

The Global Governance Readiness Index: Measuring Institutional Capacity to Deliver International Commitments

Francisca Oliveira¹, Sam Polkar¹

¹ Advocacy Unified Network

DOI: 10.62692/opuspublica.gppd.2023.02

Published: March 15, 2023

CrossMark

ABSTRACT

This article develops the Global Governance Readiness Index (GGRI) as a composite measurement framework for assessing the institutional capacity of states to translate accepted international commitments into implementation. The framework distinguishes readiness from realised performance and from perception-based governance measures, and organises readiness into five pillars: Legal-Institutional Framework, Administrative Capacity, Fiscal Readiness, Compliance Infrastructure, and Stakeholder Engagement. The design is grounded in established work on state capacity, institutional quality, international-compliance theory, and composite-indicator construction. To preserve historical integrity, the manuscript uses an explicit evidence cutoff of 15 March 2023 and does not incorporate publications or datasets released after that date. The article specifies the proposed indicators, normalisation, weighting, aggregation, missing-data and validation procedures and defines a transparent pilot design. Numerical pilot results and predictive-validity claims are not presented as established findings here because the underlying country-level dataset, reproducible code, and complete technical appendix are not part of the supplied manuscript. The GGRI is therefore presented as a methodological research framework and a pre-specified empirical design, not as a completed predictive index whose performance has already been demonstrated.

TABLE OF CONTENTS

1. Introduction

2. Literature Review: Governance Measurement, Institutional Capacity, and the Implementation Gap

2.1 The Architecture of Existing Governance Indices

2.2 Theoretical Foundations of Institutional Capacity

2.3 The Implementation Gap: From Commitment to Compliance

2.4 The AUN Mandate and the Need for a New Index

3. Theoretical Framework: Defining Governance Readiness

- 3.1 The Concept of Governance Readiness
- 3.2 The Commitment-Capacity-Outcome Continuum
- 3.3 Conceptual Distinctions: Readiness, Performance, and Perception
- 3.4 The Five-Pillar Architecture

4. Methodology: Constructing the Global Governance Readiness Index

- 4.1 Index Construction Principles
- 4.2 Data Sources and Indicator Selection
- 4.3 Normalisation, Weighting, and Aggregation

Formal composite definition

Notation used in the GGRI equations

- 4.4 Statistical Validation

Formal validation specifications

5. The GGRI Composite Model: Pillars, Indicators, and Weighting

- 5.1 Pillar 1: Legal-Institutional Framework
- 5.2 Pillar 2: Administrative Capacity
- 5.3 Pillar 3: Fiscal Readiness
- 5.4 Pillar 4: Compliance Infrastructure
- 5.5 Pillar 5: Stakeholder Engagement
- 5.6 Aggregation and Final Index

6. Pilot Application Design and Empirical Validation Protocol

- 6.1 Sample Construction
- 6.2 Pre-Specified Descriptive Analysis
- 6.3 Optional Cluster Analysis
- 6.4 Validation Against Existing Measures

7. Discussion: Counterarguments, Limitations, and Implications

- 7.1 Interpretation and Policy Implications
- 7.2 Counterarguments
- 7.3 Limitations

8. Conclusion and Future Research

References

OPUS PUBLICA

A Peer-Reviewed Open-Access Publication of the Advocacy Unified Network

The Global Governance Readiness Index:

Measuring Institutional Capacity to Deliver International Commitments

Authored by: Francisca Oliveira and Sam Polkar

Advocacy Unified Network (AUN)

Abstract

This article develops the Global Governance Readiness Index (GGRI) as a composite measurement framework for assessing the institutional capacity of states to translate accepted international commitments into implementation. The framework distinguishes readiness from realised performance and from perception-based governance measures, and organises readiness into five pillars: Legal-Institutional Framework, Administrative Capacity, Fiscal Readiness, Compliance Infrastructure, and Stakeholder Engagement. The design is grounded in established work on state capacity, institutional quality, international-compliance theory, and composite-indicator construction. To preserve historical integrity, the manuscript uses an explicit evidence cutoff of 15 March 2023 and does not incorporate publications or datasets released after that date. The article specifies the proposed indicators, normalisation, weighting, aggregation, missing-data and validation procedures and defines a transparent pilot design. Numerical pilot results and predictive-validity claims are not presented as established findings here because the underlying country-level dataset, reproducible code, and complete technical appendix are not part of the supplied manuscript. The GGRI is therefore presented as a methodological research framework and a pre-specified empirical design, not as a completed predictive index whose performance has already been demonstrated.

Keywords

global governance; institutional capacity; composite index; treaty compliance; implementation gap; sustainable development goals; readiness measurement

JEL Classification: F53, H11, K33, O19, O55 | © 2023 Opus Publica. Published under Creative Commons CC BY 4.0.

Institutional Authorship, Competing Interests, and Funding

This article is authored by Francisca Oliviera and Sam Polkar, affiliated to Advocacy Unified Network (AUN), which developed the GGRI as part of its governance-gap research programme. AUN therefore has an institutional interest in the framework's development and potential application. The manuscript distinguishes the proposed methodology from empirical validation: no predictive performance, country ranking, or superiority claim is treated as established without reproducible data and independent testing. No external funding or financial conflict of interest is identified for this historical article.

List of Tables

List of Figures

1. Introduction

The first quarter of the twenty-first century has produced the most extensive corpus of international commitments in human history. The adoption of the 2030 Agenda for Sustainable Development in September 2015, with its seventeen Sustainable Development Goals (SDGs) and one hundred and sixty-nine associated targets, established a normative framework adopted by all 193 United Nations Member States

(United Nations, 2015). Three months later, the Paris Agreement on climate change created a legally binding instrument covering nearly every economy on earth, predicated on nationally determined contributions (NDCs) whose ambition is expected to ratchet upward over five-year cycles (UNFCCC, 2015). The Addis Ababa Action Agenda, the Sendai Framework for Disaster Risk Reduction, the New Urban Agenda, and the Global Compact for Migration have further thickened the institutional texture of multilateral obligation. The rhetorical ceiling of international cooperation has rarely been higher.

As of the 15 March 2023 evidence cutoff, the most appropriate empirical context is the implementation record documented by sources already available at that date. The Sustainable Development Goals Report 2022 and the Sustainable Development Report 2022 provide contemporaneous evidence of uneven SDG progress, while the Emissions Gap Report 2022 documents the continuing gap between national climate pledges and emissions pathways (United Nations, 2022; Sachs et al., 2022; UNEP, 2022). These sources establish the problem context without importing evidence published after the article's claimed date.

This dissonance has given rise to a small but expanding literature on the implementation gap—the structural distance between the international commitments a state has formally accepted and the verifiable outcomes it actually produces (Cole, 2015; Andrews, 2013; Pritchett, Woolcock, & Andrews, 2013). Much of this literature, however, is sectoral and retrospective. It explains why a particular human rights treaty was not implemented, why a particular environmental commitment lapsed, or why a particular development goal was missed. What it does not provide is a general, comparable, and forward-looking measure of the institutional capacity that conditions implementation across the full spectrum of commitments. Existing composite governance measures—the Worldwide Governance Indicators (Kaufmann, Kraay, & Mastruzzi, 2010), the Bertelsmann Transformation Index (Bertelsmann Stiftung, 2022), the Ibrahim Index of African Governance (Mo Ibrahim Foundation, 2022), the Fragile States Index (Fund for Peace, 2022), and the Sustainable Development Report's SDG Index (Sachs et al., 2022)—offer valuable diagnostic information, but each is oriented toward observed performance rather than forward-looking readiness, and none was designed to answer the question that increasingly preoccupies treaty secretariats, development finance institutions, and the multilateral system: does this state, today, possess the institutional capacity to deliver on the international commitments it has accepted?

This article develops the Global Governance Readiness Index (hereafter, the GGRI) as a direct response to that question. The GGRI is a composite measurement model that assesses the institutional capacity of states to implement international commitments across five structurally weighted pillars: Legal-Institutional Framework, Administrative Capacity, Fiscal Readiness, Compliance Infrastructure, and Stakeholder Engagement. The model is grounded in a deliberate theoretical synthesis: the capacity-autonomy distinction advanced by Fukuyama (2013); the legal-fiscal capacity model of Besley and Persson (2009); the inclusive-institutions thesis of Acemoglu and Robinson (2012); the managerial theory of compliance articulated by Chayes and Chayes (1995); and the broader New Institutional Economics tradition associated with North (1990) and Ostrom (1990, 2005). The methodological architecture follows the ten-step protocol codified in the OECD Handbook on Constructing Composite Indicators (Nardo et al., 2008), incorporating sensitivity analysis and uncertainty quantification as integral validation components rather than post-hoc checks.

The motivation for this work is both analytical and institutional. The Advocacy Unified Network (AUN), a non-governmental, non-partisan research and advocacy organisation headquartered in The Hague, has positioned itself at the intersection of policy research and advocacy with a mandate to close governance gaps and accelerate the SDGs across every region. AUN's broader governance-gap research agenda provides the institutional context for this methodological proposal. The GGRI is conceived here as the foundational measurement theory that should precede and structurally ground that diagnostic. Without a robust, transparent, and replicable measure of governance readiness, the assertion of a governance gap remains

analytically under-specified: it can describe a deficit but cannot predict which states will incur it, in which domains, or through which institutional mechanisms. The GGRI supplies the missing analytical scaffolding.

The argument proceeds in eight stages. Section 2 reviews the architecture of existing governance indices, the theoretical foundations of institutional capacity, and the implementation-gap literature, identifying three structural limitations of the current measurement landscape. Section 3 develops the theoretical framework of the GGRI, defining governance readiness as a distinct analytical construct and articulating the five-pillar architecture. Section 4 sets out the methodology, including data sourcing, normalisation, weighting, aggregation, and validation. Section 5 presents the GGRI composite model in full, detailing the indicators, sub-pillars, and weighting rationale for each of the five pillars. Section 6 specifies the proposed pilot application and empirical validation protocol, including comparisons with the Worldwide Governance Indicators and the SDG Index. Section 7 addresses counterarguments, limitations, and the boundaries of the model's interpretive reach. Section 8 concludes with implications for the multilateral system and an agenda for future research.

The central claim of this article is narrow but consequential. Readiness—the forward-looking institutional capacity to deliver international commitments—is a distinct, measurable, and policy-actionable property of states. It is not reducible to observed performance, which is contaminated by exogenous shocks, path dependence, and the lag between institutional investment and outcome realisation. Nor is it reducible to perception-based assessments of governance quality, which are subject to halo effects, sample bias, and temporal instability (Kaufmann et al., 2010; Thomas, 2010). The GGRI proposes that readiness is best understood as a structural property of the state's institutional endowment, measurable through transparent indicators of legal architecture, administrative depth, fiscal headroom, compliance infrastructure, and stakeholder integration. If this proposition is correct, then the multilateral system possesses, for the first time, an instrument capable of moving from the descriptive language of implementation gaps to the diagnostic and predictive language of readiness differentials.

