

Formulation and Validation of a Performance Measurement Model in Governmental Organizations Using the Balanced Scorecard (A Mixed-Methods Approach)

Jafar. Heidarian¹, Gholamreza. Farsadamanollahi^{2*}, Mohammad Ali. Bidari³

¹ Department of Accounting, CT.C., Islamic Azad University, Tehran, Iran.

* Corresponding author email address: g_farsad@iauctb.ac.ir

Article Info

Article type:

Original Research

How to cite this article:

Heidarian, J. , Farsadamanollahi, G.R. , & Bidari, M. (2026). Formulation and Validation of a Performance Measurement Model in Governmental Organizations Using the Balanced Scorecard (A Mixed-Methods Approach). *Journal of Resource Management and Decision Engineering*, 5(1), 1-14.

<https://doi.org/10.61838/kman.jrmd.5.1.189>



© 2026 the authors. Published by KMAN Publication Inc. (KMANPUB). This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

The main objective of this study is to formulate and validate a performance measurement model in governmental organizations using the balanced scorecard (a mixed-methods approach). This research employed a mixed-methods design and was conducted in two qualitative and quantitative phases. In the qualitative phase, data were collected through semi-structured interviews and analyzed using thematic analysis. In the quantitative phase, data were gathered through a researcher-made questionnaire and analyzed using structural equation modeling. The statistical population in the qualitative section included academic experts and executive managers in several governmental organizations, from which 14 participants were selected through purposive sampling and theoretical saturation. The statistical population in the quantitative section consisted of employees from selected governmental organizations and institutions, of whom 400 individuals were chosen using simple random sampling. The findings of the qualitative phase produced a thematic network of the research model, including seven organizing themes and 64 basic themes. The results indicate that localization of the balanced scorecard and the definition of indicators consistent with the mission, objectives, structure, and culture of Iran's public sector have a significant effect on improving transparency, accountability, service quality, and strategic planning in these organizations. Establishing this framework, in addition to measuring financial performance, also encompasses internal processes, stakeholder satisfaction, and the enhancement of organizational learning and growth, thereby enabling better resource allocation and data-driven decision-making for managers. Moreover, challenges such as organizational resistance, weak technological infrastructures, and insufficient training were identified as obstacles to full implementation, for which practical solutions were proposed. In the quantitative section, the path analysis of the model was examined and confirmed.

Keywords: Performance Measurement, Balanced Scorecard, Governmental Organizations, Financial Dimension.

1. Introduction

Performance evaluation is a fundamental component of modern management systems, providing organizations with the tools to assess their effectiveness, efficiency, and alignment with strategic objectives. Over the past decades, the *Balanced Scorecard (BSC)* has emerged as one of the most influential frameworks for translating strategy into measurable outcomes. Initially introduced by Kaplan and Norton, the BSC emphasizes financial, customer, internal process, and learning and growth perspectives, thereby enabling managers to achieve a more comprehensive view of organizational performance (Kaplan & Norton, 2021). This multidimensional approach has gained traction across both private and public sectors, providing a systematic mechanism to align operations with mission and vision statements (Niven, 2020).

Scholars and practitioners agree that traditional performance evaluation systems overly rely on financial indicators, which only offer a retrospective view of organizational success. In contrast, the BSC incorporates both financial and non-financial measures, thereby balancing short-term and long-term objectives (Song, 2022). Empirical studies have demonstrated its effectiveness in various industries and contexts. For instance, the healthcare sector has widely adopted the BSC to evaluate service quality and efficiency (Lee et al., 2023), while governmental organizations have employed it to enhance transparency, accountability, and strategic alignment (Mohammadi et al., 2024; Sharaf-Addin & Fazel, 2020).

The BSC's significance is particularly evident in public sector organizations, where measuring performance goes beyond profitability. Public institutions are required to ensure service quality, fairness, and the equitable distribution of resources (Ghasemi Esfahlan & Khabaz Bavyl, 2020). In such contexts, performance evaluation models must reflect diverse stakeholder expectations, including those of citizens, employees, policymakers, and society at large (Ghanbari et al., 2020). Consequently, scholars have increasingly advocated the integration of sustainability, ethics, and social responsibility dimensions into the BSC framework (Nikbakht & Rahimipour, 2022).

Despite its global acceptance, BSC implementation in developing countries has faced significant challenges. Resistance to change, lack of managerial competencies, and weak technological infrastructure are common obstacles (Mansouri, 2020). In Iran, several studies have sought to adapt the BSC framework to the specific socio-cultural and

institutional environment of public organizations. For example, customized models have been designed for technical and vocational training organizations (Ghanbari et al., 2020), the Iran Health Insurance Organization (Mohammadi et al., 2024), and other governmental agencies (Aliabadi et al., 2019). These localized approaches demonstrate the necessity of tailoring global management frameworks to national contexts, ensuring their relevance and practical applicability (Iranmehr & Piriaei, 2024).

The theoretical and practical appeal of the BSC lies in its ability to provide a holistic, multidimensional lens through which performance can be measured and improved. Scholars such as Niven (Niven, 2020) emphasize its adaptability for nonprofit and governmental agencies, while empirical research in small- and medium-sized enterprises confirms the role of leadership in successful implementation (Aranda & Odriozola, 2021). More recently, the rise of positive management perspectives has further enriched the BSC framework, suggesting that integrating psychological well-being and employee engagement enhances organizational outcomes (Cignitas et al., 2022).

Internationally, BSC adoption has yielded important insights into organizational effectiveness. In China, performance evaluation models based on innovation and industry–university collaboration have highlighted the role of dynamic capabilities in sustaining competitiveness (Bai et al., 2020). In Saudi Arabia, universities have used the BSC to structure performance management, identifying the cultural and institutional factors that influence adoption (Sharaf-Addin & Fazel, 2020). Similarly, in Ethiopia, the use of BSC in public hospitals has provided evidence on both the facilitators and barriers to effective implementation (Yeshaw et al., 2025).

In addition to its widespread applications, researchers have explored methodological innovations in performance evaluation. The integration of interpretive structural modeling (Iranmehr & Piriaei, 2024), confirmatory factor analysis (Kermshahi & Salehi Tabandeh, 2024), and machine learning models (Guleria & Sood, 2023) has provided robust tools for validating and enhancing the BSC framework. These approaches ensure that the constructs measured truly capture the complex realities of organizational performance. Moreover, advances in information systems and digital technologies have introduced new methods for collecting, analyzing, and interpreting performance data (Morabito, 2016; Zimmermann, 2017).

