Strengthening good governance in low- and middle-income countries
An evidence gap map

July 2023
About 3ie

The International Initiative for Impact Evaluation (3ie) develops evidence on how to effectively transform the lives of the poor in low- and middle-income countries. Established in 2008, we offer comprehensive support and a diversity of approaches to achieve development goals by producing, synthesizing and promoting the uptake of impact evaluation evidence. We work closely with governments, foundations, NGOs, development institutions and research organizations to address their decision-making needs. With offices in Washington DC, New Delhi and London and a global network of leading researchers, we offer deep expertise across our extensive menu of evaluation services.

3ie evidence gap map reports

3ie evidence gap maps are thematic collections of information about impact evaluations or systematic reviews that measure the effects of international development policies and programmes. The maps provide a visual display of completed and ongoing systematic reviews and impact evaluations in a sector or sub-sector, structured around a framework of interventions and outcomes.

The evidence gap map reports provide all the supporting documentation for the maps, including the background information for the theme of the map, the methods and results, protocols, and the analysis of results.

About this evidence gap map report

This report presents the findings of a systematic search to identify and map the evidence base of impact evaluations and systematic reviews of interventions that aim to promote governance effectiveness in low- and middle-income countries. The EGM was developed by 3ie and made possible with generous support from the United States Agency for International Development (USAID)’s Center for Democracy, Human Rights and Governance, via a partnership with NORC at the University of Chicago. All of the content of this report is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its Board of Commissioners. Any errors and omissions are also the sole responsibility of the authors. Please direct any comments or queries to the corresponding author, Constanza Gonzalez Parrao, cgonzalez@3ieimpact.org


Executive editor: Birte Snilstveit
Managing editor: Tanvi Lal
Publications and web design: Akarsh Gupta
Production manager: Mallika Rao

© International Initiative for Impact Evaluation (3ie), 2023
Strengthening good governance in low- and middle-income countries: an evidence gap map

Constanza Gonzalez Parrao
International Initiative for Impact Evaluation (3ie)

Etienne Lwamba
3ie

Lina Khan
3ie

Ashiqun Nabi
3ie

Malte Lierl
German Institute of Global and Area Studies

Miriam Berretta
3ie

Jane Hammaker
3ie

Charlotte Lane
3ie

Katherine Quant
3ie

John Eyers
3ie

Douglas Glandon
3ie

Evidence Gap Map Report 28
July 2023
Acknowledgement

We are grateful to 3ie colleagues who provided inputs at various stages of the EGM and supported its administrative, communication and finance management. We are particularly grateful to USAID’s Center for Democracy, Human Rights, and Governance (DRG Center) and NORC at the University of Chicago for their support to the production of this review. We acknowledge and appreciate the contribution of the advisory group for this EGM: Camilla Rocca, Mo Ibrahim Foundation; Chas Cadwell, Urban Institute; Eddy Malesky, Duke University; Franklin De Vrieze, Westminster Foundation for Democracy; Kevin McLaughlin, Tetra Tech; Kristen Sample, National Democratic Institute; Luke Waggoner, National Democratic Institute; Peter van der Windt, New York University, Abu Dhabi; Sally Wallace, Georgia State University; Sarah Leddy, Tetra Tech; Susan Kemp, Creative Associates International. We also appreciate the excellent work of the research assistants of this EGM: Adedayo Olofinyehun, Aisha Shalash, Alexandra Lanier, Anika Muzib Suchi, Carolina Lozano, Finagon Hospice Appolinaire, Francis Chege, Kiara Castaman, Maciej Respekta, and Sanghwa Lee.
Executive Summary

This Evidence Gap Map (EGM) report presents the findings of a systematic search to identify and map the evidence base of interventions focusing on governance effectiveness in low- and middle-income countries (L&MICs). The principle of good governance has emerged as a priority and a driver of success in international development. Through good governance, better political decisions and quality of public services may bring impact to a majority of citizens and not to an elite minority (FCDO, 2019a). The theory of good governance is nonetheless challenged by the reality of public decision-making, policymaking, service delivery and resource management. Decreasing levels of freedom, as well as corruption, non-optimal management of public resources and services, and low levels of accountability and transparency are some of the challenges to the principles of good governance in L&MICs. Moreover, the concept of good governance itself lacks consensus around its definition, implementation, and measures of impact.

Though previous synthesis efforts have considered public administration as it relates to citizens’ input, they have not differentiated interventions across transparency, accountability, and oversight mechanisms; or have covered other aspects of good governance, such as public financial or administrative management. This EGM focuses on interventions related to government institutions and included interventions from, through and/or directed towards those institutions. This allows the EGM to primarily focus on government effectiveness, changes to the way governments work, and the architecture of public service delivery.

The EGM builds on existing theories of change as we consider interventions that support good governance through governance effectiveness to achieve longer-term impacts on public and social wellbeing and growth. Interventions targeting transparency and monitoring, the creation of participation opportunities, capacity building and administrative management, and institutional reforms shape the framework of this EGM.

Methods

We implemented a comprehensive search for quantitative impact evaluations, specific qualitative evaluations (IEs), and systematic reviews (SRs) covering four academic databases and 46 grey literature sources to identify relevant studies. We complemented this search by conducting forward and backward citation tracking of included studies and the solicitation of relevant papers from stakeholders and the public.

We used EPPI-Reviewer to manage the EGM process, including mass-deduplication of the search results, independent double screening of studies at title and abstract, full text retrieval, and independent double screening of records at full text. To improve the efficiency of the screening process, we also used machine learning at the title and abstract stage.

Using 3ie’s EGM platform, we created an online interactive map according to the interventions evaluated and outcomes reported in included studies. The platform incorporates filters to explore the evidence by regions, populations, democracy levels, among other criteria. The map can be viewed at https://developmentevidence.3ieimpact.org/egm/good-governance-through-government-effectiveness-evidence-gap-map.
Main findings

The search for evidence conducted in December 2021 and January 2022 yielded a total of 98,625 potentially relevant studies. After the removal of duplicates and screening of records, we included 504 unique studies covered by 643 papers: 465 quantitative evaluations, 19 qualitative evaluations, one study that used both quantitative and qualitative designs, and 19 systematic reviews. The EGM studies were published between 1999 and 2022, with an increase in the volume of evidence published, particularly in the last decade.

Most evaluations were conducted in East Asia and the Pacific, Latin America and the Caribbean, and Sub-Saharan Africa, which respectively covered 34%, 29% and 25% of the evidence. At the country level, included studies were mainly conducted in China, Brazil and India. Included interventions did not generally target specific population groups or settings, and 46 percent of the studies evaluated interventions implemented in electoral democracies. Ten percent of the studies evaluated interventions implemented in countries under fragile and conflict-affected situations.

Interventions within the institutional reform efforts and architecture of public service delivery group were the most common, due to the high number of studies evaluating tax policy and administrative reforms and decentralization, administrative devolution, or reorganization interventions. Other frequently evaluated intervention categories were management innovations and civil service reforms and citizen observers, monitoring of front-line service providers, and reporting mechanisms. Thirty-nine percent of included studies measured outcomes related to human and social development and economic growth and business performance. Fewer studies reported direct governance outcomes, of which the most common were access to public services, tax compliance and contributions, and public service effectiveness measures. However, there is little evidence on outcomes measuring internal governance processes and accountability and quality of policymaking.

Out of the 19 included SRs, three are ongoing, and nine were assessed as having medium or high confidence. The high and medium SRs showed that citizen engagement interventions can improve access and quality of public services and community participation in service provision governance. Moreover, institutional and community monitoring interventions can also help reduce corruption in the public sector. While the SRs present relevant findings across the four intervention groups on the map, in many cases, the evidence of the effectiveness of these interventions is modest.

Conclusion and implications

The Governance EGM provides a vast body of evidence to inform decision-making on policy and programming, particularly around tax policy, decentralization, management innovation, and citizen observers interventions. Practitioners can also draw from this evidence, particularly if interested in programs with theories of change that consider impacts on outcomes around human and social development and economic growth.

While public financial management is a critical component of good governance (USAID, 2018), this EGM is less able to provide policy and programming implications on this topic as we identified limited evidence on public budget planning, budget transparency and
tracking, public procurement, and public private partnership interventions. The governance sector should prioritize the strategic allocation of funding to build up this evidence.

The SRs included in the map cover the four intervention groups; however, the lack of SRs on tax policy interventions is the map’s main gap, considering that this category has more than 100 primary studies. The synthesis gaps identified, the number of SRs assessed as low confidence, and the overall limited and weak evidence presented in the high and medium SRs highlight a need of high confidence and up-to-date synthesis in the governance sector.

Policymakers and researchers can contribute to commissioning and conducting evaluations to fill out the primary evidence and synthesis gaps identified in the map. This can be achieved by expanding the availability and accessibility of government data and by incorporating the analysis of a broader range of outcomes, including direct internal governance measures, to understand better the effectiveness of governance interventions on government structures and processes.
List of figures and tables

Figure 1: Theory of change .................................................................................................................. 6
Figure 2: PRISMA Diagram .................................................................................................................... 11
Figure 3: Publication trend of included studies .......................................................................................... 12
Figure 4: Geographical evidence base ..................................................................................................... 13
Figure 5: Frequency of studies by targeted age, sex and setting of participants ........................................ 14
Figure 6: Frequency of intervention categories by study design ................................................................. 16
Figure 7: Frequency of outcome categories by study design ..................................................................... 20
Figure 8: Frequency of included studies by study design ....................................................................... 23
Figure 9: Equity dimensions considered in included studies ................................................................... 24
Figure 10: Levels, target and implementing stakeholders of included interventions .................................. 24
Figure 11: Types of funding agencies ........................................................................................................ 25

Table 1: EGM research questions ............................................................................................................. 4
Table 2: Summary criteria for studies to be included in the Governance EGM ...................................... 7
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3ie</td>
<td>International Initiative for Impact Evaluation</td>
</tr>
<tr>
<td>CDD / CDR</td>
<td>Community-driven development / reconstruction</td>
</tr>
<tr>
<td>CSO</td>
<td>Civil society organization</td>
</tr>
<tr>
<td>EGM</td>
<td>Evidence gap map</td>
</tr>
<tr>
<td>FCAS</td>
<td>Fragility and conflict affected situations</td>
</tr>
<tr>
<td>FCV</td>
<td>Fragility conflict and violence</td>
</tr>
<tr>
<td>HIC</td>
<td>High-income country</td>
</tr>
<tr>
<td>IE</td>
<td>Impact evaluation</td>
</tr>
<tr>
<td>L&amp;MICs</td>
<td>Low- and middle-income countries</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-government organization</td>
</tr>
<tr>
<td>PPP</td>
<td>Public private partnerships</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized control trial</td>
</tr>
<tr>
<td>SDGs</td>
<td>Sustainable Development Goals</td>
</tr>
<tr>
<td>SR</td>
<td>Systematic review</td>
</tr>
</tbody>
</table>
1. Background

1.1 Development problem being addressed

An increasing body of evidence highlights the importance of good governance for development, as development is enabled when “political decisions benefit the common good, rather than narrow elite interests, and when governments that deliver public goods and services are accountable to citizens in their spending and delivery” (FCDO, 2019a, p. 6). However, international evidence indicates that nearly one in four survey participants declare having paid a bribe when accessing public services, 57% of participants around the world do not think their government is successfully fighting corruption, and some public services, such as policing, are seen as highly corrupted for 47% of people in sub-Saharan Africa and 39% of people in Asia and the Pacific (Transparency International, 2017). Moreover, trends for institutional transparency and accountability appear to be on the decline. For example, in 2020, the number of free countries according to the Freedom House Barometer has reached its lowest level in 15 years (Freedom House, 2021), and the progress on quality of governance has registered a decline in Africa for the first time since 2010 in the Ibrahim Index of African Governance (Mo Ibrahim Foundation, 2020).