2. Literature Review: Governance Measurement, Institutional Capacity, and the Implementation Gap

2.1 The Architecture of Existing Governance Indices

The Worldwide Governance Indicators (WGI), introduced by Kaufmann, Kraay, and Zoido-Lobaton in 1999 and subsequently refined, aggregate perceptions of governance quality across six dimensions: Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. For this 15 March 2023 historical manuscript, the analysis is restricted to the WGI release and underlying information available by the cutoff, rather than later revisions. The WGI methodology uses an unobserved-components model and reports margins of error, making it useful for broad cross-country comparisons while also requiring caution about precision and comparability (Kaufmann et al., 2010; World Bank, 2022).

Notwithstanding this methodological rigour, the WGI has attracted sustained criticism on several grounds that bear directly on the design of a readiness measure. First, the WGI is perception-based: it aggregates the views of experts, survey respondents, and commercial risk providers rather than measuring the structural features of governance that produce those perceptions. Critics have argued that this design imports a business-elite bias, as commercial risk ratings are weighted heavily in the aggregation and tend to reflect investor priorities rather than the priorities of citizens or treaty beneficiaries (Thomas, 2010; Kurtz & Schrank, 2007). Second, the WGI exhibits documented 'halo effects', in which a country's overall level of development

or recent macroeconomic performance inflates perceptions of governance quality across all six dimensions, producing a spurious correlation between income and governance that may not reflect underlying institutional capacity (Kurtz & Schrank, 2007). Third, the WGI is poorly suited to over-time comparison, because changes in the underlying source data and rescaling procedures can produce apparent improvements that are statistical artefacts rather than genuine institutional change (Arndt & Oman, 2010).

The Bertelsmann Transformation Index (BTI) offers a complementary but distinct approach. The BTI analyses transformation processes toward democracy and a market economy in approximately one hundred and thirty-seven transition countries, drawing on a network of nearly three hundred country and regional experts guided by a standardised codebook (Bertelsmann Stiftung, 2022). Its dual structure—a Status Index measuring the state of political and economic transformation and a Governance Index assessing the quality of management—provides a richer diagnostic than the WGI on questions of political steering and reform management. The BTI's reliance on expert assessment, however, limits its scalability, and its deliberate focus on transition countries excludes the advanced industrial economies and the most fragile states from the comparative frame.

The Ibrahim Index of African Governance (IIAG), established in 2007, provides a detailed regional governance dataset covering African countries and multiple dimensions of governance. For historical consistency, this article relies on the 2022 IIAG information available before the manuscript cutoff rather than later editions. The IIAG is useful as a complementary regional benchmark, but its regional scope and outcome orientation constrain its use as a global, forward-looking readiness instrument (Mo Ibrahim Foundation, 2022).

The Sustainable Development Report's SDG Index provides a complementary outcome-oriented benchmark. For this historical version, the 2022 edition is the relevant source available by the manuscript cutoff (Sachs et al., 2022). The SDG Index is target-anchored and useful for measuring realised SDG performance, but it is not designed to isolate the institutional capacity that conditions future implementation. This distinction is central to the GGRI's conceptual contribution.

A further family of indices deserves brief mention for completeness. The Environmental Sustainability Index and its successor the Environmental Performance Index, developed by Esty, Levy, Srebotnjak, and de Sherbinin (2005) at the Yale Center for Environmental Law and Policy, provide composite measures of environmental sustainability and policy performance across a large sample of states. These indices share the outcome-orientation of the SDG Index and, while they supply valuable diagnostic information on environmental performance, they do not address the institutional capacity to deliver environmental commitments per se. The Human Development Index of the United Nations Development Programme, the Global Peace Index of the Institute for Economics and Peace, and the Open Budget Index of the International Budget Partnership each addresses a specific dimension of governance or development and provides complementary information, but none was designed to measure the forward-looking institutional capacity to deliver the full spectrum of contemporary international commitments.

Three structural limitations emerge from this survey. First, the dominant indices measure observed performance or perceived quality rather than the forward-looking institutional capacity that conditions future implementation. Second, the indices tend to aggregate across dimensions that are analytically distinct—capacity, autonomy, inclusion, outcome—without providing the decomposability required for diagnostic use. Third, none of the existing indices was designed to anchor its measurement frame in the international commitment architecture itself: the WGI measures governance in the abstract, the BTI measures transformation, the IIAG measures African governance, and the SDG Index measures SDG outcomes. A measure of readiness to deliver international commitments must, by contrast, take the commitment architecture as its analytical anchor.

2.2 Theoretical Foundations of Institutional Capacity

The theoretical antecedents of a readiness measure are located at the intersection of institutional economics, comparative politics, and international relations theory. North's (1990) foundational distinction between formal rules and informal constraints, and his emphasis on the transaction-cost function of institutions, established the analytical premise that institutional capacity is not reducible to the legal code on the books but inheres in the enforceability of rules and the credibility of commitments. Acemoglu, Johnson, and Robinson (2005) extended this framework to a long-run growth context, arguing that differences in economic institutions—in particular, the distinction between extractive and inclusive institutions—are the fundamental cause of cross-country differences in economic performance (Acemoglu & Robinson, 2012). The inclusive-extractive distinction is of direct relevance to readiness measurement: inclusive institutions, by dispersing political power and securing property rights across society, generate the credibility and predictability on which compliance with international commitments depends.

Fukuyama (2013) provides what is perhaps the most influential recent reframing of the governance concept. Defining governance as 'a government's ability to make and enforce rules, and to deliver services, regardless of whether that government is democratic or not,' Fukuyama argues that the quality of government is the product of an interaction between capacity—the ability of the state to enforce rules and deliver services—and autonomy—the degree to which the state is insulated from societal pressures in the exercise of that capacity. This two-dimensional framework has significant implications for index design. First, it implies that governance cannot be adequately measured through output indicators alone, because outputs confound capacity with autonomy. Second, it implies that the conventional separation of 'government effectiveness' from 'rule of law' in indices such as the WGI is analytically problematic, because both are partial expressions of the deeper capacity-autonomy interaction. The GGRI's pillar architecture, particularly its treatment of Administrative Capacity and Compliance Infrastructure as distinct but interdependent pillars, responds directly to this critique.

Besley and Persson (2009) develop a complementary framework in which state capacity is endogenously determined through the political-economy interaction of two distinct state investments: legal capacity—the institutional infrastructure that supports contract enforcement and property rights—and fiscal capacity—the institutional infrastructure that supports taxation and the financing of public goods. Their model predicts that cohesive political institutions, which constrain ruling groups from expropriating the returns to state investment, generate complementarities between legal and fiscal capacity, producing the kind of high-capacity state that can credibly commit to long-horizon international obligations. The Besley-Persson framework is of particular relevance to the GGRI because it provides an explicit theoretical foundation for two of the index's pillars: Legal-Institutional Framework, which operationalises the legal-capacity dimension, and Fiscal Readiness, which operationalises the fiscal-capacity dimension. The complementarity prediction—namely, that legal and fiscal capacity reinforce each other—also supplies a structural prior for the weighting scheme of the GGRI.

Ostrom's (1990, 2005) work on institutional diversity and the governance of commons resources introduces a further dimension that is largely absent from the dominant composite indices: the role of polycentric, multi-level governance arrangements in producing sustainable collective action. Ostrom's design principles for common-pool resource institutions—clearly defined boundaries, congruence between rules and local conditions, collective-choice arrangements, monitoring, graduated sanctions, conflict-resolution mechanisms, and recognition of organising rights—translate, with appropriate adaptation, into design principles for the multi-stakeholder implementation of international commitments. The GGRI's Stakeholder Engagement pillar draws explicitly on this Ostromian tradition, treating the inclusiveness and structural depth of stakeholder engagement as a measurable dimension of governance readiness rather than as a normative aspiration.

Two further strands of the comparative-politics literature warrant attention. The 'embedded autonomy' framework of Evans (1995) describes the configuration of state bureaucracies that combines internal coherence and professional meritocracy with dense, structured ties to industrial and civil-society actors, producing the kind of state that can pursue transformative policy agendas. Migdal's (1988) contrast between strong and weak states, indexed by the state's capacity to penetrate society, regulate social relationships, and extract and use resources in a determined fashion, provides a complementary account of the structural determinants of state capability. Both frameworks converge on a common finding: state capacity is not a unidimensional resource but a relational property of the state's embeddedness in society, and its measurement must therefore include indicators of the state-society nexus alongside indicators of internal bureaucratic capability.

The broader tradition of institutional analysis in international relations and comparative politics supplies further theoretical grounding. Keohane's (1984) seminal work on cooperation after hegemony established that international institutions persist not because they enforce compliance through coercion but because they reduce the transaction costs of cooperation, monitor behaviour, and generate the reputational capital that sustains collective action—an insight that places the domestic capacity to participate in such monitoring and reputation-building at the centre of the compliance question. Putnam's (1993) study of institutional performance in Italian regions demonstrated that governance outcomes are conditioned by the stock of social capital—the networks of civic engagement and reciprocal trust that constitute the relational substrate of institutional effectiveness—a finding that directly informs the GGRI's treatment of stakeholder engagement as a structural rather than normative dimension of readiness. March and Olsen's (1995) reframing of democratic governance as an institutional process governed by rules, routines, and normative logics of appropriateness, rather than by the aggregate logic of consequentiality, supplies the theoretical warrant for treating institutional architecture as an independent determinant of implementation capacity rather than as a mere conduit for political preferences.

2.3 The Implementation Gap: From Commitment to Compliance

The literature on treaty compliance and the implementation gap provides the third analytical pillar of the GGRI. The foundational contribution of Chayes and Chayes (1993, 1995) reframed the compliance question from a juridical enforcement problem to a managerial problem of capacity, transparency, and rule clarity. Their 'managerial theory' holds that non-compliance with international regulatory agreements is more often the product of ambiguity, capacity deficits, and temporal lags than of deliberate defection, and that the international system's compliance machinery should therefore be oriented toward capacity-building, transparency, and dispute resolution rather than toward punitive enforcement. This insight has been empirically extended by Downs, Rocke, and Barsoom (1996), who showed that the depth of cooperation—the behavioural change required by an agreement—is a more reliable predictor of non-compliance than the formal stringency of the obligation itself.

Simmons (1998) synthesises four principal theoretical perspectives on compliance: the managerial model of Chayes and Chayes, the enforcement model rooted in the work of Downs and colleagues, the normative legitimacy model associated with Franck (1990), and the domestic-transaction-cost model developed by Simmons herself, in which the cost of domestic political mobilisation against compliance conditions the willingness of governments to translate international commitments into national law. The domestic-transaction-cost model is particularly germane to the GGRI, because it implies that compliance is mediated by the domestic institutional architecture through which commitments must pass—a configuration that is itself measurable. Cole's (2015) study of compliance with the International Covenant on Civil and Political Rights provides direct empirical support for this proposition: using two-stage regression models, Cole demonstrates that state capacity, measured through indicators of bureaucratic quality and judicial

independence, is a statistically significant predictor of treaty compliance, even after controlling for selection effects into treaty commitment.

