# Evaluating Public Policy in Practice: A Critical Systematic Literature Review of Methods, Trends, and Gaps, from 2015 to 2023

**Author:**
**Rizki Ananda[1], Umanto[2]**

**Affiliation:**

Universitas Indonesia, Jl. Lingkar, Pondok Cina, Kecamatan Beji, Depok, 16424, Indonesia[1,2]

**e-Mail:**

rizki.ananda20@gmail.com[1], umanto.eko@gmail.com[2]

**\*Corresponding author**
Rizki Ananda
Universitas Indonesia, Jakarta, Indonesia
Email: rizki.ananda20@gmail.com

***ABSTRACT***

*This study conducts a systematic literature review of recent developments in policy and program evaluation from 2015 to 2023, examining methods, approaches, challenges, and future research directions. Following PRISMA guidelines, a search of the Scopus and Science Direct databases yielded 58 relevant articles. Key findings indicate a shift towards mixed-methods approaches, agent-based simulation, and stated preference techniques, although studies predominantly originate from developed countries and focus on social, environmental, and health issues. Bibliometric analysis shows that 62.9% of the articles appeared in Scopus Q1 journals. Evaluation effectiveness is significantly influenced by internal factors (e.g., program design, data quality) and external factors (e.g., political context, financial support). While evaluations impact decision-making and policy improvement, challenges remain in utilizing and generalizing results. Crucially, this review identifies gaps, including methodological limitations, a scarcity of studies on vulnerable groups, and a lack of standardized indicators. This study contributes to the advancement of policy evaluation methodologies by, for example, highlighting the need for strategic integration of qualitative and quantitative data within mixed-methods designs, going beyond mere combination to achieve nuanced understanding. It also emphasizes the critical importance of adapting systems thinking approaches, using for example, process tracing and agent-based modelling to specific developing country contexts, providing a pathway for more relevant and impactful evaluations. This research underscores the need for participatory, adaptive, and evidence-focused evaluation approaches that explicitly consider causal mechanisms, and recommends a concerted effort to develop comprehensive mixed-methods research, prioritize studies in developing countries, and build capacity for implementing these advanced methodologies.*

***Keywords:*** *Policy Evaluation Methods; PRISMA; Public Policy Evaluation; Systematic Literature Review*

## INTRODUCTION

Public policy is a vital instrument for governments to achieve developmental goals and address various social challenges. The effectiveness of public policy is highly dependent on careful and systematic evaluation to measure the impact of government interventions (Fischer et al., 2007; Goodin et al., 2006; Vedung, 1997). Policy evaluation, therefore, is not merely a final stage but also an ongoing process crucial for ensuring accountability, efficiency, and effectiveness of government programs (Dunn, 2018; Howlett, 2022; Varone et al., 2023; Freidson & Rist, 2020).

Although the field of public policy evaluation has experienced significant advancements, recent research indicates persistent methodological and contextual gaps that need addressing. While numerous studies have explored various aspects of policy evaluation, there remains a need for a more systematic and comprehensive approach that maps the evolving landscape of this field, particularly in terms of methodological trends, contextual gaps, and practical challenges. As noted by Mavrot et al. (2024), there is often a lack of in-depth reflection on how evaluation criteria are defined, chosen, and applied in practice, which can limit the validity and usefulness of evaluation findings. Most studies still focus on developed country contexts and underemphasize the unique dynamics in developing nations, including Indonesia (Freidson & Rist, 2020; Johnson & Hennessy, 2019). Furthermore, comprehensive and integrative evaluation approaches, capable of accommodating the complexities of policy in diverse socio-economic and political contexts, are still relatively uncommon (Dunn, 2018; Okamuro & Nishimura, 2021). This is a critical limitation, as highlighted by Renyaan (2023), who emphasizes the need for policy evaluation to address the dynamic

nature of policy issues and the challenges of uncertainty, causality, and data limitations. Publication trends also reveal that despite the ongoing development of various evaluation methods and approaches, the adoption and adaptation of these methods across diverse contexts, as well as the utilization of evaluation findings, still require further attention (Fynn et al., 2022; Hsieh, 2020; McGough et al., 2018; Rethlefsen et al., 2021; Dongying Sun et al., 2023b). Existing reviews often lack the specificity and depth needed to address the unique challenges faced by policymakers and researchers, especially in rapidly evolving and diverse contexts. Furthermore, as Mergoni & De Witte (2021) point out, there is a notable gap in the literature regarding the integration of efficiency and effectiveness perspectives in policy evaluation, despite the importance of both for understanding the overall impact of public interventions.

The scientific novelty of this research lies in its systematic and comprehensive approach to mapping the landscape of public policy evaluation literature, with a specific focus on methodological trends, contextual gaps, and implementation challenges. Unlike previous reviews that primarily catalogue methods or focus on specific sectors, this study provides a holistic overview of the field, identifying not only what methods are used but also where the gaps and limitations exist. This includes a critical examination of how evaluation methodologies are adapted and utilized across different geographical and socio-political contexts. Furthermore, this study emphasizes the need for more integrative approaches that combine various methods to address the complex and multidimensional nature of public policy interventions. By identifying research gaps that have not been thoroughly explored, this study contributes significantly to the

development of more adaptive and relevant theories and practices in public policy evaluation, particularly in developing country contexts. This study also provides new insights into the methods and approaches that need improvement in future evaluation efforts, while taking into consideration various contextual, social, and political factors.

Therefore, the research problem addressed in this study is how to comprehensively map and analyze publication trends, methodological approaches, and gaps in public policy evaluation literature. This study addresses this problem by systematically reviewing a wide range of recent publications, classifying them according to methodological approaches, geographical focus, and thematic areas. This approach allows for a detailed analysis of how different evaluation methods are applied, what types of policies are most frequently evaluated, and which regions or contexts are underrepresented in the current literature. This study also aims to formulate more specific and relevant research recommendations to address the various challenges in the implementation and utilization of policy evaluation. The objective of this systematic review is to provide a solid foundation for researchers and policymakers in developing more effective and impactful public policy evaluations, by considering the various contextual, methodological, and applicative factors in the future.

**METHODS**

This study employs the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure a systematic and transparent process for searching, screening, selecting, and synthesizing studies (Page et al., 2021). PRISMA, an international standard, enhances transparency and replicability in reporting

systematic literature reviews and meta-analyses (Moher et al., 2009). Its structured framework includes a 27-item checklist that covers all important aspects of the systematic review process, from research question formulation to results reporting (Page et al., 2021). Implementing this method provides a solid foundation for credible, transparent, and reproducible findings, especially in public policy evaluation (Johnson & Hennessy, 2019; Rethlefsen et al., 2021; Tam et al., 2019). A comprehensive literature search was conducted on August 3, 2024, using Scopus and Science Direct, which are reputable databases providing access to extensive collections of journal articles in social sciences, including public policy evaluation. The search strategy combined keywords such as "policy evaluation," "public policy evaluation," "policy assessment," "program evaluation," "policy analysis," and "evaluating policy performance" to broadly capture the terminology used in the literature. These keywords were chosen based on their relevance to the research focus, which is to analyze recent developments in policy and program evaluation literature.

The article selection process adhered to predefined inclusion and exclusion criteria. Inclusion criteria were: peer-reviewed articles, case studies, empirical studies, or comparative analyses published in academic journals in English between 2015 and 2024, with a specific focus on the evaluation of public policies or programs. Exclusion criteria were: opinion pieces, editorials, commentaries, book reviews, conference proceedings, non-peer-reviewed articles, studies not directly relevant to public policy evaluation or focused only on private sector organizations, and articles not accessible in full text. This ensured that the study concentrated on the most relevant and high-quality literature, reflecting recent trends.

Study quality assessment used three main criteria: journal reputation, methodological and analytical clarity, and significance of contributions. Journal reputation was assessed via Scopus Quartile rankings (Q1 being the highest). Methodological and analytical clarity were evaluated based on the transparency and detail in describing data collection and analysis. Significance of contributions was assessed by the originality of the research and the relevance of findings for theory and practice in public policy evaluation. This ensured that only high-quality studies with significant contributions were included.

