

Reassessing Effectiveness: A Critical Examination of Evaluation Strategies in International Environmental Agreements

Sihap^{1*}, Bambang W. Nugroho², Rangga A. Putra³

^{1*,2,3}Universitas Muhammadiyah Yogyakarta

Corresponding Author's e-mail : sihap.isip23@mail.umy.ac.id



e-ISSN: 2964-0962

SEIKAT: Jurnal Ilmu Sosial, Politik dan Hukum

<https://ejournal.45mataram.ac.id/index.php/seikat>

Vol. 5, No. 4, Agustus 2026

Page: 2013 - 2025

Available at:

<https://ejournal.45mataram.ac.id/index.php/seikat/article/view/3243>

DOI:

<https://doi.org/10.55681/seikat.v5i4.3243>

Article History:

Received: 06-07-2026

Revised: 11-08-2026

Accepted: 12-08-2026

Abstract : The increasing number of International Environmental Agreements (PIIs) has prompted the development of various efforts to assess the effectiveness of global environmental governance. Most research focuses on measuring agreement success through indicators of performance, compliance, and observable environmental achievements. However, academic attention to evaluation strategies used to define and measure effectiveness is still relatively limited. As a result, the epistemological assumptions and political implications inherent in evaluation practice often go unnoticed. This article aims to critically analyze the various strategies for assessing the effectiveness of International Environmental Agreements and examine how different evaluation paradigms result in different understandings of effectiveness. The research uses a qualitative approach with a meta-methodological analysis design supported by document analysis of evaluation reports, policy documents, and academic publications, with the Montreal Protocol and the Convention on Biological Diversity (CBD) used as illustrative cases. The results of the study identified three main paradigms in effectiveness evaluation, namely rationalist-positivistic, constructivist-processual, and critical-reflective. Each paradigm prioritizes different forms of evidence, indicators of success, and governance priorities, resulting in an effectiveness narrative that is not always in line. These findings suggest that effectiveness cannot be understood as a fully objective and neutral condition, but rather as a construct formed through methodological choices, epistemological assumptions, and power relations inherent in evaluation practice. As a conceptual contribution, this article offers a Reflexive Evaluation Framework that integrates the dimensions of environmental outcomes, institutional learning, and justice and transformation to produce an evaluation approach that is more reflective, multidimensional, and sensitive to the complexities of global environmental governance.

Keywords : International Environmental Agreements; the effectiveness of the regime; Global Environmental Governance; Reflexive Evaluation Framework; Politics of Evaluation.

Abstrak : Meningkatnya jumlah Perjanjian Lingkungan Hidup Internasional (PLI) telah mendorong pengembangan berbagai upaya untuk menilai efektivitas tata kelola lingkungan hidup global. Sebagian besar penelitian berfokus pada pengukuran keberhasilan perjanjian melalui indikator kinerja, kepatuhan, dan capaian lingkungan yang dapat diamati. Namun, perhatian akademis terhadap strategi evaluasi yang digunakan untuk mendefinisikan dan mengukur efektivitas masih relatif terbatas. Akibatnya, asumsi epistemologis dan implikasi politis yang melekat dalam praktik evaluasi sering kali luput dari perhatian. Artikel ini bertujuan untuk menganalisis secara kritis berbagai strategi penilaian efektivitas PLI serta

mengkaji bagaimana paradigma evaluasi yang berbeda menghasilkan pemahaman yang berbeda pula mengenai efektivitas. Penelitian ini menggunakan pendekatan kualitatif dengan desain analisis meta-metodologis yang didukung oleh analisis dokumen terhadap laporan evaluasi, dokumen kebijakan, dan publikasi akademis, dengan Protokol Montreal dan Konvensi Keanekaragaman Hayati (CBD) sebagai studi kasus ilustratif. Hasil penelitian mengidentifikasi tiga paradigma utama dalam evaluasi efektivitas, yaitu rasionalis-positivistik, konstruktivis-prosesual, dan kritis-reflektif. Setiap paradigma mengutamakan bentuk bukti, indikator keberhasilan, dan prioritas tata kelola yang berbeda, sehingga menghasilkan narasi efektivitas yang tidak selalu selaras. Temuan ini menunjukkan bahwa efektivitas tidak dapat dipahami sebagai kondisi yang sepenuhnya objektif dan netral, melainkan sebagai sebuah konstruksi yang terbentuk melalui pilihan metodologis, asumsi epistemologis, dan relasi kuasa yang melekat dalam praktik evaluasi. Sebagai kontribusi konseptual, artikel ini menawarkan Kerangka Kerja Evaluasi Refleksif yang mengintegrasikan dimensi hasil lingkungan, pembelajaran kelembagaan, serta keadilan dan transformasi, guna menghasilkan pendekatan evaluasi yang lebih reflektif, multidimensi, dan peka terhadap kompleksitas tata kelola lingkungan hidup global.