The 'looking like a state' literature of Pritchett, Woolcock, and Andrews (2013) and Andrews (2013) supplies a critical caveat. These authors argue that many states have adopted the organisational forms of modern public administration—ministries, agencies, regulatory bodies—without acquiring the functional capability that those forms are meant to deliver, producing what they term 'isomorphic mimicry' and 'premature load bearing'. The implication for readiness measurement is that indicators of organisational form (existence of an environmental ministry, existence of a human rights commission) are necessary but insufficient; they must be supplemented by indicators of functional depth (budget execution, regulatory enforcement, judicial throughput). The GGRI's indicator selection responds to this caveat by combining structural-form indicators with functional-capacity indicators at the sub-pillar level, and by weighting functional indicators more heavily than form indicators in the aggregation.

Rothstein and Teorell (2008) advance a related but distinct argument that the quality of government is best understood as the impartiality of government institutions in the exercise of public authority, rather than as the efficiency of outputs or the procedural character of democratic inputs. Their framework implies that compliance with international commitments is mediated by the impartiality of the institutions through which commitments are processed—an impartiality that is itself measurable through indicators of judicial independence, non-discrimination in public administration, and the structural insulation of regulatory agencies from political interference. The GGRI incorporates this insight through the Compliance Infrastructure pillar, which assesses not only the formal existence of compliance mechanisms but their structural impartiality as expressed in independence safeguards, transparent procedures, and appeals mechanisms.

The synthesis of these literatures yields a coherent theoretical foundation for the GGRI. Institutional capacity is multidimensional, encompassing legal, fiscal, administrative, compliance, and engagement components that interact in structurally predictable ways. Compliance with international commitments is mediated by domestic institutional architecture, not by the stringency of international obligation alone. Organisational form is a weak proxy for functional capacity. Impartiality is a constitutive element of governance quality that conditions the credibility of commitments. A composite measure that integrates these insights—distinguishing capacity from autonomy, legal from fiscal capacity, form from function, and inclusion from output—can credibly claim to measure the forward-looking readiness that existing indices leave unmeasured.

2.4 The AUN Mandate and the Need for a New Index

The Advocacy Unified Network (AUN) is a non-governmental, non-partisan research and advocacy organisation headquartered at Fluwelen Burgwal 58 in The Hague, the Netherlands, with an institutional mission dedicated to bridging the gap between politics and policy and to promoting effective and efficient democratic governance through meaningful stakeholder engagement. The Network's mandate is explicitly oriented toward the closing of governance gaps and the acceleration of the Sustainable Development Goals across every region, and its institutional architecture comprises a portfolio of flagship initiatives spanning education, justice, publishing, heritage, trust, media, and technology. Opus Publica, the Network's open-access scholarly publishing platform, provides the peer-reviewed infrastructure through which the Network's research is disseminated to the global scholarly and policy community, with formal Crossref digital object identifiers ensuring the persistent discoverability and citability of published work.

The Network's flagship analytical product, the State of Global Governance report, evaluates the architecture of global governance across multiple policy domains and frames its diagnostic through the analytical lens of the governance gap—the structural distance between the institutional design of the global governance system and the operational delivery of its commitments. The governance-gap framing is analytically

powerful because it foregrounds the implementation deficit as the central pathology of contemporary global governance, but it is analytically incomplete in one consequential respect: it can describe the gap but cannot, without a measure of readiness, predict which states will incur it, in which domains, or through which institutional mechanisms. The governance-gap diagnostic is, in this sense, a downstream application of a measurement theory that has not yet been supplied. The GGRI is designed to supply that foundational measurement theory, structurally preceding and grounding the Network's broader analytical and advocacy architecture.

The intellectual case for situating the GGRI as the foundational measurement theory of the AUN rests on three considerations. First, the Network's commitment to evidence-led advocacy requires a measurement instrument whose construction is transparent, replicable, and theoretically grounded, rather than an ad-hoc diagnostic whose methodological choices are opaque. The GGRI's adherence to the OECD Handbook on Constructing Composite Indicators protocol and its documentation of every methodological choice respond directly to this requirement. Second, the Network's positioning at the intersection of policy research and advocacy requires a measurement instrument that is policy-actionable—capable of identifying the specific institutional dimensions in which readiness is deficient—rather than merely descriptive. The GGRI's pillar-level decomposability responds to this requirement. Third, the Network's global mandate requires a measurement instrument that is applicable across the full population of states and across the full spectrum of international commitments, rather than confined to a region or a sector. The GGRI's commitment-anchored orientation and its design for global coverage respond to this requirement. The historical publication statement is therefore presented as the publisher-side record for this manuscript; it should not be used as evidence that later AUN outputs were already available at the cutoff.

3. Theoretical Framework: Defining Governance Readiness

3.1 The Concept of Governance Readiness

Governance readiness, as defined in this article, is the structural capacity of a state to translate its accepted international commitments into verifiable national outcomes through institutional mechanisms that are legally grounded, administratively capable, fiscally supported, compliance-oriented, and inclusive of stakeholders. Three features of this definition warrant elaboration. First, readiness is a structural property: it inheres in the configuration of institutions and the endowment of resources that condition implementation, rather than in the outcomes that the configuration ultimately produces. This structural orientation distinguishes readiness from performance, the measurement of realised outcomes, and from perception, the measurement of expert or popular judgement of governance quality.

Second, readiness is forward-looking. It is concerned with the capacity to deliver on commitments that have been accepted but not yet fully implemented, and with the trajectory of institutional investment that will condition future implementation. This temporal orientation distinguishes readiness from the retrospective orientation of the dominant composite indices, which measure the outcomes of past institutional investment. A state with high readiness but low current performance—for instance, a state that has recently overhauled its environmental regulatory framework but has not yet realised the pollution reductions that the framework will eventually produce—is misclassified by performance-based measures and correctly classified by a readiness measure. Conversely, a state with high current performance but declining readiness—because of regulatory rollback, fiscal deterioration, or institutional erosion—is misclassified by performance-based measures as a strong performer when its compliance trajectory is in fact deteriorating.

Third, readiness is commitment-anchored. It is measured against the architecture of international commitments that the state has accepted, rather than against an abstract standard of governance quality. This commitment-anchored orientation distinguishes readiness from the WGI's abstract conception of governance and aligns the measure with the operational reality faced by treaty secretariats and development finance institutions, which must assess the implementation capacity of states against specific, named commitments. The commitment-anchored orientation does not require the index to enumerate each of the thousands of commitments a state may have accepted; rather, it requires the index to measure the institutional capacities that condition implementation across the principal categories of contemporary international obligation—sustainable development, climate, human rights, trade, and security cooperation.

3.2 The Commitment-Capacity-Outcome Continuum

Figure 1. Commitment–capacity–outcome continuum. The GGRI is positioned as a measure of structural readiness between accepted commitments and realised outcomes; the diagram is conceptual and contains no empirical observations.

The analytical core of the GGRI is the commitment-capacity-outcome continuum, a three-stage model of the international commitment process. The first stage, commitment, captures the formal acceptance of an international obligation through signature, ratification, or accession. This stage is well documented by existing databases, including the United Nations Treaty Collection, the SDG Indicators database, and the climate NDC registry. The third stage, outcome, captures the verifiable national effects of implementation—emission reductions, poverty reductions, rights protections, regulatory outcomes—and is the principal concern of the SDG Index and the various treaty-body monitoring mechanisms. The second stage, capacity, is the institutional infrastructure that mediates between commitment and outcome, and is the analytical space that the GGRI is designed to occupy.

The continuum has three important properties. First, the relationship between commitment and outcome is not deterministic: it is conditioned by capacity, exogenous shocks, path dependence, and the time lag between institutional investment and outcome realisation. A state may ratify a treaty without possessing the capacity to implement it, producing what Cole (2015) terms decoupling between commitment and outcome. Conversely, a state may possess robust capacity but decline to commit, producing capacity without commitment. The most policy-relevant configuration is commitment without capacity, because this configuration produces the implementation gap that has come to characterise the global governance system. The GGRI is designed to identify this configuration *ex ante*, before the gap materialises in observed outcomes.

Second, the relationship between capacity and outcome may be lagged. Institutional investments in administrative capability, fiscal capacity, and compliance infrastructure may take time to translate into measurable outcomes, with the duration varying by policy domain, institutional starting point, and the nature of the investment. This lag means that contemporary outcome measures may reflect earlier institutional conditions, while contemporary capacity may be relevant to future outcomes. A readiness measure that captures contemporary capacity can therefore be evaluated as a potential leading indicator only through genuinely longitudinal empirical testing.

Third, the relationship between capacity and commitment is reciprocal. Capacity conditions the willingness of states to commit, because states are reluctant to accept obligations they cannot credibly implement, particularly under mechanisms that impose reputational or material costs for non-compliance. But commitment also conditions capacity, because the acceptance of an obligation creates a focal point around which domestic coalitions for institutional investment can mobilise. This reciprocal relationship implies that readiness is not a static endowment but a dynamic equilibrium, and that its measurement must be sensitive

to the trajectory of institutional investment as well as to its current level. The GGRI addresses this requirement through a longitudinal extension, discussed in Section 4, that tracks changes in readiness scores over time.

3.3 Conceptual Distinctions: Readiness, Performance, and Perception

A further conceptual clarification is warranted, drawing on the capability approach of Sen (1999), who conceived development not as the accumulation of resources or the realisation of outcomes but as the expansion of the substantive freedoms—the capabilities—that enable individuals and institutions to lead lives they have reason to value. Transposed to the level of the state, the capability perspective implies that governance capacity is best understood not as the stock of resources a state possesses nor as the outcomes it produces, but as the substantive freedom the state has to convert commitments into delivered results—a freedom that is constituted by the institutional architecture through which commitments are processed. This capability-theoretic framing supplies an additional justification for the structural orientation of the GGRI: the index measures the freedom-to-deliver rather than the delivery itself, and it is this freedom that constitutes the analytical core of readiness. The distinction is consequential because it implies that readiness is constitutively prior to performance, both logically and temporally, and that its measurement is therefore the precondition for any credible diagnosis of the implementation gap.

The conceptualisation of governance readiness requires explicit differentiation from three adjacent concepts: governance performance, governance perception, and governance quality. Governance performance refers to the realised outcomes of governance activity—service-delivery statistics, regulatory outcomes, rights protections, fiscal balances. Performance measures, such as the SDG Index, are essential for accountability but are not designed to diagnose the institutional determinants of performance. A state may produce strong performance through windfall resource revenues or favourable demographic transitions without possessing robust institutional capacity, and may produce weak performance despite strong capacity when subjected to exogenous shocks. Readiness is therefore analytically prior to performance: it explains why performance takes the form it does and predicts the trajectory of future performance.