The PRISMA diagram (Figure 1) illustrates the workflow. Initially, 7,462 articles were identified through database searches. After removing duplicates, 269 articles remained. The screening process involved two stages: title and abstract screening, and full-text screening. Uncertainties regarding inclusion were resolved through discussions among the research team. After full-text screening, 58 articles were assessed for eligibility, and 45 were excluded. Finally, 58 articles met the inclusion criteria and were included in the qualitative synthesis. To enhance the analysis, data extraction involved systematically gathering relevant information, such as bibliographic details, study characteristics, policy focus, evaluation approach, findings, and limitations. Thematic analysis was employed to identify recurring patterns, and comparative analysis explored similarities and differences across studies. Narrative synthesis coherently presented the findings, highlighting key trends, methodological approaches, contextual factors, and gaps.
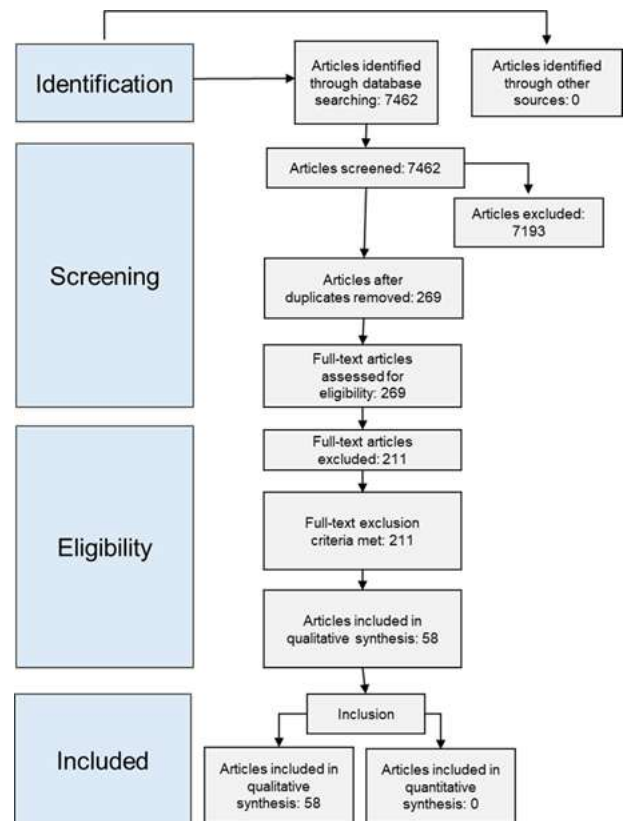


Source: Page et al., (2021)

**Figure 1.** PRISMA Diagram: A Systematic Literature Review of Policy and Program Evaluation Process

## RESULTS AND DISCUSSION
### a. Bibliographic Analysis

To provide a comprehensive overview of the research landscape relevant to public policy evaluation, a network visualization was created using the VOSviewer software. This visualization, depicted in Figure 2, is based on 269 articles collected from the Scopus and Science Direct databases and is designed to map the main themes and interconnections between concepts within the reviewed literature. This visualization not only provides a general overview but also highlights the dominant research focus in the field.
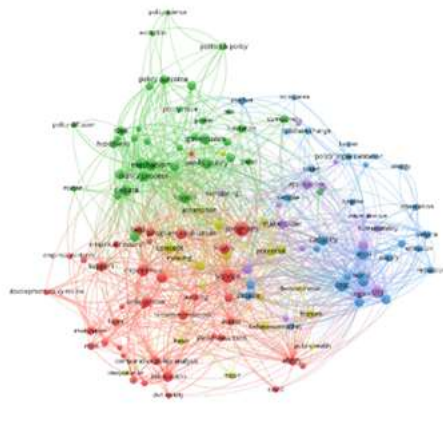
Source: processed data

**Figure 2.** Themes in Public Policy Evaluation

From the network visualization, five main clusters of keywords distinguished by color can be identified, each representing significant themes in public policy evaluation studies. The red cluster is centered on aspects of program evaluation, empirical research, experience, and support, indicating a significant focus on how the effectiveness of policies and programs is measured and influenced by empirical data and practitioners' experiences. The yellow cluster highlights the policy process, with a focus on policy formulation, mechanisms, debates, and lessons learned, emphasizing the importance of understanding policy formulation and the lessons derived from each stage of the process. The green cluster centers on policy issues, public policy, governance, power, and policy politics, indicating the significance of political and governance aspects in policy development and the influence of power structures. The blue cluster is associated with policy implementation, sustainability, and climate change, reflecting the increased attention to the practical implementation of policies and their environmental impact.

Finally, the purple cluster focuses on stakeholders, capacity, and potential, highlighting who is involved in policy formation and implementation, and how policies can create positive impacts and empower various actors. In addition to the thematic clusters, the visualization also reveals several prominent keywords in the literature, such as policy evaluation, program evaluation, policy process, public policy, implementation, and stakeholder. These keywords, which have larger circle sizes, indicate their high frequency across multiple papers.



Source: processed data

**Figure 3.** Number of Papers Published Over Years

This network visualization provides a useful visual guide to understanding the complexity of the public policy evaluation literature, and serves as a foundation for

subsequent sections of this paper. Through more in-depth analysis of selected papers (58 articles out of the initial 269), which represent the main themes from this visualization, we will explore the publication trends, journal distribution, and focus areas in more detail.

*Publication Trends in Public Policy and Program Evaluation Research*

Analysis of publication trends reveals fluctuations in the number of publications related to policy and program evaluation from 2015 to 2024 (Figure 3). Publications reached a peak in 2020 with 9 publications, followed by a decrease in 2021-2022, and an increase in 2023 with 7 publications. However, the number of publications decreased significantly in 2024 to only 3. These fluctuations reflect various factors, including developments and priorities in public policy, funding opportunities, and the level of interest among researchers regarding specific evaluation topics.

Geographic distribution of the studies within the systematic review are presented in Figure 4. Most studies come from various countries that were not specifically identified in the articles (33 studies). This could indicate a wide geographic spread, or could be due to a lack of specific location information in the articles. This is followed by studies from Europe (12 studies), with countries such as the United Kingdom, Germany, Italy, and the Netherlands. There are also studies from Asia (6 studies) including China, Japan, and India, and studies from North America (5 studies) from the United States and Canada. One study did not specify its location. This distribution indicates that public policy evaluation is a critical topic in various regions worldwide, across different economic, social, and political contexts.
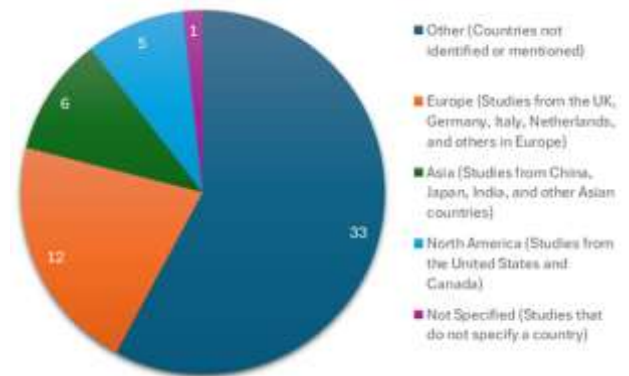


Source: processed data
**Figure 4.** Frequency Distribution of Articles

While studies from different countries offer significant contributions to the literature, there remains great potential for more in-depth and focused studies in developing country contexts, including Indonesia (Freidson & Rist, 2020). Given the complexity and unique challenges of policies in developing countries, such studies can offer novel and relevant insights for policymakers and practitioners in designing, implementing, and evaluating more effective and efficient public policy.

Moreover, the analysis of publication trends also highlights developments in methodological approaches and research focus. In recent years, there has been a rise in the use of agent-based modeling and simulation in evaluation studies, particularly in transport policy contexts (Ben-Dor et al., 2024). The stated preference (SP) method is also gaining popularity because of its ability to capture stakeholder preferences (Gatta & Marcucci, 2016). In education program evaluations, the use of randomized controlled trials (RCT) is increasing (Pham et al., 2024), demonstrating researchers' efforts to enhance the validity of findings using stronger quantitative approaches. Narrative approaches are also gaining attention McGough et al (2018) as a method to gain in-depth insights related to multi-stakeholder evaluation.

Publication trends also show a growing interest in cross-sectoral partnerships Fynn et al (2022) to support effective evaluation and public program implementation. These partnerships are seen as a way to create synergies and to ensure that evaluation results are relevant to key stakeholders. Furthermore, polycentric governance approaches are being applied to evaluation studies, which enables diverse actors to participate in decision-making and program implementation across different government levels (Omori & Tesorero, 2020). Public participation in policy evaluation, particularly on environmental issues, is also increasingly recognized as an important factor in gaining broader policy acceptance (Dongying Sun et al., 2023).