Kata kunci : Perjanjian Lingkungan Hidup Internasional; efektivitas rezim; Tata Kelola Lingkungan Hidup Global; Kerangka Kerja Evaluasi Refleksif; Politik Evaluasi.

INTRODUCTION

Over the past few decades, global environmental governance has undergone rapid institutional development, marked by the birth of hundreds of International Environmental Agreements (PIIs) that regulate various issues such as ozone depletion, climate change, biodiversity conservation, and hazardous waste management (Young, 2011). These agreements are the main instruments for countries to coordinate collective responses to cross-border environmental problems that cannot be resolved unilaterally. As the complexity and urgency of the global environmental crisis increases, the question of the extent to which these agreements are truly capable of achieving environmental improvements is becoming increasingly important for policymakers, international organizations, and the academic community. However, the effort to assess the effectiveness of international environmental agreements is not simple. Changes in environmental conditions are often influenced by a variety of interacting factors, such as technological innovations, market dynamics, domestic policies, and broader social and economic changes, making it difficult to isolate the specific contribution of an agreement to a particular environmental outcome. In addition, the objectives carried out by each agreement are also diverse, ranging from the prevention of environmental damage, sustainable resource management, to ecosystem recovery, each of which has different indicators and evaluation timeframes. This complexity is increasing because various international regimes interact with each other, so the effectiveness of an environmental agreement is often influenced by dynamics in trade, investment, development, and human rights regimes (Ward, 2006). Therefore, determining whether an International Environmental Agreement can be categorized as "effective" is a much more complex issue than simply observing changes in environmental conditions (Helm & Sprinz, 2000; Young, 2011).

To answer these challenges, various studies have developed analytical frameworks to assess the effectiveness of International Environmental Agreements. One of the most influential approaches is the output-outcome-impact model, which distinguishes between the formation of rules and institutions (output), changes in the behavior of the actors who are the target of the agreement (outcome), and the real impact on environmental conditions (impact) (Helm & Sprinz, 2000; Young, 2001). In this tradition, effectiveness is generally judged on the extent to which an agreement succeeds in achieving a set goal, changing the country's behavior, or resulting in measurable environmental improvements. The study on governance effectiveness then expands the focus by highlighting factors such as institutional design, compliance mechanisms, implementation

capacity, funding arrangements, and the quality of coordination between actors (Breitmeier et al., 2011). Although there are methodological variations among the various studies, most of these approaches still depart from the assumption that effectiveness can be identified through indicators that can be systematically observed, measured, and compared. Consequently, effectiveness evaluation is increasingly understood as a performance measurement activity that is oriented towards achieving targets and empirically verified results. This tradition has resulted in an increasingly sophisticated and influential set of evaluation instruments in global environmental governance practices. However, the dominance of this approach has also shaped a tendency to view effectiveness primarily as a matter of outcome, so that other dimensions that are more difficult to measure have the potential to be outside the evaluation space that is considered legitimate (Young, 2011).

Although the literature on the effectiveness of International Environmental Agreements is growing rapidly, academic attention is still dominated by attempts to determine the success or failure rate of an agreement based on a specific set of indicators (Helm & Sprinz, 2000; Young, 2011). Most of the research focuses on improving measurement methods, identifying factors that affect effectiveness, and developing evaluation instruments that are considered increasingly accurate and objective. However, the strong focus on the results of the evaluation leaves a relatively under-appreciated question, namely how the assessment strategy itself shapes the understanding of what is called "effectiveness". In other words, the existing literature tends to debate the degree of effectiveness of an agreement without adequately reflecting the epistemological assumptions underlying its assessment process. As a result, evaluation is often treated as a neutral technical procedure, while the political dimensions inherent in indicator selection, the determination of success standards, and the production and legitimacy of evaluative knowledge become less visible (Cox, 1981; Foucault, 1980). In fact, different assessment strategies can result in different definitions of effectiveness, as well as determine which aspects of governance are considered important, who benefits from the results of the evaluation, and what forms of change are recognized as successes. This limitation suggests that the main problem lies not only in how effectiveness is measured, but also in how effectiveness is constructed through the practice of evaluation itself. When evaluation strategies are accepted as a neutral instrument, certain dimensions of effectiveness risk being marginalized from research and policy agendas, while other dimensions gain greater legitimacy because they are easier to measure and compare. As a result, the understanding of the success of environmental governance can become narrower than the complexity of reality that it is actually trying to explain. Thus, the evaluation process itself is still an aspect that receives less attention in the study of the effectiveness of International Environmental Agreements.