Governance perception refers to the judgement of governance quality formed by experts, survey respondents, and commercial risk providers. Perception measures, such as the WGI, aggregate these judgements into composite scores that are valuable for cross-country comparison but are subject to the biases, halo effects, and temporal instabilities documented in Section 2.1. Perception is not unrelated to capacity—states with strong capacity tend to be perceived as well-governed—but the relationship is mediated by visibility, salience, and the framing effects of recent events. Readiness, by contrast, is intended to be a structural property measurable through indicators that are not dependent on the perceptions of external observers, and that are therefore less susceptible to halo effects and perception bias.

Governance quality, in the sense developed by Rothstein and Teorell (2008), refers to the impartiality of government institutions in the exercise of public authority. Quality is a constitutive dimension of governance that conditions both performance and the credibility of commitments, and is therefore a partial determinant of readiness rather than a substitute for it. A state with high-quality institutions (strong impartiality) but low capacity (limited administrative or fiscal resources) is not ready to deliver complex commitments, even though its governance is in an important sense good. Conversely, a state with high capacity but low impartiality may produce implementation outcomes in the short term but lacks the credibility to sustain them. The GGRI incorporates the quality dimension through the Compliance Infrastructure pillar, which assesses impartiality safeguards, and through the Stakeholder Engagement pillar, which assesses inclusion, while treating capacity as the structurally weight-bearing dimension of readiness.

3.4 The Five-Pillar Architecture

The GGRI is structured around five pillars, each capturing a structurally distinct dimension of the institutional capacity to deliver international commitments. The pillar architecture is grounded in the theoretical synthesis developed in Sections 2 and 3.1–3.3, and is designed to satisfy three criteria: theoretical coherence (each pillar must correspond to a theoretically distinct dimension of capacity), empirical measurability (each pillar must be operationalisable through indicators for which data are available for a broad sample of states), and policy actionability (each pillar must correspond to a domain in which policy interventions can plausibly shift the readiness score).

The first pillar, Legal-Institutional Framework, captures the formal legal architecture through which international commitments are incorporated into domestic law and the structural features of the constitutional and statutory order that condition implementation. This pillar operationalises the legal-capacity dimension of the Besley-Persson framework, the rule-clarity dimension of the Chayes-Chayes managerial model, and the formal-rules dimension of the North framework. The second pillar, Administrative Capacity, captures the bureaucratic capability of the state to design, execute, and monitor the policies required to implement commitments. This pillar operationalises the capacity dimension of the Fukuyama framework, the embedded-autonomy dimension of the Evans framework, and the strong-state dimension of the Migdal framework.

The third pillar, Fiscal Readiness, captures the fiscal headroom of the state to finance the implementation of commitments, including the structure of revenue mobilisation, the sustainability of public debt, and the alignment of budgetary allocation with commitment-related expenditure. This pillar operationalises the fiscal-capacity dimension of the Besley-Persson framework and responds to the Addis Ababa Action Agenda's emphasis on the centrality of domestic public resource mobilisation to the implementation of the 2030 Agenda (United Nations, 2015). The fourth pillar, Compliance Infrastructure, captures the institutional mechanisms through which compliance with commitments is monitored, enforced, and remediated, including the independence and capability of the judiciary, the structural position of audit and oversight institutions, and the integration of commitment-specific monitoring mechanisms into the domestic institutional architecture. This pillar operationalises the impartiality dimension of the Rothstein-Teorell framework and the transparency-and-dispute-resolution dimension of the Chayes-Chayes model.

The fifth pillar, Stakeholder Engagement, captures the structural inclusion of non-state actors—civil society organisations, private-sector entities, sub-national governments, and marginalised populations—in the design and implementation of commitment-related policy. This pillar operationalises the polycentric-governance dimension of the Ostrom framework, the embeddedness dimension of the Evans framework, and the participation norm embedded in the 2030 Agenda's commitment to 'leave no one behind'. The pillar is not a measure of democratic performance per se, but of the structural channels through which stakeholders are integrated into implementation, and is therefore applicable across regime types.

The five-pillar architecture is not the only defensible decomposition of governance readiness. Alternative architectures—distinguishing, for instance, political capacity from economic capacity, or input capacity from output capacity—have their proponents in the literature (Hanson & Sigman, 2021; Hendrix, 2010). The five-pillar architecture adopted here is justified on three grounds: its theoretical grounding in the established frameworks reviewed in Section 2.2; its alignment with the commitment-capacity-outcome continuum developed in Section 3.2; and its decomposability, which permits diagnostic use of the index in policy settings where the identification of specific capacity deficits is more valuable than an aggregate readiness score. The relative weighting of the five pillars is addressed in Section 4.3 and empirically tested in Section 6.

4. Methodology: Constructing the Global Governance Readiness Index

4.1 Index Construction Principles

Figure 2. Proposed GGRI construction and validation workflow. This is a methodological illustration; it does not imply that the empirical pipeline has already been executed.

The construction of the GGRI follows the ten-step protocol codified by Nardo, Saisana, Saltelli, and Tarantola in the OECD Handbook on Constructing Composite Indicators (Nardo et al., 2008), which provides widely used technical guidance for composite-indicator construction. The ten steps are: (i) theoretical framework development; (ii) data selection; (iii) data imputation; (iv) multivariate analysis; (v) normalisation; (vi) weighting and aggregation; (vii) uncertainty and sensitivity analysis; (viii) back to the data; (ix) links to other indicators; and (x) visualisation of the results. Each step is documented below with explicit reference to the design choices made in the construction of the GGRI.

The first step, theoretical framework development, is addressed in Section 3 and is recapitulated here only insofar as it bears on operationalisation. The five-pillar architecture, the commitment-capacity-outcome continuum, and the structural definition of governance readiness supply the analytical scaffolding within which the data-selection and aggregation choices are made. The theoretical framework is not a decoration layered on top of a data-driven exercise; it is the substantive justification for the inclusion, exclusion, and weighting of indicators, and it disciplines the interpretation of the resulting scores. The GGRI's commitment-anchored orientation, in particular, requires that each indicator be theoretically connected to the capacity to deliver international commitments, a requirement that excludes a number of indicators commonly used in governance indices on the grounds of theoretical disconnect.

The second step, data selection, applies four criteria to candidate indicators: theoretical relevance (the indicator must correspond to a theoretically distinct dimension of one of the five pillars); data quality (the indicator must be sourced from a primary or authoritative secondary source with documented methodology); coverage (the indicator must be available for at least eighty per cent of the states in the sample); and timeliness (the indicator must reflect data collected within the preceding five years, where the indicator is time-variant). These criteria are deliberately stringent and exclude a number of candidate indicators that would improve coverage at the cost of quality or theoretical relevance. The trade-off between coverage and quality is a structural feature of composite-index construction, and the GGRI errs on the side of quality, in line with the recommendations of Saltelli (2007) and Saisana, Saltelli, and Tarantola (2005).

4.2 Data Sources and Indicator Selection

The GGRI draws on primary and authoritative secondary data sources that were available by 15 March 2023. For the Legal-Institutional Framework pillar, candidate sources include the World Bank Women, Business and the Law database, OECD Indicators of Regulatory Policy and Governance where coverage is available, the V-Dem dataset (v13, March 2023), and the United Nations Treaty Collection. Later World Bank Business Ready (B-READY) material is expressly excluded from this historical version. The source-selection principle is not that a dataset is authoritative in the abstract, but that the relevant edition and data vintage were publicly available by the stated cutoff.

For Administrative Capacity, candidate sources include the UN Department of Economic and Social Affairs e-Government Survey, the V-Dem dataset, the World Bank Human Capital Index and related development indicators, the International Telecommunication Union's ICT statistics, and ILO labour-market statistics. The

manuscript treats these as source pools rather than as evidence that a completed 60-country index had already been calculated. Indicator-level selection, edition, country coverage and transformation rules must be recorded in the technical dataset accompanying any empirical implementation.

For the Fiscal Readiness pillar, the principal sources are the International Monetary Fund's Government Finance Statistics, the World Bank's World Development Indicators for tax-revenue and debt-to-GDP ratios, the OECD Revenue Statistics for tax-structure indicators, and the PEFA (Public Expenditure and Financial Accountability) Secretariat's assessments of public financial management quality. These sources provide indicators of revenue mobilisation, debt sustainability, budget execution, and fiscal transparency that are directly relevant to the capacity to finance commitment implementation. The PEFA framework, in particular, supplies indicators of the quality of public financial management that complement macroeconomic datasets and are directly relevant to implementation capacity.

For the Compliance Infrastructure pillar, the principal sources are the World Justice Project's Rule of Law Index, the V-Dem dataset for judicial-independence indicators, the International Budget Partnership's Open Budget Survey for fiscal-oversight indicators, and the UN Office on Drugs and Crime for criminal-justice-system indicators. These sources provide indicators of judicial independence, regulatory enforcement, oversight capability, and transparency that operationalise the impartiality dimension of governance quality as defined by Rothstein and Teorell (2008). For the Stakeholder Engagement pillar, the principal sources are the V-Dem dataset for civil-society-participation indicators, the Open Government Partnership's Independent Reporting Mechanism, the CIVICUS Monitor for civic-space indicators, and the SDG Indicators database for partnership and participation indicators.

Table 1. Historical source-vintage control. The table records admissibility rules rather than reporting newly calculated country data.

For an empirical release, data imputation would follow the third OECD step through multiple imputation by chained equations (MICE) for indicators with missing values below the twenty-per-cent threshold, following the recommendations of the OECD Handbook (Nardo et al., 2008). Indicators with missingness above the threshold would be excluded and the exclusion documented transparently. MICE is specified rather than single imputation because it preserves uncertainty associated with missing data and avoids the spurious precision that single imputation can introduce. The imputation model would be run separately for each pillar to respect the structural relationships between indicators within a pillar, and imputed values would be checked against the observed distribution and through sensitivity analysis of the imputation assumptions.

4.3 Normalisation, Weighting, and Aggregation

Normalisation is the fifth OECD step and addresses the heterogeneity of indicator scales and directions. The GGRI uses a distance-to-reference transformation to place directionally aligned indicators on a common 0–100 scale. For each indicator, the lower reference value, upper reference value, and any substantive benchmark used in the transformation must be frozen before empirical scoring and disclosed in the technical appendix. Where an indicator has a clear normative ceiling, that ceiling may serve as the upper reference; where no defensible normative ceiling exists, a pre-specified empirical reference such as the 95th percentile may be used as a GGRI design choice. The choice must not be changed after observing the resulting country rankings.

Directionality is handled through a pre-specified monotonic transformation so that higher transformed values consistently indicate higher readiness. A simple sign reversal is appropriate only where the scale permits it; benchmark-based indicators may require a substantively defined transformation. The general transformation is:

$$x_{ij}^* = T_j(x_{ij}) \quad (1)$$

Here x_{ij} is the raw observation for country i and indicator j ; x_{ij}^* is the directionally aligned value; and $T_j(\cdot)$ is the frozen monotonic transformation that makes higher values correspond to higher readiness.

$$z_{ij} = 100 \times (x_{ij}^* - L_j) / (R_j - L_j) \quad (2)$$

Here z_{ij} is the normalised 0–100 indicator score, L_j is the lower reference, and R_j is the upper/reference value. Reference values must be fixed and disclosed before empirical scoring.