There is also a growing trend towards ex-post evaluations, suggesting the importance of evaluator independence (van Voorst & Mastenbroek, 2019) and the use of data triangulation in order to produce high-quality evaluations. Additionally, composite indices for measuring transparency and accountability in public policy are being increasingly developed (Michener, 2015), despite ongoing debate regarding their validity. Enhancing evaluation capacity also remains a priority for public sector reforms (Pattyn & Brans, 2015). Lastly, the analysis of publication trends also indicated an increase of evaluations of specific policy sectors, including innovation policy in China (Feng & Jiang, 2021), high-speed rail policy (Li et al., 2021), and environmental policy (Chuanwang Sun et al., 2023). These specific studies demonstrate the increasing relevance and recognition of public policy and program evaluation in various policy sectors as an essential tool for enhancing efficiency and accountability in various policy areas.

*Frequency Distribution of Articles by Journals, Year, and Publication Quality*

To provide a comprehensive overview of the distribution of public policy evaluation studies within academic literature, it is necessary to analyze the frequency of article publications based on journal, year, and publication quality. Table 1 presents a list of the journals publishing the articles reviewed in this systematic literature review, along with the number of articles (N) published, the year of publication, and the journal rankings according to Scopus Quartile, CiteScore, Clarivate Analytics Impact Factor, and SCImago Journal Rank (SJR). The Evaluation Journal emerged as the journal publishing the most articles related to public policy and program evaluation, with a total of 8 articles. Furthermore, several other journals demonstrate significant contributions to this field of research, including Policy Sciences (4 articles), Science and Public Policy (5 articles), Journal of Comparative Policy Analysis: Research and Practice (2 articles), Transport Research Part A: Policy and Practice (2 articles), and Transport Reviews (2 articles).

Table 1 illustrates that public policy evaluation studies are spread across various journals that focus on diverse disciplines, including social sciences, political science, public administration, and environmental science.

**Table 1.** Frequency Distribution of Articles by journals, Year and Publication Quality

| No. | Journal | N | Year | Scopus Quartile 2023 (Category) | Scopus CiteScore (2023) | Clarivate Analytic Impact Factor (2023) | SCImago Rank (2023) |
|---|---|---|---|---|---|---|---|
| 1 | Computers in Human Behavior | 1 | 2024 | Q1 (Arts and Humanities) | 19.1 | 9.0 | 2.64 |
| 2 | Development Policy Review | 1 | 2022 | Q2 (Management, Monitoring, Policy and Law) | 3.5 | 2.0 | 0.652 |
| 3 | Education Policy Analysis Archives | 1 | 2018 | Q3 (Education) | 1.2 | 0.6 | 0.313 |
| 4 | Environmental Impact Assessment Review | 1 | 2023 | Q1 (Management, Monitoring, Policy and Law) | 12.6 | 9.8 | 1.96 |
| 5 | Environmental Science and Policy | 1 | 2017 | Q1 (Management, Monitoring, Policy and Law) | 10.9 | 4.9 | 1.6 |
| 6 | Evaluation Journal | 8 | 2015, 2016, 2017, 2020, 2021, 2022 | Q1 (Sociology and Political Science) | 3.6 | 2.4 | 0.658 |
| 7 | Food Policy | 2 | 2016, 2020 | Q1 (Sociology and Political Science) | 11.4 | 6.8 | 2.12 |
| 8 | Government Information Quarterly | 1 | 2020 | Q1 (Sociology and Political Science) | 15.7 | 7.8 | 2.17 |
| 9 | Information Processing and Management | 1 | 2023 | Q1 (Management Science and Operations Research) | 17.0 | 7.4 | 2.13 |
| 10 | International Journal of Drug Policy | 1 | 2016 | Q1 (Health Policy) | 7.8 | 4.4 | 1.36 |
| 11 | International Journal of Management Education | 1 | 2018 | Q1 (Strategy and Management) | 10.3 | 6.0 | 1.26 |
| 12 | Journal of Adolescent Health | 1 | 2019 | Q1 (Public Health, Environmental and Occupational Health) | 10.4 | 5.5 | 2.27 |
| 13 | Journal of Air Transport Management | 1 | 2015 | Q1 (Management, Monitoring, Policy and Law) | 12.4 | 3.9 | 1.51 |

| 14 | Journal of Comparative Policy Analysis: Research and Practice | 2 | 2020, 2022 | Q1 (Public Administration) | 6.9 | 3.9 | 1.043 |
|----|----|---|----|----|----|----|----|
| 15 | Journal of Environmental Management | 1 | 2023 | Q1 (Management, Monitoring, Policy and Law) | 13.7 | 8.0 | 1.77 |
| 16 | Land Use Policy | 1 | 2018 | Q1 (Management, Monitoring, Policy and Law) | 13.7 | 6.0 | 1.85 |
| 17 | Policy Sciences | 4 | 2019, 2020, 2022 | Q1 (Public Administration) | 9.7 | 3.8 | 1.635 |
| 18 | Policy Studies | 1 | 2023 | Q1 (Political Science and International Relations) | 5.4 | 2.2 | 0.59 |
| 19 | Policy Studies Journal | 2 | 2018, 2020 | Q1 (Public Administration) | 11.6 | 4.1 | 1.88 |
| 20 | Policy and Society | 1 | 2015 | Q1 (Public Administration) | 18.0 | 5.7 | 2.22 |
| 21 | Politics & Policy | 3 | 2016, 2021 | Q2 (Political science and international relations) | 2.5 | 1.4 | 0.353 |
| 22 | Public Health | 1 | 2019 | Q1 (Public Health, Environmental and Occupational Health) | 7.6 | 3.9 | 1.2 |
| 23 | Qualitative Research Journal | 1 | 2021 | Q2 (Education) | 2.4 | 0.80 | 0.46 |
| 24 | Regional Science Policy & Practice | 1 | 2022 | Q2 (Management, Monitoring, Policy and Law) | 3.6 | 1.7 | 0.59 |
| 25 | Review of Policy Research | 1 | 2021 | Q2 (Public Administration) | 4.5 | 2.3 | 0.699 |
| 26 | Science and Public Policy | 5 | 2019, 2020, 2021, 2023 | Q1 (Public Administration) | 4.5 | 2.6 | 0.823 |
| 27 | Social Policy and Administration | 2 | 2016, 2019 | Q1 (Public Administration) | 6.5 | 2.6 | 1.16 |
| 28 | Social Science & Medicine | 1 | 2018 | Q1 (History and Philosophy of Science) | 9.1 | 4.9 | 1.95 |
| 29 | Technology in Society | 1 | 2024 | Q1 (Sociology and Political Science) | 17.9 | 10.1 | 2.25 |
| 30 | The Asian Journal of Shipping and Logistics | 1 | 2019 | Q2 (Management Science and Operations Research) | 7.8 | 3.3 | 0.78 |

| 31 | Transport Policy | 1 | 2021 | Q1 (Geography, Planning and Development) | 12.1 | 6.3 | 1.74 |
|---|---|---|---|---|---|---|---|
| 32 | Transport Reviews | 2 | 2016, 2023 | Q1 (Transportation) | 17.7 | 9.5 | 3.016 |
| 33 | Transportation Research Procedia | 1 | 2020 | Q3 (Transportation) | 9.1 | 1.54 | 0.384 |
| 34 | Transportation Research, Part A: Policy and Practice | 2 | 2020, 2024 | Q1 (Management Science and Operations Research) | 13.2 | 6.3 | 2.182 |
| 35 | Transportmetrica A: Transport Science | 1 | 2017 | Q1 (Engineering) | 8.1 | 3.6 | 1.1 |
| 36 | World Development | 1 | 2015 | Q1 (Sociology and Political Science) | 12.7 | 5.4 | 2.25 |

Source: processed data 2025

Several journals, such as Computers in Human Behavior (1 article), Environmental Science and Policy (1 article), and Technology in Society (1 article), show that research related to public policy evaluation is not limited to the social sciences alone. Furthermore, the presence of articles in journals related to health policy highlights that research in this area extends beyond just the social sciences. Additionally, the presence of articles across various journals focusing on transport, logistics, shipping, and planning demonstrates that policy evaluation studies also extend to the economic and technical sectors.

Beyond journal distribution, Table 1 also indicates that the evaluation studies in this systematic literature review were published between 2015 and 2024, with most publications occurring from 2018 to 2023. An analysis of publication quality using the Scopus Quartile rankings shows that a majority of articles (22 articles, or approximately 62.9%) were published in journals indexed in Scopus Quartile 1 (Q1). This suggests that research on public policy and program evaluation is largely dominated by high-quality studies with significant methodological and theoretical contributions. Meanwhile, 10 articles (28.6%) were published in journals indexed in Scopus Quartile 2 (Q2), and 1 article (2.9%) in a Scopus Quartile 3 (Q3) journal. The remaining 2 articles (5.7%) were published in journals that were not Scopus-indexed.