The academic discourse on performance evaluation is not limited to organizational frameworks alone but also extends to sector-specific applications. In the field of information technology and networking, performance evaluation methods have been applied to storage systems, RF propagation models, and container technologies (Chowdhury et al., 2019; Morabito, 2016; Shakir et al., 2022). In material sciences, evaluations have addressed the functional capabilities of innovative materials (Shen et al., 2019), while in the energy sector, studies have focused on electroreduction processes (Wang et al., 2021). These diverse applications underscore the universality of performance evaluation principles across academic and practical domains.

The BSC framework is not without its criticisms. Some scholars argue that implementing the BSC can become a bureaucratic exercise if not integrated into the daily routines of organizations (Armstrong, 2020). Others highlight the risk of focusing too narrowly on quantitative indicators, neglecting qualitative dimensions such as employee satisfaction, culture, and innovation (Martínez-Caro et al., 2015; Sindhu et al., 2019). Furthermore, the absence of adequate feedback mechanisms and weak alignment between indicators and strategic goals may undermine the effectiveness of BSC adoption (Samei et al., 2019).

Nevertheless, evidence suggests that when properly adapted and executed, the BSC significantly enhances decision-making, strategic alignment, and stakeholder satisfaction (Song, 2022). For example, in higher education, the application of performance evaluation models has strengthened the quality of e-learning systems (Martínez-Caro et al., 2015), while in the financial sector, banks have utilized the sustainable BSC to align profitability with social and environmental objectives (Nikbakht & Rahimipour, 2022). In public cultural organizations, prioritization of performance indicators has been used to guide senior management decisions (Aliabadi et al., 2019).

The expansion of BSC applications also reflects the increasing complexity of organizational environments. With globalization, digitalization, and growing stakeholder expectations, organizations are compelled to adopt frameworks that can capture both tangible and intangible performance outcomes (Zimmermann, 2017). Moreover, the interplay between organizational learning, innovation, and accountability is now recognized as a critical factor in ensuring sustainable performance (Armstrong, 2020; Cignitas et al., 2022).

A growing body of research also underscores the contextual dependency of BSC success. In some cases, organizational culture and leadership commitment determine the effectiveness of the framework (Aranda & Odriozola, 2021), while in others, resource availability and institutional structures act as enablers or barriers (Yeshaw et al., 2025). These findings highlight the importance of situating performance evaluation models within the socio-political and economic realities of each country (Ghasemi Esfahlan & Khabaz Bavyl, 2020; Iranmehr & Piriaei, 2024).

Another important contribution of recent studies is the emphasis on adaptability. In Iran, research on auditing firms has investigated the dimensions of an effective BSC through factor-analytic approaches (Kermshahi & Salehi Tabandeh, 2024). In military institutions, BSC models have been adjusted to reflect jihadi management principles (Iranmehr & Piriaei, 2024). Similarly, performance transparency has been emphasized as a critical outcome in public organizations, linking evaluation systems to the broader goals of accountability and good governance (Ghasemi Esfahlan & Khabaz Bavyl, 2020).

Given these insights, it is evident that the BSC framework is more than a performance measurement tool; it is a strategic management system capable of integrating organizational vision with operational execution (Kaplan & Norton, 2021). Its applications across different contexts—ranging from healthcare (Lee et al., 2023), insurance (Mohammadi et al., 2024), and banking (Nikbakht & Rahimipour, 2022) to education (Martínez-Caro et al., 2015) and military organizations (Iranmehr & Piriaei, 2024)—confirm its adaptability and enduring relevance.

Despite its strengths, challenges remain. Resistance to change, inadequate training, and insufficient integration of qualitative measures continue to limit the transformative potential of the BSC (Mansouri, 2020; Samei et al., 2019). Moreover, cultural and institutional barriers often impede the localization of the framework in non-Western contexts (Sharaf-Addin & Fazel, 2020). These barriers underline the necessity for empirical research that contextualizes the BSC within specific organizational and national settings (Aliabadi et al., 2019; Ghanbari et al., 2020).

In light of the gaps identified, the present study aims to design and validate a performance measurement model for governmental organizations using the balanced scorecard approach, adapted to the socio-cultural and institutional context of Iran.

2. Methods and Materials

In this study, the methodology was designed and implemented based on *Saunders' Research Onion Model*. This model, which encompasses layers from research philosophy to data collection and analysis methods, provides a systematic framework for advancing the study. The present research adopted a mixed-methods approach (quantitative–qualitative), and in order to achieve the main objective, both qualitative and quantitative phases were employed as complementary to each other.

In the qualitative phase, the research had an exploratory nature and, relying on the interpretivist philosophy and inductive approach, aimed to identify and deeply understand the underlying components of the phenomenon under investigation. This phase was conducted using a single-case study strategy and focused on a population consisting of academic experts in the fields of finance and accounting, as well as senior managers of governmental organizations. Sampling was carried out purposefully, and 14 semi-structured interviews were conducted until theoretical saturation was reached. The main data collection tool consisted of semi-structured interviews guided by open and key questions. The qualitative data were analyzed using the *thematic analysis* method, through stages including familiarization with the data, open coding, identification of initial themes, review, and final definition of themes. Ultimately, this process led to the extraction of the main dimensions of the study.

In the quantitative phase, the research was designed and executed based on positivist philosophy and a deductive approach. The statistical population included experts from selected departments in Tehran (Tax Administration, Department of Education, Municipalities, and the Social Security Insurance Organization) who collaborated in the study. Sampling was conducted using convenient random sampling, and since the population was unlimited, the sample size was estimated at 384 individuals using Cochran's formula; ultimately, 400 individuals participated in this study. The data collection tool was a researcher-made questionnaire derived from the themes identified in the qualitative phase, covering the main variables of the research. The validity of the questionnaire was confirmed by experts in the field of finance, and its reliability was verified through Cronbach's alpha coefficient higher than 0.7.

Quantitative data were analyzed using SPSS software at two levels: descriptive (demographic characteristics) and inferential (hypothesis testing). To evaluate and confirm the

conceptual model, confirmatory factor analysis was performed using AMOS software, which demonstrated that the factor structure of the questionnaire and the relationships between the variables were consistent with the collected data.

Overall, the integration of qualitative and quantitative methods made it possible to present a practical and valid model for performance measurement in governmental organizations using the balanced scorecard approach. In this way, the qualitative phase identified the key dimensions and concepts, while the quantitative phase evaluated and confirmed them within the target population. The main role of confirmatory factor analysis was to stabilize and finalize the structure of the model.

3. Findings and Results

Most of the interviewees were male (72%), while the lowest percentage was female (28%). Regarding work experience, 7% had 10–15 years, 50% had 15–20 years, and 43% had more than 20 years. Based on education level, 64% held doctoral degrees, while 36% held master's or bachelor's degrees.

