

Mapping the Evaluation Landscape: A Systematic Review of Research Design, Instruments, and Methods in Public Policy Implementation Studies

Intan Suciati Daeng Baji, Sidik Pramono

Universitas Indonesia, Indonesia

Email: intan.daengbaji@gmail.com, sidik.pramono18@ui.ac.id

Abstract:

The need for evidence-based evaluation of public policy implementation has grown increasingly urgent amid the complexities of contemporary governance and demands for accountability. This research aims to identify and analyze the characteristics of public policy implementation evaluation research through a systematic literature review (SLR) approach. By analyzing 119 articles from Scopus-indexed journals, this study maps six main dimensions: publication trends, research design, evaluation topics, evaluation approaches, data collection instruments, and data analysis techniques. Findings reveal a dominance of qualitative approaches—particularly thematic analysis and interviews—as well as a concentration of topics in the health, education, and social welfare sectors. The study also highlights limitations in the use of advanced quantitative methods and the need to strengthen cross-sectoral evaluative instruments. The implications of this research underscore the urgency of enhancing evaluative capacity, diversifying methodological approaches, and institutionalizing evaluation within the policy cycle to strengthen adaptive and sustainable public governance.

Keywords: *Public Policy Implementation, Policy Evaluation Studies, Evaluation Instruments and Techniques*

Corresponding: Nama author

E-mail: Email author



INTRODUCTION

Systematic evaluation of public policies is essential to ensure accountable and effective governance, especially amid increasingly complex social, economic, and environmental challenges. Evaluation serves as a vital tool for assessing whether policies are implemented in accordance with their objectives and achieve their desired outcomes, thereby identifying strengths and weaknesses to improve governance and strengthen stakeholder support. The importance of evaluation is even more prominent in sensitive political contexts, where resource allocation is critical (Mendes & Sordi, 2013). Integrating evaluation into the policy cycle is crucial for promoting evidence-based policymaking, as it ensures that decisions are based on factual data and responsive to community needs. Thus, governance effectiveness can be improved in a more measurable and sustainable manner.

The concept of “evidence-based policy” emphasizes the need for rigorous evaluation practices to produce efficient policies, although in reality, the use of evidence still varies significantly among public institutions and policy arenas (Head, 2016). Therefore, although integrating evaluation into the policy cycle promises to improve responsiveness and accountability, there is an urgent need to address various barriers to evidence use and ensure that evaluation findings are effectively translated into actionable policy insights. Ultimately, research that adopts a systematic approach to evaluation will make a significant contribution to

developing policies that are not only effective but also fair and sustainable in responding to the complex challenges of contemporary governance.

The crucial role of evaluation has become increasingly clear amid global events that test the capacity of policy responses and their ability to adapt. The COVID-19 pandemic has underscored the need for adaptive and context-sensitive policy evaluation practices, highlighting the importance of integrating evaluation throughout the policy cycle to address complex global challenges such as climate change and social inequality. The pandemic has shown that governments can implement rapid evidence-based policies, setting precedents for future crises. However, traditional evaluation methods may not suffice in urgent situations, requiring a shift toward more flexible and transformative evaluation approaches. Evaluators must adapt to new realities without compromising their principles, as the pandemic has introduced non-traditional actors into the evaluation space, offering similar expertise and meeting evolving demands (Naidoo, 2022). Furthermore, evaluation should be viewed as a strategic resource for public management and social control, integrating qualitative and normative dimensions to enhance accountability and participation (Laisner & Mario, 2015). By embedding evaluation throughout the policy cycle, governments can ensure systematic responses to complex issues, promote accountability and learning to identify socially fair solutions, and encourage transformational approaches (Mavrot & Malandrino, 2023).

Recent developments in policy implementation evaluation studies highlight a dynamic landscape characterized by diverse methodological approaches and significant changes in research focus. The literature underscores the critical role of policy implementation in achieving sustainable development, yet many policies fail to meet their objectives due to inadequate theoretical frameworks and evaluation methods. In Portugal, for example, policy implementation studies are still evolving, with noticeable gaps in procedural approaches that are crucial for effective governance (Mota, 2019). In environmental policy, effectiveness evaluation faces conceptual, methodological, and empirical challenges—such as defining effectiveness and addressing questions of cause and effect—that are particularly important for policy evaluation at the European Union (EU) level (Gysen et al., 2002). Overall, these trends reflect a growing understanding of policy implementation informed by global perspectives and collaborative governance, emphasizing the need for comprehensive evaluation frameworks to improve policy outcomes. To understand the latest developments in policy implementation evaluation studies, a literature review is needed that not only describes the variety of approaches but also maps methodological trends and tendencies developing globally. Therefore, this study is organized as a *Systematic Literature Review (SLR)* that specifically analyzes scientific articles indexed in the Scopus database.

Previous studies have attempted to map aspects of policy evaluation, but they have often been limited to specific sectors or approaches. For example, health policy evaluation studies tend to focus on clinical indicators and outcomes, while environmental policy evaluations emphasize sustainability and compliance aspects. Meta-review or *systematic literature review (SLR)* research on the patterns and evaluation instruments of public policy implementation across sectors remains relatively limited. The majority of systematic studies focus more on substantive issues (such as the impact of a particular program) than on comprehensive mapping

of the research designs, measurement instruments, and analytical techniques widely used in this field. This gap demonstrates the need for a more holistic and structured review to identify patterns, methodological trends, and gaps in policy implementation evaluation practices.