To achieve positive and sustainable change, governance relies on government’s willingness and effectiveness in the delivery of their core functions (FCDO, 2019a). In that sense, governance is not enough to reach development, but countries should aim to promote good governance:

The ability of governments to develop efficient and effective public management processes, to meet the basic security and service needs of the population, and, more broadly, to carry out public sector responsibilities at any and all levels of government level. This is often a key determinant of a country’s ability to sustain democratic reform. Priority areas include legislative strengthening, public policy development and implementation, decentralization and local capacity, anti-corruption initiatives, and security sector reform — (USAID, 2020).

Although governance is a long-known concept, it emerged as a key priority for international development in the 1990s following the failure of the Structural Adjustment Program (SAP) and the publication of the World Bank’s report “Sub-Saharan Africa: From crisis to sustainable growth” (World Bank, 1989). The report highlights the “deteriorating quality of government, epitomized by bureaucratic obstruction, pervasive rent seeking, weak judicial systems, and arbitrary decision making” (World Bank, 1989, p. 3), and calls for more investment in good governance to allow sustainable development changes in the region. While bad governance is generally associated with poverty, good governance would be a route toward development. Donors and companies have increasingly based their support on governance-related conditions and plans for reform (Mercy Corps, 2011). Good governance has been qualified as “the single most important factor in eradicating poverty and promoting development” by the former UN Secretary-General Kofi Annan (United Nations, 2012a). Good governance has also become a recurring theme in recent calls for democratization and governance changes, through examples such as the Arab Spring (Diwan, 2011), Myanmar (UNDP, 2015), and Belarus (Chatham House, 2021).
However, the concept of good governance raises a number of issues and questions for international development policymakers, practitioners, and researchers. The first issue is around its definition. Although there is relative consensus about the concept of governance as the use of power and authority for the management of a country, defining good governance is more complex, as different organizations and actors can assign different meanings (United Nations, 2012b). In the absence of an established set of criteria to define what good governance is, the concept tends to be defined by its objectives and characteristics (FCDO, 2009):

- Effectiveness: the ability and authority of leaders, governments, and public organizations to get things done. This is based on the principle of viable and transparent public administration that acts effectively on its core functions.
- Responsiveness: how leaders, governments and public organizations behave in responding to the needs and rights of citizens. This is based on the principle of a governance directed towards the common good and enhancing human capital.
- Accountability: the ability of citizens to hold leaders, governments, and public organizations to account. This is in contrast to the practice of corruption, which can lead to an inefficient and unequitable system.

Secondly, in the absence of a clear definition of good governance, there are as many different pathways to achieve good governance as there are ways to govern. Each country’s governance is shaped by its geography, its history and social characteristics, and by its international environment. What works in a given context might not work in another (FCDO, 2009).

The diversity of pathways leads to a third issue, which is the difficulty of measuring good governance. In the absence of a common set of criteria, measuring or determining good governance can become very complex and will often be based on the outcomes of governance rather than on its inputs (Ashiku & Krypa, 2016).

Lastly, the complexity of the concept of good governance can also lead to a diverse range of actors involved in good governance. Although governments are at the core of a good governance system, they are not the only stakeholders. Public and private organizations, companies, citizens, the judicial sector, among others, are all actors that have a key role in setting the pathway to good governance.

### 1.1.1 The funding landscape

As donors consider the ability of governments to use and distribute aid effectively, governance and anti-corruption represent key priorities for the global aid agenda (Cheney, 2019; World Bank, 2020). For example, a higher proportion of OECD Development Assistance Committee (DAC) funding for governance has gone to low-income countries or countries affected by conflict, which reflects a belief that improved governance is essential to growth and development (OECD, 2014). Regarding anti-corruption efforts, in 2019 FCDO launched the Global Anti-Corruption Programme Prosperity Fund. Committing £45 million over four years, the fund aims to “disincentivise corruption, design out opportunities for corruption, recover illegally gained assets and help end impunity” (FCDO, 2019b, p.2).

From the 1990s and following the end of the Cold War, governance aid has grown (OECD, 2014). Governance funding supports core government functions in L&MICs,
such as public sector and administrative management, public financial management, and resource mobilization. In addition, this sector supports efforts such as decentralization, anti-corruption, and democratic participation (OECD, n.d.). OECD DAC official commitments for government and civil society rose from $2.5 billion in 1995, the first year for which data is available, to $12.7 billion in 2019 (OECD, n.d.).

Governance interventions aim to achieve complex changes that can be difficult to quantify. This could prompt bilateral donors to shift their efforts to changes that are easier to measure and interventions that produce more immediate results (Carothers & de Gramont, 2011; Cheney, 2019). Indicators of improved governance, such as a better functioning public financial management office or parliament, may be case-specific or not easily compared across countries (Carothers & de Gramont, 2011; Rodden & Wibbels, 2019). The wide variation in governing contexts complicates the development of standardized governance indicators (Carothers & de Gramont, 2011). The complexity or individuality of governance interventions could contribute to the variation in definitions or best practices to carry them out.

Good governance is a complex outcome to achieve, and approaches may vary by context. Questions have been raised about the extent to which governance interventions are the most cost-effective approach to development or to reducing poverty; and whether interventions that seem logical, such as decentralizing authority, necessarily lead to the desired results (Carothers & de Gramont, 2011; Cheney, 2019; Grindle, 2004; Rodden & Wibbels, 2019). Strengthening governance may also require donor flexibility, a willingness to experiment, and a longer timeframe to collaborate with governments, build trust, and adapt approaches to local contexts (Rodden & Wibbels, 2019; Cheney, 2019; Carothers & de Gramont, 2011; Grindle, 2004).

Moreover, in the context of the Sustainable Development Goals (SDGs), good governance has also been subject to numerous debates, including on their place within the SDGs. For example, SDG 16: “Peace, justice and strong institutions” includes a commitment to effective, accountable and inclusive institutions (United Nations, 2015), all of which are concepts relevant to good governance. While some authors highlight the importance of having an SDG specifically focusing on good governance (Bierman et al., 2014), others see good governance as a prerequisite to achieving the SDGs (Figueiredo, 2021; Glass & Newig, 2019).

1.1.2 Why it is important to do this EGM

Recent studies have reviewed evidence about citizen engagement in the public sphere, though with a focus on accountability, transparency and oversight of a broad spectrum of government activities, or on internal efforts to strengthen the governments’ public service delivery approaches. For example, 3ie conducted an EGM that compiled evidence about inclusive political processes and accountable governments, with a focus on relations between citizens and the state (Phillips et al., 2017). The International Rescue Committee has also carried out a group of relevant evidence maps of peer-reviewed research (IRC, 2016), among which one centered on interventions measuring power-related outcomes and another focused on cross-cutting service delivery interventions. Likewise, the BRAC Institute of Governance and Development has recently conducted an EGM on social accountability interventions in L&MICs (Zahan, 2021), with a focus on citizens’ participation in policy processes, responsive services, and transparency.
Though these EGMs considered public administration as it relates to citizen input, such as in community driven development initiatives, they do not differentiate interventions across various transparency, accountability and oversight mechanisms; or consider other aspects of good governance, such as public financial or administrative management. In addition, a recent systematic review synthesized evidence about citizen engagement in public service management, including participation, inclusion, transparency and accountability initiatives; however, because citizens were its focal point, the review excluded interventions that did not involve citizen engagement (Waddington et al., 2019).

An evidence gap map that considers a broader range of topics can inform our collective understanding about a more diverse set of governance approaches and contexts. In the absence of an established set of criteria to define good governance, we focused on aspects of governance effectiveness, including effectiveness of public administration, responsiveness, and accountability. Thus, the purpose of this map was to determine the distribution of the evidence regarding interventions to strengthen government effectiveness in L&MICs and provide easy access to this literature.

1.2 Study objectives and questions

By bringing together and mapping evidence, we highlight under-researched areas of governance effectiveness and facilitate critical thinking about the methods of evaluation used in the field. This EGM provides stakeholders with the information required to make evidence informed decisions, and to support future allocation of resources towards under-researched areas of good governance.

This project aimed to improve access to evidence on the effects of interventions to promote good governance in L&MICs among policymakers, researchers, and the development community. It does so by identifying, describing, and summarizing the available evidence in a clear and structured way. The EGM facilitates the use of evidence to inform future research and policy decisions.

To meet this aim, the specific objectives of this EGM were twofold:
- Identify and describe the evidence on the effects of interventions to promote good governance in L&MICs;
- Identify potential primary evidence and synthesis gaps.

To meet these objectives, this EGM addressed the research questions shown in Table 1.

Table 1: EGM research questions

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What is the extent and what are the characteristics of empirical evidence on the effects of interventions to strengthen good governance through improved government effectiveness in L&amp;MICs?</td>
<td>Coverage</td>
</tr>
<tr>
<td>2. What are the major primary and synthesis evidence gaps in the literature?</td>
<td>Gaps</td>
</tr>
<tr>
<td>3. What intervention/outcome areas should be prioritized for primary research and/or evidence synthesis?</td>
<td>Research needs</td>
</tr>
</tbody>
</table>
2. Scope

In this map we included studies which measured the impact of interventions to promote good governance through governance effectiveness, accountability and responsiveness. It included interventions that, as per the definition of the World Bank, focus on “the perception of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government’s commitment to such policies” (World Bank, 2021).

This EGM exclusively focused on interventions related to government and only included interventions from, through and/or directed towards those governments. In addition, the EGM primarily focused on government effectiveness, changes to the way the government works, and the architecture of public service delivery.

2.1 Conceptual framework

2.1.1 Definition
The definition of good governance used in this EGM, based on USAID’s Programmatic Approaches Inventory, recognizes those challenges in a holistic approach:

Governance refers to the exercise of political, economic, and administrative authority to manage a country’s affairs at all levels, including the capacity to formulate, implement, and enforce public policies and deliver services. Good governance includes modes of administrative authority that are inclusive, participatory, transparent, responsive, effective, and accountable. This includes support for strengthening legislatures and local governments; capacity-building; security sector reform; separation of powers through institutional checks and balances, between the executive and legislative branches, in particular; and avenues within government structures for accountability, oversight and meaningful public participation. In conflict or post-conflict environments, this includes support for reconstruction of state institutions and systems — (USAID, 2020).