Based on these gaps, this article aims to critically analyze the various strategies used in assessing the effectiveness of International Environmental Agreements. Rather than questioning whether or not an agreement is effective, this study focuses on how different evaluation approaches shape, interpret, and prioritize different meanings of effectiveness. To achieve this goal, this research utilizes the perspective of the Sociology of Knowledge which views that knowledge is not solely found objectively, but is formed through certain social processes (Berger & Luckmann, 1966). Perspektif tersebut dipadukan dengan Teori Kritis yang menyoroti hubungan antara pengetahuan, kekuasaan, dan legitimasi dalam praktik tata kelola global (Cox, 1981; Habermas, 1971). In addition, this study also adopts the concept of Dynamic Effectiveness which views effectiveness as a process that continues to evolve through learning, adaptation, and institutional transformation, not just the end result that can be measured at a specific point in time (Young, 2011, 2018). The

combination of these three perspectives provides an analytical foundation for examining the assumptions and implications inherent in various effectiveness assessment strategies.

This article argues that effectiveness assessment is not a completely neutral technical process, but rather a practice shaped by certain epistemological and political assumptions (Cox, 1981; Foucault, 1980). Each evaluation paradigm produces a different definition of what is considered a success, failure, evidence, or value change, allowing for different narratives of effectiveness to emerge of the same agreement. In particular, the dominance of outcome-oriented evaluation approaches has narrowed the understanding of effectiveness by placing measurable environmental outcomes as key indicators of success, while other dimensions such as institutional learning, procedural and distributive justice, and discursive change are often overlooked (Young, 2001, 2011). As a result, evaluation practices have the potential to legitimize certain forms of governance while overriding alternative understandings of environmental success. By critically comparing various valuation paradigms and examining their implications, this article contributes in the form of an epistemological critique of evaluation practices in international environmental governance. In addition, this study also offers a reflective evaluation framework that integrates the dimensions of environmental outcomes, institutional learning, and justice and transformation as a basis for understanding effectiveness more comprehensively.

METHODOLOGY

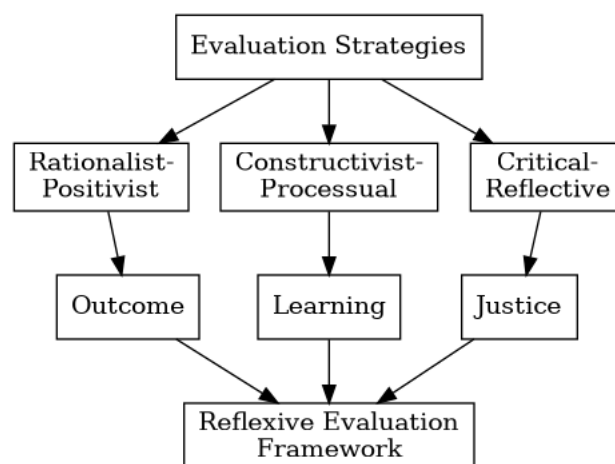
This study uses a qualitative approach with a meta-methodological analysis design to examine how the effectiveness of International Environmental Agreements is constructed through different evaluation strategies. In contrast to conventional effectiveness research that seeks to measure the success rate of an agreement based on certain indicators, this research focuses on the assumptions, logic, and implications that underlie the assessment process itself. Thus, the main unit of analysis of the study is not the performance of international environmental agreements, but rather the evaluative approach used to assess such performance. This approach was chosen because it allows for a critical exploration of how various assessment strategies define the concept of effectiveness, determine the forms of evidence that are considered valid, and establish standards of success used in global environmental governance practices (Creswell & Poth, 2018; Yanow & Schwartz-Shea, 2014). To clarify how evaluation strategies work in an empirical context, this study utilizes comparative illustrative cases, i.e. the use of cases that function as analytical illustrations to show the differences in perspectives produced by different evaluation paradigms (Gerring, 2007). Through this design, the research seeks to reveal that effectiveness is not only an objectively observable outcome, but also a product of methodological and epistemological choices that shape the evaluation process.

This study relies on secondary data from three main source groups. First, assessment reports, which are official assessment reports issued by international organizations, treaty secretariats, and scientific bodies involved in the process of evaluating the implementation and achievement of international environmental agreements. Second, policy reports that contain indicators, policy recommendations, and monitoring and evaluation frameworks used in global environmental governance. Third, academic evaluations, namely scientific articles and publications that assess the effectiveness of international environmental agreements using various methodological approaches. To illustrate how evaluation strategies produce different interpretations of effectiveness, this study uses two cases that are often placed in contrasting positions in the literature, namely the Montreal Protocol and the Convention on Biological Diversity. The Montreal Protocol is widely seen as one of the most successful examples of international environmental cooperation because it is able to demonstrate relatively clear and measurable progress towards the reduction of ozone-depleting

substances (Andersen & Sarma, 2002; Young, 2011). In contrast, the Convention on Biological Diversity is often categorized as a case with a more debated level of effectiveness due to the complexity of targets, indicators, and expected outcomes (Secretariat of the Convention on Biological Diversity, 2020). The selection of the two cases is not intended to compare the substantive successes of each agreement, but rather to show how the narrative of "success" and "failure" can be formed through the use of different evaluation strategies. Thus, both cases serve as empirical anchors that help uncover the characteristics and implications of various effectiveness assessment paradigms.