This study aims to answer six main questions: (1) How has the number of policy implementation evaluation studies trended over time? (2) What research designs were used in these studies? (3) What topics were most frequently studied? (4) What interventions were used to improve evaluation? (5) What instruments were used to measure policy implementation? (6) What data analysis techniques were applied? This study makes a unique contribution by mapping six key dimensions in the study of public policy implementation evaluation: publication trends, research design, evaluation topics, evaluation approaches, data collection instruments, and data analysis techniques. Unlike previous reviews that tend to focus on sectoral issues or specific policies, this study presents a holistic, structured *SLR*-based approach to identify patterns, gaps, and methodological trends across sectors. As such, this research not only expands understanding of existing evaluative practices but also provides an empirical foundation that can strengthen evidence-based policy evaluation capacity more broadly.

RESEARCH METHOD

This study employed a Systematic Literature Review (SLR) approach to examine trends in the evaluation of public policy implementation based on scientific articles published in the Scopus database. The SLR approach was chosen because it enabled studies to be conducted in a structured, transparent, and replicable manner, ensuring that results were trustworthy and usable as academic and policy references. The entire process was carried out through stages of literature search using related keywords, article screening based on inclusion and exclusion criteria, content coding, and thematic analysis of the obtained data.

Scopus was selected as the single data source due to its status as the largest and most comprehensive bibliometric database globally, which ensured the credibility and scientific quality of the analyzed information. Scopus covers a wide range of high-reputation journals across disciplines, making it ideal for multidisciplinary policy evaluation (Moya-Anegón et al., 2007). The database's rigorous peer-review process further ensured the empirical relevance of indexed articles, maintaining high scientific standards (Moya-Anegón et al., 2007). Additionally, Scopus facilitated structured and systematic data mining, aligning with SLR principles, particularly in interdisciplinary fields (Moro & Rita, 2022).

Scopus's extensive coverage is evident in studies such as a bibliometric analysis of occupational therapy publications, which used Scopus to track trends and identify leading journals and countries (Sau & Nayak, 2023). Similarly, its comprehensive indexing is highlighted in research on urban waste management, providing a robust dataset for bibliometric analysis and SLRs across fields (Mayes-Ramírez et al., 2023). Overall, Scopus's broad coverage, strict indexing standards, and support for systematic data analysis made it indispensable for comprehensive literature reviews and policy evaluation.

This study analyzed articles from Scopus using the following filters. The first filter applied keywords on January 19, 2025: (("public policy implementation" OR "policy implementation" OR "implementation of public policies" OR "evaluation of public policies"

OR "policy evaluation" OR "implementation evaluation" OR "performance evaluation" OR "effectiveness evaluation" OR "impact evaluation") AND ("qualitative methods" OR "quantitative methods" OR "mixed methods" OR "case studies" OR "comparative policy analysis" OR "evaluation methods")), yielding 17,037 articles. These keywords combined terms such as “public policy implementation,” “policy implementation,” and “implementation of public policies” to capture variations in describing the policy implementation process. Meanwhile, terms like “policy evaluation,” “implementation evaluation,” “performance evaluation,” “effectiveness evaluation,” and “impact evaluation” ensured articles explicitly focused on evaluation.

To cover methodological diversity, terms such as “qualitative methods,” “quantitative methods,” and “mixed methods” were added, capturing studies aligned with policy research complexity. Terms like “case studies” and “comparative policy analysis” addressed single-case or cross-policy analyses common in implementation studies. Boolean operators (OR within groups, AND between groups) ensured broad yet focused results, enabling comprehensive searches across contexts while staying relevant. The second filter limited the time range to 2020–January 19, 2025, yielding 4,803 articles. This five-year limit ensured up-to-date, contextually relevant data reflecting recent developments in evaluation methods.

The third filter applied social sciences and decision sciences limits, yielding 1,249 articles, as these fields form the foundation for public policy evaluation in social, institutional, and analytical contexts. The fourth filter restricted document types to articles, yielding 1,064 articles, prioritizing peer-reviewed publications with complete methodological details over proceedings or reviews. The fifth filter (final publication stage) yielded 1,019 articles, ensuring complete, stable information. The sixth filter (keywords: Policy Implementation, Performance Evaluation, Performance Assessment) yielded 473 articles, focusing on process and performance evaluation while excluding formulation or normative topics. Including “performance evaluation” and “performance assessment” emphasized effectiveness, efficiency, success, or failure analyses, ensuring evaluative depth quantitatively and qualitatively. This improved screening accuracy, transparency, and alignment with research questions on trends, designs, and frameworks.

The seventh filter limited sources to scientific journals (still 473 articles), as they provide rigorous outlets for tested research. The eighth filter restricted language to English, yielding 461 articles, for international reach and terminological uniformity. The final filter selected open-access articles, yielding 253 articles, supporting open science principles of transparency and accessibility. Of these, 220 were downloadable; one was excluded as a case report, leaving 219 for analysis.

This study applied systematic screening with pre-established inclusion and exclusion criteria. Inclusion criteria covered articles explicitly addressing one or more research questions or providing relevant empirical/conceptual insights on public policy implementation evaluation. Studies on government programs were included, as Zahariadis et al. (2023) emphasized systematic approaches to national policy styles, inherently including programs, and Chriqui et al. (2023) highlighted implementation science for evidence-based policies achieving public health outcomes. Exclusion criteria eliminated articles unable to address research

questions or provide relevant information. Selection involved manual abstract review, followed by full-text analysis if needed. Of 219 articles, 119 met inclusion criteria and were analyzed.

The analytical instruments classified and extracted information systematically across six dimensions, each addressing a research question. Articles were manually coded from abstracts, methods, and discussion sections. These dimensions were: (1) Publication Year, tracking temporal distribution (2020–early 2025); (2) Research Design (e.g., case study, qualitative, quantitative, mixed methods, experimental, comparative), based on methodological statements; (3) Main Research Topic (e.g., education, health, environment, digital governance, public service), mapping frequent evaluation issues; (4) Treatment/Intervention, identifying strategies or programs improving implementation/evaluation; (5) Evaluation Measurement Instruments (e.g., surveys, interviews, document analysis, monitoring systems, official data); and (6) Data Analysis Techniques (e.g., regression, content analysis, statistical analysis, SEM, MCDA). Data were categorized by these dimensions to answer research questions and presented graphically.