Weighting is the sixth OECD step and is among the most consequential decisions in composite-index construction, because the weighting scheme determines the contribution of each indicator, sub-pillar, and pillar to the final score. The GGRI uses a hybrid weighting architecture. Structural pillar weights are theory-driven: Legal-Institutional Framework 0.20, Administrative Capacity 0.25, Fiscal Readiness 0.20, Compliance Infrastructure 0.20, and Stakeholder Engagement 0.15. Within pillars, PCA is proposed as the statistical basis for indicator weights.

Within each pillar, the PCA is to be performed on the correlation matrix of the normalised indicator set. A loading below 0.40 on the first component is a review flag rather than an automatic exclusion rule; any exclusion must be theoretically justified and documented. Because the historical manuscript does not contain the country-level matrix, no loading, eigenvalue, or derived indicator weight is reported as an observed result. The loading-to-weight transformation specified below is a GGRI design rule rather than an OECD-required formula.

For pillar or sub-pillar s , let R_s denote the correlation matrix of the normalised indicators. The PCA eigenproblem is:

$$R_s q_{sk} = \lambda_{sk} q_{sk} \quad (3)$$

Here λ_{sk} is the eigenvalue and q_{sk} is the corresponding eigenvector for component k . The first component ($k = 1$) is the proposed statistical basis for the indicator-weighting step.

$$w_{js} = |\ell_{js1}| / \sum_j |\ell_{js1}| \quad (4)$$

Here ℓ_{js1} is the loading of indicator j on the first principal component in s . This loading-to-weight transformation is a GGRI design rule and must be frozen before the empirical release; it is not presented as an OECD-required formula.

Aggregation is specified as linear additive aggregation. This preserves an interpretable compensatory structure while recognising that compensability is a substantive assumption that must be stress-tested against defensible alternatives in the empirical release. The mathematical hierarchy is defined below; the equations are design specifications, not evidence that the historical index was calculated.

$$S_{is} = \sum_j w_{js} z_{ij} \quad (5)$$

Sub-pillar score, with $\sum_j w_{js} = 1$.

$$P_{ip} = \sum_s v_{sp} S_{is} \quad (6)$$

Pillar score, with $\sum_s v_{sp} = 1$.

$$GGRI_i = \sum_p W_p P_{ip} \quad (7)$$

Final index, with $\sum_p W_p = 1$ and $W = (0.20, 0.25, 0.20, 0.20, 0.15)$.

Formal composite definition

$$GGRI_i = \sum_p W_p [\sum_s v_{sp} (\sum_j w_{js} z_{ij})] \quad (8)$$

This is the central mathematical definition of the proposed GGRI and shows the complete indicator → sub-pillar → pillar → index hierarchy without asserting that country-level scores have already been calculated.

Notation used in the GGRI equations

4.4 Statistical Validation

Because the supplied manuscript does not contain the underlying country-level dataset or reproducible code, this historical article specifies the validation protocol rather than reporting numerical validation results as established facts. The intended validation design includes global sensitivity analysis across indicator selection, normalisation, weighting and aggregation; uncertainty propagation for missing and imputed values; convergent-validity tests against contemporaneous governance and SDG measures; and discriminant-validity tests against conceptually adjacent constructs. Any reported coefficient, confidence interval, ranking or significance level in a future empirical release must be traceable to a versioned dataset and analysis script.

Formal validation specifications

For a country i , define the missingness rate among eligible indicators as:

$$m_i = M_i / J_i \quad (9)$$

M_i is the number of missing eligible observations and J_i is the total number of eligible observations. Under the proposed design, $\tau = 0.20$ is the missingness threshold.

Eligibility is then defined as:

$$\text{Eligible}_i = 1 \text{ if } m_i \leq \tau; 0 \text{ otherwise} \quad (10)$$

Sensitivity analysis should perturb the principal methodological choices over substantively defensible ranges and report rank correlations, rank-reversal rates, uncertainty intervals and variance-based sensitivity measures. No numerical threshold or finding is treated as validated merely because the methodology anticipates it. The purpose of this protocol is to make the eventual empirical implementation auditable rather than to imply that the results have already been obtained.

Convergent validity is specified as a hypothesis: GGRI scores should be positively associated with contemporaneous governance-performance and SDG-performance measures while remaining conceptually distinguishable from them. The appropriate comparison variables and their exact data vintages must be frozen at the historical cutoff. Pearson or Spearman correlations may be reported as appropriate to the distribution, accompanied by confidence intervals and a clear statement of sample size and missing-data treatment.

For a continuous comparison variable X , Pearson convergent-validity association may be represented as:

$$r(GGRI, X) = \text{Cov}(GGRI, X) / (\sigma_{GGRI} \sigma_X) \quad (11)$$

Discriminant validity is likewise specified prospectively. Candidate comparators include GDP per capita, the Human Development Index and other adjacent constructs. The analysis must not infer validity from an arbitrary universal correlation threshold; instead, it should justify expected relationships theoretically, report uncertainty, and explain whether the observed association is compatible with the claim that governance

readiness is related to, but not identical with, economic development, human development, peace or governance perception.

For predictive validity, a temporally ordered specification may be written as:

$$Y_{i,t+1} = \alpha + \beta \text{GGRI}_{i,t} + \gamma' X_{i,t} + \varepsilon_{i,t+1} \quad (12)$$

β is interpreted only after the outcome definition, time ordering, controls, model diagnostics, and reproducible data/code are frozen. No coefficient is reported here.

For robustness analysis, an alternative specification a can be compared with the baseline specification 0 using:

$$\Delta \text{GGRI}_i^{(a)} = \text{GGRI}_i^{(a)} - \text{GGRI}_i^{(0)} \quad (13)$$

Rank robustness can be evaluated analogously as $\Delta \text{Rank}_i^{(a)} = \text{Rank}_i^{(a)} - \text{Rank}_i^{(0)}$. Alternative specifications may perturb indicator inclusion, reference values, imputation, weights, or aggregation, with all choices documented.

5. The GGRI Composite Model: Pillars, Indicators, and Weighting

This section presents the proposed specification of the Global Governance Readiness Index, documenting the five pillars, their sub-pillars, principal candidate data sources, and structural weights. The presentation is designed to support replication of the framework; however, a complete empirical replication additionally requires the indicator-level dataset, exact source vintages, transformation log, PCA outputs, imputation files, and analysis code, none of which are included in the supplied historical manuscript.

5.1 Pillar 1: Legal-Institutional Framework

The Legal-Institutional Framework pillar (weight: 0.20) captures the formal legal architecture through which international commitments are incorporated into domestic law and the structural features of the constitutional and statutory order that condition implementation. The pillar is organised into three sub-pillars: Treaty Incorporation, Regulatory Quality, and Constitutional Architecture. The Treaty Incorporation sub-pillar measures the formal integration of international commitments into the domestic legal order, including the constitutional or statutory mechanism through which treaties acquire domestic legal force, the existence of implementing legislation for major treaty regimes, and the timeliness of treaty-body reporting. The principal indicators are the ratification status of the seventeen principal international human rights instruments, the incorporation status of the principal multilateral environmental agreements, and the timeliness of state-party reports to the treaty bodies.

The Regulatory Quality sub-pillar measures the structural features of the regulatory architecture that condition the implementation of commitment-related policy, including the use of regulatory impact assessment, the transparency of regulatory consultation, and the existence of ex-post evaluation mechanisms. The principal indicators are drawn from the OECD's Indicators of Regulatory Policy and Governance (iREG), which provide internationally comparable measures of regulatory-management practices, and from the pre-2023 World Bank regulatory-governance evidence; B-READY is excluded because its first report was not available by the manuscript cutoff for regulatory-enforcement indicators. For states outside the OECD sample, the V-Dem dataset's regulatory-quality indicators are used as a complementary source, with the substitution documented transparently.

The Constitutional Architecture sub-pillar measures the structural features of the constitutional order that condition the credibility of commitments, including the constitutional entrenchment of rights, the independence of constitutional review, and the existence of constitutional constraints on executive overreach. The principal indicators are drawn from the Comparative Constitutions Project and the V-Dem dataset. The sub-pillar is weighted lower than the Treaty Incorporation and Regulatory Quality sub-pillars, on the grounds that constitutional architecture is a slow-moving structural feature that conditions the long-horizon credibility of commitments but does not directly condition the implementation of specific commitments in the short term. The sub-pillar weights are 0.40 for Treaty Incorporation, 0.35 for Regulatory Quality, and 0.25 for Constitutional Architecture.

Empirical validation of this pillar is specified as a pre-specified hypothesis for the future pilot dataset. No numerical correlation is reported in this historical manuscript because the underlying country-level observations and reproducible analysis files are not included in the supplied source.

5.2 Pillar 2: Administrative Capacity

The Administrative Capacity pillar (weight: 0.25) captures the bureaucratic capability of the state to design, execute, and monitor the policies required to implement commitments. The pillar is weighted highest of the five, on the theoretical grounds that administrative capacity is the most directly operational dimension of readiness: it is the dimension through which commitments are translated into implemented programmes, and its deficit is the most proximate cause of implementation failure. The pillar is organised into three sub-pillars: Bureaucratic Professionalism, Digital Infrastructure, and Human Capital Depth.

The Bureaucratic Professionalism sub-pillar measures the structural features of the civil service that condition its capability, including meritocratic recruitment, career stability, and functional specialisation. The principal indicators are drawn from the V-Dem dataset's bureaucratic-quality indicators, the Quality of Government Institute's civil-service indicators, and the World Bank's Bureaucratic Effectiveness framework. The sub-pillar weights functional indicators (such as the proportion of senior officials recruited through competitive examination) more heavily than form indicators (such as the existence of a civil-service statute), in line with the 'looking like a state' critique of Pritchett, Woolcock, and Andrews (2013) and the recommendation that functional depth be weighted above organisational form.

The Digital Infrastructure sub-pillar measures the digital capability of the public administration, including the sophistication of e-government services, the integration of digital systems across government functions, and the digital literacy of the population. The principal indicators are drawn from the UN Department of Economic and Social Affairs' e-Government Survey, the Telecommunications Infrastructure Index, and ITU ICT statistics available by the historical cutoff. The inclusion of digital infrastructure as a distinct sub-pillar reflects the contemporary recognition that digital capability is a constitutive element of administrative capacity in the implementation of commitments that require large-scale data management, real-time monitoring, and citizen-facing service delivery, including the SDG monitoring framework and the Paris Agreement's enhanced transparency framework.

