorship map in Figure 2, developed with the software VOSViewer. It shows the clusters relative to the groups of research on the subject.
Figure 2 – Authorship map of articles of BP

This analysis allows the visualization of the existence of prominent groups as regards research on performance evaluation. Centrally, the map shows the existence of four large clusters, connected among themselves, by means of outstanding authors, such as Andy Neely, Mike Bourne, Ken Platts, Mike Kennerley, Monica Franco-Santos, Veronica Martinez and Umit Bititci. Together, this large group is responsible for 37 articles of the fragment selected from the literature, and it represents 32% of the total. Other clusters, led by David Otley, Leonardo Ensslin, Marc Wouters and Paolo Taticchi, also feature highlights.

The visualization of the most representative research groups in the area allows the expansion in the understanding of the evolution of the performance evaluation research. From this understanding it is possible to provide researchers of the theme with understanding about
new directions, trends and emerging themes, which are emerging as topics of already renowned researchers and subjects suggested as a focus of future investigations.

This analysis contributes to better understanding the integration among researchers in the area, the connection of different expertise and the direction of new studies. Are shown the relationships between different researchers and how these relationships promote new studies on fields, phenomena and behaviours that are still poorly explored. As an example, it is appropriate to highlight the growing need for social research related to performance evaluation systems, promoted by Bititci et al. (2012). In this study, the authors emphasize the inevitability of the discussion about concepts of influence of the systems in the individuals and how this influence relates to changes in the performance evaluation system, in a movement of autopoiesis (adapting and evolving). From the publication of Bititci et al. (2012), it is possible to identify that the group led by Umit Bititci has applied efforts to develop new studies related to social phenomena promoted by the use of performance evaluation systems, which emerged in the last years. Thus, we highlight the importance of this analysis as a supplier of subsidies for new studies.

An analysis was also made of the studies according to their nature, and the authors of theoretical and empirical studies were identified. Figure 3 shows the highlights found in this analysis. As a result, the following authors of theoretical articles stand out: Andy Neely, who authored 9 theoretical articles, and Mike Bourne, who authored 6 articles. Umit Bititci can also be cited as author of theoretical studies. However, he stands out from other authors, especially for the total number of empirical works he has authored: 7 articles.

Another factor which was analyzed for outstanding authors is their research path. It was confirmed that among the most prominent authors of the BP, there is a line of specific and continuous research in Performance Evaluation, which currently involves the integration of multiple institutions.
Figure 3 – Outstanding authors in BP

Table 4 shows the journals that have published more articles on performance evaluation.

<table>
<thead>
<tr>
<th>Journal</th>
<th>T*</th>
<th>E*</th>
<th>C***</th>
<th>Location</th>
<th>JCR</th>
<th>SJR</th>
<th>H-index</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. J. of Operations &amp; Production Management</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>UK</td>
<td>2.252</td>
<td>2.198</td>
<td>94</td>
</tr>
<tr>
<td>Management Accounting Research</td>
<td>6</td>
<td>5</td>
<td>11</td>
<td>USA</td>
<td>-</td>
<td>1.913</td>
<td>56</td>
</tr>
<tr>
<td>I. J. of Productivity &amp; Performance Management</td>
<td>3</td>
<td>7</td>
<td>10</td>
<td>UK</td>
<td>-</td>
<td>0.785</td>
<td>31</td>
</tr>
<tr>
<td>Measuring Business Excellence</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>UK</td>
<td>-</td>
<td>0.338</td>
<td>19</td>
</tr>
<tr>
<td>I. J. of Production Economics</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>Netherlands</td>
<td>2.782</td>
<td>2.749</td>
<td>144</td>
</tr>
<tr>
<td>Production Planning &amp; Control</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>UK</td>
<td>1.532</td>
<td>1.295</td>
<td>50</td>
</tr>
<tr>
<td>I. J. of Business Performance Management</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>UK</td>
<td>-</td>
<td>0.194</td>
<td>15</td>
</tr>
<tr>
<td>Computers &amp; Industrial Engineering</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>UK</td>
<td>2.086</td>
<td>1.63</td>
<td>88</td>
</tr>
<tr>
<td>British Accounting Review</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>USA</td>
<td>1.340</td>
<td>0.711</td>
<td>42</td>
</tr>
<tr>
<td>I. J. of Contemporary Hospitality Management</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>UK</td>
<td>2.176</td>
<td>1.329</td>
<td>35</td>
</tr>
<tr>
<td>Accounting, Organizations and Society</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>UK</td>
<td>2.464</td>
<td>2.515</td>
<td>90</td>
</tr>
<tr>
<td>Benchmarking: An International Journal</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>UK</td>
<td>-</td>
<td>0.556</td>
<td>38</td>
</tr>
<tr>
<td>Journal of the Operational Research Society</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>UK</td>
<td>1.225</td>
<td>1.026</td>
<td>75</td>
</tr>
<tr>
<td>Management Decision</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>UK</td>
<td>1.134</td>
<td>0.909</td>
<td>48</td>
</tr>
<tr>
<td>I. J. of Production Research</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>UK</td>
<td>1.693</td>
<td>1.445</td>
<td>91</td>
</tr>
<tr>
<td>Journal of Manufacturing Technology Management</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>UK</td>
<td>-</td>
<td>0.605</td>
<td>45</td>
</tr>
</tbody>
</table>