RESULTS AND DISCUSSION

Analysis of the publication year dimension shows that the distribution of articles meeting the inclusion criteria in this systematic literature review is concentrated in the period from 2022 to 2025. No articles meeting the inclusion criteria for this study were found in 2020 and 2021. A total of 43 articles were published in 2022, followed by 40 articles in 2023, and 35 articles in 2024. Meanwhile, as of the last data collection period, only 1 article had been published in 2025. This distribution has several important implications. First, the absence of relevant articles meeting the inclusion criteria for this study in 2020 and 2021 reflects that an explicit focus on the evaluation of public policy implementation, with clear methodologies and open access, has not yet become mainstream in the Scopus scientific literature at the beginning of this decade. This can be linked to the global crisis caused by the COVID-19 pandemic, which, although it triggered a great need for policy reform and governance evaluation, has not yet been directly responded to in the form of scientific publications that meet the inclusion standards in this study (Korzeb & Niedziółka, 2020; Ozili, 2021; Song & Zhou, 2020). Second, the increase in the number of publications since 2022 marks an important turning point in public policy literature, with increased academic attention to data- and theory-based policy implementation evaluation (Arias et al., 2025; Cairney et al., 2021; van den Dool & Schlauffer, 2024). The year 2022 marks the beginning of the consolidation of literature examining the effectiveness, efficiency, and dynamics of policy implementation in various sectors. This also marks a shift in focus from normative evaluation to a more systematic and measurable evaluative approach. Third, the relatively stable trend between 2022 and 2024 reflects the beginning of the establishment of a policy implementation evaluation research community, with a significant number of articles published each year. Although there was only one article in 2025, this is reasonable considering that the year is still ongoing, and scientific publications require a long cycle of data collection, writing, review, and publication.

Overall, these findings justify the selection of the last five years (2020–2025) as the study period. Although not all years produced articles that met the criteria, this period allowed

the study to capture the latest dynamics, emerging trends, and methodological shifts in public policy implementation evaluation recorded in the Scopus scientific literature. These findings also provide an argumentative basis that evaluative public policy studies based on methods and theory are rapidly developing as a scientific field that is increasingly important in supporting accountable and adaptive governance.

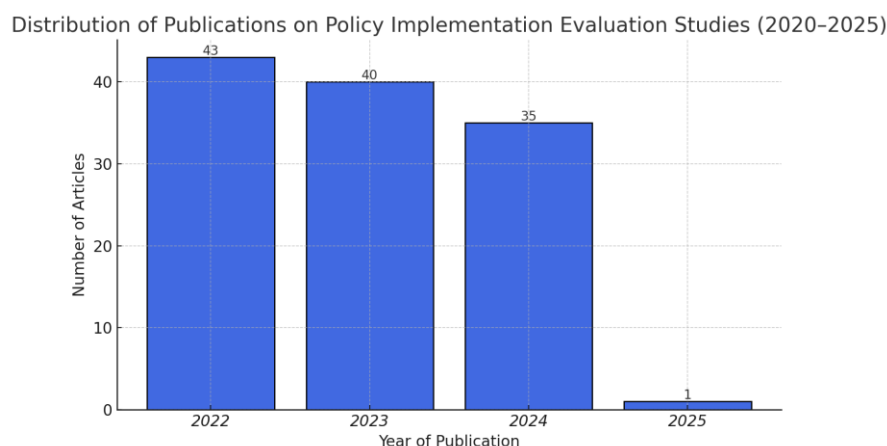


Figure 1 Trends in the Publication of Public Policy Implementation Evaluation Studies (2010–2023)

Source: Compiled by the authors based on 119 Scopus-indexed articles (2025).

The analysis shows that the most dominant approach used is Qualitative Case Study, which was found in 52 articles. This approach shows a strong preference among researchers for in-depth exploration in the context of specific policy implementation. Studies in this category generally use techniques such as document analysis, interviews with stakeholders, and institutional process tracing. This approach is well suited to capturing the complexity of policy implementation dynamics, power relations, and actor interactions that are often contextual and difficult to generalize. Mixed Methods research design was the second most widely used category, found in 36 articles. This approach reflects a methodological trend that is increasingly moving towards the integration of qualitative and quantitative techniques. Generally, studies in this category combine surveys with interviews or triangulate quantitative performance indicators with qualitative narratives from policy actors. This integration allows researchers to examine both the context and the outcomes of policies more comprehensively.

Quantitative design was used in 9 articles, indicating the existence of a statistics-based approach, although in relatively limited numbers. Studies in this category generally apply techniques such as regression analysis, structural modelling, or quantitative data-based impact evaluation. Although not dominant, this approach is important in producing findings that are generalizable and empirically testable. In addition, there are 7 articles classified as Qualitative and Case Study, both of which often overlap with qualitative case studies but are classified separately because they do not explicitly state an exploratory or integrative orientation. These studies typically use a one-dimensional approach without triangulation methods or in-depth contextual descriptions. Comparative design was found in 4 articles, which compared policy

implementation in different regions or institutions with the aim of identifying contextual or institutional variations. Although relatively few in number, this approach has strategic value in understanding differences in implementation across policy environments. The Experimental category (including quasi-experimental) appeared in only 2 articles, indicating that experimental and causal inference approaches are still rarely used in policy implementation evaluation studies, despite their strong potential for identifying policy impacts more accurately. Finally, 2 articles were categorized as Unclassified. These articles did not provide explicit information about the research design used or provided methodological descriptions that were too general and ambiguous to be classified into one of the design categories. The existence of this category reflects this study's commitment to methodological transparency and caution in classification. Thus, each article was included in the total analysis, but only categorized when sufficient information was available. Overall, this breakdown reinforces the finding that qualitative and context-based approaches still dominate the policy implementation evaluation literature. Although trends toward mixed methods and quantitative approaches are developing, the adoption of experimental, comparative, and systems-based designs remains very limited. Diversifying research designs in the future is essential to strengthen methodological robustness and enhance the transferability of findings in public policy evaluation studies.