The analysis was carried out through four interrelated stages to identify and compare effectiveness assessment strategies in the International Environmental Agreement. The first stage is the identification of evaluation strategies, which is to examine various assessment reports, policy documents, and academic publications to identify indicators, criteria, and evaluative logic used in assessing the effectiveness of an agreement. At this stage, attention is focused on how effectiveness is defined, the types of evidence used, as well as the relationships built between governance and environmental outcomes. The second stage is paradigm classification, which is to group the various evaluation strategies found into three main paradigms: rationalist-positivistic, constructivist-processual, and critical-reflective. This grouping is carried out based on the epistemological assumptions that underlie each approach, especially related to how to understand knowledge, evidence, and governance success. The third stage is a Critical Discourse Analysis of the selected documents to reveal how language, categories, indicators, and knowledge claims are used in shaping a particular narrative of effectiveness. Through this analysis, the research seeks to explore assumptions that are often not explicitly stated but affect the way effectiveness is understood and evaluated. The fourth stage is a cross-paradigm comparison that aims to identify the differences, common points, and implications of each assessment strategy on understanding the effectiveness of the International Environmental Agreement. This comparison allows research to examine how different paradigms can produce different interpretations of the same agreement, as well as the basis for formulating a more reflective and multidimensional approach to evaluation. In the classification process, the rationalist-positivistic paradigm is identified through an emphasis on measurable indicators, target achievement, and causal relationships between policy interventions and environmental outcomes. The constructivist-processual paradigm is recognized for its focus on social learning, norm diffusion, capacity building, and the dynamics of institutional processes that develop over time. Meanwhile, the critical-reflective paradigm is characterized by attention to power relations, procedural and distributive justice, exclusion of certain actors, and the political implications arising from governance practices and environmental evaluations (Fairclough, 2013; Yanow & Schwartz-Shea, 2014).

Figure 1. Analytical Structure of Evaluation Paradigms



Source: Processed by the author, (2026)

RESULTS AND DISCUSSION

Rationalist–Positivist Evaluation

In the rationalist-positivistic paradigm, the effectiveness of the International Environmental Agreement is assessed primarily through indicators that can be observed, measured, and empirically verified. This approach places the achievement of results as the main basis in determining the success of an environmental regime (Helm & Sprinz, 2000; Young, 2011). In the case of the Montreal Protocol, the evaluation of effectiveness is largely based on relatively clear quantitative indicators, such as the decrease in the production and consumption of Ozone-Depleting Substances (ODS), the level of state compliance with the ODS phase-out schedule, as well as the development of ozone layer recovery that can be monitored through atmospheric observations and satellite technology (Gonzalez et al., 2015; Pyle et al., 2022). The consistent availability of data allows the relationship between international policy and environmental change to be traced more systematically. In contrast, the application of the same approach to the Convention on Biological Diversity (CBD) faces much greater challenges. Unlike the ozone layer which has relatively standardized biophysical indicators, biodiversity encompasses a wide range of dimensions that are difficult to reduce to a single size (Secretariat of the Convention on Biological Diversity, 2020). Assessments of CBD's effectiveness rely on diverse indicators, such as species loss rates, ecosystem conditions, conservation area area, and national capacity in biodiversity management, which often indicate non-uniform trends. Data limitations, variations in measurement methodologies between countries, and the complexity of the relationship between policies and ecological change make evaluating the effectiveness of CBD much more difficult than the Montreal Protocol. In this context, the rationalist-positivistic paradigm shows how the availability of measurable indicators plays an important role in shaping perceptions of the success or failure of an international environmental agreement.