The Clarivate Analytics Impact Factor data indicates that the journals publishing these articles have varying impact factors, with a few, such as Computers in Human Behavior (9.0) and Technology in Society (10.1), having a high impact factor. This indicates that public policy evaluations are published in journals that have a significant impact and high visibility within the academic community. In addition, the SCImago Journal Rank (SJR) data also shows that journals publishing these articles have varied rankings. This shows that articles included in this review were selected from a range of highly reputable journals which have significant influence in their respective academic fields.

In addition to general analyses based on journals and year, it's important to note the distribution of articles based on policy area. Findings suggest that health is often a central focus in partnership-based program evaluations, as shown in the Get Healthy Get Active study (Fynn et al., 2022), reflecting the efforts of health sectors to collaborate in achieving policy effectiveness and health

outcomes. Analysis of literature on urban transportation policy reveals a concentration of studies in journals focusing on logistics and transportation (Gatta & Marcucci, 2016), while evaluations of online security programs (Pham et al., 2024) and environmental policies (van Sluisveld et al., 2017) have specific distributions in relevant journals. Simulation-based studies (Ben-Dor et al., 2024) have demonstrated increasing quality of publications in highly-ranked journals, indicating strong interest in data-driven policies. Studies that focus on climate change policy (Trosvik et al., 2023) highlight that methodological quality influences the reliability of the evaluation results.

### *Distribution of Articles Based on Policy and Evaluation Focus*

As shown in Figure 5, This classification is aimed at mapping the areas that have been most explored and providing a comprehensive overview of current research trends. The figure shows the most frequent focus is on the Evaluation of Social Programs and Policies, totaling 12 articles (20.7%). Studies in this category cover diverse topics, including evaluations of the effectiveness of poverty alleviation programs, implementation of social assistance, and the protection of vulnerable groups. Based on a review of these articles, the studies emphasize the importance of a systemic perspective, which considers the interactions of different elements and local contexts within each policy (Caffrey & Munro, 2017). Moreover, these studies highlight the significance of collaboration among stakeholders as a key factor in the success of program evaluation and implementation (Fynn et al., 2022).
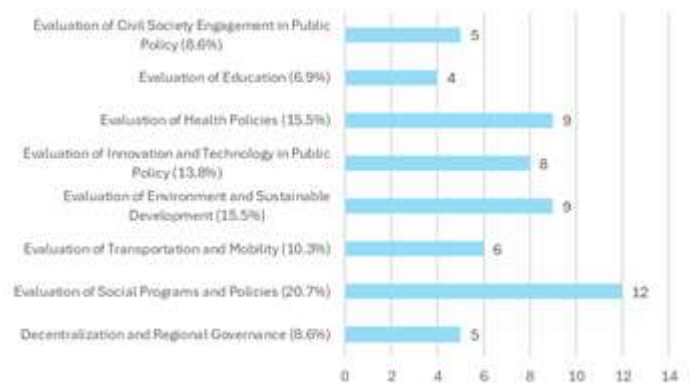


Source: processed data

**Figure 5.** Distribution of Articles Based on Policy and Evaluation Focus

The second most frequently discussed category is the Evaluation of Environment and Sustainable Development, with 9 articles (15.5%). These studies focus on assessing environmental impacts of development policies, the implementation of green public procurement, and actions taken to achieve sustainability. Various approaches were used including, carbon footprint analysis to measure environmental impacts, as well as agent-based models to examine complex interactions in water resources management (Cerutti et al., 2016; Chuanwang Sun et al., 2023).

Following closely, the Evaluation of Health Policies category also has 9 articles (15.5%), reflecting the importance of evaluations within this sector. Studies in this category discuss various topics such as evaluations of public health programs, the impact of health interventions on vulnerable populations, and the role of cross-sectoral collaborations in achieving policy effectiveness (Fynn et al., 2022; Lakerveld et al., 2020). The category of Evaluation of Innovation and Technology in Public Policy contains 8 articles (13.8%). Studies in this category focus on the role of technology in public policy and innovations such as the application of technology and artificial intelligence (AI) in supporting decision-

making in public policy (Bi et al., 2023; Xu & Hu, 2024).

The category of Evaluation of Civil Society Engagement in Public Policy includes 5 articles (8.6%), highlighting the important role of civil society organizations (CSOs) in promoting transparency, accountability, and public participation in policy cycles. Various methods were used to explore the role of CSOs in public policy advocacy and how this interaction can influence the evaluation process (Bundi & Trein, 2022; Uddin, 2023). Then, the category Decentralization and Regional Governance encompasses 5 articles (8.6%), and focuses on the role of local capacities in policy implementation at the regional level, highlighting how decentralization can impact the allocation of resources and decision-making (Adams, 2016).

Furthermore, the Evaluation of Transportation and Mobility is represented by 6 articles (10.3%), covering policies and case studies related to urban transportation, mobility, and logistic systems, using empirical data and simulations to assess the effectiveness of policy interventions on traffic and emissions reduction (Ben-Dor et al., 2024; Gatta & Marcucci, 2016; Hsieh, 2020). Finally, the Evaluation of Education category includes 4 articles (6.9%), focusing on participant engagement and the use of innovative evaluation methods to improve education quality, through the application of both qualitative and quantitative research methods for evaluating educational interventions (Griggs & Crain-Dorough, 2021; McGough et al., 2018). This distribution highlights that public policy evaluation currently addresses various sectors and themes relevant to development challenges. However, the majority of articles in this review focused on social issues, environment, and health, with less emphasis on areas such as transportation and education.

**b. Content Analysis**

The analysis of 58 selected articles reveals a diverse range of theories and methods employed in public policy evaluation research. Various theoretical perspectives, such as the Capability Approach (CA), which emphasizes individual freedoms and improved quality of life, are used as conceptual frameworks for evaluating the impact of development policies (Garcés-Velástegui, 2022). In the context of transportation policy, stated preference (SP) and agent-based simulation are employed to capture diverse stakeholder preferences and analyze their responses to different policies (Ben-Dor et al., 2024; Gatta & Marcucci, 2016). These methods facilitate analyses that are more specifically tailored to each stakeholder's viewpoint, considering their preferences and responses.

Several studies investigating policy cohesion utilize contribution analysis and qualitative comparative analysis (QCA) to assess small-scale policy interventions with limited data. These methods are frequently combined with quantitative approaches like difference-in-differences and regression discontinuity to enhance the robustness of evaluation results (Koudoumakis et al., 2022). Simultaneously, narrative approaches are employed to explore the depth of stakeholder experiences and perceptions, particularly in multi-stakeholder evaluations, offering a more contextual and nuanced understanding (McGough et al., 2018). Furthermore, text mining and the Policy Modeling Consistency (PMC) index are utilized to analyze policy coherence, focusing on the identification of key themes and the alignment of various policy elements (Xu & Hu, 2024).

In order to evaluate the effectiveness of public policy more rigorously, several studies use quantitative methods like randomized controlled trials (RCTs) (Pham et

al., 2024). Our analysis of these articles reveals that, while quantitative methods such as regression analysis (Hsieh, 2020; Xu & Hu, 2024) and RCTs (Pham et al., 2024) are commonly employed, a substantial proportion (approximately 40%) utilize mixed-methods approaches. These approaches combine quantitative and qualitative data to offer a more holistic understanding of policy impacts. Examples include studies integrating qualitative interviews with quantitative surveys (Fynn et al., 2022), combining case studies with statistical analysis (Caffrey & Munro, 2017), and employing QCA alongside contribution analysis (Koudoumakis et al., 2022). This emphasis on mixed methods aligns with the findings of Mergoni and De Witte (2021), who observed a growing trend towards nonparametric frontier analysis in policy evaluation, which often necessitates the integration of diverse data sources.

However, we also note significant limitations. The application of realist evaluation, a valuable approach for understanding underlying causal mechanisms (Adams, 2016), remains infrequent. This scarcity is also highlighted by Mavrot et al. (2024), who emphasize the need for more in-depth reflection on evaluation criteria and their alignment with specific policy contexts. Furthermore, many studies, particularly those employing purely quantitative methods, often overlook the importance of contextual factors and stakeholder perspectives, potentially leading to an incomplete or biased assessment of policy effectiveness. This concern resonates with Renyaan's (2023) call for more people-centered evaluations that consider the broader social and psychological impacts of policies.