2.1.2 The theory behind these interventions
The concept of good governance is based on the normative theory that governments’ responsiveness, effectiveness, inclusiveness and accountability are drivers of growth, human development and stability. Positive changes would happen through the work of the government for the benefit of their citizens. As highlighted by FCDO:

Open, inclusive, accountable governance is fundamental to delivering sustainable development and tackling global challenges. And it supports our national interest by contributing to international prosperity, security, and the rules-based international system — (FCDO, 2019a, p. 3).

This theory of change emphasizes the importance of considering good governance in a system based on the interaction between government bodies and civil society: for good governance to happen, government bodies need to respond effectively to the needs of civil society through accountable, transparent and inclusive decision-making processes. The recognition of this interconnection between government and civil society enables
better targeting of good governance interventions by focusing on those aiming at positive and beneficial change in the architecture and way of working of governments. The interventions primarily target governments bodies but also require integrating civil society as a key stakeholder for decision-making processes. These interventions target responsiveness, effectiveness, inclusiveness and accountability to address the needs of civil society and improve the delivery of services, protect the poor, and increase trust and legitimacy (IDA, 2021; Mercy Corps, 2011). Lastly, these interventions are both about driving change and building the state capability to meet those objectives (FCDO, 2009).

As shown in Figure 1, interventions to improve governance can have a positive effect on a number of outcomes: legitimate authority, compliance with rules, inclusive decision-making, accountability of public decision-makers, and quality of policy making, competence and performance of public officials, stewardship of public resources, and public service delivery. These outcomes would contribute to the achievement of longer-term impacts as per the definition of good governance: prosperity, stability, and sustainability through public and social wellbeing and growth (FCDO, 2019a, p. 3).

**Figure 1: Theory of change**

![Figure 1: Theory of change](image)

**2.2 Criteria for including or excluding studies**

Table 2 summarizes the inclusion/exclusion criteria we adopted for this EGM. The screening criteria, built upon the intervention-outcome framework, set the scope of the review in a way that is comprehensive yet manageable within the limits of time and resources. Section 3 describes how the intervention-outcome framework was developed, and further information on the inclusion-exclusion criteria, including the full framework, are presented in Appendix A.
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>We included studies targeting any population type, implemented in any L&amp;MICs (Appendix A1).</td>
</tr>
<tr>
<td>Interventions</td>
<td>This EGM exclusively focused on interventions related to government institutions and only included interventions from, through and/or directed towards those institutions. This included, for example, interventions related to at least one of the following dimensions: Transparency, monitoring and oversight; Creation of participation opportunities, as part of institutional processes and decision-making; Capacity building and administrative management; and Institutional reform efforts (Appendix A2). We included studies that evaluated the impact of at least one of these interventions. For studies that evaluated multi-component interventions, they were included if at least one of the subcomponents matched one of the intervention categories. We excluded interventions in the areas of corporate governance, informal governance, governance in the non-profit sector, and international governance.</td>
</tr>
<tr>
<td>Outcomes</td>
<td>We looked at intermediate and final outcomes (Appendix A3). The outcome groups we looked at were Legitimate authority, Compliance with rules, Inclusive decision-making, Accountability of public decision-makers and quality of policymaking, Competence and performance of public officials, public servants and decision-makers, Stewardship of public resources, Public service delivery, and Public and social wellbeing and growth.</td>
</tr>
<tr>
<td>Study designs</td>
<td>We included impact evaluations and systematic reviews that measured the effects of a relevant intervention on outcomes of interest (Appendix A4). For impact evaluations, we included counterfactual study designs that used an experimental or quasi-experimental design and/or analysis method to measure the net change in outcomes that were attributed to an intervention (i.e. policy, program, project). We included randomized designs, as well as non-randomized studies that were able address confounding and selection bias. We also included qualitative evaluations that followed specific methodological approaches to account for effectiveness. For systematic reviews, we included effectiveness reviews that synthesized the effects of an intervention on outcomes of interest. We excluded reviews that only described programmatic approaches or synthesized findings on barriers and facilitators to implementation.</td>
</tr>
<tr>
<td>Other</td>
<td>Studies published in any language were included, although the search terms used were in English only. Studies published in 1990 and onwards were eligible. We included ongoing and completed impact evaluations and systematic reviews. For on-going studies, we included prospective study records, protocols, and trial registries (Appendix A5).</td>
</tr>
</tbody>
</table>
3. Methods

3.1 Overall approach

EGMs aim to establish what we know, and do not know, about the evidence evaluating the effects of interventions in a thematic area (Snilstveit et al., 2016). They present existing evidence within specific thematic areas or sectors in a structured framework of interventions and outcomes. In this way, EGMs serve as effective tools for policymakers, donors, practitioners and researchers in making evidence informed decisions within specific sectors and help prioritize efforts in specific thematic areas by identifying evidence gaps. To collect and collate the existing evidence, we have followed the standards and methods for EGMs developed by 3ie (Snilstveit et al., 2016, Snilstveit et al., 2017). Appendix B includes additional information on the development, interpretation, and use of EGMs.

3.2 Conceptual framework development

We developed the intervention-outcome framework by consulting relevant literature cited in the previous sections. We also received feedback on the proposed framework from stakeholders within USAID and an external Advisory Group of experts. Malte Lierl, the subject matter expert for this project, provided essential input to develop the intervention categories and the theory behind the interventions we have reviewed.

3.3 Search strategy

We adopted a systematic search strategy following published guidelines (Kugley et al., 2017). This strategy was designed to address potential publication bias issues by systematically searching academic bibliographic databases and implementing additional searches for grey literature in specialist organizational websites, websites of bilateral and multilateral agencies, and repositories of research in international development.

We conducted searches within four academic databases and 46 organizational websites in December 2021. The full list of these sources and the search strings employed are presented in Appendix C. The precise strings and logic (e.g., index terms and truncation operators) were adapted for each database and platform.

In addition, we contacted key experts and organizations through an Advisory Group to identify additional studies that met the inclusion criteria. We also conducted backward and forward citation tracking for included studies in January 2022 to minimize the possibility of missing relevant evidence.

3.4 Screening protocol

We used the EPPI-Reviewer 4 software to manage the EGM implementation process (Thomas et al., 2020). We imported studies into EPPI-Reviewer and, following the removal of duplicates, two team members independently screened the titles and abstracts against our inclusion criteria. We utilized EPPI-Reviewer’s machine learning tool ‘Classifier’ to identify the studies that were more likely to be included, and therefore assigned first for screening, streamlining the EGM process. The studies identified as potentially includable at the title and abstract stage were subsequently screened by two independent reviewers at full text. Further details of each screening step can be found in Appendix D.
3.5 Data extraction and critical appraisal

We systematically extracted data from all included studies directly on 3ie’s specialized online platform designed for EGMs – the Development Evidence Portal. For the data extraction, we followed the DEP protocol and refined it based on the intervention-outcome framework and stakeholder feedback. A random subset of 10 percent of the single-coded data was reviewed by a member of the core team for quality assurance. The tool used for the data extraction process is presented in Appendix E, which covered the following broad areas:

- **Basic study and publication information**: general characteristics of the study including authors, publication date and status, study location, intervention category, outcomes reported, definition of outcome measures, population of interest, study and program funders, time periods for delivery and analysis.
- **Topical cross-cutting issues**: including, among others, democratic/autocratic context, equity considerations including sex, and cost information.
- **Critical appraisal**: All included systematic reviews were critically appraised following Lewin and colleagues (2009). The critical appraisal process assessed the quality of the systematic reviews including activities related to the search, screening, data extraction, and synthesis, and covered the common areas that are prone to biases. Drawing on guidance provided by Snilstveit and colleagues (2017), each systematic review was rated as high, medium or low confidence, indicating the level of confidence we have in the findings of the review based on the methods the authors used. A review classified as high-confidence used methods that align with best practices: the search process was sufficient to identify all potentially relevant studies, bias was avoided in the selection of studies, and appropriate methods were applied to assess risks of bias in included impact evaluations and synthesize the findings on effects. We extracted and summarized the findings of the high and medium confidence systematic reviews. However, we did not critically appraise impact evaluations as this is beyond the scope of EGMs. The tool used for the critical appraisal of SRs is presented in Appendix F.

3.6 Dealing with multi-component interventions

Multi-component interventions were defined as activities with components across several intervention categories of the intervention-outcome framework, which are jointly evaluated in one study. These were differentiated from multi-arm interventions, which are activities with components in multiple intervention categories but for which a study provides an independent effect estimate for each component. Studies evaluating multi-arm interventions were categorized into each of the intervention categories for which they provided an effect estimate. In turn, the project team adopted the following approach to consistently code studies evaluating multi-component interventions:

1. The team determined the main intervention category of focus (e.g., based on the authors’ presentation of the intervention) and coded the study under such main component.
2. For studies where the identification of a main component was not possible (e.g., multi-components implemented as a package of activities), the team coded the study as multi-component, indicating the combination of components if feasible.
The team recognizes the limitations of this approach as it required a strict prioritization of the main component of an intervention over other components which may also play a part in the findings of an evaluation. However, in order to avoid double-counting studies in the map which do not present independent effect estimates, the team selected the above approach to allow a consistent categorization of the body of evidence that reflects more accurately the evaluation of such interventions.

3.7 Analysis and reporting

To answer Research Question 1 regarding the extent and characteristics of the evidence base, we present the distribution of studies by date of publication, intervention(s) studied, outcomes reported, and population considered, including regions, countries, and specific population groups. For the high and medium confidence SRs, we further extracted summaries of the key findings for policy implications.

To answer Research Question 2 regarding gaps in the evidence, we combined knowledge of the evidence distribution with sectoral knowledge to identify meaningful primary evidence gaps, where no IEs exist, and synthesis gaps, where no up-to-date or high confidence SRs exist despite a cluster of IE evidence.

To answer Research Question 3 regarding priority areas for primary research and synthesis, we shared the draft findings with stakeholders at USAID and the Advisory Group and solicited input regarding policymakers and practitioner priorities for future research.

4. Findings

Following the PRISMA guidelines for reporting (Page et al., 2021), this section presents the EGM’s key findings, including the search results, characteristics of included studies, interventions and outcomes captured, study designs and types. It both presents the clusters and the gaps of evidence on governance effectiveness interventions in L&MICs. The full list of included IEs and SRs studies is available in the References section and the online map can be viewed at https://developmentevidence.3ieimpact.org/egm/good-governance-through-government-effectiveness-evidence-gap-map.

4.1 Volume of the evidence

The evidence search identified 98,625 potentially relevant studies (Figure 2). These included 74,398 studies from the search in academic databases, 7,020 records from the search in grey literature sources, and 17,207 records from citation tracking.

A total of 643 evaluation papers were ultimately included, covering 504 unique studies accompanied by 139 linked papers.1 Of the 504 unique studies, 486 were primary evaluations (including one study that was included for both its quantitative and qualitative evaluation designs) and 19 were systematic reviews.