The dominance of the rationalist-positivistic paradigm in the study of the effectiveness of the International Environmental Agreement is inseparable from its ability to provide evaluation standards that are considered objective, systematic, and replicable (Helm & Sprinz, 2000; Young, 2001). Relying on quantitative indicators and observable causal relationships, this approach offers a solid basis for policymakers and international organizations to assess whether an intervention results in the desired change. This advantage makes this paradigm very influential in monitoring, reporting, and evaluation (M&E) practices because it is able to produce effectiveness claims that look more credible and easy to communicate to various stakeholders. However, this power is also

the source of its limitations. The high reliance on measurable indicators drives a tendency to prioritize aspects of governance that can be converted into quantitative data, while other dimensions that are more difficult to measure tend to receive less attention. In this context, effectiveness is often reduced to the issue of achieving targets and changing indicators, even though the social process, institutional learning, or political transformation that occurs may be of equal significance (Young, 2011). This paradigm also tends to favor agreements that have clear objectives, available substitution technologies, and relatively stable indicators, as seen in the Montreal Protocol (Gonzalez et al., 2015). On the other hand, issues that have a multidimensional character and are difficult to measure, such as biodiversity, are at risk of being perceived as less effective not only because of governance failures, but because of the limitations of the evaluation instruments used. Therefore, the rationalist-positivistic paradigm not only provides a tool for assessing effectiveness, but also indirectly determines the types of effectiveness that can be recognized and recognized in international environmental governance practices.

Constructivist–Processual Evaluation

In contrast to the rationalist-positivistic paradigm that focuses on achieving measurable results, the constructivist-processual paradigm views effectiveness as a social process that develops through interaction, learning, and normative change among the actors involved (Haas, 1992; Risse et al., 2013). In this perspective, effectiveness is not solely determined by the success of achieving specific environmental targets, but also by the ability of an agreement to create mechanisms that encourage long-term cooperation and collective behavior change. In the context of the Convention on Biological Diversity (CBD), for example, effectiveness is often attributed to its ability to build national capacity through the development of National Biodiversity Strategies and Action Plans (NBSAPs), expand knowledge exchange between governments and the scientific community, and strengthen cooperation networks that support biodiversity management (Secretariat of the Convention on Biological Diversity, 2020). In addition, CBD also plays a role in disseminating conservation norms and the sustainable use of biodiversity into the national policy agendas of various countries. A similar phenomenon can also be found in the Montreal Protocol which not only resulted in a reduction in the use of the SDGs, but also created a shared learning mechanism that encourages the dissemination of technological innovations and compliance practices among participating countries (Andersen & Sarma, 2002). From a constructivist-procedural point of view, the process of capacity building, knowledge exchange, norm diffusion, and institutional learning is an important indicator that shows that the effectiveness of an agreement can be realized through social and institutional transformation that takes place gradually.

The constructivist-processual paradigm offers a broader understanding of effectiveness by shifting the focus from the achievement of results to the process of enabling those outcomes to be formed (Biermann & Pattberg, 2008; Haas, 1992). In this approach, effectiveness is seen as a dynamic phenomenon, where behavior change, collective learning, and institutional adaptation are integral to the success of international environmental governance. Treaties serve not only as a set of rules that govern state actions, but also as an arena that facilitates the exchange of knowledge, the formation of a common identity, and the internalization of norms that can influence policies at different levels of government. Thus, the success of a regime is not solely measured by the environmental achievements seen at a given point in time, but also by its ability to build capacity to respond to new challenges and sustain long-term cooperation. This perspective allows different forms of change that were previously difficult to recognize by results-based approaches to become part of effectiveness evaluation. However, the expansion of the meaning of effectiveness also presents its own challenges. A strong emphasis on the learning and adaptation process has the

potential to make evaluations too permissive of real ecological failures. Under certain conditions, the existence of a dialogue forum, capacity building, or the development of a cooperative network can be seen as an indicator of success even if the environmental conditions that are the main target of the agreement do not show significant improvement. As a result, this paradigm risks blurring the distinction between productive processes and substantive outcomes, so that effectiveness evaluations can shift from questions about what has been achieved to simply acknowledging that the process of cooperation is ongoing. Therefore, although the constructivist-processual paradigm has succeeded in broadening the understanding of effectiveness, it still faces challenges in linking social and institutional processes to environmental change that is the main goal of international governance (Young, 2011).

Critical–Reflective Evaluation

In contrast to the previous two paradigms that focused on results and processes, the critical-reflective paradigm directs attention to the power relations and socio-political structures that shape international environmental governance practices (Bernstein, 2005; Newell, 2008). In this perspective, effectiveness is not only understood through environmental achievements or institutional learning, but also through the political consequences and distribution of benefits generated by an agreement. In the case of the Montreal Protocol, a number of studies have shown that its success in reducing the use of SDGs is not entirely independent of the inequality of economic and technological capacity between developed and developing countries (Andersen & Sarma, 2002). Although funding and technology transfer mechanisms have been provided, the transition process to alternative technologies still creates a certain level of dependence on technology manufacturers that are mostly in developed countries. Meanwhile, in the context of the Convention on Biological Diversity (CBD), critical attention is often directed to the limited participation of indigenous peoples and local communities in decision-making processes that directly affect the territories and resources they manage (Zurba & Papadopoulos, 2021). In addition, various conservation targets often face structural economic constraints that encourage the exploitation of natural resources as part of national development strategies (Okereke, 2008). Under such conditions, the success of an agreement is influenced not only by institutional design or compliance levels, but also by economic inequality, access to technology, political representation, and the broader structure of global development. Therefore, the critical-reflective paradigm places the issues of North–South inequality, technology dependence, indigenous participation, and structural economic constraints as important elements in understanding the effectiveness of International Environmental Agreements.