Compared to these previous reviews (Mavrot et al., 2024; Mergoni & De Witte, 2021; Renyaan, 2023), our study contributes a broader, cross-cutting analysis of public policy evaluation methodologies. We move beyond cataloging methods to assess their effectiveness in generating valid and relevant evidence. This allows us to pinpoint specific strengths, such as the ability of mixed-methods approaches to capture complexities, and to identify persistent gaps, such as the underutilization of realist evaluation and the neglect of contextual factors.

Additionally, health policy evaluations often adopt frameworks from the Centers for Disease Control and Prevention (CDC) and the Consolidated Framework for Implementation Research (CFIR) to assess policy implementation and impacts (Lakerveld et al., 2020).

Approaches and frameworks in policy and program evaluation reflect a sophisticated understanding of public policy dynamics. These various approaches focus on causal mechanisms, complex systems, stakeholder perspectives, and data quality and transparency. For instance, Adams (2016) proposed a framework for decentralization by identifying causal mechanisms influencing resource availability, constraints, and incentives. Fontaine (2020) realist policy design approach emphasized causal mechanisms within open systems, highlighting interactions among actors, policy instruments, and administrative contexts. To address complexity, system dynamics (SD) and agent-based modeling (ABM) were also frequently employed for assessing the impact of water resources and transportation management policies (Ben-Dor et al., 2024; Chuanwang Sun et al., 2023; Zenezini & Marco, 2020). Fuzzy Set Qualitative Comparative Analysis (fsQCA) is also utilized to analyze multiple causalities (Garcés-Velástegui, 2022), and evaluations of policies are sometimes combining QCA and process tracing (PT) to enrich understanding of

effectiveness factors and causal mechanisms (Pattyn et al., 2022; Rothgang & Lageman, 2021). To explore stakeholder perspectives and social values, the Appreciative Inquiry (AI) framework is often used, promoting collaboration among stakeholders (Griggs & Crain-Dorough, 2021). Furthermore, Narrative Policy Framework (NPF) is used to support active stakeholder engagement in policy processes (McGough et al., 2018).

Content analysis of the reviewed articles revealed several factors influencing the effectiveness of policy evaluation. Internal factors, which are inherent to the policy or program itself, such as the capacity of local governments, local data quality, and program design, play a crucial role. For instance, in decentralized contexts, government capacity, local data quality, accountability, and public participation are crucial for the success of evaluations (Adams, 2016). Meanwhile, strong program designs, clear program theories, the quality of program data, and cross-sector partnerships are critical in the evaluation of social policies (Caffrey & Munro, 2017; Fynn et al., 2022). The use of robust quantitative methods, such as agent-based simulation and modified elasticity analysis, were found to be influential in producing accurate evaluations of transport policies (Ben-Dor et al., 2024; Hsieh, 2020). In the context of environmental and sustainability policy evaluation, relevant data and indicators, in-depth ecological knowledge, analysis of product life cycles, and the engagement of local communities are recognized as key internal factors (Cerutti et al., 2016; Chuanwang Sun et al., 2023). In technology and innovation policies, technology competence, academia-industry partnerships, and relevant research facilities are important internal factors (Bi et al., 2023). Internal factors for evaluating health policy include the capacity of health systems, health data quality, collaboration among health professionals, patient engagement, and community involvement (Fynn et al., 2022; Lakerveld et al., 2020). Internal factors for the evaluations of public participation in policy-making include capacity building in civil society organizations, public participation, access to information and transparent decision-making processes.

External factors also play a crucial role, often stemming from the broader socio-political context. In decentralized governance, these factors include the political climate, national policy backing, regulatory frameworks, and the existing institutional setting (Adams, 2016). For social policies, factors include financial support, political bias, civil society engagement, and the socioeconomic conditions of program beneficiaries. The transport sector is influenced by factors like transport regulations, infrastructure, technological progress, and social change. Environmental and sustainability policies are impacted by energy policies, public pressure and global opinion, the promotion of green technology, and robust regulations (Cerutti et al., 2016; Chuanwang Sun et al., 2023). Similarly, for innovation and technology policy, external factors include funding policies, technology regulations, the innovation climate, organizational culture, availability of skilled labor, and global technology developments (Bi et al., 2023). Finally, evaluations of public participation are influenced by government support, regulations of civil society involvement, the existence of public space, a participatory culture, and access to accurate information. These internal and external factors, which interact with each other, emphasize the necessity of using a comprehensive perspective when designing and implementing evaluations to create the most effective and impactful policy evaluation within various public policy domains.
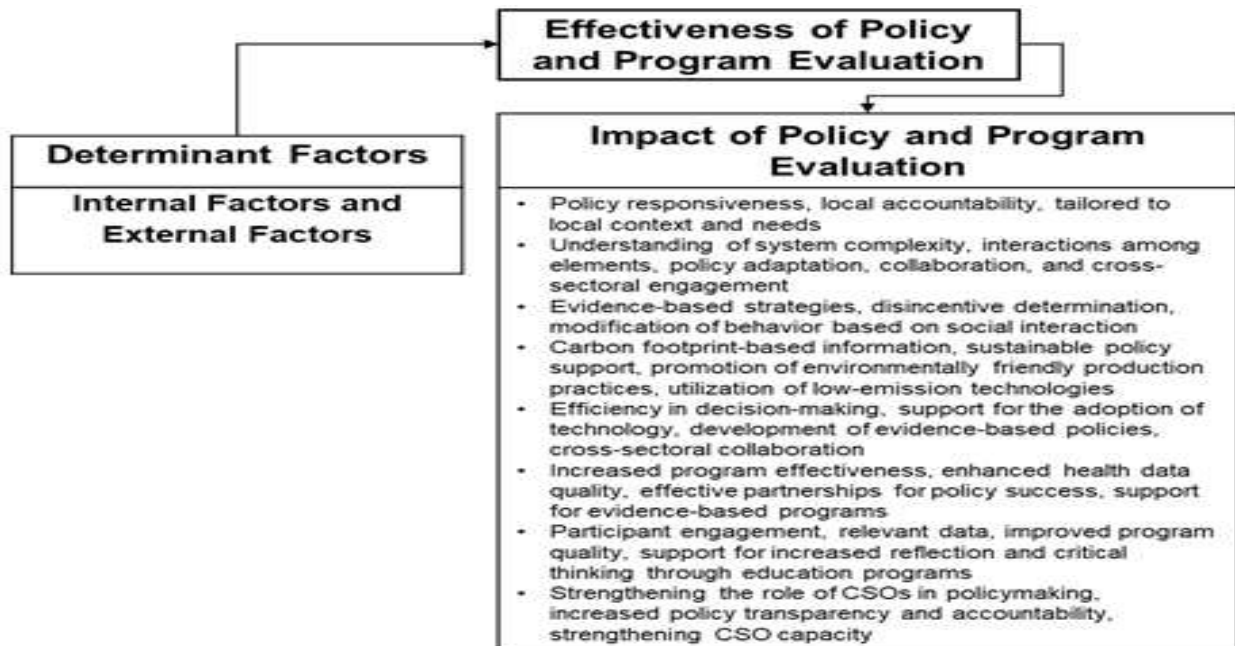
Source: processed data

**Figure 6.** Relationship Between Determinants of Evaluation Effectiveness and the Impact of Policy and Program Evaluation Utilization

The impact and utilization of policy and program evaluations, depicted in Figure 6, are a result of intricate interactions between internal and external factors. The impact of evaluation is reflected in improved policy responsiveness and local accountability, greater understanding of system complexity, and the development of evidence-based strategies. The utilization of evaluation leads to increased program effectiveness, data quality, stakeholder engagement, and a stronger role for civil society organizations. In decentralized contexts, evaluations facilitate more responsive policies and increase governmental accountability (Adams, 2016). The study of social policies also highlights the importance of a system-based approach emphasizing interconnections and understanding complex contexts (Caffrey & Munro, 2017). Adaptable and collaborative partnerships among stakeholders are also identified as crucial for effective policy implementation (Fynn et al., 2022).