1 Linked records are study versions by the same authors, studying the same intervention and research question(s). For instance, a working paper would serve as the linked record of a journal article. We identified the latest version of a study as the main record, and all older versions as linked records.
4.2 Characteristics of the evidence base

4.2.1 Publication trend over time

While our search looked for studies from 1990 onwards, included studies in the EGM were published between 1999 and 2022 (Figure 3). Half of the studies in the map were...
published between 2018 and 2022. Until 2008, between one to five studies were published per year. Publications increased from 2009, with one to four SRs published per year between 2010 and 2020, and one to three qualitative studies published per year since 2012. Quantitative designs, both experimental and quasi-experimental, represented the majority of published papers each year. Due to the date of the search (December and January 2021) and the publication of our EGM (March 2022), the EGM only includes three studies published in 2022.²

**Figure 3: Publication trend of included studies**

![Publication trend of included studies](image)

Source: 3ie (2022).

### 4.2.2 Geographic distribution

East Asia and the Pacific was the main region of implementation of impact evaluations (n = 173, 34%; Figure 4) with the majority of these studies conducted in China (n = 114).³ Latin America and the Caribbean was the region with the second highest number of studies (n = 144, 29%), due to the studies conducted in Brazil (n = 56), Colombia (n = 32), and Mexico (n = 23). Sub-Saharan Africa was the third highest region of implementation (n = 127, 25%), with Uganda (n = 37), Ethiopia (n = 15), and the Democratic Republic of the Congo (n = 11) having the largest number of studies.

---

² We identified these studies at the citation tracking stage. An additional search of evidence would be needed to cover the full spectrum of publications from 2022.

³ All percentages presented in this report are based on the number of interventions divided by the total number of studies (n = 504). Because the map includes multi-component interventions, for certain analyses the total of percentages can run over 100 percent.
China was the country with the highest number of primary studies ($n = 114$), followed by Brazil ($n = 56$) and India ($n = 47$). Based on the World Bank (n.d.) classification, most of the studies were conducted in lower middle-income ($n = 209$) and low-income countries ($n = 170$), while the rest of the studies took place in upper middle-income countries ($n = 146$). Ten percent of the evaluations were conducted in fragile, conflicted, or violent settings ($n = 49$).

For countries with V-Dem’s regime categorizations data available (Coppedge et al., 2022), a high number of studies were conducted in countries with electoral democracies ($n = 232$, 46%), followed by electoral autocracies ($n = 140$, 28%), closed autocracies ($n = 137$, 27%), and liberal democracies ($n = 23$, 5%).

---

4 The total number of studies shown in this figure does not equal the total number of included studies because of studies that focused on more than one country, which were counted more than once. However, studies including more than 15 countries have been coded as “multi-country”, and hence are not included in the figure.

5 To code the country income level, we considered the first implementation year of each intervention. If this was not reported, the study’s publication year was taken as reference. World Bank data are not available after 2020, hence, for interventions implemented after this year, 2020 was used as reference.

6 We considered the intervention’s first year of implementation to code the country electoral democracy category. If this was not available, the publication year was taken as reference. V-Dem data are not available after 2020, hence, for interventions implemented after this year, 2020 was used as reference.
4.2.3 Population and settings

Data were extracted on the population and setting of the interventions evaluated in included studies. These characteristics were coded based on the authors' own words. Most of the studies did not target a specific age group (n = 339, 67%; Figure 5). Similarly, one-fourth of the studies focused on interventions that targeted population of all ages (n = 133, 26%). The same was true for the targeted sex of the interventions, which was unspecified for 61 percent of the studies (n = 309), while 35 percent (n = 175) of the interventions targeted the whole population (both women and men). The setting of the included studies was also mainly unspecified (n = 218, 43%) or targeted both urban and rural settings (n = 155, 31%). The prevalence of “unspecified” or “whole population” across these three filters can be related to the body of studies focusing on public institutions, which work across populations and settings, and/or within public institutions.

Figure 5: Frequency of studies by targeted age, sex and setting of participants

Source: 3ie (2022).

4.2.4 Interventions

Included studies were not evenly distributed across the 21 intervention categories of this EGM (Figure 6). The institutional reform efforts and architecture of public service delivery intervention group is the most prevalent in this EGM. Within this group, a high proportion of studies focused on tax policy and administrative reforms, and management of non-tax revenues (n = 108, 21%), and decentralization, administrative devolution, or reorganization (n = 77, 15%). Other intervention categories frequently evaluated included management innovations and civil service reforms (n = 62, 12%); citizen observers, monitoring of front-line service providers, and reporting mechanisms (n = 58, 12%); and participatory consultations and deliberative democracy (n = 36, 7%).

---

7 For example, when reports included an explicit reference to the whole population as the beneficiary of the intervention, the “whole population” code was used. If there was no mention of a specific population, the “unspecified” code was used.
Transparency, monitoring, and oversight interventions

Transparency, monitoring, and oversight interventions were represented in 24 percent of all included studies (n = 121), making it the second largest intervention group in the EGM. While there was variation across interventions, the intended recipients of these interventions were generally citizens (e.g. citizen reporting, citizens’ observation, and audit of civil services), in addition to government officials or staff (e.g. audits of governmental department, integration of new compliance standards, and publication of governmental performance reports). Some interventions also included an element of working with community members to understand or interpret the information generated or consider the next steps to act on the information (Ngunyi et al., 2010).

Most of the studies in this group focused on citizen observers, monitoring of front-line services, and reporting mechanisms (hereafter citizen observers; n = 58). This was followed by interventions focusing on compliance of government agencies and staff through financial, compliance and performance audits, or internal control and inspections (n = 26), and compliance management and reform (hereafter compliance management; n = 19). The implementation of interventions from these two categories was sometimes related. For example, the Anti-corruption Country Threshold Program implemented in Uganda included a specific focus on preventing procurement related corruption. Following a series of audits, the program then transitioned to quarterly audits at the district level (Ngunyi et al., 2010).

Less evidence was available on interventions focusing on the monitoring and availability of governmental data: 14 studies reported on performance standards and dissemination, 12 studies focused on open data and freedom of information, and three studies covered interventions on budget transparency and expenditure tracking. Examples of these interventions include the work of Khan and colleagues (2020) in Punjab, who tested the impact of open data and information on different forms of government-based guidelines against COVID-19; the work of Sexton (2017), which assessed the impact of a transparency workshop at the community level in Peru including information about budget disbursement and its impact on the participation of citizens in participatory budgeting activities; and the work of Zhang and colleagues (2021), who evaluated a performance management reform in the Guangdong province in China.
Figure 6: Frequency of intervention categories by study design

Source: 3ie (2022). Note: One study may evaluate interventions with multiple arms, thus the total number of studies in this figure is greater than the number of included studies.
Creation of participation opportunities

Interventions within the creation of participation opportunities group were represented in 16 percent of the studies on the map (n = 80). Most of these studies focused on participatory consultations and deliberative democracy interventions through participatory budgets, townhalls, or consultations of citizens (hereafter participatory consultations; n = 36). For example, Hamilton (2014) analyzed the impact of citizen participation in local budgetary decisions in Brazil, and the Participatory and Responsive Governance project in Niger implemented multi-stakeholder dialogues between community leaders, municipal and regional councilors, and citizens (Benyishay et al., 2017). These interventions often followed or were complemented with a citizen observers component.

Community-driven development and community-driven reconstruction (hereafter CDD; n = 32) or the co-production of public services (n = 15) were interventions often complemented with participatory consultation components. An example of this is the work led by the Forum for Public Health in Bangladesh, where community members were first invited to attend community meetings to discuss the importance of safe drinking water and, in a second phase, the community was invited to take part in the decision process on the allocation of resources for safe water (Madajewicz et al., 2021). Quotas for appointed positions was the only intervention category within the EGM for which we did not identify eligible studies.

Capacity building and administrative management

Capacity building and administrative management interventions were covered by 17 percent of studies in the EGM (n = 86). This evidence was unevenly spread across intervention categories: more than half of the studies in this group focused on management innovations and civil service reforms (hereafter management innovation; n = 62), which included programs targeting salaries of civil servants, training and capacity building of government staff, incentives, and sectoral reforms. For example, in China, the Health VIII program included infrastructure investments and improved planning and management between townships and county hospitals (Wagstaff & Yu, 2007). In India, the Rajasthan Police Department tested new managerial approaches through the limitation of arbitrary transfers, the rotation of duty and days off, and capacity building activities for staff and management (Banerjee et al., 2021).

The management innovation intervention category was sometimes combined with the adoption of e-government and digital transformation of administrative processes (n = 14), including the adoption of biometric smartcards, electronic procurement, or the use of text-messaging for information provision. For example, Chong and colleagues (2014) analyzed the impact of the introduction of a national identification card by the Bolivian Police, both in regard to the technology integration and the management innovation it led to.

Less evidence was available on capacity building and information for public decision-makers (hereafter capacity building, n = 12). This intervention category directly targeted decision-makers and government officials to build their capacities, including the adoption of new processes and procedures, the use of technology-based tools, or the support in the leadership of local programs. For example, to support the responsiveness of local governments to issues raised by civil society organizations, the Millennium Challenge Corporation trained district government officials in Rwanda to promote civic participation and information sharing (Nichols-Barrer et al., 2015).
Institutional reform efforts and architecture of public service delivery

Interventions focused on institutional reform efforts and architecture of public service delivery were covered by 233 studies (46% of the EGM). This was both the group with the highest number of studies and also the most unbalanced group, ranging from four studies in the public budget planning reforms category to 108 studies in the tax policy, administrative reforms, and management of non-tax revenues category (hereafter, tax policy).

The tax policy intervention category was highly driven by China-based evidence, which represented over half of the studies in this category. Examples include the Sin Tax Law on cigarette consumption in the Philippines (Preciados & Carcajente, 2017), the adoption of advanced technologies by the Chinese Tax Bureau to increase tax compliance (Li et al., 2020), or the use of emails for tax payment reminders in Costa Rica (Brockmeyer et al., 2016).

Decentralization, administrative devolution, or reorganization (hereafter decentralization) was the second most frequent category of this intervention group, with 77 studies. These interventions were very prevalent in Brazil and China. For example, Fritscher and Zamora (2016) analyzed the impact of the 1997 Mexico state-level decentralization of health funding, and Li and Yang (2021) studied the 2008 establishment of the Ministry of Environment Protection in China, which centralized environmental policymaking.

The EGM also identified evidence on design and targeting of public services interventions (n = 17), reflecting an interest in methods to maximize the impact of public services for beneficiaries. For example, Alatas and colleagues (2014) studied the differential impact of targeting poor beneficiaries in Indonesia using a proxy means test approach versus a $2 poverty threshold.

Fifteen studies were identified on public budget implementation and expenditure reforms, and four on public budget planning reforms. These interventions mainly focused on debt policies, intergovernmental and regional transfers, and the budgetization and management of public resources. For example, Litschig and Morrison (2012) used Brazil’s discontinuities in federal funding to local governments to analyze their impact on government spending.