The critical-reflective paradigm departs from questions that are fundamentally different from the rationalist and constructivist paradigms, namely effective for whom and effective according to whose standards (Bernstein, 2005; Schlosberg, 2004). The question shifts the focus of evaluation from just achieving results or the sustainability of the process to the power relationship that determines how success is defined and measured. In this perspective, effectiveness is never completely neutral because it is always linked to actors who have a greater ability to set the agenda, set indicators, and produce knowledge that is considered legitimate in global environmental governance (Cox, 1981; The question shifts the focus of evaluation from just achieving results or the sustainability of the process to the power relationship that determines how success is defined and measured. In this perspective, effectiveness is never completely neutral because it is always linked to actors who have a greater ability to set the agenda, set indicators, and produce knowledge that is considered legitimate in global environmental governance Foucault, 1980). Thus, the issue of power becomes a central element in evaluation because it determines who benefits from a

regime, who bears the cost of its implementation, and whose voice is accommodated in the decision-making process. From this point of view, environmental success achieved through an exclusive mechanism or resulting in an unequal distribution of benefits cannot be separated from the issue of justice, both in the form of procedural justice and distributive justice (Schlosberg, 2004). In addition, the critical-reflective paradigm also questions the legitimacy of claims of effectiveness generated by international institutions when the standards of success used do not fully represent the experiences of the group. Thus, the issue of power becomes a central element in evaluation because it determines who benefits from a regime, who bears the cost of its implementation, and whose voice is accommodated in the decision-making process. From this point of view, environmental success achieved through an exclusive mechanism or resulting in an unequal distribution of benefits cannot be separated from the issue of justice, both in the form of procedural justice and distributive justice directly affected (Bernstein, 2005). Therefore, evaluation is no longer understood as an attempt to measure how well an agreement works, but rather as a process that examines how claims of success are constructed, who gains authority to define them, and what political consequences arise from those definitions. Although often considered too normative and less operational than other paradigms, the critical-reflective approach opens up space to look at dimensions of effectiveness that are often overlooked in results-based and process-based evaluations, especially as they relate to inequalities, representations, and structural transformations in international environmental governance.

The Politics of Measuring Effectiveness

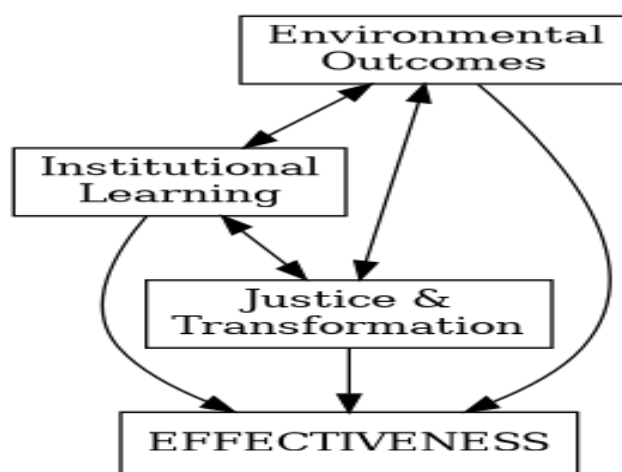
Comparisons between the rationalist-positivistic, constructivist-processual, and critical-reflective paradigms show that effectiveness is never present as a completely objective fact and independent of the evaluation process used to identify it. What is referred to as effectiveness always depends on the indicators chosen, the type of evidence that is considered valid, as well as the assumptions underlying the measurement process. Thus, indicators cannot be understood as completely neutral instruments because each indicator implicitly reflects decisions regarding what aspects are considered important to observe and what aspects can be ignored (Fukuda-Parr & McNeill, 2019; Merry, 2011). In this context, data is not merely a passive representation of reality, but rather the product of conceptual, methodological, and political choices that determine how a phenomenon is categorized, measured, and interpreted (Sihap et al., 2026). In this context, data is not merely a passive representation of reality, but rather the product of conceptual, methodological, and political choices that determine how a phenomenon is categorized, measured, and interpreted (Merry, 2011). When success is defined through the achievement of environmental indicators, environmental outcomes become the focus of the evaluation; When institutional learning is made the main focus, adaptive capacity and changing norms gain a more prominent position; And when justice is the basis for evaluation, then the distribution of benefits, political representation, and power relations become an inseparable part of the assessment of effectiveness. Therefore, evaluation is not only a mechanism for assessing environmental governance, but also an instrument that shapes governance priorities themselves (evaluation shapes governance priorities) (Cox, 1981; Fukuda-Parr & McNeill, 2019). By determining what counts as success and what is considered relevant to measure, evaluation practices directly influence policy direction, resource allocation, as well as prioritized forms of change in international environmental governance.