For the qualitative data analysis, thematic analysis was applied, and among different methods, the *thematic network analysis* method was used. Initially, the preliminary theoretical codes were extracted to identify themes. At this stage of the study, the concepts and key points obtained regarding “developing a model for applying the balanced scorecard framework as a tool for performance measurement in governmental organizations” were listed from the interview process. Accordingly, statements, concepts, and items extracted from the interviews were subjected to precise analysis and harmonization (choosing more accurate wording and eliminating common concepts), which resulted in 190 items. The derived themes were arranged in a checklist for conducting interviews, and after further interviews with experts, some items were removed or revised. In the following section, several examples of interviews are referenced.

In the next stage, thematic network analysis was carried out. After comparing the extracted concepts, related concepts were categorized into overarching categories, and based on titles from relevant theories or concepts derived from the study, general titles were assigned to the categories. In this way, after constant comparison of the interview responses, similar responses were aligned and similar concepts were extracted from them. Additionally, closely

related items were merged, and the themes were classified into seven main categories:

- **Category One: Strengths and Weaknesses of Systems** – including 39 basic themes.
- **Category Two: Dimensions of the Balanced Scorecard Framework** (covering desirable characteristics of the financial perspective, internal processes, etc.) – including 31 basic themes.
- **Category Three: Drivers of the Balanced Scorecard Framework** (reasons and necessities for its creation, such as the need for accountability and transparency) – including 15 basic themes.
- **Category Four: Strategies and Practical Recommendations for Establishing the Balanced Scorecard Framework** – including 20 basic themes.
- **Category Five: Consequences of Applying the Framework** (positive results of implementing this model) – including 12 basic themes.
- **Category Six: Current Situation** (strengths and weaknesses of governmental systems) – including 15 basic themes.

- **Category Seven: Desired Situation** – including 27 basic themes.

Subsequently, the thematic network was presented. The purpose of the thematic network is to establish relationships among the generated categories. This process is usually carried out based on a paradigmatic model and assists the theorist in facilitating the process of theory building. In the thematic network stage of the present study, the relationship between the main category and other categories was identified. At this stage, the main and subcategories were interconnected so that theoretical concepts could be compiled to identify the factors influencing “the performance measurement model in governmental organizations using the balanced scorecard.” These steps enabled the researcher to integrate the concepts obtained in earlier stages and employ them to present the thematic network.

In the next stage, the screening of the generated themes is conducted.

In this section, in order to examine the importance of each organizing theme within each basic theme, expert surveys were carried out, and the most important categories were selected and placed in the thematic network model.

Table 1

Thematic Analysis

Organizing Themes	Main Themes	Basic Themes
Strengths and Weaknesses of Systems	Positive Points	Increasing employee participation; Greater transparency and accountability; Providing feedback to employees and managers; Creating a clear framework for performance evaluation; Monitoring progress and continuous assessment
	Weak Points	Lack of comprehensiveness of indicators and excessive focus on quantitative indicators; Inconsistency of indicators with the tasks and responsibilities of some positions; Lack of incentive and disciplinary mechanisms aligned with evaluation results; Lack of use of evaluation results in performance improvement and decision-making; Employee resistance to evaluation
Dimensions of the Balanced Scorecard Framework (including desirable characteristics of the financial perspective, internal processes, etc.)	Financial Perspective	Cost reduction; Increasing financial productivity and value creation for society; Transparency and accountability in the optimal allocation and utilization of financial resources
	Customer (Citizens) Perspective	Increasing citizen satisfaction; Enhancing public trust in governmental organizations; Responding to citizen demands
	Internal Process Perspective	Standardization of processes; Development of e-government; Improving the quality of provided services
	Learning and Growth Perspective	Human resource development; Improving employee motivation and job satisfaction; Attracting and retaining top talent
	Social and Environmental Perspective	Observance of social and ethical responsibilities; Preservation of the environment and natural resources; Creating equal opportunities for all members of society
Drivers of the Balanced Scorecard Framework (causes and necessities, such as the need for accountability and transparency, etc.)	Drivers	Accountability and transparency; Improved performance management; Enhancement of service quality; Optimal allocation of resources; Reduction of corruption and assurance of administrative integrity; Increased public oversight and citizen participation in performance monitoring; Improvement of planning and budgeting processes
Strategies and Practical Recommendations for Establishing the Balanced Scorecard Framework	Strategies	Defining measurable and assessable indicators; Training and empowering public sector employees; Developing mission, vision, and strategic objectives; Designing performance indicators; Establishing data collection and analysis systems; Developing a

Consequences of Applying the Balanced Scorecard Framework (positive results of implementing this model)	Consequences	strategic roadmap and linking strategy to operational actions; Creating a steering committee; Utilizing information technology and performance management software; Facilitating stakeholder participation in process design
Current Situation (existing strengths and weaknesses)	Weak Points	Enhancing transparency and accountability; Improving service quality; Increasing citizen satisfaction; Focusing on strategic objectives; Improving data-driven decision-making; Strengthening the culture of learning and development
Desired Situation (Recommendations)	Recommendations	Sole focus on quantitative indicators and lack of attention to qualitative indicators; Superficial and formal evaluations; Lack of use of evaluation results in decision-making; Need for specialized skills in design and application; Misalignment of indicators with strategic objectives; Difficulty in measuring some qualitative indicators; Possibility of manipulating results to present a positive image
		Developing comprehensive and integrated performance evaluation systems; Balanced attention to qualitative and quantitative indicators; Using evaluation results to improve performance and decision-making; Defining clear and measurable strategic objectives; Designing transparent and simple processes for performance measurement and monitoring; Establishing effective links between strategic objectives and operational actions; Utilizing modern technologies, IT, and smart tools for data collection and analysis; Designing appropriate reward and incentive systems based on performance; Emphasizing organizational culture and values; Paying attention to sustainability indicators and social responsibility

In this section, the process of data analysis and the extraction of basic, organizing, and overarching themes from the raw data obtained from the interviews is demonstrated in the table above. After transcribing the interviews, quotations that explicitly or implicitly referred to the research questions were selected, and then the basic, organizing, and overarching themes were extracted from them. The thematic network structure consists of 64 basic themes organized under seven main organizing themes that had been identified from the outset. The thematic network is presented in Figure 1. As can be seen, in constructing the thematic network, only the overarching themes and organizing themes were retained, which are arranged under the overarching themes.