Eight studies focused on public private partnerships (hereafter PPP) and six on public procurement reforms. These intervention categories covered different types of contracting schemes, the introduction of procurement systems or reform of procurement rules, and the development of PPPs. Examples of public procurement reforms include the work of Blancas and colleagues (2011) assessing the reform of procurement rules in Brazil called inversao das fases, which facilitated the review of bids by evaluating the price of proposals before their technical aspects. In turn, included studies assessing PPPs were particularly focused on the health sector. For example, Bakibinga and colleagues (2014) analyzed the impact of a PPP in Kenya aimed to facilitate the access to healthcare for mothers and children living in two Nairobi slums.

Multi-component interventions

The EGM identified six multi-component studies (1% of the studies in the map). All of these studies covered at least one of three intervention categories: citizen observers, management innovations, or capacity building. For example, in the Democratic Republic
of Congo, USAID’s Integrated Governance Activity implemented two activities: the capacity building of healthcare personnel and leadership, and a community scorecard of health services (NORC, 2019). In Colombia, Blattman and colleagues (2022) analyzed a reform of municipal services including two components: the creation of an inter-agency task force, and the creation of a liaison office in charge of the gathering and reporting of local concerns on public services.

4.2.5 Outcomes

Figure 7 shows the distribution of studies across the 28 outcome categories. Public and social wellbeing and growth was the most represented group on the map (n = 272, 54%), due to the high number of studies that measured outcomes around human and social development (n = 129, 26%) and economic growth and business performance (n = 79, 16%). Other outcomes often measured were access to public services or government benefits (n = 78, 15%), public service effectiveness (n = 81, 16%), and tax compliance and contributions (n = 74, 15%).

Studies measuring outcomes within the legitimate authority group were represented by six percent of the evidence; this is the group with the lowest number of studies (n = 30). Although outcomes on public trust are measured in 28 studies, there is less evidence available measuring the consent to being governed and perceptions of legitimacy (n = 11). Examples of the latter category include attitudes towards state directives or the perception of state capacity (Khan et al., 2020), or the acceptance of central government authority and the perceptions of government (Beath et al., 2015).

The compliance with rules outcomes group was covered by 19 percent of the EGM studies (n = 97), mainly driven by the large number of studies focused on tax policy interventions which commonly measured outcomes related to tax compliance and contributions (n = 74). In contrast, 24 studies reported outcomes around compliance with laws and rules. Examples of outcome measures within this latter category include changes in environmental regulations (Ding et al., 2021), and measures of compliance of governments in the implementation of new regulations (Koleros & Stein, 2015) or with the principles of transparency and accountability (Anderson et al., 2019).

The inclusive decision-making outcome group accounted for eight percent of the EGM (n = 40). Inclusive and equitable interest articulation and representation outcomes were measured in 34 studies, including indicators such as the number of participants from a given community, their attendance at committees and meetings, or the reporting of their opinion during decision-making processes (e.g. Komorowska, 2016; Bakonyi et al., 2015; Casey et al., 2013). These outcomes were often measured in interventions from the creation of participation opportunities group and are usually measured through direct observation. Conversely, 16 studies reported outcomes on political self-efficacy and 10 on independence from political pressure. These measures usually included less objective indicators, such as the perception of political independence, the interaction of a group with the decision-maker, or a group’s influence on the decision-making process (e.g. Beath et al., 2018; Bakonyi et al., 2015; Sheely, 2015).
Figure 7: Frequency of outcome categories by study design

Source: 3ie (2022). Note: Studies may report multiple outcomes, therefore the total number of studies in this figure is greater than the number of included studies.
Outcomes within the accountability of public decision-makers and quality of policymaking group were covered in nine percent of the studies (n = 47). On the accountability side, studies mainly reported on executive oversight (n = 19) and voter preferences and election results (n = 9) through indicators of internal checks and balances, the reporting of complaints and actions deriving from complaints (Pieri & Lafuente, 2020), and the ability of citizens to hold their governments accountable through their party preference during elections or election turnout (Sato & Imai, 2010). On the quality of policymaking side, studies measured the quality of policy choices (n = 11) in comparison to the needs identified or expressed by citizens, the allocation of resources and investment, or the ranking of priorities (Olken, 2010). In turn, studies that measured regulatory burden outcomes (n = 9) included indicators of delays in the award of certifications (Yanez-Pagans & Machicado-Salas, 2014), resources and financial costs associated with the formalization of firms (Galiani et al., 2017), or the length of processes and costs associated with the payment of taxes (Fang et al., 2017).

The competence and performance of public officials, public servants and decision-makers outcome group was represented in 14 percent of the evidence (n = 70). The outcome most frequently measured of this group was the performance of legislators, public decision-makers, or public servants (n = 49), often measured in evaluations of citizen observers interventions. For example, Pretari (2019) looked at the capacities of mobilization driven by local governments in Tanzania, and their willingness to facilitate access to social services following the integration of technologies. Less evidence is available on measures related to motivation, honesty, or work ethic (n = 23), which included the time allocated to specific tasks, the level of autonomy of government staff (Bandiera et al., 2020), or measures of asymmetry of assessment and favoritism (de Janvry et al., 2021). Outcomes related to the knowledge and informedness of public decision-makers were covered in 14 studies. For example, in Uganda, a scorecard program measured both the monitoring of politicians’ performance and the politicians’ monitoring of services, allowing them to be better informed (Grossman & Michelitch, 2018). Four studies reported measures around the quality of intra-institutions relations. For example, Komorowska (2016) looked at the how much the committees of a CDD program collaborated with each other on common problems.

The stewardship of public resources outcome group was measured in 19 percent of the studies on the map (n = 98). This was driven by the high number of studies reporting on corruption (n = 47) and public spending measures (n = 42), which were common outcomes in the evaluation of interventions around audit and compliance, participatory processes, and decentralization. For example, Olken (2010) looked at the impact of government audits on discrepancies between project costs, both estimates and official spendings, and the reduction of missing expenditures. Less evidence was available on cost-effectiveness (n = 12) and monetary policy performance measures (n = 7), which were often reported in evaluations of the impact of policies on the public budgetary deficit (Arbeláez et al., 2021; An & Hou, 2020).

Outcomes within the public service delivery group were reported in 35 percent of the studies in the EGM (n = 174), possibly due to the fact that these outcomes are often based on publicly available data and are usually available for several years. Within this group, the most common measures reported related to public service effectiveness (n = 81), access to public services or government benefits (n = 78), public satisfaction (n =
public service efficiency (n = 41). For example, in Sierra Leone, Christensen and colleagues (2021) studied the impact of accountability interventions in the Ebola context, such as on the general utilization of health services, user satisfaction, health service delivery, the frequency of service offered, the number of staff to deliver services, and the hours clinics were open.

The public and social wellbeing and growth outcome group was reported in 54 percent of the EGM studies (n = 272). This was mainly driven by the high number of studies that measured human and social development outcomes (n = 129) across a range of development sectors. For example, studies in this category measured access to agricultural resources (Abate et al., 2019), the effect of decentralization on educational outcomes (Leer, 2016), or measures of infant mortality and other health related outcomes (e.g. Christensen et al., 2021; Malik et al., 2017; Fritscher & Zamora, 2016). Outcomes around economic growth and business performance were reported in 79 studies, which were often evaluations of the impact of tax policy interventions (e.g. Yu & Qi, 2022; Corral et al., 2019; Scherer, 2015). In turn, measures of public knowledge and engagement (n = 60) were often reported in studies of participatory consultations. For example, Beramendi and colleagues (2018) looked at how an infrastructure program in West Bank and Gaza increased capacities of service users to apply and pay for permits online, while Benyishay and colleagues (2019) looked at how the Niger Participatory and Responsive Governance Project incentivized citizens’ participation and engagement with the government to voice their concerns. Fewer studies reported on measures of social cohesion (n = 27) and externalities (n = 26). For the latter, most of the studies reported indicators of environment-related externalities (e.g. Zhang et al., 2020; Persha & Meshack, 2016; Tang et al., 2016).

4.2.6 Study design

Systematic reviews

The EGM included 19 systematic reviews, of which three are ongoing. Among the 16 completed reviews, 15 discussed the effectiveness of the interventions through either meta-analysis (n = 4) or a synthesis of quantitative findings (n = 11). One qualitative systematic review used a realist synthesis approach.

The critical appraisal of the 16 completed systematic reviews rated four studies as high confidence, five as medium confidence, and seven as low confidence. The main pitfalls of low confidence SRs were related to their search strategies (a lack of searching for grey literature and/or for literature cited by included studies), their screening processes (non-independent screening by two reviewers), and their reporting (a lack of presenting the characteristics and/or the results of the risk of bias assessment of included studies).

Impact evaluations

The EGM included 465 quantitative evaluations, 19 qualitative evaluations, and one evaluation that used eligible quantitative and qualitative evaluation designs. The majority of the included impact evaluations used a quasi-experimental design (n = 306, 61%; Figure 8). Among the 159 (31%) studies using experimental designs, 71 complemented their analysis with a quasi-experimental or qualitative design. Among quasi-experimental methods, the design most commonly used was fixed-effects, which includes difference-in-difference estimations (n = 206, 41%). Fixed-effects and difference-in-difference estimations were also often used in experimental evaluations. The EGM included 20
studies (4%) that used specific qualitative designs to account for effectiveness. Process tracing was the most common qualitative evaluation design (n = 11, 2%), followed by contribution analysis (n = 4, 1%).

**Figure 8: Frequency of included studies by study design**

We identified 75 studies (15%) that reported cost data of the intervention evaluated. Two-thirds of these studies (n = 49) provided detailed cost information, including itemized costs such as administrative and program costs, or estimates of the intervention benefits relative to its costs. This is compared to 26 studies reporting the overall cost of the intervention or the cost per participant. In addition, 12 percent (n = 60) of included studies reported having received approval from an independent ethics review board to conduct the evaluation. Although 80 percent of these studies were experimental evaluations, 31 percent of all RCTs in the EGM reported having ethical clearance.

### 4.2.7 Equity dimensions and focus

The majority of the studies in the EGM did not include an equity approach when evaluating interventions (n = 432, 86%). Among the studies that did, the most common equity approach used was a focus on interventions that targeted vulnerable populations (n = 50, 10%), which included populations in FCAS, populations recovering from violence, rural populations, or populations with high levels of poverty. The second most adopted approach was the consideration of equity through sub-group analysis, both by sex (n = 16, 3%) and other attributes (n = 16, 3%). Examples of the latter group include salary caps (Ferraz & Finan, 2009), committee membership (Pandey et al., 2009), and age of local leaders (Wu et al., 2020).