Furthermore, quantitative analysis derived from simulation can offer insights about the effects of disincentives in transportation (Ben-Dor et al., 2024), while elasticity analysis in parking policies contributes to the understanding of spillover effects (Hsieh, 2020). Using carbon footprint evaluations informs green procurement policies (Cerutti et al., 2016), while agent-based modeling facilitates informed decisions in water management (Chuanwang Sun et al., 2023). In technological innovation, evaluations can lead to effective decision-making and cross-sector collaborations (Bi et al., 2023; Xu & Hu, 2024). In healthcare policies, evaluations are shown to increase program effectiveness and to support practices that improve service delivery (Fynn et al., 2022; Lakerveld et al., 2020). Finally, approaches such as Appreciative Inquiry (AI) and narrative approaches have facilitated

constructive engagement among stakeholders (Griggs & Crain-Dorough, 2021; McGough et al., 2018). Nevertheless, various challenges to utilizing evaluation findings still persist, including variations in local capacity, data accuracy, and the complexity of policy systems (Adams, 2016; Ben-Dor et al., 2024; Caffrey & Munro, 2017; Cerutti et al., 2016).

### c. Literature Gaps and Future Research Directions

The analysis of the existing literature on public policy evaluation highlights significant gaps that need to be addressed in order to advance the field. These gaps can be broadly categorized into three main areas: (1) methodological limitations and approaches; (2) a lack of contextual and specific studies, particularly in developing countries; and (3) limitations in measurement and indicators.

Regarding methodological approaches, public policy evaluation studies remain limited in the use of integrative and comprehensive methods. The effective use of mixed methods, combining both qualitative and quantitative approaches, is not sufficiently applied to gain a deeper understanding of policy complexities (Green et al., 2015; Parr & Churchill, 2020). Longitudinal studies and long-term evaluations, which are critical for assessing the sustained impact of policies, are also lacking (Sandberg, 2016). While realist evaluation offers a promising avenue for understanding causal mechanisms, it remains underutilized, particularly in developing countries where unique characteristics and resource limitations often render existing methodologies inadequate (Adams, 2016; Parr & Churchill, 2020). Complex policy evaluation studies should also more fully utilize simulation and agent-based modeling to analyze dynamic systems, and natural experiments that provide insight into policy effects within real-world conditions (Ben-Dor et al., 2024; Green et al., 2015; Zenezini & Marco, 2020). Furthermore, evaluation approaches often emphasize formal results, neglecting broader dimensions like stakeholder experiences and the complexities of social and cultural contexts (Caffrey & Munro, 2017).

In terms of contextual and specific studies, most existing research focuses on developed country contexts, with limited studies in developing nations (Uddin, 2023). Studies that specifically consider institutional constraints, resource limitations, and political hurdles in developing countries remain scarce (Adams, 2016; Uddin, 2023). There is also a significant gap in studies that evaluate the impacts of policy on minority, vulnerable, and marginalized groups, and where inequities in policy outcomes are often overlooked (Lakerveld et al., 2020). Furthermore, there is a need for studies that explore how policy contexts influence implementation and acceptance of policies, including the effects of cultural context, local wisdom, and socio-political factors (Hsieh, 2020; Marra & McCullagh, 2018). Evaluations that integrate psychological, emotional, and well-being aspects of policy recipients are also lacking, limiting a comprehensive understanding of the non-material dimensions of policy impact (Qi & Wu, 2018).

Limitations also exist in terms of measurement and indicators for public policy evaluation. The lack of standardized indicators for cross-country and cross-context comparisons presents a barrier to more comprehensive evaluations, especially for transparency, accountability, and policy responsiveness (Lakerveld et al., 2020). Data limitations also create obstacles for long-term and comprehensive evaluation efforts, hindering the measurement of policy impacts over time (Sandberg, 2016). Measuring policy "spillover effects," which refer to policy impacts on unintended groups, remains

challenging (Pham et al., 2024). Limitations related to data validity, reliability, and the lack of long-term data, as well as secondary data gaps related to various relevant policy aspects also need addressing. There is also limited research exploring the application of the open science principles in the design of policy evaluations and the dissemination of policy results that can enhance transparency and accountability (Moradi & Abdi, 2023).

Future research should address these identified gaps. Methodologically, future research needs to explore and embrace more innovative approaches. The increased use of case studies, grounded theory, and process tracing (PT) can help to understand causal mechanisms and complex policy dynamics (Rothgang & Lageman, 2021). Furthermore, data mining techniques and network analysis can be used to explore patterns and to uncover actor relations within policy networks (Xu & Hu, 2024). The use of agent-based modeling (agent-based modeling) and system dynamics are encouraged for analyzing complex systems and interactions among policy actors (Ben-Dor et al., 2024; Chuanwang Sun et al., 2023; Zenezini & Marco, 2020). More adaptive evaluation approaches are needed, especially to address measurement bias and to enhance the validity of findings, using methods like user simulation and list experiments (Michener, 2015; Pham et al., 2024). Finally, the adoption of comparative approaches is recommended to compare evaluation studies across various contexts and countries, enriching analyses and strengthening generalizations.

Future studies also need to pay more attention to contextual diversity, particularly in developing countries (Bundi & Trein, 2022; Uddin, 2023). Explicit investigation of policy challenges in developing countries is necessary, along with how local contexts influence policy design, implementation, and

evaluation outcomes (Uddin, 2023). More research is also required to better understand policy impacts on minority, vulnerable, and marginalized groups in order to address inequalities (Lakerveld et al., 2020). Furthermore, studies need to explore how specific policy issues affect the psychological dimensions of policy recipients to achieve more holistic and people-centered evaluations (Qi & Wu, 2018). Research on the influence of stakeholder perspectives, local communities (local rationalities), and the effects of socio-political contexts on the implementation and reception of policies are also warranted (Adams, 2016; Hsieh, 2020; Marra & McCullagh, 2018; McConnell et al., 2020). Future research also needs to further develop methodologies for multi-dimensional and integrative studies, incorporating social, economic, environmental, and psychological dimensions to better capture the complex impact of policies (Cerutti et al., 2016; Fynn et al., 2022; Lin et al., 2020). Research that integrates cross-sectoral perspectives is also needed to achieve better policy design and implementation (Fynn et al., 2022), as well as explorations into how existing social structures and power relations influence policy outcomes, to better address inequality.

Finally, future research needs to focus on strengthening capacity for evaluation, while also improving engagement with stakeholders and civil society. Involvement of stakeholders in evaluation processes and a need to provide evaluators with the capacity and skills, including for the integration of horizontal and multi-dimensional perspectives, is critical (Bundi & Trein, 2022; Omori & Tesorero, 2020; Sjöö & Callerstig, 2023; Uddin, 2023). Future studies should also explore the role and operational mechanism of the "facilitating state" in policy-making (Xu & Hu, 2024), and should

assess how context (political and technical factors) influences the use of policy evaluation findings, as well as the implementation of person-centered evaluations and Appreciative Inquiry (AI) to enhance the evaluation of policy effectiveness (Bundi & Trein, 2022; Griggs & Crain-Dorough, 2021; Vasilenko et al., 2019). Furthermore, it is recommended that research focuses on developing more adaptable models and frameworks to address the challenges posed by data bias and that promotes the more effective use of mixed method designs to achieve a more comprehensive understanding of the policy context (Green et al., 2015; Koudoumakis et al., 2022; Michener, 2015; Parr & Churchill, 2020). The importance of using evaluation results to generate policy making based on empirical evidence should also be highlighted in future research.

## CONCLUSION

This systematic literature review reveals a significant shift in public policy evaluation, moving beyond a sole focus on quantifiable outcomes. While traditional quantitative methods like randomized controlled trials (RCTs) remain relevant, there is a clear and growing trend towards incorporating mixed-methods approaches, systems-based thinking, and qualitative inquiry. This evolution emphasizes a more holistic understanding of policy, encompassing not just what the outcomes are, but also how and why policies function (or fail) within specific processes, contexts, and considering the diverse perspectives of stakeholders. Crucially, our findings underscore a critical need for more comprehensive research that adapts these evolving methodologies to the unique challenges and opportunities present in developing country contexts.

The effectiveness of any policy evaluation is contingent upon a complex interplay of internal and external factors. Internal factors, such as the design of the program being evaluated, the quality of available data, and the capacity of local institutions, interact dynamically with external factors like the prevailing political climate, the availability of funding, and broader socio-economic conditions. Successful evaluations, therefore, must acknowledge and account for this complexity. This includes integrating considerations of contextual nuances, the diversity of affected populations, and even the psychological and emotional well-being of individuals impacted by the policies under scrutiny. Moreover, active engagement with stakeholders throughout the evaluation process, coupled with transparency in both methodology and data handling, are increasingly recognized as essential components of rigorous and impactful policy evaluation.