The Human Capital Depth sub-pillar measures the depth and quality of human capital in the public administration, including the educational attainment of public servants, the existence of specialised training in commitment-relevant domains, and the retention of skilled personnel. The principal indicators are drawn from the World Bank's Human Capital Index, the UNESCO Institute for Statistics for educational-attainment data, and the ILO for labour-market indicators relevant to public-sector human capital. The sub-pillar weights are 0.40 for Bureaucratic Professionalism, 0.35 for Digital Infrastructure, and 0.25 for Human Capital Depth. The weighting reflects the theoretical priority of bureaucratic professionalism as the most direct operational dimension of administrative capacity, with digital infrastructure as an increasingly consequential enabling

dimension, and human capital depth as a structural precondition that conditions the ceiling of bureaucratic capability.

Empirical validation of this pillar is specified as a pre-specified hypothesis for the future pilot dataset. No numerical correlation is reported in this historical manuscript because the underlying country-level observations and reproducible analysis files are not included in the supplied source.

5.3 Pillar 3: Fiscal Readiness

The Fiscal Readiness pillar (weight: 0.20) captures the fiscal headroom of the state to finance the implementation of commitments, including the structure of revenue mobilisation, the sustainability of public debt, and the alignment of budgetary allocation with commitment-related expenditure. The pillar is grounded in the Besley-Persson fiscal-capacity framework, which treats the state's capacity to raise revenues as a structural investment in the institutional infrastructure of taxation, and in the Addis Ababa Action Agenda's emphasis on domestic public resource mobilisation as the central pillar of SDG financing (United Nations, 2015). The pillar is organised into three sub-pillars: Revenue Mobilisation, Debt Sustainability, and Budget Quality.

The Revenue Mobilisation sub-pillar measures the capacity of the state to raise revenues domestically, including the tax-to-GDP ratio, the share of direct taxes in total revenue, and the effectiveness of tax administration. The principal indicators are drawn from the IMF's Government Finance Statistics, the World Bank's World Development Indicators for tax-revenue ratios, and the OECD Revenue Statistics for tax-structure indicators. The sub-pillar gives particular weight to the share of direct taxes in total revenue, on the theoretical grounds that direct taxation is the most demanding form of revenue mobilisation and the most consequential for the development of fiscal capacity, as emphasised by Besley and Persson (2009). The sub-pillar also incorporates indicators of tax administration effectiveness, including the tax-compliance gap and the efficiency of tax-collection procedures.

The Debt Sustainability sub-pillar measures the sustainability of the public-debt trajectory, including the debt-to-GDP ratio, the share of external debt in total debt, and the cost of debt servicing as a proportion of revenue. The principal indicators are drawn from the IMF's World Economic Outlook database, the World Bank's International Debt Statistics, and the IMF's Debt Sustainability Framework for low-income countries. The sub-pillar weights debt-to-GDP more heavily than external-debt share, on the grounds that the debt-to-GDP ratio is the more directly relevant indicator of fiscal headroom for commitment implementation, while the external-debt share captures the additional vulnerability associated with foreign-currency denomination. The cost-of-debt-servicing indicator captures the crowding-out effect of high debt servicing on the fiscal space available for commitment-related expenditure.

The Budget Quality sub-pillar measures the quality of public financial management, including the credibility of the budget, the comprehensiveness of fiscal reporting, and the alignment of budget allocation with commitment-related priorities. The principal indicators are drawn from the PEFA Secretariat's assessments, which provide internationally comparable measures of public financial management quality across seven dimensions of performance. The sub-pillar gives particular weight to the budget-credibility dimension, which measures the gap between budgeted and executed expenditure, on the theoretical grounds that this dimension is the most directly relevant to the implementation of commitments, which depends on the conversion of budgeted allocations into executed programmes. The sub-pillar weights are 0.40 for Revenue Mobilisation, 0.30 for Debt Sustainability, and 0.30 for Budget Quality.

Empirical validation of this pillar is specified as a pre-specified hypothesis for the future pilot dataset. No numerical correlation is reported in this historical manuscript because the underlying country-level

observations and reproducible analysis files are not included in the supplied source.

5.4 Pillar 4: Compliance Infrastructure

The Compliance Infrastructure pillar (weight: 0.20) captures the institutional mechanisms through which compliance with commitments is monitored, enforced, and remediated, including the independence and capability of the judiciary, the structural position of audit and oversight institutions, and the integration of commitment-specific monitoring mechanisms into the domestic institutional architecture. The pillar operationalises the impartiality dimension of the Rothstein-Teorell framework and the transparency-and-dispute-resolution dimension of the Chayes-Chayes managerial model. The pillar is organised into three sub-pillars: Judicial Independence, Oversight Institutions, and Monitoring Integration.

The Judicial Independence sub-pillar measures the structural features of the judicial system that condition its capacity to enforce compliance with commitments, including the constitutional guarantees of judicial independence, the structural insulation of the judiciary from executive interference, and the procedural framework for judicial review of executive action. The principal indicators are drawn from the World Justice Project's Rule of Law Index, the V-Dem dataset's judicial-independence indicators, and the Comparative Constitutions Project for constitutional-entrenchment indicators. The sub-pillar weights structural indicators (such as the existence of constitutional guarantees of judicial independence and the structural insulation of judicial appointments) more heavily than procedural indicators (such as the existence of judicial-review procedures), on the grounds that structural features are more robust to political erosion than procedural features.

The Oversight Institutions sub-pillar measures the structural position and capability of audit institutions, anti-corruption agencies, human-rights commissions, and other independent oversight bodies whose function is to monitor compliance and remediate violations. The principal indicators are drawn from the International Budget Partnership's Open Budget Survey, the International Organisation of Supreme Audit Institutions (INTOSAI) for audit-institution indicators, and the V-Dem dataset for oversight-body indicators. The sub-pillar gives particular weight to the independence and capability of supreme audit institutions, on the grounds that these institutions are the most consequential for fiscal-oversight compliance, and to the structural position of human-rights commissions, on the grounds that these institutions are the most consequential for human-rights compliance.

The Monitoring Integration sub-pillar measures the integration of commitment-specific monitoring mechanisms into the domestic institutional architecture, including the existence of national SDG monitoring frameworks, the integration of NDC monitoring into national climate reporting, and the structural channels through which treaty-body recommendations are transmitted into domestic policy processes. The principal indicators are drawn from the SDG Indicators database, the UNFCCC's national communications and biennial reports, and the UN human-rights mechanisms' Universal Periodic Review database. The sub-pillar is included to capture the institutional integration of commitment-specific monitoring, which is a structural feature of compliance infrastructure that is not captured by general governance measures. The sub-pillar weights are 0.40 for Judicial Independence, 0.35 for Oversight Institutions, and 0.25 for Monitoring Integration.

Empirical validation of this pillar is specified as a pre-specified hypothesis for the future pilot dataset. No numerical correlation is reported in this historical manuscript because the underlying country-level observations and reproducible analysis files are not included in the supplied source.

5.5 Pillar 5: Stakeholder Engagement

The Stakeholder Engagement pillar (weight: 0.15) captures the structural inclusion of non-state actors—civil-society organisations, private-sector entities, sub-national governments, and marginalised populations—in the design and implementation of commitment-related policy. The pillar is grounded in the Ostrom framework's emphasis on polycentric governance and the Evans framework's emphasis on embedded autonomy, and in the 2030 Agenda's commitment to leave no one behind. The pillar is weighted lowest of the five, on the theoretical grounds that stakeholder engagement is a constitutive but not a sufficient condition for readiness, and that its contribution to implementation is mediated through the other pillars rather than operating directly. The pillar is organised into three sub-pillars: Civic Space, Partnership Architecture, and Inclusion.

The Civic Space sub-pillar measures the structural conditions for civil-society participation, including the legal framework for civil-society organisation, the freedom of association and expression, and the practical conditions for civic mobilisation. The principal indicators are drawn from the CIVICUS Monitor, the V-Dem dataset's civil-liberties indicators, and the World Justice Project's fundamental-rights dimension. The sub-pillar weights structural indicators (such as the legal framework for civil-society organisation) more heavily than outcome indicators (such as the number of civil-society organisations), on the grounds that structural conditions are more robust to measurement variation and more directly relevant to the capacity for sustained engagement.

The Partnership Architecture sub-pillar measures the structural channels through which non-state actors are integrated into commitment-related policy processes, including the existence of formal consultation mechanisms for SDG implementation, the integration of private-sector actors into climate-policy processes, and the structural channels for sub-national-government participation in national policy processes. The principal indicators are drawn from the Open Government Partnership's Independent Reporting Mechanism, the SDG Indicators database for partnership indicators, and the V-Dem dataset for consultation-process indicators. The sub-pillar gives particular weight to the existence of formal, institutionalised consultation mechanisms, on the grounds that ad-hoc consultation is more susceptible to political manipulation and less reliable as an indicator of structural inclusion.

The Inclusion sub-pillar measures the structural inclusion of marginalised populations in commitment-related policy, including the legal framework for non-discrimination, the existence of special measures for marginalised populations, and the disaggregation of monitoring data by relevant demographic dimensions. The principal indicators are drawn from the World Bank's Women, Business and the Law database, the SDG Indicators database for disaggregation indicators, and the V-Dem dataset for inclusion indicators. The sub-pillar is included to operationalise the 2030 Agenda's commitment to leave no one behind, which is understood in this framework as a structural feature of readiness rather than as a normative aspiration. The sub-pillar weights are 0.35 for Civic Space, 0.35 for Partnership Architecture, and 0.30 for Inclusion.

5.6 Aggregation and Final Index

Figure 3. Proposed structural weights of the five GGRI pillars. The values are the weights explicitly specified in the manuscript and sum to 1.00.

The aggregation of the five pillar scores into the final GGRI score uses the structural weights specified above: 0.20 for Legal-Institutional Framework, 0.25 for Administrative Capacity, 0.20 for Fiscal Readiness, 0.20 for Compliance Infrastructure, and 0.15 for Stakeholder Engagement. The weights sum to unity. For a completed empirical release, the normalisation reference values, directionality rules, treatment of outliers, and any percentile-based reference points must be fixed before scores are calculated and disclosed in the technical appendix. The present manuscript specifies the rule but does not claim that a final 60-state score set has been independently reproduced.

With the five structural pillar weights specified in the manuscript, Equation (7) can also be written explicitly as:

$$\text{GGRI}_i = 0.20P_{i1} + 0.25P_{i2} + 0.20P_{i3} + 0.20P_{i4} + 0.15P_{i5} \quad (14)$$

This is a definitional restatement of the proposed structural weights, not a reported empirical score calculation.

The composite architecture of the GGRI is designed to permit diagnostic decomposition at multiple levels. At the pillar level, the index can be decomposed to identify the structural dimensions in which a state's readiness is concentrated or deficient; at the sub-pillar level, it can identify the institutional features conditioning readiness; and at the indicator level, it can identify the data points driving a completed score. These are design properties of the proposed framework. A technical appendix containing the indicator-level dataset, source-vintage matrix, transformation log, imputation outputs and code would be required for an empirical release and is not part of the supplied historical manuscript. The formal hierarchy is defined in Equations (5)–(8), while Equation (14) makes the five structural pillar weights explicit.