Among the studies that considered equity (Figure 9), the dimension most studied was socio-economic status (n = 42), which can be explained in part by the high proportion of studies focusing on tax and revenue interventions. This was followed by sex (n = 39), place of residence (n = 21), and age (n = 11).
4.2.8 Target and implementer of the interventions

The included studies covered interventions at all levels of implementation, from focusing on individuals to national level interventions (Figure 10). We identified a large group of studies implemented at the local level (n = 181, 36%), mainly related to participatory interventions or the piloting of reforms. This was followed by interventions implemented at the individual level (n = 140, 28%), mostly as a consequence of the numerous tax policy interventions targeting individuals and firms. Interventions implemented at the national (n = 120, 24%) and subnational levels (n = 121, 24%) were also covered by included studies, especially those conducted in China. Interventions implemented at the household level were the least common in this body of evidence (n = 36, 7%).
Interventions evaluated by included studies mainly targeted citizens (n = 260, 52%). Similarly, public institutions (n = 241, 48%) and public officials (n = 169, 34%) represented a considerable proportion studies and were often targeted together (n = 107, 21%). Public institutions were also the main implementing stakeholder of included interventions, such as through public reforms and changes in ways of working within government entities (n = 430, 85%). In this endeavor, interventions implemented by public institutions sometimes also received the support from domestic and international organizations (n = 62, 12% and n = 39, 8%, respectively).

4.2.9 Research and implementation funding
Of all included studies in the EGM, 284 (56%) reported the funding source of the research and 168 (33%) reported the source that funded the program (Figure 11). Government agencies were the main source of funding of both the interventions evaluated by included studies (n = 99) and the evaluations themselves (n = 72). International aid agencies and financial institutions were also reported as funders of programs (n = 37 and n = 20, respectively) and research (n = 39 and n = 34, respectively). Examples of these funding organizations include the Asian Development Bank, the UK Foreign, Commonwealth, and Development Office, the World Bank Group, and the Millennium Challenge Corporation. For research specifically, academic institutions (n = 51) were another important source of funding. Examples of academic institutions supporting the evaluation of studies include the International Growth Centre, Yale University, and Evidence in Governance and Politics.

Figure 11: Types of funding agencies

Source: 3ie (2022).

4.3 Report for ongoing studies
Four percent (n = 21) of included studies are ongoing studies or protocols. Documentation of these 21 ongoing studies has been published between 2009 and 2021, including three systematic reviews, 14 experimental evaluations, and four quasi-experimental studies. The main focus of these studies is on management innovations (n = 7), citizens observers (n = 5), and participatory consultations (n = 4).
4.4 Findings from high and medium confidence SRs

4.4.1 Characteristics of the body of evidence
We identified nine medium or high confidence SRs (four SRs were assessed as high confidence and five as medium confidence). These SRs were published between 2010 and 2020, and covered 28 countries across Sub-Saharan Africa, Europe and Central Asia, East Asia and the Pacific, South Asia, and Latin America and the Caribbean.

These nine SRs covered all intervention groups of the EGM. Two intervention categories were covered by more than one SR: citizen observers (Squires et al., 2020; Waddington et al., 2019; Molina et al., 2017; Hanna et al., 2011) and management innovations (Rockers & Bärnighausen, 2013; Carr et al., 2011; Hanna et al., 2011). In addition, these SRs reported outcomes across five of the outcome groups, with studies most commonly measuring outcomes related to public and social wellbeing and growth, public service delivery, and stewardship of public resources. Particularly, the outcome categories more frequently reported were around human and social development (n = 5) and access to public services (n = 3).

4.4.2 Main findings of SRs
The following sections present the main findings of the nine high and medium confidence SRs. For reviews that synthesized evidence across primary studies using meta-analysis, we detail the size of the effects. For reviews that provided narrative synthesis, we provide a summary of the findings, as reported by their authors. The description of the reviews’ findings is arranged by the four main intervention groups of the EGM.

Transparency, monitoring and oversight
Four reviews analyzed interventions focusing on the transparency, monitoring and oversight group: the reviews of Waddington and colleagues (2019) and Molina and colleagues (2017) were both appraised as having high confidence, while the reviews of Squires and colleagues (2020) and Hanna and colleagues (2011) were assessed as having medium confidence. All four reviews focused on the impact of citizen observers interventions. Two of them also covered interventions with a compliance management component (Molina et al., 2017) or a performance standards design and dissemination component (Waddington et al., 2019). As described more in detail below, the reviews reported positive effects of social accountability mechanisms on reducing corruption and improving public service delivery.

Based on 35 included studies, Waddington and colleagues (2019) highlighted that direct engagement between service users and providers can improve some intermediate or final service delivery outcomes. While the authors found a small increase in physical access to services and service quality (Standardized Mean Difference (SMD) = 0.08, 95% Confidence Interval (CI) = 0.00, 0.15) and SMD = 0.10, 95% CI = 0.03, 0.18, respectively), they found no changes in the reduction of absenteeism, leakages from embezzlement, or cost of services (SMD = 0.02, 95% CI = -0.19, 0.24; SMD = 0.02, 95% CI = -0.18, 0.21; and SMD = 0.07, 95% CI = -0.11, 0.24, respectively). Authors stated, however, that this is not enough to guarantee impact, especially in a situation of supply chain bottlenecks.

Comparable results were observed in Molina and colleagues’ (2017) review, as the study underscored the positive effect of community monitoring interventions on measures of occurrence of corruption (SMD = 0.15, 95% CI = 0.01, 0.29) and corruption perception (Risk
Difference (RD) = 0.08, 95% CI = 0.02, 0.13). Nevertheless, the authors recognized that the review had limited evidence to generalize results because it included 15 studies conducted primarily in Africa and Asia and only a few measured similar corruption indicators.

The other two reviews analyzed the evidence using narrative synthesis. Based on a narrative synthesis of six reviews, Squires and colleagues (2020) noted that, in the health sector, community monitoring approaches can be promising interventions to improve both the quality of health services and community education and empowerment. This review of reviews reported that the involvement of community members in health audit cycles could reduce delays in the decision to seek care, facilitate better access to information, and improve the supply of drugs and equipment for health services, among other intermediary outcomes. However, the authors also acknowledged that there is less evidence on the health impacts of community monitoring, such as mortality.

Hanna and colleagues (2011) used textual narrative synthesis to describe the evidence of anti-corruption interventions. Based on 11 evaluations conducted in different regions of the world and various sectors, the authors observed that institutional and community monitoring interventions could potentially reduce corruption in the short term. The review proposed that for monitoring strategies to work best, programs should have the objective of decreasing corruption, and they should be implemented in combination with adequate incentives that establish the consequences of being corrupt.

Creation of participation opportunities
Two reviews analyzed interventions related to the creation of participation opportunities: Waddington and colleagues (2019) and King and colleagues (2010), which were assessed as having high and medium confidence, respectively. These reviews focused on distinct aspects of participation: King and colleagues (2010) focused on CDDs, while Waddington and colleagues (2019) analyzed interventions where citizens were engaged in the planning, management, or oversight stages, including, among others, CDDs.

Waddington and colleagues’ (2019) review included 35 studies. The authors looked at the direct consultation of citizens and found mixed results on immediate outcomes. Citizen engagement interventions were effective in improving some measures of participation in service delivery governance; for example, for meeting attendance (SMD = 0.69, 95% CI = 0.22, 1.15) and knowledge about the services provided (SMD = 0.09, 95% CI = 0.01, 0.17). However, they found no effects in terms of provider responsiveness, including, for instance, measures of politician performance (SMD = 0.06, 95% CI = −0.17, 0.05) and staff motivation (SMD = 0.23, 95% CI = −0.08, 0.54).

In turn, King and colleagues (2010) measured the impact of CDD interventions on social cohesion in Sub-Saharan Africa. The review included eight primary studies and reported on 15 pro-social outcomes. However, seven of these measures were reported in more than one study, and two findings were replicable across evaluations: CDD interventions showed a weak positive effect on collective trust (RD = 0.35, Standard Errors (SE) = 0.14) and a negative effect on inter-group relations (RD = -0.20, SE = 0.10). The authors concluded that the evidence of the effect of CDD interventions is modest due to the limited amount of comparable evidence and called for more studies on interventions promoting social cohesion, especially regarding inter-group relations.
Capacity building and administrative management
Three SRs analyzed interventions focusing on capacity building and administrative management: Rockers and Bärnighausen's (2013) review was assessed as high confidence, and the work of Hanna and colleagues (2011) and Carr and colleagues (2011) were both assessed as medium confidence. While Hanna and colleagues (2011) found evidence that anti-corruption interventions may hold promise for reducing corruption in specific contexts, Carr and colleagues (2011) and Rockers and Bärnighausen (2013) found a paucity of evidence around interventions to hire, train, and remunerate public servants in the health and education sectors.

Rockers and Bärnighausen (2013) looked at interventions for hiring, retaining, and training managers in health systems. The review included two primary studies, one conducted in Cambodia and the other in Mexico, Colombia, and El Salvador. While authors were not able to pool evidence across studies, the evaluations showed that providing private contracts with international NGOs in Cambodia may improve access and use of health systems, and that proving training courses may have positive effects on the performance of health system managers. However, both studies were assessed as low quality under the GRADE approach, and the authors appealed for more evidence to understand the effectiveness of these interventions across different contexts.

Carr and colleagues (2011) focused on interventions including a change in salaries for public servants on their performance and the quality of public services within the education and health sectors. The review identified one eligible study conducted in Brazil, which reported a positive effect of teachers' pay raises on students' performance. The authors described the review results as inconclusive and highlighted the striking dearth of empirical evidence connecting salary changes to performance measures of civil servants.

Based on a narrative synthesis of 11 studies, Hanna and colleagues (2011) noted that anti-corruption policies that use explicit incentives can potentially reduce corruption within the public sector in the short term. The authors observed that incentives – which can be financial or non-financial and can be used as punishments or rewards, such as wage reductions (or increases) or publishing corruption activities of elected officials in the media – may be more effective when implemented in conjunction with monitoring strategies to increase the probability of detecting corrupt activities.

Institutional reform efforts and architecture of public service delivery
Three reviews analyzed institutional reform efforts and architecture of public service delivery: Qin and colleagues (2019), whose SR was assessed as high confidence, and the reviews of Aboal and colleagues (2012) and Hanna and colleagues (2011), which were both assessed as medium confidence. All three SRs provided narrative syntheses of interventions focused on various aspects of this group, including public private partnerships, public procurement reforms, design of public services, and decentralization policies.

Aboal and colleagues (2012) conducted a narrative review of 22 primary studies. It evaluated interventions to enforce public contracts for the investment of resources and political capital, including through PPPs and public procurement reforms. The authors observed that more effective contract enforcement may increase investment rates, and that included studies seemed to have based their evaluations on at least some of the hypothesized causal mechanisms connecting contract enforcement with investment,
such as better infrastructure and access to financial institutions, reduced uncertainty, and increased rentability. However, the authors concluded that the evidence on this topic was weak, underdeveloped, did not meet strong quality standards, and was potentially subject to publication bias.