Based on data collected from 400 respondents, the gender composition of the sample shows that a substantial majority of participants were male (80%, $n = 320$), while females constituted only 20% ($n = 80$). In terms of age distribution, the largest group fell within the 25–35 year range, comprising 201 individuals (50.6%) of all respondents. This was followed by the 36–45 year group with 91 individuals (24%), the 46–50 year group with 55 individuals (13.8%), and finally those aged 50 and over with 46 individuals

(11.6%). This distribution indicates that most participants were young and middle-aged employees. From the perspective of educational attainment, bachelor's degrees accounted for the highest share with 310 individuals (77.5%), while 90 individuals (22.5%) held a master's degree or higher. This suggests that the respondents' educational level was primarily concentrated at the bachelor's level, although a considerable proportion had postgraduate education.

To determine the validity of the variables in this section, confirmatory factor analysis (CFA) was used. The AMOS output indicates that all factor loadings exceed 0.60. According to the AMOS output, the calculated χ^2/df is 1.52; a χ^2/df less than 5 indicates an acceptable model fit. The root mean square error of approximation (RMSEA) should be less than 0.10, and in the presented model this value equals 0.069. The indices GFI, AGFI, CFI, and NFI should also be greater than 0.90, and in the model under review they are each above the specified thresholds. Therefore, the data of this study exhibit an acceptable fit with the factorial structure of this scale, indicating alignment of the items with the variables in this section.

Table 2

Fit Indices for the Variables "Strengths and Weaknesses of Systems"

Category	Index Name	Abbreviation	Acceptable Fit	Study Value
Absolute Fit	Degrees of Freedom	DF	–	398
	Significance Level	P	< .05	.000
	Chi-Square/df Ratio	CMIN/DF	1–5	1.52
	Chi-Square Coverage Level	Chi-Square	> 5%	.66
	Comparative Fit Index	CFI	> .90	.911
Comparative Fit	Adjusted Goodness of Fit Index	AGFI	> .90	.925
	Non-Normed Fit Index	NNFI	> .90	.921
	Normed Fit Index	NFI	Close to 1	.969

Parsimonious Fit	Comparative Fit Index	CFI	> .90	.911
	Relative Fit Index	RFI	> .50	.64
	Incremental Fit Index	IFI	0–1	.71
	Parsimonious Normed Fit Index	PNFI	> .50	.66
	Parsimonious Goodness of Fit Index	PGFI	> .50	.852
	Root Mean Square Error Approximation	RMSEA	< .10	.069
	Normalized Chi-Square	CMIN	1–3	2.3

To determine the validity of this section, CFA was employed. All factor loadings exceed 0.60. As shown in the AMOS output, the calculated χ^2/df is 1.59; a value less than 5 indicates acceptable fit. The RMSEA should be less than 0.10, and in the presented model it equals 0.041. The indices GFI, AGFI, CFI, and NFI should be greater than 0.90, and

in the model under review they each surpass the specified thresholds. Therefore, the data of this study exhibit an acceptable fit with the factorial structure of this scale, indicating alignment of the items with the variables in this section.

Table 3

Fit Indices for the Variables “Model Dimensions”

Category	Index Name	Abbreviation	Acceptable Fit	Study Value
Absolute Fit	Degrees of Freedom	DF	–	398
	Significance Level	P	< .05	.000
	Chi-Square/df Ratio	CMIN/DF	1–5	1.59
	Chi-Square Coverage Level	Chi-Square	> 5%	.66
	Comparative Fit Index	CFI	> .90	.937
Comparative Fit	Adjusted Goodness of Fit Index	AGFI	> .90	.958
	Non-Normed Fit Index	NNFI	> .90	.911
	Normed Fit Index	NFI	Close to 1	.936
	Comparative Fit Index	CFI	> .90	.912
	Relative Fit Index	RFI	> .50	.68
Parsimonious Fit	Incremental Fit Index	IFI	0–1	.89
	Parsimonious Normed Fit Index	PNFI	> .50	.901
	Parsimonious Goodness of Fit Index	PGFI	> .50	.922
	Root Mean Square Error Approximation	RMSEA	< .10	.041
	Normalized Chi-Square	CMIN	1–3	1.59

The numbers on the paths represent factor loadings, and for the drivers they all exceed 0.60. The calculated χ^2/df is 2.69; a value less than 5 indicates acceptable fit. The RMSEA should be less than 0.10, and in the presented model it equals 0.021. The indices GFI, AGFI, CFI, and NFI should

be greater than 0.90, and in the model under review they each surpass the specified thresholds. Therefore, the data of this study exhibit an acceptable fit with the factorial structure of this scale.

Table 4

Fit Indices for the Variables “Drivers”

Category	Index Name	Abbreviation	Acceptable Fit	Study Value
Absolute Fit	Degrees of Freedom	DF	–	398
	Significance Level	P	< .05	.000
	Chi-Square/df Ratio	CMIN/DF	1–5	2.69
	Chi-Square Coverage Level	Chi-Square	> 5%	.23
	Comparative Fit Index	CFI	> .90	.925
Comparative Fit	Adjusted Goodness of Fit Index	AGFI	> .90	.914
	Non-Normed Fit Index	NNFI	> .90	.905
	Normed Fit Index	NFI	Close to 1	.968
	Comparative Fit Index	CFI	> .90	.950
	Relative Fit Index	RFI	> .50	.74

Parsimonious Fit	Incremental Fit Index	IFI	0–1	.65
	Parsimonious Normed Fit Index	PNFI	> .50	.78
	Parsimonious Goodness of Fit Index	PGFI	> .50	.88
	Root Mean Square Error Approximation	RMSEA	< .10	.021
	Normalized Chi-Square	CMIN	1–3	2.2

To determine the validity of the variables in this section, CFA was used. The numbers on the paths are factor loadings, and all factor loadings exceed 0.60. The findings in Table 5 show that CFI, GFI, NFI, RMR, and RMSEA are all within

acceptable ranges. These goodness-of-fit characteristics indicate that the data of this study align well with the factorial structure of this scale.

Table 5

Fit Indices for the Variables “Strategies and Recommendations”

Category	Index Name	Abbreviation	Acceptable Fit	Study Value
Absolute Fit	Degrees of Freedom	DF	–	398
	Significance Level	P	< .05	.000
	Chi-Square/df Ratio	CMIN/DF	1–5	1.99
	Chi-Square Coverage Level	Chi-Square	> 5%	.59
	Comparative Fit Index	CFI	> .90	.932
Comparative Fit	Adjusted Goodness of Fit Index	AGFI	> .90	.914
	Non-Normed Fit Index	NNFI	> .90	.960
	Normed Fit Index	NFI	Close to 1	.913
	Comparative Fit Index	CFI	> .90	.920
	Relative Fit Index	RFI	> .50	.87
Parsimonious Fit	Incremental Fit Index	IFI	0–1	.59
	Parsimonious Normed Fit Index	PNFI	> .50	.931
	Parsimonious Goodness of Fit Index	PGFI	> .50	.924
	Root Mean Square Error Approximation	RMSEA	< .10	.028
	Normalized Chi-Square	CMIN	1–3	2.12

To determine the validity of the variables in this section, confirmatory factor analysis (CFA) was employed. All factor loadings exceeded 0.60. The fifth-level goodness-of-fit indices reported in Table 6 indicate that CFI, GFI, NFI, RMR, and RMSEA were within acceptable ranges, and these

fit characteristics show that the data of this study exhibit an appropriate fit with the factorial structure of this scale, indicating alignment of the items with the latent constructs in this section.