Despite the advancements identified in this review, several methodological and practical limitations persist. These include the need for more adaptive and contextual evaluation methods capable of capturing the intricacies of real-world policy implementation. Further development and refinement of measurement and indicator systems are also needed, particularly to address issues of data bias and to ensure relevance across diverse cultural and socio-economic settings. We also identified a crucial role for multidisciplinary collaboration, bringing together expertise from various fields to tackle the multifaceted challenges of policy evaluation. The appropriate and ethical application of emerging technologies—including big data analytics, agent-based simulation, and artificial intelligence (AI)—holds significant promise for enhancing the accuracy, efficiency, and depth of future evaluations. Addressing the identified research gaps necessitates a multi-pronged approach.

Future research should prioritize methodological innovation, with a focus on strategically integrating qualitative and quantitative data within mixed-methods designs. This integration should go beyond simple combination, aiming for a nuanced understanding of causal mechanisms and context-specific policy impacts. The expanded use of realist evaluation, which explicitly seeks to uncover how and why policies work (or fail) in particular settings, is particularly warranted, especially in developing countries where it remains underutilized. The development and validation of context-specific indicators that are sensitive to local realities are also paramount. Future evaluations should embrace a systems thinking, including through the use of agent-based modeling, system dynamics, and process tracing to better capture complexities and interaction within policy systems.

A concentrated focus on developing country contexts is essential. This includes conducting more empirical studies that directly address the unique challenges faced by these nations, such as limited resources, institutional constraints, and diverse cultural contexts. Research should explicitly investigate the impacts of policies on vulnerable and marginalized populations, ensuring that issues of equity and inclusion are central to the evaluation agenda. The incorporation of local knowledge and perspectives is crucial for ensuring the relevance and promoting the ownership of evaluation findings.

Strengthening evaluation capacity and fostering meaningful stakeholder engagement are also critical. This requires investing in evaluator training, equipping practitioners with the skills to implement advanced methodologies effectively. Promoting participatory evaluation approaches, which actively involve stakeholders in all stages of the evaluation process, is vital for ensuring that evaluations are relevant, credible, and ultimately useful. Enhanced transparency and data sharing, guided by open science principles, will further improve the credibility and accessibility of evaluation findings. Finally, dedicated efforts are needed to translate evaluation knowledge into actionable policy recommendations, fostering evidence-based decision-making. Future research should explore the utilization of machine learning, AI in policy evaluation, and conduct studies on technological implementation in developing countries.

## REFERENCES

1. Adams, B. E. (2016). Assessing the Merits of Decentralization: A Framework for Identifying the Causal Mechanisms Influencing Policy Outcomes. *Politics and Policy*, *44*(5), 820–849. https://doi.org/10.1111/polp.12172

2. Ben-Dor, Golan., Ogulenko, Aleksey., Klein, Ido., Ben-Elia, Eran., Benenson, Itzhak. (2024). Simulation-based policy evaluation of monetary car driving disincentives in Jerusalem. *Transportation Research Part A Policy and Practice.* 183. 104061. https://doi.org/10.1016/j.tra.2024.104061

3. Bi, X., Nie, H., Zhang, G., Hu, L., Ma, Y., Zhao, X., Yuan, Y., Wang, G. (2023). Boosting question answering over knowledge graph with reward integration and policy evaluation under weak supervision. *Information Processing & Management*, *60*(2), 103242. https://doi.org/10.1016/j.ipm.2022.103242

4. Bundi, P., & Trein, P. (2022). Evaluation use and learning in public policy. *Policy Sciences*, *55*(2), 283–

309. https://doi.org/10.1007/s11077-022-09462-6

5.  Caffrey, L., Munro, E. (2017). A systems approach to policy evaluation. *Evaluation*, *23*(4), 463–478. https://doi.org/10.1177/1356389017730727

6.  Cerutti, A. K., Contu, S., Ardente, F., Donno, D., Beccaro, G. L. 2016. Carbon footprint in green public procurement: Policy evaluation from a case study in the food sector. *Food Policy*, *58*, 82–93. https://doi.org/10.1016/j.foodpol.2015.12.001

7.  Dunn, William N., (2018). *Public Policy Analysis: An Integrated Approach*, 6th Edition, Routledge

8.  Feng, K., & Jiang, Z. (2021). State Capacity and Innovation Policy Performance: A Comparative Study on Two Types of Innovation Projects in China. *Review of Policy Research*, *38*(4), 427–453. https://doi.org/10.1111/ropr.12420

9.  Fischer, F., Miller, G. J., Sidney, M. S., (2007). *Handbook of Public Policy Analysis: Theory, Politics, and Methods*. CRC Press

10. Fontaine, G. (2020). The contribution of policy design to realist evaluation. *Evaluation*, *26*(3), 296–314. https://doi.org/10.1177/1356389020902496

11. Fynn, Judith F., Milton, Karen., Hardeman, Wendy., Jones, Andy P., (2022). A model for effective partnership working to support programme evaluation. *Evaluation*. 28(3), 284–307. https://doi.org/10.1177/13563890221096178

12. Freidson E., Rist, R., (2020). *Program Evaluation and The Management of Government: Patterns and Prospects across Eight Nations* (Comparative Policy Evaluation). 1st Edition, Routledge

13. Garcés-Velástegui, P. (2022). Using the Capability Approach and Fuzzy Set Qualitative Comparative Analysis in Development Policy Evaluation. *Journal of Comparative Policy Analysis: Research and Practice*, *24*(2), 179–197. https://doi.org/10.1080/13876988.2019.1699277

14. Gatta, V., Marcucci, E. (2016). Stakeholder-specific data acquisition and urban freight policy evaluation: Evidence, implications and new suggestions. *Transport Reviews*, *36*(5), 585–609. https://doi.org/10.1080/01441647.2015.1126385

15. Green, J., Roberts, H., Petticrew, M., Steinbach, R., Goodman, A., Jones, A., & Edwards, P., (2015). Integrating quasi-experimental and inductive designs in evaluation: A case study of the impact of free bus travel on public health. *Evaluation*, *21*(4), 391–406. https://doi.org/10.1177/1356389015605205

16. Griggs, D. M., Crain-Dorough, M. (2021). Appreciative inquiry's potential in program evaluation and research. *Qualitative Research Journal*, *21*(4), 375–393. https://doi.org/10.1108/QRJ-06-2020-0059

17. Goodin, R., Moran, M., Rein, M., (2006). *The Oxford Handbook of Public Policy*. Oxford University Press

18. Howlett, M., (2022). *The Routledge Handbook of Policy Tools*, 1st Edition. Routledge

19. Hsieh, H. (2020). Transport policy evaluation based on elasticity analysis

with social interactions, *Transportation Research Part A-policy and Practice*, 139, 273-296. https://doi.org/10.1016/j.tra.2020.07.011

20. Johnson, Blair T., Hennessy, Emily A., (2019). Systematic reviews and meta-analyses in the health sciences: Best practice methods for research syntheses. S*ocial Science & Medicine*. 233(4), 237-251. https://doi.org/10.1016/j.socscimed.2019.05.035

21. Koudoumakis, P., Botzoris, G., Protopapas, A. (2022). Cohesion policy evaluation: Guidelines for selection of appropriate methods. *Regional Science Policy & Practice*, *14*(5), 1062–1085. https://doi.org/10.1111/rsp3.12524

22. Lakerveld, J., Woods, C., Hebestreit, A., Brenner, H., Flechtner-Mors, M., Harrington, J. M., Kamphuis, C. B. M., Laxy, M., Luszczynska, A., Mazzocchi, M., Murrin, C., Poelman, M., Steenhuis, I., Roos, G., Steinacker, J. M., Stock, C. C., van Lenthe, F., Zeeb, H., Zukowska, J., & Ahrens, W., (2020). Advancing the evidence base for public policies impacting on dietary behaviour, physical activity and sedentary behaviour in Europe: The Policy Evaluation Network promoting a multidisciplinary approach. *Sustainable Food Systems for Healthy Diets in Europe and Central Asia*, *96*, 101873. https://doi.org/10.1016/j.foodpol.2020.101873

23. Li, H., Dong, X., Jiang, Q., & Dong, K. (2021). Policy analysis for high-speed rail in China: Evolution, evaluation, and expectation. *Transport Policy*, *106*, 37–53. https://doi.org/10.1016/j.tranpol.2021.03.019

24. Lin, Q.-S., Hu, T.-J., & Zhou, X.-H. (2020). Estimating the daily trend in the size of the COVID-19 infected population in Wuhan. *Infectious Diseases of Poverty*, *9*(1), 69. https://doi.org/10.1186/s40249-020-00693-4

25. Marra, M., & McCullagh, C. (2018). Feeling able to say it like it is: A case for using focus groups in programme evaluation with international cohorts. *The International Journal of Management Education*, *16*(1), 63–79. https://doi.org/10.1016/j.ijme.2017.12.006