Focusing on the design of public services, Qin and colleagues (2019) conducted a narrative review of studies evaluating the relationship between user charges and health outcomes. The review included 17 primary studies covering 12 developing countries and identified a modest association between reducing user charges and the improvement of health outcomes, which was especially marked in studies focused on children and lower-income populations. The authors suggested that increased access to healthcare may be a potential mechanism conducting to improved health outcomes when expenditures for individuals are cut.

Hanna and colleagues (2011) conducted a narrative synthesis of decentralization interventions as strategies that introduce structural changes to reduce corruption. Based on three studies, the authors found limited evidence that decentralization can work to reduce corruption in certain contexts. These interventions could be more successful when implemented in settings with established infrastructure and staffing capacity, and when they involve community participation to increase accountability. The authors noted that more research and cost information is needed to generalize findings across contexts.

**Cross-cutting recommendations from high and medium confidence SRs**
While these SRs aimed to synthesize different topics around governance, they advocate for at least two approaches when designing and implementing interventions: that they are tailored to the local context, considering its structures, complexities and values, and that they aim to build local capacities to facilitate their success.

For example, Waddington and colleagues (2019) recommended designing citizen engagement interventions in collaboration with the local service providers that will be implementing the intervention to ensure buy-in, as well as with local groups of interest to develop their social capital and capacity for collective action. Similarly, Hanna and colleagues (2011) advised implementing monitoring strategies along with incentives to deter corruption, with both elements tailored to the actors and market structures involved in the intervention. Molina and colleagues (2017) also identified that providing appropriate information and tools to citizens can facilitate their involvement in community monitoring interventions.

**4.5 Gaps analysis**
Almost all intervention and outcome categories listed in the EGM framework were covered by at least one study. However, several gaps in the evidence indicate an uneven prioritization and focus of research and funding on some aspects of good governance. We organize this section into primary evidence gaps, where little or no impact evaluation evidence were identified for a particular intervention, outcome, or population, and synthesis gaps, where we identified a cluster of primary study evidence but a lack of high confidence and up-to-date SRs (defined loosely as published in the last five years). Finally, we highlight methodological gaps in the evidence. More information on how to interpret EGMs is available in Appendix B.
4.5.1 Primary study evidence gaps

*Quotas for appointed positions* is the only intervention category without an eligible study. Although the evaluation of quotas for elected positions is prevalent, as presented in the Political Competition EGM (Gonzalez Parrao et al., 2022), we found no study on quotas for non-elected positions that met the inclusion criteria. This discrepancy may relate to fewer quotas for appointed positions being implemented in L&MICs, or to difficulties in rigorously evaluating such interventions.

There are gaps in evaluations of interventions related to government budget and partnership management. We identified four studies in the *public budget planning reforms*, 15 studies in the *public budget implementation and expenditure reforms*, and three studies in the *budget transparency and expenditure tracking* categories. This might be due to the difficulty to access data and/or run rigorous impact evaluations on these types of interventions. While the intervention categories with the greatest number of studies (*tax policy* and *decentralization*) usually relied on publicly available panel data to conduct rigorous analyses, interventions on budget planning and transparency may be run internally by governments and may require additional data collection beyond what is readily available for researchers. The challenge of having less data publicly available might also explain the scarcity of evidence observed in the partnership management area, where we identified eight studies on *public private partnerships* and six on *public procurement reforms*.

Outcomes focusing on government processes and the accountability and quality of policymaking, are less frequently studied. A small number of studies reported outcomes related to internal governance measures, such as *quality of intra-institution relations* (n = 4), *monetary policy performance* (n = 7), and *independence from political pressure* (n = 10). Several of these outcomes rely on data that may not be publicly available, which could make it more complex to measure these outcomes consistently and accurately. Likewise, accountability outcomes were reported in few studies, including *executive oversight* (n = 19), *voter preferences and election results* (n = 9), *quality of policy choices* (n = 11), and *regulatory burden* (n = 9). Measures of the extent of scrutiny of government decisions, accountability of government officials, or legality and fairness of policy choices require an agreed measurement framework and continued data collection over time. These outcome groups may be less prioritized when compared to more readily available human and social development data. This could make links between governance interventions and changes in government processes harder to assess or quantify, which is a challenge for understanding the mechanisms under which governance interventions are successful.

Studies tend to report outcomes that are directly related to the intervention type. Half of the studies on *tax policy* interventions reported outcomes on tax compliance; similarly, citizen participation interventions usually report outcomes on social development, engagement and access to services. However, less evidence is available on the impact of these interventions on additional, more indirect outcomes. For example, we found no studies evaluating interventions in the *creation of participation opportunities* group that included outcome measures on *cost effectiveness* or *regulatory burden*, and no studies assessing the impact of *institutional reform efforts* that reported outcomes on *consent to being governed and perception of legitimacy*, *quality of intra-institutions relations*, or *knowledge and informedness of public decision-makers*. 
The evidence on tax policy and decentralization is cut by half if we do not account for China, Brazil, and India. The high number of primary studies conducted in China (n = 114), Brazil (n = 56) and India (n = 47) drive the results for these two interventions and also for some outcomes, such as tax compliance and economic growth. Although this cut does not lead to an absolute gap of evidence in these intervention and outcome categories, the prevalence of studies in these three countries, especially in China, has great influence in body of evidence available around governance effectiveness.

There are gaps of evidence on interventions implemented in fragile contexts and liberal democracies. Based on V-Dem data, only five percent of studies (n = 23) were conducted in liberal democracies, whereas the rest of the studies are roughly distributed across electoral democracies, electoral autocracies, and closed autocracies. This could be explained by the correlation between liberal democracies and HICs in the V-dem regimes index. In addition, only 49 studies focused on interventions implemented in FCAS. This may relate to the difficulty of implementing governance effectiveness interventions in a fragile and conflict-affected context and the necessity to wait for the stabilization of the political situation to be able to act. Considering the high number of studies that relied on panel data, this gap could also be due to the greater difficulty in collecting and accessing data in FCAS. For example, the gap of evidence observed in studies conducted in the Middle East and North Africa (n = 7) and Europe and Central Asia (n = 19) can be at least partly explained by the democracy and/or fragility levels in these regions.

4.5.2 Synthesis gaps
Of the 19 SRs on the map, only half were assessed as high or medium confidence. Between one and four systematic reviews on governance issues were published each year between 2010 and 2020. However, only four of the SRs identified were appraised as high confidence and five as medium confidence. While there are no apparent trends over time in regard to the confidence level of these SRs, based on the critical appraisal results, there is an overall gap of high confidence synthesis of governance evaluations.

Citizen observers is the intervention category with the highest number of SRs but only four were appraised as high or medium confidence. The transparency, monitoring and oversight intervention group includes almost half of the SRs of the map (n = 9), of which two are ongoing and the seven completed SRs covered at least one citizens observers component. One of these SRs, published in 2013, also focused on budget transparency and expenditure tracking and compliance management interventions, but it was assessed as low confidence.

There is a synthesis gap for co-production of public services interventions. The creation of participation opportunities intervention group has the lowest number of high or medium confidence SRs as only two of the six completed SRs identified in the EGM were assessed as such. We identified two SRs analyzing the evidence on the co-production of public services, but both were appraised as low confidence and were published between 2013 and 2014.

E-government and capacity building interventions are not covered by high or medium confidence SRs. The capacity building and administrative management intervention group includes over 100 IEs and four completed SRs; however, three of these SRs were rated as high or medium and only focused on management innovation interventions. Moreover, the
three SRs were published between 2011 and 2013. It is encouraging that we have also identified seven ongoing primary studies evaluating management innovation interventions, which could add to an up-to-date synthesis of these interventions.

There are synthesis gaps for the interventions with the highest number of primary studies on the map. The institutional reforms intervention group includes six completed SRs but none of these SRs focused on tax policy interventions, covered by 109 individual evaluations. Additionally, only three SRs focused on decentralization interventions, two of which were assessed as low confidence and the third SR was published in 2011. Within this group, we also did not identify SRs covering interventions on public budget planning reforms or public budget implementation and expenditure reforms.

Forty percent of the outcomes identified for this EGM are not studied by SRs. While 12 out of the 28 outcome categories were not included in any of the SRs, four additional categories were only reported in SRs assessed as low confidence, accounting for 57 percent of all the outcome categories in the EGM. These understudied outcomes concentrate on five groups, which are directly related to governance measures: legitimate authority, compliance with rules, inclusive decision-making, accountability of public decision-makers and quality of policy-making, and competence and performance of public officials, public servants and decision-makers.

4.5.3 Methodological gaps
A third of the studies used a randomized evaluation design. In contrast, the majority of the primary studies on the map (61%) used a quasi-experimental design. The paucity of studies with an experimental design may illustrate the difficulty of implementing this evaluation approach to analyze the impact of a range of governance interventions. While randomized evaluations were more commonly used to study transparency and participation interventions, this design was less used to evaluate the most popular interventions in the EGM: only four experimental studies focused on decentralization interventions (compared to 69 quasi-experimental studies), and 28 experimental evaluations focused on tax policy interventions (compared to 79 quasi-experimental studies). Similarly, public budgeting interventions (public budget planning reforms and public budget implementation and expenditure reforms) were not evaluated using experimental designs.

There is a lack of qualitative evaluations on governance interventions. We identified 19 studies using an eligible qualitative evaluation design, and one that included eligible quantitative and qualitative designs. Qualitative studies were usually conducted to evaluate interventions related to citizen observers and participatory consultations, and were often published as institutional reports. Another possible alternative for understanding this gap is that researchers may use variations of these qualitative designs but do not label them as such. If this is the case, our search strategy may have missed those studies. Qualitative studies may be informative when large-scale experiments or observational studies are not feasible, but they can still try to approximate counterfactual scenarios methodically and transparently.

Equity was considered in only 14 percent of included studies. The most commonly used approaches for considering equity were targeting vulnerable populations and conducting sub-group analyses. Socio-economic status and sex of participants were the most common dimensions being considered. Systematically incorporating equity
approaches can help identify heterogeneous treatment effects, consider not only average effects but also potential distributional consequences, and safeguard against discriminatory effects in governance programming.

There is a gap in the reporting of ethical approvals among included studies. Only 12 percent of studies reported having received approval from an independent ethics review board. While most of the studies that reported this information were experimental evaluations, which account for a third of the studies in the EGM, only 31 percent of all randomized evaluations reported having ethical clearance. This is a key research stage to ensure the protection of study participants and their communities; however, this gap may also be due to a lack of reporting this information rather than not going through an ethical clearance process.

5. Conclusions and implications

5.1 Conclusions

There is a large evidence base studying the effect of governance interventions in low- and middle-income countries (L&MICs). The EGM covers 504 unique studies, including 465 quantitative evaluations, 19 qualitative evaluations, one study that used both quantitative and qualitative designs, and 19 systematic reviews.

Half of the studies in the map were published between 2008 and 2022 and a third were conducted in the East Asia and the Pacific region. There is an increase in the number of studies published between 1999 and 2020, particularly in the last decade. The majority of the studies were conducted in East Asia and the Pacific (34%), Latin America and the Caribbean (29%), and Sub-Saharan Africa (25%). This is driven by the large number of primary studies conducted in three countries: China (n = 114), Brazil (n = 56) and India (n = 47).