Table 6

Fit Indices for the Variables “Outcomes”

Category	Index Name	Abbreviation	Acceptable Fit	Study Value
Absolute Fit	Degrees of Freedom	DF	–	398
	Significance Level	P	< .05	.000
	Chi-Square/df Ratio	CMIN/DF	1–5	2.26
	Chi-Square Coverage Level	Chi-Square	> 5%	.21
	Comparative Fit Index	CFI	> .90	.918
Comparative Fit	Adjusted Goodness of Fit Index	AGFI	> .90	.914
	Non-Normed Fit Index	NNFI	> .90	.933
	Normed Fit Index	NFI	Close to 1	.945
	Comparative Fit Index	CFI	> .90	.921
	Relative Fit Index	RFI	> .50	.65
Parsimonious Fit	Incremental Fit Index	IFI	0–1	.74
	Parsimonious Normed Fit Index	PNFI	> .50	.69
	Parsimonious Goodness of Fit Index	PGFI	> .50	.88
	Root Mean Square Error of Approximation	RMSEA	< .10	.066
	Normalized Chi-Square	CMIN	1–3	1.15

To determine the validity of the social factors, confirmatory factor analysis was used. All factor loadings exceeded 0.60. The sixth-level goodness-of-fit indices reported in Table 7 show that CFI, GFI, NFI, RMR, and RMSEA were within acceptable ranges. These fit

characteristics indicate that the data of this study have an appropriate fit with the factorial structure of this scale, demonstrating alignment of the items with the constructs representing the current situation.

Table 7

Fit Indices for the Variables "Current Situation"

Category	Index Name	Abbreviation	Acceptable Fit	Study Value
Absolute Fit	Degrees of Freedom	DF	–	398
	Significance Level	P	< .05	.000
	Chi-Square/df Ratio	CMIN/DF	1–5	2.36
	Chi-Square Coverage Level	Chi-Square	> 5%	.39
	Comparative Fit Index	CFI	> .90	.911
Comparative Fit	Adjusted Goodness of Fit Index	AGFI	> .90	.936
	Non-Normed Fit Index	NNFI	> .90	.945
	Normed Fit Index	NFI	Close to 1	.940
	Comparative Fit Index	CFI	> .90	.911
	Relative Fit Index	RFI	> .50	.77
Parsimonious Fit	Incremental Fit Index	IFI	0–1	.48
	Parsimonious Normed Fit Index	PNFI	> .50	.85
	Parsimonious Goodness of Fit Index	PGFI	> .50	.88
	Root Mean Square Error of Approximation	RMSEA	< .10	.069
	Normalized Chi-Square	CMIN	1–3	2.11

To determine the validity of the variables in this section, confirmatory factor analysis was employed. All factor loadings exceeded 0.60. The sixth-level goodness-of-fit indices reported in Table below indicate that CFI, GFI, NFI,

RMR, and RMSEA were within acceptable ranges, and these characteristics show that the data of this study are suitably fitted to the factorial structure of this scale, indicating alignment of the items with the constructs in this section.

Table 8

Fit Indices for the Variables "Desired Situation"

Category	Index Name	Abbreviation	Acceptable Fit	Study Value
Absolute Fit	Degrees of Freedom	DF	–	398
	Significance Level	P	< .05	.000
	Chi-Square/df Ratio	CMIN/DF	1–5	3.14
	Chi-Square Coverage Level	Chi-Square	> 5%	.29
	Comparative Fit Index	CFI	> .90	.910
Comparative Fit	Adjusted Goodness of Fit Index	AGFI	> .90	.928
	Non-Normed Fit Index	NNFI	> .90	.922
	Normed Fit Index	NFI	Close to 1	.964
	Comparative Fit Index	CFI	> .90	.910
	Relative Fit Index	RFI	> .50	.61
Parsimonious Fit	Incremental Fit Index	IFI	0–1	.54
	Parsimonious Normed Fit Index	PNFI	> .50	.79
	Parsimonious Goodness of Fit Index	PGFI	> .50	.83
	Root Mean Square Error of Approximation	RMSEA	< .10	.048
	Normalized Chi-Square	CMIN	1–3	1.69

With respect to the magnitude of the significance coefficients, for rejecting or confirming the relationships, the CR (critical ratio) must be greater than 1.96 or less than

–1.96. Parameter estimates that fall between these two thresholds are not considered significant. Values within this

range indicate that the computed regression weights do not differ significantly from zero at the 95% confidence level.

Accordingly, the research model was finally assessed using AMOS software, and as can be seen, all relationships—given the magnitude of the path coefficients—are confirmed at the 95% confidence level.

The model related to “applying the balanced scorecard framework as a tool for performance measurement in governmental organizations” is presented in the table above. Based on the obtained results, the study components exerted pairwise effects within the final research model.

Figure 1

Results of the simulation of four furnaces lined with local refractory bricks based on the defined input parameters.