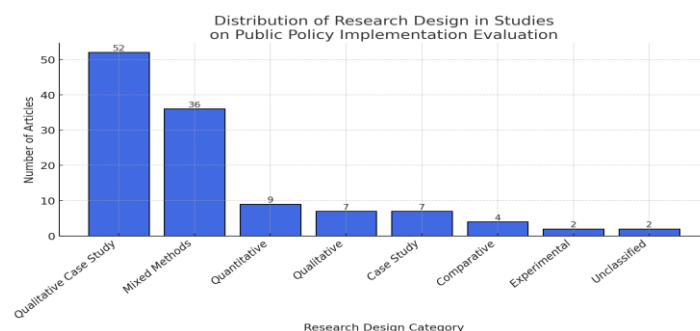


Figure 2 Distribution of Research Designs in Public Policy Implementation Evaluation Studies

Source: Authors' synthesis of systematically reviewed articles (2025).

The classification of 119 articles evaluating public policy implementation shows that the objects of evaluation are spread across ten thematic categories, reflecting the broad scope of public policy and the scientific response to sectoral and cross-sectoral dynamics. These findings confirm that policy evaluation today not only measures formal performance but also responds to value transformations, institutional structural changes, and the dynamics of public participation. The category of Environmental Policy and Climate Change is the dominant topic with 30 articles. These articles evaluate policies on natural resource management, climate change, and the transition to a green economy. Evaluations are conducted through collaborative approaches such as network governance and adaptive frameworks. This topic demonstrates the converge between global demands such as the Paris Agreement and evaluative practices that are cross-sectoral and contextual. Education Policy (17 articles) shows that evaluation in the education sector does not only focus on curriculum or accessibility, but also examines the gap between national design and local capacity. Evaluations are conducted using mixed methods to

produce a holistic understanding of the social and spatial context of education policy implementation.

The Health and Social Protection category (13 articles) assesses the sustainability of public policies in crisis situations, including the COVID-19 pandemic. These evaluations combine aspects of institutional capacity, the effectiveness of social safety networks, and the involvement of cross-sectoral actors in responding to the emergency needs of communities. A total of 9 articles are included in Local Participation and Community-Based Policy, which evaluates participatory mechanisms such as development planning forums, community forums, and community collaboration in policy implementation. This type of evaluation expands the focus from policy outcomes to the deliberate process as part of successful implementation. Infrastructure and Public Services (8 articles) evaluates the performance of public projects such as clean water, housing, and transportation. The evaluation was conducted using an indicator-based approach and performance reporting, emphasizing budget effectiveness and program realization. Employment and Labor Policy (6 articles) covers the evaluation of employment programs, vocational training, and labour productivity enhancement. The main focus is on the effectiveness of interventions in reducing unemployment and strengthening access to the formal labour market. Entrepreneurship and Economic Policy (4 articles) assesses policies for local economic empowerment such as support for MSMEs, cooperatives, and entrepreneurship initiatives. Evaluations in this category reflect the importance of community-based economic development and the role of policy in supporting grassroots economies. Although it consists of only 1 article, Social Integration and Marginalized Groups is classified separately due to its theoretical weight and policy significance. The issue of marginalization in public policy evaluation is complex and diverse, requiring different approaches depending on the nature of the situation.

Marginalization often arises from systemic exclusion based on social norms, resulting in significant losses for the affected groups. This is evident in various fields, such as ethnicity, where exclusion is reflected in access to the labour market, housing, and education systems that superficially promote inclusion but continue assimilation practices (Bracic, 2022). Therefore, maintaining marginalization as an independent category in policy evaluation is crucial to address specific and urgent social justice implications. In another context, the issues of Digital Governance and E-Government, although only reflected in one article, are maintained as separate categories due to their strategic relevance to bureaucratic reform and modern governance. Digitalization has significantly transformed public administration through increased efficiency, transparency, and citizen participation.

This transformation is reflected in the transition from traditional e-government to a more integrated digital governance model, which leverages big data and cloud computing to improve public service delivery and citizen interaction (Yukhno, 2022). Furthermore, the integration of user-generated content and social media technology in e-government platforms promotes public values through e-democracy, enabling citizens to actively participate and provide feedback on government policies and programs (Manoharan et al., 2022). Thus, combining these issues into the “Others” category would overlook significant developments in policy practice and new directions in public policy research. The “Others” category, which includes 30 articles, more

accurately represents highly contextual, cross-thematic, or hybrid evaluative issues that cannot be included in existing categories without sacrificing classification accuracy. This category reflects an inclusive coding approach to maintain thematic and methodological integrity while accommodating the complexity and diversity of non-sectoral public policy research..

These findings show that the evaluation of public policy implementation currently spans a very broad spectrum (from the environment, education, social, economic, to digital technology) with an increasingly specific, adaptive, and participatory approach. This thematic classification not only explains what is being evaluated, but also why and how evaluation is conducted in an increasingly complex and dynamic policy context. This narrative enriches the development of public administration science, strengthens cross-sectoral evaluation methodologies, and helps identify research gaps for the future policy evaluation agenda.