Caption: * Theoretical; ** Empirical; *** Consolidated. UK=United Kingdom; USA=United States of America.
Source: Developed by the authors.
The scope of most of the journals that publish the largest number of articles on performance evaluation (62.5%) was oriented to the field of operations, production and productivity. This is the case of the International Journal of Operations & Production Management, which accounts for most publications on the subject, both in empirical and in theoretical articles, and the International Journal of Productivity & Performance Management. However, other journals, e.g., Management Accounting Research, are oriented to the publication of research on managerial accounting. It was also found that the vast majority (82%) of the selected articles was published by journals based in the United Kingdom.

Figure 4 shows the distribution of keywords used by the articles.

Figure 4 – Keywords most used by the articles

Source: Developed by the authors.
It was found that the words “performance management” and “performance measurement” are the most commonly used. It is also possible to identify the set of topics relative to performance evaluation, aligned with the various fields of research on this theme. They were found to be aligned with themes such as local government, strategic management, small and medium-sized enterprises, strategy and competitiveness, among others, in addition to the tools used in research, such as Balanced Scorecard and AHP.

The keywords analysis allows the recognition of the existing dismemberment in the literature on performance evaluation, with an important finding. As Lebas (1995) describes, the performance evaluation is composed of measurement and management of the performance. However, the literature related to the theme has developed, for a long time, with the distinct visualization between these two processes. The concern with "what to measure" was a subject frequently discussed throughout the evolution of the theme, but few efforts were made to understand "how to use measures to manage performance in organizations". The management of the information promoted by the measurement system gives a better understanding of the factors of success and failure in an organization, as well as an understanding of the informational needs of the managers, to be supplied by the measurement system. In this sense, the integration between the two processes is fundamental for the advancement in the field, together with the understanding of the social phenomena involved in the processes made possible by the performance evaluation systems. Thus, the keywords analysis contributes to identify the emphasis adopted by the articles, allowing the view of the literature panorama and relationships between its keywords.

With respect to the areas that use performance evaluation, it is widely accepted that organizational performance is a multifaceted concept and, therefore, it is not surprising that, more than once, the issue of how the performance of organizations can best be measured was approached by a variety of researchers from different disciplines (NEELY, 1999). Field research was approached by a diverse group of people (FRANCO-SANTOS et al., 2007), as can be seen with the authors identified previously.

As regards the field of development of the study, the classification of areas is aligned with Franco-Santos, Lucianetti and Bourne (2012). Based on a wide variety of disciplines, including accounting, strategic management and business strategy, human resources management, production and operations management, marketing, service management, industrial engineering, facilities management, public sector management, psychology, change management and organizational behaviour (NEELY, 1999; FRANCO-SANTOS et al., 2007;
WAGGONER; NEELY; KENNERLEY, 1999; BERRY et al., 2009; BITITCI et al., 2012; YADAV; SUSHIL; SAGAR, 2014; VAN CAMP; BRAET, 2016), they sorted the fields into Accounting, Strategy and Business Operations and Engineering (FRANCO-SANTOS et al., 2012).

It was found that the field of operations covers 50% of the selected studies, i.e., 35 empirical articles, followed by the field of administration and strategy, which concentrates 42% of the studies (29 articles). The field of accounting has only 8% of empirical studies (5 articles) on performance evaluation, which is surprising, considering that accounting paved the way for studies in the field (OTLEY, 1999; BITITCI et al., 2012).

As regards the emphasis placed on performance evaluations, the works were analyzed for their concern with performance measurement and effective performance management. Performance measurement includes procedures for definition of objectives, data collection, analysis and interpretation of data on performance, while performance management involves evaluating differences between actual and desired results, identifying and signalling differences that are critical (thus ensuring the intervention of management), understanding why deficiencies have occurred, and, when necessary, introducing and monitoring corrective measures to bridge significant gaps in performance (MELNYK et al., 2014).

There is a predominant emphasis in the literature on performance measurement (55%) (38 articles). The focus of the other 29% (20 articles) lies only on performance management. There is still a small portion of works that are concerned with the integration of these two fields, considering that 15% of the works (11 articles) being analyzed were focused on measuring performance, i.e., they were oriented towards the effective use of information produced for management of an organization.