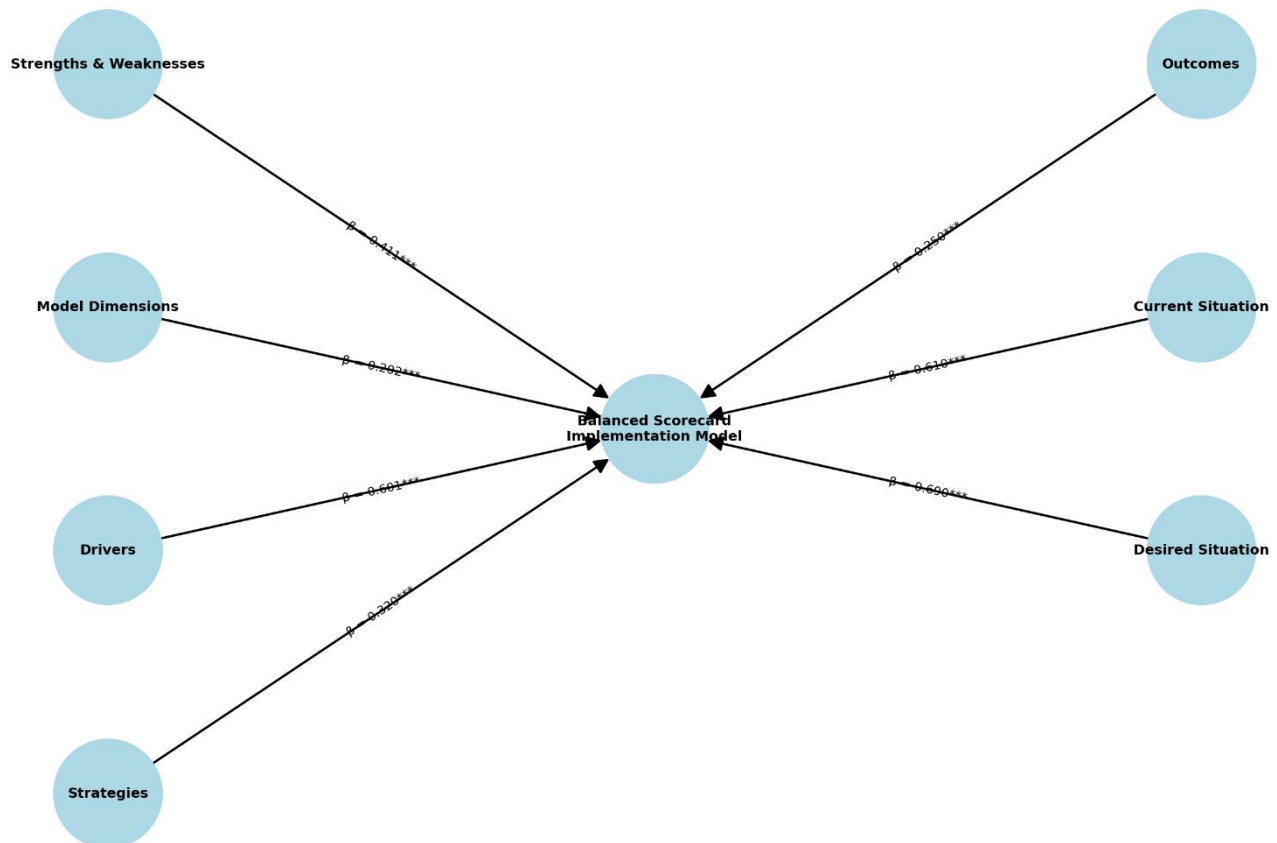


Table 9

Results of Implementing the Model of Applying the Balanced Scorecard Framework as a Tool for Performance Measurement in Governmental Organizations

Relationship	Standardized Estimate	Standard Error	Critical Ratio	Significance Level
Strengths and Weaknesses of Systems → Model of Applying the Balanced Scorecard Framework	.411	.024	3.14	.000
Model Dimensions → Model of Applying the Balanced Scorecard Framework	.202	.011	3.71	.000
Drivers → Model of Applying the Balanced Scorecard Framework	.601	.041	4.18	.000
Strategies → Model of Applying the Balanced Scorecard Framework	.320	.039	2.75	.000
Outcomes → Model of Applying the Balanced Scorecard Framework	.250	.025	3.61	.000
Current Situation → Model of Applying the Balanced Scorecard Framework	.610	.012	4.24	.000
Desired Situation → Model of Applying the Balanced Scorecard Framework	.690	.034	3.28	.000

4. Discussion and Conclusion

The findings of this study highlight the critical role of the Balanced Scorecard (BSC) as an effective framework for evaluating and enhancing organizational performance in governmental institutions. By identifying seven central dimensions—strengths and weaknesses of systems, model dimensions, drivers, strategies, outcomes, current situation, and desired situation—the research confirmed that all these factors significantly contribute to the adoption and institutionalization of the BSC framework. The statistical validation through confirmatory factor analysis (CFA) and the final structural model provided robust evidence that the adapted BSC can function as a comprehensive performance evaluation tool that reflects the unique institutional and socio-cultural requirements of public organizations in Iran. These results resonate with the growing body of literature that emphasizes the multidimensionality of performance management frameworks and their role in bridging strategy with execution (Kaplan & Norton, 2021; Niven, 2020).

One of the key findings of this study was the significance of system strengths and weaknesses as predictors of BSC implementation. The results demonstrated that clear structures, transparent mechanisms, and effective feedback processes act as enablers, while overreliance on quantitative indicators and resistance to evaluation constitute major barriers. These findings align with the work of Ghasemi Esfahlan (Ghasemi Esfahlan & Khabaz Bavyl, 2020), who argued that performance transparency in public institutions requires not only robust frameworks but also cultural readiness. Similarly, Mansouri (Mansouri, 2020) identified that weaknesses in organizational performance management systems often stem from inadequate integration of qualitative measures and misalignment between strategic goals and operational indicators. Together, these studies emphasize that recognizing systemic strengths and weaknesses is foundational for designing a functional BSC tailored to the public sector.

Another important contribution of the results is the validation of the “model dimensions” construct, which reflects the classical four perspectives of the BSC—financial, customer, internal processes, and learning and growth—as well as extended perspectives such as social and environmental responsibilities. This aligns with the argument made by Nikbakht (Nikbakht & Rahimipour, 2022), who advocated for a sustainable BSC that incorporates social and ecological measures alongside traditional financial and operational indicators, particularly

in financial institutions. Similarly, Song (Song, 2022) emphasized that listed companies increasingly require performance frameworks that reflect broader stakeholder concerns beyond profitability. The findings of the present study confirm that in governmental organizations, inclusion of social and environmental dimensions is not optional but rather an essential element for accountability and long-term legitimacy.

The results also highlighted the strong impact of drivers such as accountability, transparency, and optimal resource allocation on successful BSC implementation. This finding corroborates the conclusions of Yeshaw (Yeshaw et al., 2025), who reported that accountability and transparency are decisive enablers of BSC adoption in public hospitals in Ethiopia. Similarly, Sharaf-Addin (Sharaf-Addin & Fazel, 2020) observed that in Saudi public universities, the push for accountability and service quality drove the adaptation of the BSC as a performance management system. The current study adds to this body of evidence by demonstrating that similar drivers are equally critical in Iranian governmental institutions, confirming the universality of accountability and transparency as prerequisites for performance evaluation frameworks across diverse contexts.