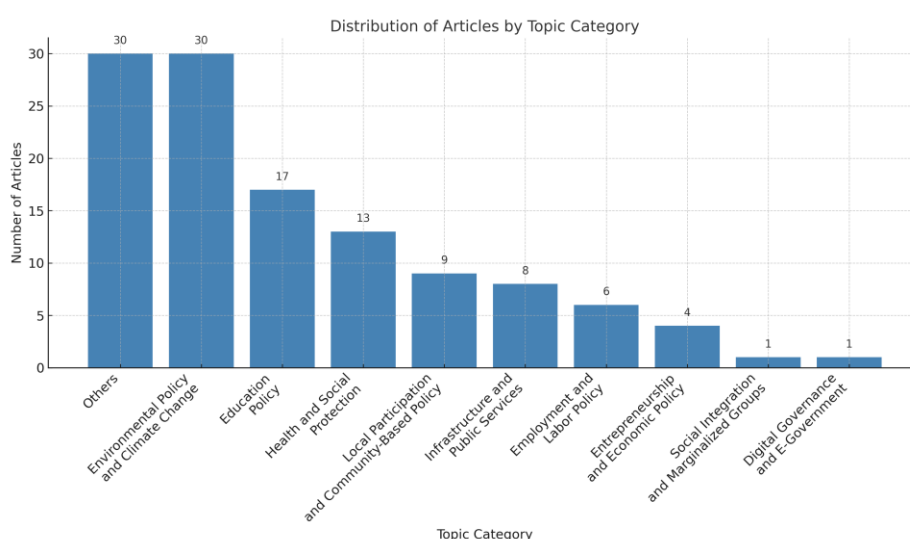


Figure 3 Evaluation Topics in Public Policy Implementation Studies

Source: Generated by the authors based on content classification of 119 Scopus articles (2025).

An analysis shows that there is no dominant evaluation model. Instead, there are 11 categories of intervention approaches that are used in various ways according to the policy focus, implementation context, and epistemological framework adopted by the researchers. This classification reflects the maturity of evaluative approaches that are adaptive, contextual, and continuously developing in public administration practice. The category with the most articles is Impact Assessment or Quantitative Evaluation, which includes 19 articles. Studies in this category use quantitative approaches such as regression, Difference-in-Differences (DiD), and other statistical estimation techniques to objectively measure policy impact. Articles such as “Policy and Fertility: A Case Study of the Quebec Family Policy Using Difference-in-Differences” are examples of the explicit use of this methodology.

The methodological contribution of this approach lies in its strength in producing evidence-based policy findings, strengthening the testing of causal relationships, and generating evaluative data that can be compared across regions and sectors. The second largest category is Policy Evaluation Framework Use, found in 18 articles. These studies explicitly adopt

evaluative models such as CIPP (Context, Input, Process, Product), BS 8001, or Sustainable Development Goals (SDGs) indicators as an analytical framework for evaluating policies. This framework is used to ensure that evaluations do not only focus on final outcomes but also consider fundamental elements in policy design and implementation. The advantage of this approach is that it provides consistency in the analytical structure, expands the transferability of evaluative methods across cases, and strengthens the integration between evaluation theory and policy practice. A total of 15 articles were included in the Digital Tools and Decision Support Systems category.

Studies in this category demonstrate the use of information technology such as spatial modelling systems GIS, interactive dashboards, digital data collection platforms, and data-driven decision support systems. Evaluations in this category mark the digital transformation in the implementation and monitoring of public policies. The integration between evaluative approaches and governance digitalization contributes to the efficiency, accountability, and responsiveness of increasingly complex public policy systems.

The Barrier Identification and Strategic Recommendations category consists of 12 articles. Studies in this category aim to map barriers that arise in the policy implementation process, whether structural, institutional, technical, or social, and propose relevant solutions based on field findings. This approach emphasizes the local context and dynamics of policy actors. Its contribution to evaluative research is to strengthen the problem-driven approach, which focuses on identifying the root causes of problems in public policy implementation and formulating strategic recommendations to improve its effectiveness. Capacity Development and Training was found in 8 articles. These studies highlight the importance of strengthening human and institutional capacity through training, mentoring, or competency enhancement programs as part of strategies to improve policy implementation success.

These interventions are particularly relevant in the context of decentralization and low-capacity regions, where the quality and readiness of implementers are key factors in policy success. A total of 8 articles also fall under the category of Community Participation and Engagement. Studies in this category emphasize the importance of active involvement of communities, local stakeholders, and vulnerable groups in policy planning, implementation, and evaluation. This approach contributes to the establishment of collaborative governance and more democratic and inclusive evaluation practices. Governance and Institutional Reform was recorded in 5 articles. The main focus of studies in this category is on redesigning institutions, reforming bureaucracy, and strengthening coordination mechanisms between actors and sectors in policy implementation. Evaluations in this category contribute to improving institutional arrangements and strengthening accountability systems in public policy implementation. There are 3 articles classified as Policy Adoption and Implementation Study, which evaluate how a policy is adopted and implemented in a local or sectoral context.

This approach is important for understanding the gap between policy design and its implementation in the field, as well as tracing the socio-political factors that influence the success of policy adoption. Institutional Roles and Actor Dynamics also appeared in 3 articles, focusing on policy actor interactions, leadership, and role distribution in implementation.

These studies expanded our understanding of power configurations, conflicts of interest, and coordination dynamics as determinants of policy implementation success or failure. Thematic/Issue-Specific Intervention Analysis includes 3 articles focusing on policies in specific thematic issues such as migration, gender, trade, or sports policy. Although these issues are sectoral and specific, the evaluative approach in this category opens up space for exploration of the unique characteristics of policy implementation in sectors that are often neglected in the mainstream literature. Finally, the Others category includes 25 articles that do not explicitly describe their evaluative approach or have intervention descriptions that are too general or hybrid to be classified into the other ten categories. The use of this category is not a methodological weakness but an analytical precaution to maintain the accuracy and integrity of the classification. Its existence also reflects the potential for developing new categories and the need for more explicit methodological clarification in future public policy evaluation practices.