The most commonly evaluated interventions were tax policies, decentralization, management innovations, and citizens observers. All but one intervention category are covered by at least one study, but these categories are not distributed evenly: tax policy interventions were evaluated by 108 studies, while no eligible study evaluated quotas for appointed positions. The institutional reforms intervention group was the most frequently studied, driven by evaluations of tax policy and administrative reforms, and management of non-tax revenues (n = 108) and decentralization, administrative devolution, or reorganization (n = 77). Evaluations of management innovations and civil service reforms were also prevalent (n = 62), targeting salaries, incentives, and capacity building of government staff and sectoral management reforms. Within the transparency, monitoring, and oversight intervention group, a substantial portion of included studies were also conducted to evaluate citizens observers, monitoring of front-line service providers, and reporting mechanisms (n = 58).

Outcomes on human and social development and growth were more frequently studied than more direct indicators of governance quality. Two-thirds of included studies measured outcomes related to the public social well-being and growth group, mainly through two types of categories: human and social development and economic growth and business performance. The prevalence of these development outcomes could
be related to the fact that these indicators can rely on publicly available or more easily accessible panel data. In turn, the most commonly studied outcomes that measured governance more directly focused on access to public services, tax compliance and contributions, and public service effectiveness measures. There is little evidence on outcomes measuring internal governance processes and accountability and quality of policymaking, such as regulatory burden (n = 9), voter preferences and election results (n = 9), monetary policy performance (n = 7), and quality of intra-institution relations (n = 4).

Similarly, studies that evaluated similar interventions also measured similar outcomes, usually those closely related to the intervention type. For example, 94 percent of the studies that evaluated tax policies reported outcomes on tax compliance or economic growth, and 75 percent of the studies that evaluated decentralization policies reported outcomes on human and social development, economic growth, or public spending.

There is an overall lack of high confidence and up-to-date systematic reviews in the governance sector. Although the EGM includes 19 systematic reviews, only four were assessed as having high confidence and five as medium confidence. The transparency, monitoring, and oversight intervention group has the highest number of high or medium SRs, which particularly synthesized interventions around citizen observers. In turn, the institutional reforms interventions group has the largest synthesis gap on the map. Despite the large number of primary studies available, we did not identify SRs covering tax policy interventions.

While the SRs present relevant findings across the four intervention groups of the map, in many cases the evidence of the effectiveness of these interventions is weak and the reviews do not tend to synthesize effect sizes. The main conclusions for each group and cross-cutting findings are summarized below:

1. Within the transparency, monitoring and oversight group, interventions that promote the direct engagement between public service providers and their users can have a positive effect on improving access to services (SMD = 0.08, 95% CI = 0.00, 0.15) as well as the quality of public services (SMD = 0.10, 95% CI = 0.03, 0.18), but are not effective in reducing other intermediate outcomes, such as the cost of services (SMD = 0.07, 95% CI = −0.11, 0.24; Waddington et al., 2019). In turn, community monitoring interventions can have a positive effect on reducing episodes of corruption (SMD = 0.15, 95% CI = 0.01, 0.29; Molina et al., 2017), particularly when combined with incentives that establish consequences to corruption activities (Hanna et al., 2011). In the health sector, community audits may also help improve the quality of services and the knowledge and empowerment of communities (Squires et al., 2020).

2. In the creation of participation opportunities group, citizen engagement interventions can have a positive effect on increasing some measures of participation in the governance of service provision, such as meeting attendance (SMD = 0.69, 95% CI = 0.22) and knowledge about the services provided (SMD = 0.09, 95% CI = 0.01, 0.17), but they do not seem to improve measures of provider responsiveness, such as politicians’ performance (SMD = 0.06, 95% CI = −0.17, 0.05) and staff motivation (SMD = 0.23, 95% CI = −0.08, 0.54; Waddington et al., 2019). However, the evidence on CDD interventions is weak in showing a positive effect on collective trust (RD = 0.35, SE = 0.14) and a negative effect on inter-group relations (RD = -0.20, SE = 0.10; King et al., 2010).
3. Within the capacity building and administrative management group, anti-corruption interventions that use financial and non-financial incentives may hold promise for reducing corruption in the short term (Hanna et al., 2011). The evidence on the impact of interventions to hire, train, and remunerate public servants in the health and education sectors in L&MICs is inconclusive and/or of low quality (Rockers & Bärnighausen, 2013; Carr et al., 2011).

4. In the institutional reform efforts and architecture of public service delivery group, interventions to ensure contract enforcement, such as in public procurement reforms or PPPs, may promote higher levels of investment, although the evidence is weak (Aboal et al., 2012). In terms of changes in the design of healthcare systems, there is a modest association between reducing user charges and improvements in health outcomes, potentially through increased access to health services (Qin et al., 2019). Finally, there is limited evidence that decentralization policies may help reduce corruption, particularly when involving community accountability mechanisms and in settings with infrastructure and staffing capacity (Hanna et al., 2011).

5. Cross-cutting recommendations across these SRs include designing and implementing interventions that consider the local structures, complexities and values, and incorporate capacity-building components of local actors and groups.

5.2 Implications for policy

The Governance EGM provides a vast body of primary evidence to inform policymakers and practitioners, who can consult the map considering the following:

- The evidence is clustered around tax policy and decentralization interventions. While most of these evaluations were conducted in China, Brazil, or India, the evidence offers a substantial base to inform decision-making on policy and programming in different contexts.
- Policymakers can draw from the evidence available to inform their programming, particularly if their theories of change consider impacts on human and social development and economic growth.
- Decisions around governance programming can draw on the findings from systematic reviews of high and medium confidence. Takeaways are available across the four intervention groups analyzed in the EGM; however, these results are often nuanced by the limited evidence identified in each SR.
  - Citizen engagement interventions can be effective in improving the access and quality of public services and increasing the community’s participation in the governance of service delivery.
  - Community and institutional monitoring interventions can also be effective in reducing corruption in the public sector.
- Most of the interventions included in this EGM were implemented by public institutions and targeted public institutions or citizens. Policymakers should consider collaborating with researchers to support the production, collection, and monitoring of data that can be used for rigorous research. Data availability and accessibility are especially relevant for reform-based interventions, where data may only be produced or collected by government entities. More data on internal governance could build government capacity to track this information, enabling a broader assessment of the impact of interventions on good governance.
5.3 Implications for future research

The number of studies identified the EGM reflects a shared interest in the field of evaluating the effectiveness of governance interventions. However, there are sharp differences in the availability of studies across the categories. These discrepancies may illustrate the prioritization of research investment and the challenges in measuring the effect of some governance areas using rigorous methods. Future research may consider the following:

- Upcoming research should focus on evaluating interventions for which no studies were identified. Rigorous evaluations of the only blank category, quotas for appointed positions, should be prioritized.
  - For intervention categories for which studies were identified, yet the evidence is concentrated on specific groups or geographical regions, future research should aim to expand the evidence base. For example, evaluations on tax and decentralization policies are abundant, but these were mainly conducted in three countries: China, Brazil, and India.
  - To address the scarce evidence around public financial management, researchers should prioritize opportunities to help fill the gap on the effectiveness of interventions on public budget planning, implementation, and transparency.

- Future synthesis work may be useful in intervention categories with sufficient studies but without systematic reviews or high/medium confidence reviews, such as tax policy interventions, open data and freedom of information initiatives, or e-government developments. Researchers and practitioners could also consider updating the evidence of the current reviews, as five of the nine high or mediums confidence SRs on the map were published between 2009 and 2012.

- Upcoming research in Europe and Central Asia and the Middle East and North Africa should be encouraged to help fill in the evidence gaps in these regions. Both regions integrate countries and contexts relevant to governance effectiveness regarding their democracy levels and governance transitions.

- New practical and logistical approaches could be developed to conduct rigorous evaluations of governance interventions within FCAS, for example, to access accurate and reliable data and collaborate with communities and local and national governments.

- More research should be done to develop reliable metrics of governance and make them available for use in program evaluation, as current studies are often not measuring directly what they want to influence. Researchers can contribute to building the evidence base by identifying creative ways to evaluate governance interventions and outcomes and advocating for these evaluations to be embedded in governance programming efforts.

- The findings of the EGM raise additional research questions that researchers could consider as an invitation to further develop the evidence base, thus contributing to informed decision-making on governance. Examples of these questions include:
  - What is the impact of good governance interventions on internal governance processes and measures of accountability and quality of policymaking?
  - How can interventions targeting budget planning, monitoring, and disbursement improve good governance outcomes?
  - Which indicators are the best fit for purpose to measure good governance?
  - Which patterns in good governance interventions can be identified to better understand what makes governance more or less good?
Online appendices

Online appendix A: Criteria for including or excluding studies

Online appendix B: About Evidence Gap Maps

Online appendix C: Search strategy

Online appendix D: Screening and data extraction protocol

Online appendix E: Critical appraisal tool

Online appendix F: Data extraction codebook
References

Report

https://doi.org/10.2499/p15738coll2.133323

https://assets.publishing.service.gov.uk/media/57a08a6bed915d3cfd000756/Contract_enforcement_2012Aboal__2__.pdf


https://doi.org/10.1016/j.jclinepi.2017.02.013

https://doi.org/10.1007/s10797-019-09561-w

https://doi.org/10.1111/ajps.12428


https://doi.org/10.26417/ejser.v6i2.p155-160


https://hummedia.manchester.ac.uk/institutes/gdi/publications/workingpapers/bwpi/bwpi-wp-13610.pdf


https://doi.org/10.1016/j.worlddev.2014.10.024


https://doi.org/10.1371/journal.pone.0238776

https://doi.org/10.3390/su8090931


https://images.transparencycdn.org/images/GCB_Citizens_voices_FINAL.pdf


Included studies

This list presents the references for the 504 included studies in the EGM. Main studies are aligned to the left, and linked studies are shown with bullet points below the main record.


65


https://www.nber.org/system/files/working_papers/w20751/revisions/w20751.rev0.pdf


https://doi.org/10.2139/ssrn.1703543


https://ssrn.com/abstract=3219383

https://www.nber.org/papers/w24879


https://openknowledge.worldbank.org/handle/10986/19146


https://openknowledge.worldbank.org/handle/10986/26246


Other publications in the 3ie Evidence Gap Map Report Series

The following papers are available from https://www.3ieimpact.org/evidence-hub/publications/evidence-gap-maps


An increasing body of evidence highlights the importance of good governance for development. Through good governance, better political decisions and quality of public services may bring impact to a majority of citizens and not to an elite minority. However, decreasing levels of freedom, as well as corruption, mismanagement of public resources and services, and low levels of accountability and transparency are some of the challenges to the principles of good governance in low-and middle-income countries. Authors of this report present findings of an evidence gap map that primarily focuses on interventions on government effectiveness, changes to the way governments work, and the architecture of public service delivery.