The strategies validated in this study—including the definition of measurable indicators, training and capacity-building, IT utilization, and stakeholder participation—were also found to significantly predict BSC adoption. These findings support the insights of Aranda (Aranda & Odriozola, 2021), who highlighted leadership and strategic alignment as key factors in successful BSC implementation in small- and medium-sized enterprises. In a governmental context, however, the emphasis on stakeholder participation underscores the political and social nature of public administration, where multiple constituencies influence policy-making and service delivery. As Cignitas (Cignitas et al., 2022) noted, integrating positive management principles into the BSC fosters employee engagement and stakeholder trust, thereby ensuring the long-term sustainability of the system.

The outcomes dimension of the model, confirmed in this study, demonstrated that implementing the BSC improves transparency, accountability, decision-making, and citizen satisfaction. These findings echo the conclusions of Mohammadi (Mohammadi et al., 2024), who developed performance evaluation indicators for the Iran Health Insurance Organization and found that such frameworks enhance accountability and service quality. Similarly, Ghanbari (Ghanbari et al., 2020) confirmed that adopting

performance evaluation models with strong financial and operational perspectives significantly improved transparency and planning in vocational training organizations. These converging results indicate that well-designed performance frameworks yield positive outcomes across different governmental sectors.

The results further validated the importance of examining the current and desired states of organizational systems. Identifying weaknesses in current systems—such as superficial evaluations, inadequate use of results in decision-making, and misalignment of indicators with strategy—provides a baseline for improvement. This observation is consistent with the analysis of Aliabadi (Aliabadi et al., 2019), who showed that prioritizing evaluation indicators for senior managers in cultural organizations enables targeted interventions. The desired situation emphasized in this study—comprehensive evaluation systems, integration of qualitative and quantitative indicators, and alignment with strategic goals—corresponds with the perspectives of Kermshahi (Kermshahi & Salehi Tabandeh, 2024), who argued that effective BSC frameworks in auditing firms require precisely defined and measurable strategic indicators.

The findings also resonate with international studies. In China, Bai (Bai et al., 2020) demonstrated the use of performance evaluation models to measure innovation outcomes in industry–university collaboration, emphasizing the necessity of multi-dimensional measures. Similarly, Lee (Lee et al., 2023) found that BSC adoption in healthcare firms enhanced both financial and operational performance, supporting the adaptability of the framework across industries. Martínez-Caro (Martínez-Caro et al., 2015) also provided evidence from higher education, showing that performance evaluation models improve the quality of e-learning systems, underscoring the applicability of BSC principles in knowledge-based organizations.

Another dimension worth emphasizing is the methodological robustness of the study, as it employed both qualitative and quantitative methods. This is in line with the recommendations of Armstrong (Armstrong, 2020), who argued for integrating learning and development insights into organizational evaluation systems. It also parallels the work of Zimmermann (Zimmermann, 2017) and Morabito (Morabito, 2016), who emphasized the importance of rigorous performance evaluation methods, particularly in complex technological environments. The incorporation of confirmatory factor analysis and structural modeling further strengthens the generalizability and credibility of the

findings, aligning with the statistical approaches advocated in previous BSC validation studies (Samei et al., 2019; Sindhu et al., 2019).

Taken together, the findings of this study reinforce the theoretical and practical relevance of the BSC as a comprehensive performance evaluation framework that enhances decision-making, accountability, and transparency in public institutions. By situating these results in alignment with both local and international research, this study contributes to the ongoing scholarly dialogue on how performance evaluation frameworks can be effectively adapted to diverse contexts. The evidence demonstrates that although challenges persist, the BSC continues to evolve as a dynamic tool capable of integrating financial, social, and environmental dimensions of organizational success (Cignitas et al., 2022; Song, 2022).

This study is not without limitations. First, the sample was drawn primarily from governmental organizations within a specific national context, which may restrict the generalizability of the findings to other countries or organizational types. Second, although the mixed-methods design enhanced validity, reliance on self-reported data may have introduced social desirability bias, particularly in sensitive areas such as accountability and transparency. Third, while the confirmatory factor analysis and structural modeling provided strong evidence for the proposed framework, cross-sectional data limit the ability to capture dynamic changes over time. Finally, the study did not explore in depth the cultural and political dynamics that may moderate the relationship between BSC implementation and organizational outcomes.

Future studies could build on these findings by conducting longitudinal research to examine how BSC implementation influences organizational performance over time. Comparative studies across different countries and cultural settings would also be valuable for identifying the contextual factors that facilitate or hinder successful adoption. Researchers could also expand the model by incorporating digital transformation and artificial intelligence tools, which increasingly shape performance evaluation in modern organizations. Furthermore, experimental or quasi-experimental designs could be employed to establish causal relationships between BSC adoption and performance outcomes.

From a practical standpoint, the findings suggest that policymakers and managers in governmental organizations should prioritize the development of comprehensive performance evaluation systems that balance quantitative

and qualitative indicators. Training and capacity-building programs are essential to overcome resistance and build managerial competencies for BSC adoption. Integrating social and environmental perspectives into performance frameworks is also critical to enhance accountability and public trust. Finally, managers should ensure that evaluation results are systematically incorporated into decision-making processes, thereby reinforcing the strategic alignment and sustainability of organizational initiatives.

Authors' Contributions

Authors contributed equally to this article.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethics Considerations

In this research, ethical standards including obtaining informed consent, ensuring privacy and confidentiality were considered.