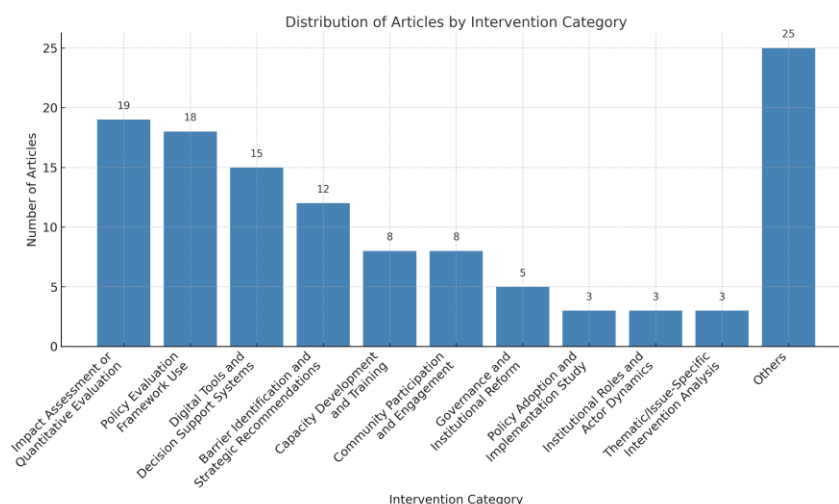


Figure 4 Treatment and intervention methods used in public policy implementation evaluation studies

Source: Authors' classification of intervention approaches based on thematic coding of selected literature (2025).

The Measurement Instrument dimension in public policy implementation evaluation studies refers to the tools, techniques, and data sources used by researchers to collect information, evaluate processes, and assess the level of policy implementation success. The selection of instruments is not just a technical aspect, but a fundamental methodological decision in determining the quality, validity, and depth of the analysis produced. Therefore, this dimension directly answers the fifth question in this systematic review, namely how the policy evaluation process is applied operationally in scientific practice. Of the 119 articles analysed, it was found that interviews were the most dominant instrument, used in 81 articles. Interviews—whether in-depth, semi-structured, or Focus Group Discussions (FGD)—were used to explore the subjective experiences, perceptions of implementing actors, and interpretations of policy dynamics at the local level. This dominance reflects the trend toward an actor-centred evaluation paradigm, in which in-depth exploration is prioritized over

quantitative measurement alone. A total of 14 articles used policy document analysis, which is a systematic review of official documents such as policy papers, strategic plans, implementing regulations, and program reports. This instrument is crucial in assessing the coherence between normative and implemented policies, as well as in tracing the continuity of policy logic from formulation to implementation.

Meanwhile, questionnaires were used in 12 articles as part of a survey approach or mixed methods. Questionnaires play an important role in capturing the perceptions or satisfaction levels of a broad target group and enable statistical analysis to identify trends, correlations, or differences between regions. However, their number is still far below that of interviews, indicating that many policy evaluation studies in this region place greater emphasis on qualitative exploration. Only three articles used official performance data, such as performance indicator reports from government agencies or program implementation statistics. The low utilization of this instrument indicates limitations in access to reliable and standardized administrative data. In fact, in results-based evaluation, credible administrative data is essential for objectively measuring policy effectiveness. This highlights structural challenges in public data governance, where reporting is often not integrated or difficult to access for scientific purposes. Nine articles were categorized into the Others group, which includes non-conventional or complex combined approaches such as field observation, policy system modelling, use of big data, and integration of spatial and administrative data. Some articles in this category do not explicitly mention the instruments used, but they can be identified through the description of alternative data collection methods or techniques employed. This approach reflects that in contemporary evaluation practices, mixed and experimental methods are increasingly used to address dynamic and multidimensional policy challenges.

During the classification process, some articles were found to use more than one instrument. To maintain systematic coherence and avoid double counting, each article was categorized into only one type of instrument that was considered most dominant. This principle is known as dominant instrument classification, where the selection of categories is based on the instrument that is most methodologically described, contributes the most to the main data in the analysis, or underlies the interpretation of the research results. This single classification strategy is in line with common practices in systematic literature reviews to ensure the accuracy of data aggregation. Considering all the above findings, it can be concluded that the evaluation of public policy implementation still heavily relies on perception-based data and policy actors' narratives. Meanwhile, quantitative and administrative data-based instruments remain underutilized. Therefore, future efforts should focus on strengthening the integration of policy information systems, enhancing technical capacity in data collection, and improving the accessibility of government performance data to support stronger evidence-based evaluation.

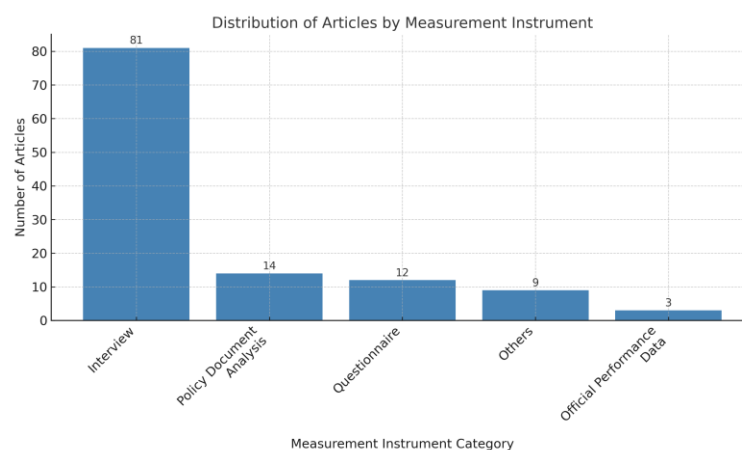


Figure 5 Data collection instruments used in public policy implementation evaluation

Source: Compiled by the authors using inductive coding of Scopus-indexed studies (2025).