Another analysis was performed for the performance evaluation tools. As mainly results, the Balanced Scorecard is the tool that predominates in most studies (23%), whether used alone or in combination with another tool. Yet, research in 30% of the works was developed by proposing models based on the literature. The other 24% was based on proposals developed by the authors.
4 FINAL REMARKS

Performance evaluation is crucial to the management of any organization. Over time, it has been gaining interest from the academy. However, it is clearly necessary to reflect on research conducted on performance evaluation, in order to give scientific contributions to identify and seek solutions to practical problems experienced in the organizational context.

Thus, the objective of this research was to identify the development of literature on performance evaluation, in order to identify the articles with greater scientific recognition, which are the most relevant, the most cited and the most referenced as well as authors, journals, keywords in use, fields of development of the research studies, emphasis on performance measurement and management and tools used by the scientific community which is devoted to the theme.

The analyses showed that the main featured article is Neely, Gregory and Platts (1995), with “Performance measurement system design: A literature review and research agenda”, published in 1995 and republished, upon invitation of the journal in 2005, to celebrate 25 years of the International Journal of Operations and Production Management, because of its relevance and timeliness, even 10 years after its publication. The article of greater influence within the BP is “Designing, implementing and updating performance measurement systems”, of Bourne, Mills, Wilcox, Neely and Platts (2000), with the highest number of betweenness, having been cited by 38 works in the BP. The co-citation analysis of the articles reinforces the influence of these two studies on the performance evaluation field. The studies with the highest number of co-citations denote that, besides the general recognition, through the total number of citations, it also has its recognition among the main manuscripts of the area, recognized in this fragment of literature.

There were four large clusters of authors, connected among themselves, by means of outstanding authors, such as Andy Neely, Mike Bourne, Ken Platts, Mike Kennerley, Monica Franco-Santos, Veronica Martinez and Umit Bititci, who represent leading researchers of the theme of performance evaluation. Together, this large group is responsible for 37 articles of the fragment selected from the literature, and it represents 32% of the total.

Most of the journals which were most receptive to the theme (62.5%) are from the field of operations, as is the case of International Journal of Operations & Production Management and International Journal of Productivity & Performance Management. The
second most receptive journal was Management Accounting Research, geared towards managerial accounting. The vast majority (82%) was published by journals based in the United Kingdom.

The keywords “performance management” and “performance measurement” are the most commonly used. However, there is a wide range of fields related to the theme. The area of operations covers 50% of the studies; management and strategy cover 42% of the works while accounting covers only 8% of empirical studies on performance evaluation, which is surprising, considering that accounting paved the way for studies in the field (OTLEY, 1999; BITITCI et al., 2012).

In the literature, there is great emphasis on performance measurement (55%). Only 29% of research was focused on performance management. There is still a small portion of works that are concerned with the integration of the two processes, demonstrating that, despite the seminal concepts, such as Lebas (1995), there is a lack of literature regarding a complete performance evaluation system that deals with the measurement and management of the performance. This lack may also be associated with the subjectivity of understanding the terms adopted by the main works of the area, which sometimes use different terms to designate the same object. Still, it was noted that the Balanced Scorecard is the tool that predominates in most studies (23%), used either individually or in combination with another tool; in 30% of the works, research was developed by proposing models based on the literature and, in 24% of them, research was based on a proposal developed by the authors.

It should be emphasized that this article sought to highlight the literature about the theme aiming to allow for an overview of such literature, in order to promote the development of new research studies to align performance evaluation with organizational needs. Because of the volume of literature on this subject, a representative fragment had to be selected for the proposed analysis. This selection was performed with the aid of ProKnow-C. Moreover, the processes of the representativeness test and feedback at the time of selection, whose aim was to eliminate the eminent bias of the research, were an attempt to ensure that no important article was left out of the selection.

Overall, the main contributions of the work reside in providing a general overview of the literature on performance evaluation, presenting the highlights found. We seek to broaden the knowledge about the theme from the collection of seminal articles on the theme and the
main studies related to the development and practical application of the concepts of performance evaluation, making the article a reference to guide new research in the area.

Still, in conceptual terms, it is appropriate to emphasize that the study offers a vision of opportunities for evolution to the field, especially with regard to the emerging need to deal with the impacts and social influences from the use of the performance evaluation systems, raised from studies such as Bititci et al. (2012). Social bias is fundamental for the progress of concepts and the development of a theory for the area. The process of performance evaluation influences the behaviour of involved people. This influence can change the way people deal with system-driven learning and other behavioural aspects. In this sense, the literature lacks the development of empirical and theoretical studies that advance in this knowledge.

REFERENCES


DEVELOPMENT OF PERFORMANCE EVALUATION THEME: A SYSTEMATIC ANALYSIS OF THE LITERATURE


APPENDIX A: Codes of the works composing the bibliographic portfolio used during research

Theoretical studies


Empirical studies


[E54] BEHERY, M.; JABEEN, F.; PARAKANDI, M. Adopting a contemporary performance management system: A fast-growth small-to-medium enterprise (FGSME) in the


APPENDIX B: Network of co-citation among articles in the BP and their references

Source: Developed by the authors.