The weighting scheme reflects a deliberate analytical judgement about the relative contribution of each pillar to the overall construct of governance readiness. The assignment of the highest weight to Administrative Capacity (0.25) reflects the theoretical priority, established in the Fukuyama and Evans frameworks, of bureaucratic capability as the most directly operational dimension of implementation. The equal weighting of Legal-Institutional Framework, Fiscal Readiness, and Compliance Infrastructure (each 0.20) reflects the structural complementarity between legal and fiscal capacity emphasised by Besley and Persson and the constitutive role of compliance infrastructure in sustaining implementation over time. The lowest weighting of Stakeholder Engagement (0.15) reflects the theoretical judgement that inclusion is a constitutive but mediating rather than directly operational dimension of readiness. Alternative weighting schemes and their effects on country rankings should be tested only after the empirical dataset and code are frozen; no such robustness result is claimed here.

Table 2 provides the structural architecture of the proposed GGRI and makes the weighting logic auditable without implying that an empirical country score set has been calculated.

Table 2. Proposed GGRI pillar architecture and source pools. Source pools are candidate inputs identified in the manuscript; they are not evidence that a completed country-level score set exists.

6. Pilot Application Design and Empirical Validation Protocol

6.1 Sample Construction

The proposed pilot is a stratified sample designed to test whether the GGRI can be calculated consistently across different institutional, geographic, income and regime contexts. The sample should be frozen before calculation, with an explicit country list, inclusion criteria, source-vintage matrix and missing-data rule. The manuscript does not treat the previously stated sixty-state numerical results as verified because the country-level dataset and selection file are not included in the supplied evidence.

A defensible pilot should document the inclusion of major economies together with smaller and data-constrained states, while avoiding claims of statistical representativeness unless the sampling design and population weights justify them. Income group, geographic region, regime classification and commitment

profile should be recorded as descriptive stratification variables rather than used to imply causal representativeness.

6.2 Pre-Specified Descriptive Analysis

The empirical implementation should report the number of observations, country coverage by pillar, missingness by indicator, imputation rates, distribution of each indicator after normalisation, pillar means and dispersion, and the resulting GGRI distribution. Country scores, rankings and cluster assignments should be reported only from a versioned dataset that can be independently reconstructed. The numerical range, mean, standard deviation and country-level rankings stated in the earlier manuscript are therefore withdrawn from this historical version.

Descriptive interpretation should distinguish observed score differences from theoretical interpretation. Statements that particular countries form 'high-readiness', 'transitional' or 'deficit' groups should follow the empirical results rather than precede them. Any thresholds used for classification must be pre-specified and sensitivity-tested.

Pillar-level relationships should be examined using the pre-specified statistical procedure and reported with sample size and uncertainty. The previous manuscript's numerical pillar-to-index and GDP correlations are withdrawn because their underlying calculation files are not available for audit.

Inter-pillar correlations may be used to assess whether the five dimensions are empirically related while remaining sufficiently distinguishable. These correlations should be interpreted as descriptive evidence about the proposed measurement architecture, not as proof of causal relationships among institutional dimensions.

6.3 Optional Cluster Analysis

If cluster analysis is retained in a future empirical release, the number of clusters, distance metric, standardisation, initialisation procedure and stability checks must be specified in advance. Cluster labels must be data-derived and should not be assigned before examining the results. The previously reported five-cluster solution and country counts are withdrawn from this historical version because the underlying observations are unavailable.

Any cluster interpretation should be accompanied by robustness checks, including alternative cluster counts and sensitivity to influential observations. Cluster membership must not be treated as a causal typology or as evidence that countries with similar scores share identical institutional mechanisms.

6.4 Validation Against Existing Measures

Convergent-validity testing should compare the GGRI with contemporaneous governance and SDG measures available by 15 March 2023, using frozen source vintages and a documented common country sample. The previously reported correlations of 0.78 and 0.82 are withdrawn from this version because they cannot be independently reproduced from the supplied manuscript.

Predictive validity requires a genuinely longitudinal design. A defensible test would freeze a GGRI baseline year, specify a subsequent outcome window, define the compliance outcome independently of the GGRI construction, and pre-specify controls and model diagnostics. The earlier manuscript's claim that a 2015 GGRI predicted 2015–2022 compliance is not retained, because the manuscript does not supply evidence that the GGRI was actually constructed in 2015 or that the historical scores and compliance dataset are reproducible.

Country-level diagnostic examples should be added only after the underlying scores are audited. Illustrative country narratives cannot substitute for a reproducible empirical dataset and must not be presented as observed GGRI findings.

Sensitivity and uncertainty analysis should be run after the complete indicator-level dataset is frozen. The analysis should report how rankings and scores change under alternative normalisation, weighting, imputation and aggregation assumptions. The numerical sensitivity indices and confidence bands previously reported are withdrawn because their source calculations are not included in the supplied manuscript.

Accordingly, this section establishes a reproducible empirical protocol rather than asserting completed findings. This distinction is essential to preserve the integrity of the claimed 15 March 2023 publication date.

Table 3. Empirical validation protocol. No row constitutes a completed result in the historical manuscript.

Reproducibility gate for any empirical release. Before any GGRI country ranking, correlation, regression, cluster result, sensitivity statistic, uncertainty interval or predictive-validity claim is published, the release should contain: (1) a versioned country-level dataset; (2) a source-vintage matrix; (3) an indicator dictionary with directionality and transformations; (4) imputation outputs and diagnostics; (5) PCA/loadings and weighting outputs; (6) aggregation code; (7) sensitivity and uncertainty code; (8) a machine-readable results file; and (9) a frozen environment specification. This gate is necessary to preserve the distinction between a historical methodological article and a later empirical release.

7. Discussion: Counterarguments, Limitations, and Implications

7.1 Interpretation and Policy Implications

The methodological contribution of the article is the explicit definition and operational architecture of governance readiness. The GGRI is proposed as a complement to performance and perception measures, not as a demonstrated superior predictor. Its policy usefulness therefore depends on successful empirical validation, replication and longitudinal testing in future releases. This distinction prevents the methodological proposal from being presented as an already-validated forecasting instrument.

For development finance institutions, the GGRI provides a measurement of the institutional capacity that conditions the absorption and effective use of development finance, supporting the differentiation of financing modalities across the readiness spectrum. If future validation supports the readiness construct, higher-scoring states could be considered for greater use of programmatic financing, while lower-scoring states might warrant financing packages paired with capacity-building; such applications would remain contingent on empirical validation and context-specific assessment. The differentiation of financing modalities along the readiness spectrum has the potential to improve the development effectiveness of concessional finance, a concern that has acquired renewed urgency in the aftermath of the COVID-19 pandemic and the associated deterioration of fiscal headroom in many low- and middle-income states.

For states and other users, the GGRI is intended to support structured diagnosis once an audited score set exists. Pillar and sub-pillar decomposition can identify candidate institutional constraints, but such diagnoses should be treated as measurement outputs rather than causal explanations unless supported by additional evidence.

The implications for the multilateral system extend beyond the operational domains of treaty secretariats and development finance institutions. The High-Level Political Forum on Sustainable Development, the principal UN platform for the follow-up and review of the 2030 Agenda, has repeatedly emphasised the importance of capacity-building as a means of implementation, but has lacked a robust instrument for the diagnosis of capacity deficits. A validated future GGRI release could provide such a complementary instrument, supporting the Forum's review function with a readiness-based diagnosis alongside the outcome-based review of the SDG Indicators. Similarly, the enhanced transparency framework of the Paris Agreement, which requires states to report on their implementation of nationally determined contributions, can be strengthened by the integration of readiness indicators into the reporting architecture, permitting the Conference of the Parties to identify states whose compliance trajectories are at risk before the deficits materialise in the biennial reports. The multilateral system could potentially benefit from a readiness measure that is structurally prior to the outcome measures on which it currently relies, subject to empirical validation and institutional uptake.

7.2 Counterarguments

Three counterarguments to the GGRI warrant consideration. The first, advanced from the perception-measurement tradition, holds that structural indicators cannot capture the dimension of governance quality that perception measures capture, because governance quality is partly constituted by the lived experience of citizens and the operational experience of firms, neither of which is fully captured by structural indicators of legal architecture or administrative capability. This counterargument has merit: perception measures capture a dimension of governance that structural measures do not, and the GGRI does not claim to replace perception measures but to complement them. Whether the two measures capture related but distinct constructs is an empirical question for the validation protocol specified in Section 4.4; the historical manuscript does not report a completed correlation analysis. The appropriate response to the counterargument is not to abandon structural measurement but to use it alongside perception measurement, with each measure contributing its distinctive signal.

The second counterargument, advanced from the construct-validity tradition, holds that the GGRI's pillar architecture is a theoretically convenient decomposition of governance capacity rather than an empirically validated structure, and that alternative decompositions would produce materially different scores. This counterargument is partially supported by the literature on composite-index construction, which has documented the sensitivity of composite scores to the choice of decomposition (Becker et al., 2017; Saltelli, 2007). The GGRI's response to the counterargument is twofold. First, the pillar architecture is grounded in a deliberate theoretical synthesis, as documented in Section 3, and is not an arbitrary decomposition. Second, the sensitivity of GGRI scores and rankings to alternative weighting schemes is an explicit test in the validation protocol. No robustness result is claimed in this historical manuscript because the underlying empirical calculations are not supplied. The robustness of the rankings to alternative decompositions is a question for future empirical investigation, and is acknowledged as a limitation in Section 7.3.

The third counterargument, advanced from the post-colonial and critical-development tradition, holds that composite indices of governance are instruments of Northern epistemic authority that impose a universalised standard of governance on Southern states and that legitimate interventionist agendas under the guise of measurement (Rajak, 2011; Sachs, 2019). This counterargument raises substantive concerns that the GGRI takes seriously. The index's commitment-anchored orientation is a partial response: by measuring readiness against the international commitments that states have themselves accepted, rather than against an external standard of governance quality, the GGRI reduces the imposition of external normative standards. The index's reliance on primary and authoritative secondary data sources, rather than on commercial risk ratings that may reflect investor priorities, is a further response. The index's transparency and replicability, including the full documentation of the indicator selection, weighting, and aggregation choices, is a further response,

permitting critical scrutiny of the index's construction. The GGRI does not claim to be normatively neutral, but it does claim to be methodologically transparent and theoretically grounded, and to be usable as a diagnostic instrument by states themselves rather than only by external actors.

7.3 Limitations

First, the proposed indicators are imperfect proxies for institutional features. Several sources rely partly or wholly on expert assessments, and some administrative and public-finance measures are infrequent. The GGRI therefore cannot eliminate measurement error merely by combining indicators. Second, the supplied manuscript does not include the complete country-level dataset, transformation log, imputation output or executable analysis code. Consequently, the numerical findings in the earlier version cannot be certified as reproducible and are not retained in this rectified historical version.