Data analysis techniques are an important element in public policy implementation evaluation studies because they determine how data are interpreted and how evidence is developed to assess the success or failure of policy implementation. Within the framework of this systematic literature review, this dimension specifically addresses the sixth research question: What are the most frequently used analytical approaches in policy implementation evaluation studies, and how do they differ based on the research design applied? An analysis of 119 articles shows that the most dominant technique used is Thematic/Content Analysis, which is recorded in 82 articles. This method is well suited for evaluation studies that focus on contextual and narrative exploration, such as in research on local policy implementation. In qualitative research, thematic analysis is also useful because of its ability to handle data holistically and explore deep nuances, such as the experiences of patients and caregivers, decision-making processes, and the impact of interventions (Barrett & Twycross, 2018). Furthermore, Regression/Statistical Analysis was used in 14 articles, including linear regression, ANOVA, and correlation tests. These techniques are grouped together due to their numerical approach and ability to test relationships between variables quantitatively. Although direct citations are not used in this section, these techniques have become standard in quantitative data-based evaluation studies that emphasize hypothesis testing and statistical inference. Multi-Criteria Decision Analysis (MCDA) appears in 3 articles, particularly in the evaluation of technical policies such as environment and infrastructure.

MCDA is a powerful framework for decision-making in complex contexts involving diverse objectives and indicators. For example, Szabo et al. used MCDA with the Analytic Hierarchy Process (AHP) approach to evaluate economic, social, and environmental factors in sustainable development, which then produced a composite index for resource allocation and sustainability assessment (Szabo et al., 2021). In contrast, Structural Equation Modelling (SEM) was found in only one article and classified separately due to its technical complexity and theoretical model-based approach. SEM enables the modelling of latent variables and accounts for measurement errors, making it a powerful tool for comprehensive theory testing, as commonly used in marketing and information systems (Henseler, 2018). However, its

application in public policy evaluation remains limited due to technical constraints and the scarcity of available data. A total of 19 articles fall into the category of Others, covering analytical techniques that cannot be classified under a single category, such as agent-based modelling, network analysis, or combinations of techniques that are not explicitly described. The use of this category follows the principle of inclusive coding, which maintains flexibility without imposing unrepresentative classifications, especially in studies based on complex systems that rely on non-traditional approaches. Overall, the distribution of analytical techniques shows a dominance of narrative and qualitative approaches, while the use of advanced quantitative techniques remains limited. Going forward, there is a need to strengthen methodological capacity and integrate exploratory and inferential approaches. Diversifying analytical techniques not only broadens the scope of evaluation but also enhances the validity of policy recommendations—making them more evidence-based and sensitive to real-world complexities.

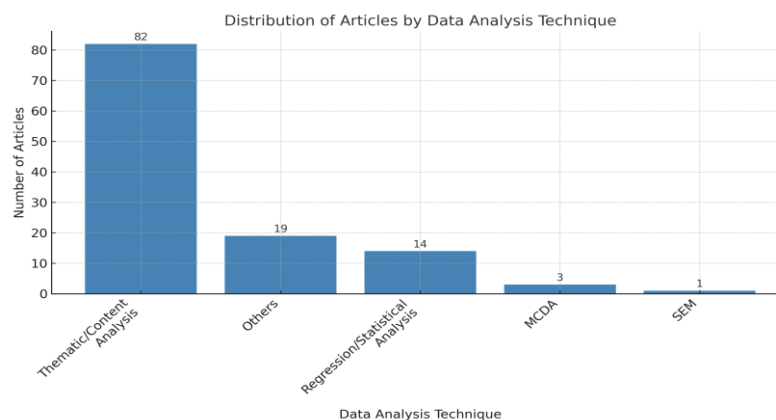


Figure 6 Data analysis techniques in public policy implementation evaluation studies

Source: Authors' systematic analysis of analytical methods used in the selected literature (2025).

This review confirms that public policy implementation evaluation practices are still dominated by contextual qualitative approaches, with a primary focus on the health, education, and social welfare sectors. Meanwhile, strategic topics such as environmental policy, infrastructure, and digital governance have not been evaluated extensively, indicating a disparity in attention to policy studies. Other limitations include the low use of standardized evaluation instruments and the minimal application of advanced quantitative analysis techniques. This hinders the accumulation of comparative knowledge across studies. On the other hand, the findings also reflect the importance of methodological flexibility in dealing with complex and dynamic policy contexts. The main contribution of this study lies in the systematic mapping of six dimensions of evaluation that provide an empirical foundation for the development of more structured, responsive, and evidence-based evaluative practices. Looking ahead, strengthening evaluation capacity, expanding the scope of policies evaluated, and integrating methodological approaches are critical steps to enhance the quality and effectiveness of evaluation in public governance. Although this study provides a systematic contribution to mapping public policy implementation evaluation practices, there are several

limitations that should be noted. First, the scope of the articles analysed is limited to Scopus-indexed journals that are available in open access or restricted to a specific search period, potentially missing relevant publications outside this scope. Second, the categorization coding is inductive based on the content of the articles, which may result in interpretations that are limited to the original context of the first author of the article. Third, this study does not explicitly conduct a methodological quality analysis (critical appraisal) of each article, which could be the focus of future studies.