26. Mavrot, C., Potluka, O., Balzer, L., Eicher, V., Haunberger, S., Heuer, C., & Viallon, F.-X. (2025). What evaluation criteria are used in policy evaluation research: A cross-field literature review. Evaluation and Program Planning, 108, 102512. https://doi.org/10.1016/j.evalprogplan.2024.102512

27. McConnell, A., Grealy, L., & Lea, T. (2020). Policy success for whom? A framework for analysis. *Policy Sciences*, *53*(4), 589–608. https://doi.org/10.1007/s11077-020-09406-y

28. McGough, David., Bedell, Claudine., Tinkler, Barri., (2018). Building a Dangerous Outpost in the Green Mountain State: a Case Study of Educator Preparation Policymaking. *Education Policy Analysis Archives*, 26, 37. https://doi.org/10.14507/epaa.26.2848

29. Mergoni, A., & De Witte, K. (2022). Policy evaluation and efficiency: A systematic literature review.

International Transactions in Operational Research, 29(3), 1337–1359. https://doi.org/10.1111/itor.13012

30. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

31. Michener, G. 2015. Policy Evaluation via Composite Indexes: Qualitative Lessons from International Transparency Policy Indexes. *World Development*, *74*, 184–196. https://doi.org/10.1016/j.worlddev.2015.04.016

32. Moradi, S., & Abdi, S. (2023). Open science-related policies in Europe. *Science and Public Policy*, *50*(3), 521–530. https://doi.org/10.1093/scipol/scac082

33. Okamuro, Hiroyuki., Nishimura, Junichi., (2021). Effects of multilevel policy mix of public R&D subsidies: Empirical evidence from Japanese local SMEs. *Science and Public Policy*. 48(6), 829-840. https://doi.org/10.1093/scipol/scaa075

34. Omori, S., & Tesorero, B. S. (2020). Why Does Polycentric Governance Work for Some Project Sites and Not Others? Explaining the Sustainability of Tramline Projects in the Philippines. *Policy Studies Journal*, *48*(3), 833–860. https://doi.org/10.1111/psj.12299

35. Page, Matthew., Mckenzie, Joanne., Bossuyt, Patrick., Boutron, Isabelle., Hoffmann, Tammy., Mulrow, Cynthia., Shamseer, Larissa., Tetzlaff, Jennifer., Akl, Elie., Brennan, Sue., Chou, Roger., Glanville, Julie., Grimshaw, Jeremy., Hróbjartsson, Asbjørn., Lalu, Manoj., Li, Tianjing., Loder, Elizabeth., Mayo-Wilson, Evan., Mcdonald, Steve., Moher, David., (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *Journal of Clinical Epidemiology*. 134(1), 178-189. https://doi.org/10.1016/j.jclinepi.2021.03.001

36. Parr, S., & Churchill, H. (2020). The Troubled Families Programme: Learning about policy impact through realist case study research. *Social Policy and Administration*, *54*(1), 134–147. https://doi.org/10.1111/spol.12529

37. Pattyn, V., Álamos-Concha, P., Cambré, B., Rihoux, B., Schalembier, B., (2022). Policy Effectiveness through Configurational and Mechanistic Lenses: Lessons for Concept Development. *Journal of Comparative Policy Analysis: Research and Practice*, *24*(1), 33–50. https://doi.org/10.1080/13876988.2020.1773263

38. Pattyn, V., Brans, M. (2015). Organisational analytical capacity: Policy evaluation in Belgium. *Special Issue on The Dynamics of Policy Capacity*, *34*(3), 183–196. https://doi.org/10.1016/j.polsoc.2015.09.009

39. Pham, Truong., Goto, Daisaku., Tran, Duc. (2024). Child online safety education: A program evaluation combining a randomized controlled trial and list experiments in Vietnam. *Computers in Human Behavior*. 156(100795), 108225.

https://doi.org/10.1016/j.chb.2024.1
08225

40. Qi, D., Wu, Y. (2018). Does welfare stigma exist in China? Policy evaluation of the Minimum Living Security System on recipients' psychological health and wellbeing. *Social Science & Medicine*, *205*, 26–36. https://doi.org/10.1016/j.socscimed.2018.03.041

41. Renyaan, D. (2023). Exploring a Whole Understanding of Public Policy Evaluation. Formosa Journal of Science and Technology, 2(4), Article 4. https://doi.org/10.55927/fjst.v2i4.3902

42. Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., Koffel, J. B., & PRISMA-S Group, (2021). PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic reviews*, *10*(1), 39. https://doi.org/10.5195/jmla.2021.962

43. Rothgang, M., & Lageman, B. (2021). The unused potential of process tracing as evaluation approach: The case of cluster policy evaluation. *Evaluation*, *27*(4), 527–543. https://doi.org/10.1177/13563890211041676

44. Sandberg, J. (2016). Between Poor Relief and Human Capital Investments—Paradoxes in Hybrid Social Assistance. *Social Policy and Administration*, *50*(3), 316–335. https://doi.org/10.1111/spol.12111

45. Sjöö, K., & Callerstig, A.-C. (2023). The challenges in integrating horizontal perspectives in sectoral policy evaluation. *Policy Studies*, *44*(2), 174–194. https://doi.org/10.1080/01442872.2021.2015312

46. Sun, Chuanwang., Jialin, Min., Li, Jinqiang., Cai, Weiyi., (2023). Public participation and policy evaluation in China's smog governance. *Environmental Impact Assessment Revie*w. 100. https://doi.org/10.1016/j.eiar.2023.107052

47. Sun, Dongying., Zhang, Miao., Jung, Donghwi., (2023). Policy evaluation of economic – environmental tradeoffs in regulating industrial water use: An agent-based model. *Journal of Environmental Management*. 364. https://doi.org/10.1016/j.eiar.2023.107052

48. Tam, Wilson., Tang, Arthur., Woo, Brigitte., Goh, Yong-Shian. (2019). Perception of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement of authors publishing reviews in nursing journals: A cross-sectional online survey. *BMJ Open*. 9(4). 26271. https://doi.org/10.1136/bmjopen-2018-026271

49. Trosvik, L., Takman, J., Björk, L., Norrman, J., & Andersson-Sköld, Y. (2023). A meta-evaluation of climate policy evaluations: Findings from the freight transport sector. *Transport Reviews*, *43*(5), 867–887. https://doi.org/10.1080/01441647.2023.2175275

50. Uddin, M. K. (2023). NGOs' approach to human rights and the challenges in Bangladesh. *Development Policy Review*, *41*(3). https://doi.org/10.1111/dpr.12667

51. Varone, F., Jacob, S., Bundi, P., (2023). *Public Policy Evaluation*. Edward Elgar Publishing Limited

52. van Sluisveld, M. A. E., Hof, A. F., van Vuuren, D. P., Boot, P., Criqui, P., Matthes, F. C., Notenboom, J., Pedersen, S. L., Pfluger, B., & Watson, J. (2017). Low-carbon strategies towards 2050: Comparing ex-ante policy evaluation studies and national planning processes in Europe. *Environmental Science & Policy*, *78*, 89–96. https://doi.org/10.1016/j.envsci.2017.08.022

53. van Voorst, S., Mastenbroek, E. (2019), Evaluations as a decent knowledge base? Describing and explaining the quality of the European Commission's ex-post legislative evaluations. *Policy Sciences*, 52(4), 625–644. https://doi.org/10.1007/s11077-019-09358-y

54. Vasilenko, S. A., Glassman, J. R., Kugler, K. C., Peskin, M. F., Shegog, R., Markham, C. M., Emery, S. T., Coyle, K. K. (2019). Examining the Effects of an Adolescent Pregnancy Prevention Program by Risk Profiles: A More Nuanced Approach to Program Evaluation. *Journal of Adolescent Health*, *64*(6), 732–736. https://doi.org/10.1016/j.jadohealth.2018.12.003

55. Vedung, E., (1997). *Public Policy and Program Evaluation*, Routledge

56. Xu, T. L., & Hu, Y. (2024). Towards sustainable prosperity? Policy evaluation of Jiangsu advanced manufacturing clusters. *Technology in Society*, *77*, 102583. https://doi.org/10.1016/j.techsoc.2024.102583

57. Zenezini, G., Marco, A. D., (2020). City logistics policy evaluation with system dynamics. *The 11th International Conference on City Logistics, Dubrovnik, Croatia, 12th - 14th June 2019*, *46*, 253–260. https://doi.org/10.1016/j.trpro.2020.03.188