Toward a Reflexive Evaluation Framework

Based on the limitations and contributions of each paradigm, this article proposes a Reflexive Evaluation Framework as a more comprehensive approach to assessing the effectiveness

of International Environmental Agreements. This framework departs from the assumption that effectiveness cannot be reduced to a single dimension, but must be understood as an interaction between environmental outcomes, institutional learning processes, and dimensions of justice and transformation (Bernstein, 2005; Young, 2011). The first dimension, Environmental Outcomes, maintains the importance of measuring changes in environmental conditions as a substantive indicator of the success of an agreement. The second dimension, Institutional Learning, emphasizes the capacity of actors and institutions to adapt, exchange knowledge, build common norms, and respond to new challenges that arise in the implementation process (Haas, 1992). Meanwhile, the third dimension, Justice and Transformation, focuses on the distribution of benefits and burdens, the quality of stakeholder participation, and the extent to which environmental governance is able to drive more inclusive and sustainable change (Okereke, 2008; Schlosberg, 2004). These three dimensions are not positioned as mutually reinforcing indicators, but as complementary components in understanding effectiveness as a whole. The framework does not view effectiveness as a condition that can be determined through a single measure, but rather as the result of a dynamic interaction between environmental achievement, institutional learning capacity, and socio-political legitimacy that underpin the long-term sustainability of environmental governance. Positive environmental outcomes will be difficult to maintain without adequate learning capacity and institutional adaptation, while environmental and institutional success risks losing legitimacy if achieved through processes that ignore justice and representation. Instead, attention to justice and transformation requires strong institutional support as well as the ability to bring about real environmental change. Thus, the three dimensions form an interdependent and mutually reinforcing relationship, so that no single dimension can be used separately to determine the effectiveness of an International Environmental Agreement. By integrating these three dimensions in a single reflective evaluation framework, effectiveness is no longer understood as merely achieving targets, but rather as a multidimensional process that connects environmental performance, collective learning, and socio-political transformation in international environmental governance.

Figure 2. Reflexive Evaluation Framework



Source: Processed by the author, (2026)

CONCLUSION

This study shows that the effectiveness of International Environmental Agreements cannot be understood as an objective condition that can be identified through a single universal evaluation

strategy. Analysis of the rationalist-positivistic, constructivist-processual, and critical-reflective paradigms shows that each approach produces a different understanding of what is meant by effectiveness, what indicators are considered relevant, and what forms of change are recognized as success. The rationalist-positivistic paradigm places measurable outcomes as the primary basis for evaluation, the constructivist-processual paradigm highlights the importance of learning and institutional adaptation, while the critical-reflective paradigm directs attention to the relationships of power, justice, and legitimacy that shape environmental governance practices. These differences show that effectiveness is not just a characteristic inherent in an agreement, but rather the result of an interpretation process influenced by epistemological assumptions and certain methodological choices. As such, no assessment strategy is completely neutral as each approach simultaneously determines which aspects are considered important, what forms of evidence are judged valid, and how success is defined. These findings confirm that effectiveness evaluation should be understood as a process that is not only technical, but also political and epistemological, as it helps shape the way academics, policymakers, and international institutions understand and respond to global environmental challenges.

This article makes three main contributions. First, for the study of International Environmental Governance, this study expands the understanding of effectiveness by showing that the success of environmental governance cannot be reduced to environmental achievements alone, but needs to be understood through the interaction between outcomes, processes, and dimensions of justice. Second, with regard to Regime Effectiveness Studies, this article shifts the focus of analysis from the question of whether a regime is effective to how effectiveness is constructed and evaluated through different paradigms. Third, regarding Evaluation Studies, this study emphasizes that evaluation practices not only function as a measurement instrument, but also as a mechanism that affects knowledge production, policy priority formation, and governance legitimacy. Based on these findings, future research needs to develop a more integrative evaluation approach through the use of hybrid indicators that combine environmental achievements, institutional learning capacity, and justice dimensions in one coherent analytical framework. In addition, strengthening participatory evaluation is important to ensure that the perspective of the directly affected groups gains greater space in the assessment process. Further research also needs to encourage the development of justice-oriented assessments that not only measure what is achieved, but also consider how those successes are achieved, who benefits, and who bears the consequences. Through this development direction, the effectiveness evaluation is expected to become a more reflective, inclusive, and able instrument to capture the complexity of international environmental governance more fully.