Third, coverage constraints are likely to be greatest for small, fragile and data-sparse states. Any future expansion toward all UN member states should report coverage, missingness and imputation uncertainty explicitly rather than implying uniform measurement quality across countries.

Fourth, the structural orientation of the GGRI does not capture every short-term political or contextual determinant of implementation. Political instability, elite capture, conflict, exogenous shocks and policy discontinuity may alter implementation even where structural readiness is comparatively strong. The GGRI should therefore be interpreted alongside, not instead of, complementary governance and outcome measures.

Fifth, the five-pillar decomposition and weighting scheme are substantive analytical choices. They are transparent and theoretically motivated but not uniquely determined by theory. Their empirical adequacy must be tested through sensitivity analysis, stakeholder scrutiny and longitudinal validation. The present article does not claim that its chosen weights are the only or objectively correct weights.

8. Conclusion and Future Research

This article develops the Global Governance Readiness Index as a composite measurement framework for the institutional capacity of states to deliver accepted international commitments. Its contribution is to define readiness as a distinct analytical construct and to specify a five-pillar architecture covering legal-institutional, administrative, fiscal, compliance and stakeholder dimensions. The framework is grounded in established scholarship and follows the logic of transparent composite-indicator construction.

The central proposition is methodological rather than predictive: governance readiness is a potentially measurable property that complements observed performance and perception. The rectified historical version does not claim that the GGRI has already been demonstrated to be a superior predictor of compliance, because the supplied manuscript lacks the underlying dataset and reproducible analysis required to substantiate that conclusion.

The GGRI is positioned as an analytical foundation for subsequent AUN governance research. That institutional positioning is distinct from empirical validation: future AUN publications may apply or extend the framework, but such later work must not be used as evidence for the historical empirical claims of this 15 March 2023 article.

Future research should proceed in four directions: first, release a complete country-level dataset and codebook; second, publish the transformation, imputation, weighting and uncertainty scripts required for exact replication; third, extend the index longitudinally and test predictive validity against independently

defined compliance outcomes; and fourth, expand coverage while quantifying data-quality differences across countries and regions.

The GGRI is therefore offered as a transparent, testable measurement framework rather than as a completed claim of predictive superiority. Its scientific value will depend on reproducibility, open scrutiny, appropriate historical data vintages and successful out-of-sample validation. Those requirements are deliberately stated here so that subsequent empirical editions can be assessed against a clear methodological baseline.

References

- Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity, and poverty*. Crown Publishers.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1A, pp. 385–472). Elsevier.
- Andrews, M. (2013). *The limits of institutional reform in development: Changing rules for realistic solutions*. Cambridge University Press.
- Arndt, C., & Oman, C. (2010). Use and misuse of governance indicators. In *The politics of governance indicators* (pp. 1–28). OECD Publishing.
- Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, 80, 12–22.
- Bertelsmann Stiftung. (2022). *Transformation Index BTI 2022: Governance in international comparison*. Verlag Bertelsmann Stiftung.
- Besley, T., & Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, 99(4), 1216–1252.
- Chayes, A., & Chayes, A. (1993). On compliance. *International Organization*, 47(2), 175–205.
- Chayes, A., & Chayes, A. (1995). *The new sovereignty: Compliance with international regulatory agreements*. Harvard University Press.
- Cole, W. M. (2015). Mind the gap: State capacity and the implementation of international treaties. *International Organization*, 69(3), 559–599.
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., et al. (2023). *V-Dem Codebook v13*. Varieties of Democracy Institute, University of Gothenburg.
- Downs, G. W., Rocke, D. M., & Barsoom, P. N. (1996). Is the good news about compliance good news about cooperation? *International Organization*, 50(3), 379–406.
- Esty, D. C., Levy, M., Srebotnjak, T., & de Sherbinin, A. (2005). *Environmental sustainability index: Benchmarking national environmental stewardship*. Yale Center for Environmental Law and Policy.
- Evans, P. (1995). *Embedded autonomy: States and industrial transformation*. Princeton University Press.
- Franck, T. M. (1990). *The power of legitimacy among nations*. Oxford University Press.
- Fukuyama, F. (2013). What is governance? *Governance*, 26(3), 347–368.

- Fund for Peace. (2022). *Fragile States Index 2022: Annual report*. Fund for Peace.
- Hanson, J. K., & Sigman, R. (2021). Measuring state capacity: Theoretical and empirical considerations. *Journal of Peace Research*, 58(4), 587–600.
- Hendrix, C. S. (2010). *Measuring state capacity: A theoretical and empirical exploration*. Paper prepared for the ISA Annual Convention.
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The Worldwide Governance Indicators: Methodology and analytical issues* (World Bank Policy Research Working Paper 5430). World Bank.
- Kaufmann, D., Kraay, A., & Zoido-Lobaton, P. (1999). *Aggregating governance indicators* (World Bank Policy Research Working Paper 2195). World Bank.
- Keohane, R. O. (1984). *After hegemony: Cooperation and discord in the world political economy*. Princeton University Press.
- Kurtz, M. J., & Schrank, A. (2007). Growth and governance: Models, measures, and mechanisms. *Journal of Politics*, 69(2), 538–554.
- March, J. G., & Olsen, J. P. (1995). *Democratic governance*. Free Press.
- Migdal, J. S. (1988). *Strong societies and weak states: State-society relations and state capabilities in the Third World*. Princeton University Press.
- Mo Ibrahim Foundation. (2022). *2022 Ibrahim Index of African Governance: Index report*. Mo Ibrahim Foundation.
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing.
- North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Ostrom, E. (2005). *Understanding institutional diversity*. Princeton University Press.
- Pritchett, L., Woolcock, M., & Andrews, M. (2013). Looking like a state: Techniques of persistent failure in state capability for implementation. *Journal of Development Studies*, 49(7), 889–909.
- Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy*. Princeton University Press.
- Rajak, D. (2011). *In good company: An anatomy of corporate social responsibility*. Stanford University Press.
- Rothstein, B., & Teorell, J. (2008). What is quality of government? A theory of impartial government institutions. *Governance*, 21(2), 165–190.
- Sachs, J. D., Schmidt-Traub, G., Kroll, C., Lafortune, G., & Fuller, G. (2022). *Sustainable development report 2022*. Cambridge University Press.
- Sachs, W. (2019). *The development dictionary: A guide to knowledge as power* (2nd ed.). Zed Books.
- Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society: Series A*, 168(2), 307–323.

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, 81(1), 65–77.

Sen, A. (1999). *Development as freedom*. Oxford University Press.

Simmons, B. A. (1998). Compliance with international agreements. *Annual Review of Political Science*, 1, 75–93.

Thomas, M. A. (2010). What do the Worldwide Governance Indicators measure? *European Journal of Development Research*, 22(1), 31–54.

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development (A/RES/70/1)*. United Nations.

United Nations. (2022). *The Sustainable Development Goals report 2022*. United Nations Department of Economic and Social Affairs.

UNEP. (2022). *Emissions gap report 2022: The closing window*. United Nations Environment Programme.

UNFCCC. (2015). *Paris Agreement*. United Nations Framework Convention on Climate Change.

World Bank. (2022). *Worldwide Governance Indicators: 2022 update*. World Bank.

Evidence domain	Historical source-vintage rule	Status in this version
Governance / regulation	Use editions publicly available by 15 March 2023; later B-READY material excluded.	Framework source pool only
Democracy / institutional structure	V-Dem v13 (March 2023) is within the cutoff.	Framework source pool only
Governance comparison	Use the WGI release available by the cutoff; later revisions are not used as historical evidence.	Validation comparator specified, not calculated
SDG performance	Use contemporaneous SDG evidence available by the cutoff; later 2023 reports are excluded.	Validation comparator specified, not calculated
Climate implementation context	Use contemporaneous climate-gap evidence available by the cutoff; later 2023 report is excluded.	Context only

Symbol	Meaning	Level
i	country/state	observation unit
j	indicator	indicator

s	sub-pillar	sub-pillar
p	pillar	pillar
x_{ij}	raw indicator observation	indicator
z_{ij}	normalised 0–100 indicator score	indicator
w_{js}	indicator weight; $\sum w_{js} = 1$	sub-pillar
v_{sp}	sub-pillar weight; $\sum v_{sp} = 1$	pillar
W_p	structural pillar weight; $\sum W_p = 1$	GGRI
S_{is}	sub-pillar score	sub-pillar
P_{ip}	pillar score	pillar
$GGRI_i$	overall GGRI score	index
$R_s, \lambda_{sk}, q_{sk}$	PCA correlation matrix, eigenvalue, eigenvector	PCA
ℓ_{js1}	first-component loading	PCA/weighting

Pillar	Pillar weight	Sub-pillars (weights)	Principal candidate source pools
Legal-Institutional Framework	0.20	Treaty Incorporation (0.40); Regulatory Quality (0.35); Constitutional Architecture (0.25)	Women, Business and the Law; OECD iREG; V-Dem v13; UN Treaty Collection; Comparative Constitutions Project
Administrative Capacity	0.25	Bureaucratic Professionalism (0.40); Digital Infrastructure (0.35); Human Capital Depth (0.25)	UN e-Government Survey; V-Dem; World Bank HCI/development indicators; ITU; ILO
Fiscal Readiness	0.20	Revenue Mobilisation (0.40); Debt Sustainability (0.30); Budget Quality (0.30)	IMF Government Finance Statistics; World Bank WDI; OECD Revenue Statistics; PEFA

Compliance Infrastructure	0.20	Judicial Independence (0.40); Oversight Institutions (0.35); Monitoring Integration (0.25)	World Justice Project; V-Dem; International Budget Partnership; UNODC; SDG/UNFCCC/UN human-rights monitoring
Stakeholder Engagement	0.15	Civic Space (0.35); Partnership Architecture (0.35); Inclusion (0.30)	V-Dem; OGP IRM; CIVICUS Monitor; SDG Indicators; Women, Business and the Law

Validation question	Pre-specified test	Required evidence before a numerical claim
Construct dimensional coherence /	Pillar and sub-pillar diagnostics; inspect loading structure and conceptual overlap.	Versioned indicator matrix and PCA outputs
Convergent validity	Compare with contemporaneous governance and SDG measures using Pearson or Spearman correlations as appropriate.	Common-country sample, source vintages, coefficients, confidence intervals, missing-data treatment
Discriminant validity	Compare with adjacent constructs such as GDP per capita and HDI without imposing an arbitrary universal threshold.	Theoretical expectation, coefficients, uncertainty and interpretation
Sensitivity robustness /	Perturb indicator selection, normalisation, weighting and aggregation; report rank correlations and rank reversals.	Complete sensitivity specification and reproducible outputs
Uncertainty	Propagate missing-data and imputation uncertainty and report uncertainty intervals.	Imputation files, assumptions and uncertainty calculations
Predictive validity	Test whether readiness predicts later implementation/compliance outcomes in a temporally ordered design.	Pre-specified outcome definition, time ordering, model code and results