CONCLUSION

This systematic review analyzed 119 articles on public policy implementation evaluation across six dimensions—publication trends, research design, evaluation topics, approaches (treatments), data collection instruments, and data analysis techniques—revealing a field dominated by qualitative methods (e.g., interviews and thematic analysis), with primary focus on health, education, and social welfare sectors while neglecting infrastructure, environment, and digital transformation. Conventional formative and summative frameworks prevailed, but advanced quantitative techniques, participatory approaches, and standardized instruments remained limited, hindering cross-study synthesis, inter-sectoral comparisons, and overall evaluative capacity due to methodological and institutional gaps. For future research, scholars should prioritize mixed-methods integration, expanded topic coverage for underrepresented areas, standardized cross-contextual instruments, public sector capacity-building via training and systems, and institutionalized evaluation within policy cycles through regulations and incentives to foster adaptive, evidence-based governance.

REFERENCE

- Arias, N., Marco-Fondevila, M., & Yetano, A. (2025). From theory to practice: A literature review of theoretical frameworks in public sector sustainability reporting. *Journal of Public Affairs*, 25(3), e70063.
- Barrett, D., & Twycross, A. (2018). Data collection in qualitative research. *Evidence-Based Nursing*, 21(3), 63–64. <https://doi.org/10.1136/EB-2018-102939>
- Bracic, A. (2022). Ethnicity and social exclusion. *Nationalities Papers*, 50(6), 1045–1056. <https://doi.org/10.1017/nps.2022.72>
- Cairney, P., St Denny, E., & Mitchell, H. (2021). The future of public health policymaking after COVID-19: A qualitative systematic review of lessons from Health in All Policies. *Open Research Europe*, 1, 23.
- Chriqui, J. F., Asada, Y., Smith, N. R., Kroll-Desrosiers, A. R., & Lemon, S. C. (2023). Advancing the science of policy implementation: A call to action for the implementation science field. *Translational Behavioral Medicine*. <https://doi.org/10.1093/tbm/ibad034>
- Head, B. (2016). Toward more “evidence-informed” policy making? *Public Administration Review*, 76(3), 472–484. <https://doi.org/10.1111/puar.12475>
- Henseler, J. (2018). Partial least squares path modeling: Quo vadis? *Quality & Quantity*, 52(1), 1–8. <https://doi.org/10.1007/s11135-018-0689-6>
- Korzeb, Z., & Niedziółka, P. (2020). Resistance of commercial banks to the crisis caused by the COVID-19 pandemic: The case of Poland. *Equilibrium: Quarterly Journal of Economics and Economic Policy*, 15(2), 205–234.
- Laisner, R. C., & De Mario, C. G. (2015). Os desafios da avaliação de políticas públicas como

- instrumento estratégico de gestão e de controle social. *Revista de Políticas Públicas*, 18(2), 619–630. <https://doi.org/10.18764/2178-2865.v18n2p619-630>
- Manoharan, A. P., Choi, Y., & Holzer, M. (2022). Digital governance in cities: An international comparison of e-government performance. *Government Information Quarterly*, 39(1), 101635. <https://doi.org/10.1016/j.giq.2021.101635>
- Mavrot, C., & Malandrino, A. (2023). Three years of COVID-19 pandemic: Coping with crisis governance in the long term. *European Policy Analysis*, 9(2), 96–100. <https://doi.org/10.1002/epa2.1175>
- Mayes-Ramírez, M. M., Gálvez-Sánchez, F. J., Ramos-Ridao, Á., & Molina-Moreno, V. (2023). Urban waste: Visualizing the academic literature through bibliometric analysis and systematic review. *Sustainability*, 15(3), 1846. <https://doi.org/10.3390/su15031846>
- Moro, S., & Rita, P. (2022). Data and text mining from online reviews: An automatic literature analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(3), e1448. <https://doi.org/10.1002/widm.1448>
- Mota, L. F. (2019). Estudos de implementação de políticas públicas: Uma revisão de literatura. *Revista do Serviço Público*, 92, 133–150. <https://doi.org/10.7458/spp20209211728>
- Naidoo, I. (2022). Evaluation for transformational change: Learning from practice (pp. 17–26). Springer. https://doi.org/10.1007/978-3-030-78853-7_2
- Ozili, P. K. (2021). COVID-19 pandemic and economic crisis: The Nigerian experience and structural causes. *Journal of Economic and Administrative Sciences*, 37(4), 401–418.
- Sau, K., & Nayak, Y. (2023). Scopus-based bibliometric and scientometric analysis of occupational therapy publications from 2001 to 2020. *F1000Research*, 11, 155. <https://doi.org/10.12688/f1000research.108772.2>
- Song, L., & Zhou, Y. (2020). The COVID-19 pandemic and its impact on the global economy: What does it take to turn crisis into opportunity? *China & World Economy*, 28(4), 1–25.
- Szabo, Z. K., Szádóczi, Z., Bozóki, S., Stănciulescu, G. C., & Szabo, D. (2021). An analytic hierarchy process approach for prioritisation of strategic objectives of sustainable development. *Sustainability*, 13(4), 2254. <https://doi.org/10.3390/su13042254>
- van den Dool, A., & Schlauffer, C. (2024). Policy process theories in autocracies: Key observations, explanatory power, and research priorities. *Review of Policy Research*, 41(6), 865–891.
- Yukhno, A. (2022). Digital transformation: Exploring big data governance in public administration. *Public Organization Review*, 22(4), 977–996. <https://doi.org/10.1007/s11115-022-00694-x>
- Zahariadis, N., Karokis-Mavrikos, V., Exadaktylos, T., Kyriakidis, A., Sparf, J., & Petridou, E. (2023). Advancing the operationalization of national policy styles. *European Policy Analysis*. <https://doi.org/10.1002/epa2.1172>