DAFTAR PUSTAKA

- Andersen, S. O., & Sarma, K. M. (2002). *Protecting the Ozone Layer: The United Nations History*. Earthscan.
- Berger, P. L., & Luckmann, T. (1966). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Anchor Books.
- Bernstein, S. (2005). Legitimacy in Global Environmental Governance. *Journal of International Law and International Relations*, 1(1–2), 139–166.
- Biermann, F., & Pattberg, P. (2008). Global Environmental Governance: Taking Stock, Moving Forward. *Annual Review of Environment and Resources*, 33(1), 277–294. <https://doi.org/10.1146/annurev.enviro.33.050707.085733>
- Breitmeier, H., Underdal, A., & Young, O. R. (2011). The Effectiveness of International Environmental Regimes: Comparing and Contrasting Findings from Quantitative Research. *International Studies Review*, 13(4), 579–605.

- Cox, R. W. (1981). Social forces, states and world orders: Beyond international relations theory. *Millennium: Journal of International Studies*, 10(2), 126–155.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (4th ed.). SAGE Publications.
- Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings 1972–1977*. Pantheon Books.
- Fukuda-Parr, S., & McNeill, D. (2019). Knowledge and Politics in Setting and Measuring the SDGs. *Global Policy*, 10(S1), 5–15.
- Gerring, J. (2007). *Case Study Research: Principles and Practices*. Cambridge University Press.
- Gonzalez, M., Taddonio, K., & Sherman, N. (2015). The Montreal Protocol: How Today's Successes Offer a Pathway to the Future. *Journal of Environmental Studies and Sciences*, 5(2), 122–129.
- Haas, P. M. (1992). Introduction: Epistemic Communities and International Policy Coordination. *International Organization*, 46(1), 1–35. <https://doi.org/10.1017/S0020818300001442>
- Habermas, J. (1971). *Knowledge and Human Interests*. Beacon Press.
- Helm, C., & Sprinz, D. F. (2000). Measuring the Effectiveness of International Environmental Regimes. *Journal of Conflict Resolution*, 44(5), 630–652.
- Merry, S. E. (2011). Measuring the World: Indicators, Human Rights, and Global Governance. *Current Anthropology*, 52(S3), S83–S95.
- Newell, P. (2008). The Political Economy of Global Environmental Governance. *Review of International Studies*, 34(3), 507–529.
- Okereke, C. (2008). Equity Norms in Global Environmental Governance. *Global Environmental Politics*, 8(3), 25–50.
- Pyle, J. A., Keeble, J., Abraham, N. L., Chipperfield, M. P., & Griffiths, P. T. (2022). Integrated Ozone Depletion as a Metric for Ozone Recovery. *Nature*, 608, 719–723.
- Risse, T., Ropp, S. C., & Sikkink, K. (2013). *The Persistent Power of Human Rights: From Commitment to Compliance* (2nd ed.). Cambridge University Press.
- Schlosberg, D. (2004). Reconceiving Environmental Justice: Global Movements and Political Theories. *Environmental Politics*, 13(3), 517–540.
- Secretariat of the Convention on Biological Diversity. (2020). *Global Biodiversity Outlook 5*. CBD Secretariat. <https://www.cbd.int/gbo5>
- Sihap, S., Fahrudin, A. F., & Djajasukanta, R. N. (2026). Dilema Otonomi dalam Tata Kelola Iklim Global pada Desentralisasi Subnasional Indonesia. *Jurnal Ragam Pengabdian*, 3(2), 1713–1723. <https://doi.org/10.62710/g2am4x91>
- Ward, H. (2006). International Linkages and Environmental Sustainability: The Effectiveness of the Regime Network. *Journal of Peace Research*, 43(2), 149–166.
- Yanow, D., & Schwartz-Shea, P. (2014). *Interpretation and Method: Empirical Research Methods and the Interpretive Turn* (2nd ed.). Routledge.
- Young, O. R. (2001). Inferences and Indices: Evaluating the Effectiveness of International Environmental Regimes. *Global Environmental Politics*, 1(1), 99–121.
- Young, O. R. (2011). *Effectiveness of international environmental regimes: Existing knowledge, cutting-edge themes, and research strategies*.
- Young, O. R. (2018). Research Strategies to Assess the Effectiveness of International Environmental Regimes. *Nature Sustainability*, 1(9), 461–465. <https://doi.org/10.1038/s41893-018-0132-y>

Zurba, M., & Papadopoulos, A. (2021). Indigenous Participation and the Incorporation of Indigenous Knowledge and Perspectives in Global Environmental Governance Forums: A Systematic Review. *Environmental Management*, 72(1), 84–99.