References

- Aliabadi, M., Bagheri Fard, M. H., & Goodarzi, G. (2019). Prioritizing performance evaluation indicators for senior managers of cultural organizations. *Strategic Management Thought (Andishe Modiriyat)*, 13(2 (No. 26)), 303-326. https://smt.isu.ac.ir/article_2742.html
- Aranda, A., & Odriozola, M. (2021). Implementing the Balanced Scorecard in SMEs: The Role of Leadership. *Journal of Small Business Management*. [https://rep-](https://rep-dspace.uminho.pt/bitstreams/51935012-70f2-4dca-9584-c76ceb75da41/download)
- [dspace.uminho.pt/bitstreams/51935012-70f2-4dca-9584-c76ceb75da41/download](https://rep-dspace.uminho.pt/bitstreams/51935012-70f2-4dca-9584-c76ceb75da41/download)
- Armstrong, M. (2020). *Armstrong's Handbook of Learning and Development*. Kogan Page. <https://www.amazon.com/Armstrongs-Handbook-Learning-Development-Practice/dp/139860190X>
- Bai, X. J., Li, Z. Y., & Zeng, J. (2020). Performance evaluation of China's innovation during the industry-university-research collaboration process-an analysis basis on the dynamic network slacks-based measurement model. *Technology in Society*, 62, 101310. <https://doi.org/10.1016/j.techsoc.2020.101310>
- Chowdhury, F., Zhu, Y., Heer, T., Paredes, S., Moody, A., Goldstone, R., & Yu, W. (2019). I/o characterization and performance evaluation of beegfs for deep learning. Proceedings of the 48th International Conference on Parallel Processing,
- Cignitas, C. P., Torrents Arevalo, J. A., & Crusells, J. V. (2022). Positive Management and the Balanced Scorecard: A Successful Strategy for Organizations. *Journal of Positive School Psychology*, 6(3), 2606-2627. <https://www.journalppw.com/index.php/jpsp/article/view/2021>
- Ghanbari, H., Ebrahimi Saravolia, M. H., Amiri, M., Bolou, G., & Ghorbani Zadeh, V. (2020). Designing a performance evaluation model for the Technical and Vocational Training Organization of Iran with an emphasis on the financial approach. *Industrial Management Outlook*, 10(4 (No. 40)), 9-40. https://jimp.sbu.ac.ir/article_87556.html
- Ghasemi Esfahlan, L., & Khabaz Bavy, S. (2020). A model for performance transparency in governmental organizations. *Public Organizations Management*, 9(1 (No. 33)), 107-122. https://ipom.journals.pnu.ac.ir/article_7336.html
- Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. *Education and Information Technologies*, 28(1), 1081-1116. <https://doi.org/10.1007/s10639-022-11221-2>
- Iranmehr, K., & Piriaei, Z. (2024). Designing a performance evaluation model for commanders of major NAJA units based on Jihadi management with an interpretive structural approach. *Military Science and Techniques*, 20(67), 271-293. <https://doi.org/10.22034/qjmst.2024.2010564.1940>
- Kaplan, R. S., & Norton, D. P. (2021). *The Balanced Scorecard: Strategy 2.0*. Harvard Business Review Press. <https://leadernetworks.com/wp-content/uploads/2014/06/BSC-1.pdf>
- Kermshahi, B., & Salehi Tabandeh, A. Z. (2024). Investigating the dimensions of an effective balanced scorecard approach on performance evaluation of auditing firms: Using confirmatory factor analysis. *Auditing Knowledge*, 23(93), 180-204. <https://danesh.dmk.ir/article-1-3056-fa.html>
- Lee, B., Tsui, A. S. C., & Yau, O. H. M. (2023). Impact of Balanced Scorecard implementation on company performance of PRC listed companies in the healthcare industry. *Journal of Transnational Management*, 35-73. <https://doi.org/10.1080/15475778.2023.2191379>
- Mansouri, H. (2020). Identifying factors and components affecting the performance management system of governmental organizations. *Productivity Management (Beyond Management)*, 14(53), 97-118. <https://sid.ir/paper/400793/fa>
- Martínez-Caro, E., Cegarra-Navarro, J. G., & Cepeda-Carrión, G. (2015). An application of the performance-evaluation model for e-learning quality in higher education. *Total Quality Management & Business Excellence*, 26(5-6), 632-647. <https://doi.org/10.1080/14783363.2013.867607>

- Mohammadi, E., Bakhtiari Ahad, N. M. M., Efatpanah, M., Rezaei, M., & Shahali, Z. (2024). Performance evaluation indicators of Iran Health Insurance Organization based on upstream documents. *Iran Health Insurance*, 7(1), 22-41. <http://journal.ihio.gov.ir/article-1-303-fa.html>
- Morabito, R. (2016). A performance evaluation of container technologies on internet of things devices. 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS),
- Nikbakht, M. R., & Rahimpour, A. (2022). Organizational performance evaluation using the sustainable balanced scorecard model (Case study: Bank Sarmayeh). *Accounting and Management Auditing Knowledge*, 11(41), 63-79. https://www.jmaak.ir/article_19152_8d42f01f968ca40050a5b20fc100cd8a.pdf
- Niven, P. R. (2020). The Balanced Scorecard: Step-by-Step for Government and Nonprofit Agencies. <https://books.google.com/books?hl=fa&lr=&id=7kuoDwAAQBAJ&oi=fnd&pg=PR9&dq=The+Balanced+Scorecard:+Step-by-Step+for+Government+and+Nonprofit+Agencies&ots=kuVTLILG1F&sig=Bf1uoamoY6vfwBZ1MFj1Yno4yWo>
- Samei, E., Bakalyar, D., Boedeker, K. L., Brady, S., Fan, J., Leng, S., & Wang, J. (2019). Performance evaluation of computed tomography systems: summary of AAPM Task Group 233. *Medical Physics*, 46(11), e735-e756. <https://doi.org/10.1002/mp.13763>
- Shakir, Z., Al-Thaadan, A., Alsabah, R., Al-Sabbagh, A., Salah, M. E. M., & Zec, J. (2022). Performance evaluation for RF propagation models based on data measurement for LTE networks. *International Journal of Information Technology*, 14(5), 2423-2428. <https://doi.org/10.1007/s41870-022-01006-8>
- Sharaf-Addin, H. H., & Fazel, H. (2020). Balanced Scorecard Development as a Performance Management System in Saudi Public Universities: A Case Study Approach. *Asia-Pacific Journal of Management Research and Innovation*, 17(1-2), 57-70. <https://doi.org/10.1177/2319510X211048591>
- Shen, Y., Wu, X., Tao, J., Zhu, C., Lai, Y., & Chen, Z. (2019). Icephobic materials: Fundamentals, performance evaluation, and applications. *Progress in Materials Science*, 103, 509-557. <https://doi.org/10.1016/j.pmatsci.2019.03.004>
- Sindhu, I., Daudpota, S. M., Badar, K., Bakhtyar, M., Baber, J., & Nurunnabi, M. (2019). Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. *IEEE Access*, 7, 108729-108741. <https://doi.org/10.1109/ACCESS.2019.2928872>
- Song, X. (2022). Application of Balanced Scorecard in Performance Management and Evaluation of Listed Companies. *Hindawi*, 1-11. <https://doi.org/10.1155/2022/2247890>
- Wang, Y., Wang, C., Li, M., Yu, Y., & Zhang, B. (2021). Nitrate electroreduction: mechanism insight, in situ characterization, performance evaluation, and challenges. *Chemical Society Reviews*, 50(12), 6720-6733. <https://doi.org/10.1039/D1CS00116G>
- Yeshaw, S., Asrade, G., Hagos, A., Chanie, M. G., & Worku, N. (2025). Practice of balanced scorecard implementation and its contributing factors among public primary hospital professionals in Central Gondar zone, Northwest Ethiopia. *Frontiers in Public Health*, 12, 1424133. <https://doi.org/10.3389/fpubh.2024.1424133>
- Zimmermann, A. (2017). Modelling and performance evaluation with timenet 4.4. Quantitative Evaluation of Systems: 14th International Conference, QEST 2017, Berlin, Germany, September 5-7, 2017, Proceedings 14,