



Peer-Review Report Form

Manuscript Title:	Who is Fit to Govern? Citizen Perceptions of Leadership Qualifications in Governance: A Systematic Review
Type of Article:	Review Article

Editor

<u>General Comments and Recommendations</u>
The reviewers acknowledge the significance of your manuscript and recommend several revisions to further improve its clarity, coherence, organization, and overall scholarly quality. In addition, you are advised to thoroughly review and validate all references included in the manuscript. Please ensure that all cited sources are accurate, accessible, and retrievable through Google Scholar to uphold the credibility and reliability of the study. Additionally, make sure to include the DOI or website link for every reference to facilitate verification and accessibility of sources. Furthermore, only sources that are cited within the manuscript should appear in the References section, and all entries listed in the References section must also be cited in the text. For example, if 20 unique sources are cited in the manuscript, then only those 20 sources should be included in the References section. Please note that repeated citations of the same source within the text are counted as a single reference entry; therefore, the total number of references may be lower than the total number of in-text citations.

Reviewer 1

<u>General Comments and Recommendations</u>
The authors attempt to synthesize empirical evidence on citizen evaluations of public leaders. The effort to consolidate this fragmented literature is commendable. Nevertheless, the manuscript requires substantial revision to meet the rigorous standards of the journal. The analysis remains far too descriptive. The authors enumerate themes but fail to develop a novel theoretical model to explain the causal mechanisms at play. The synthesis lacks profound critical depth, and the authors must elevate the discourse beyond mere observation.

<u>Specific Comments and Recommendations</u>
Title and Abstract-The title is adequate but slightly redundant. The abstract provides a clear summary of the research, yet it misses a definitive statement of the theoretical contribution. The authors state that integrity acts as a threshold qualification. The abstract, however, needs to clarify how this review departs from prior meta analyses. The authors must clarify if their procedures are strictly PRISMA protocol compliant.
Introduction-The introduction establishes the context well and outlines four research questions. Yet, the rationale for why the authors chose these specific four questions remains vague. The text notes that evidence remains fragmented. The authors must explain the exact nature of this fragmentation. Furthermore, the reliance on recent citations is appropriate for modern context,



but the historical foundation is exceptionally weak. Fifty years of scholarship on political legitimacy precedes the year 2010. To omit foundational texts severely weakens the theoretical premise.

Materials and Methods-The choice to include non peer reviewed policy briefs and dissertations alongside experimental data introduces severe heterogeneity. The authors must justify this choice more robustly. The search strategy relies on a narrow set of databases. The exclusion of regional databases limits the global applicability of the results. The quality appraisal relies on MMAT protocols. The authors admit that they did not exclude weak studies, yet they claim zero weak studies were retained. This statistical anomaly requires a thorough explanation. Did the reviewers suffer from positive bias? Furthermore, the phrase mixed methods appears, and the authors must clarify the exact parameters of these varied approaches.

Results-The results section reads like a simple list rather than a synthesized narrative. The five themes identified are intuitive but lack profound analytical depth. For example, the claim that service delivery serves as a practical test of competence is elementary. The authors need to dissect the anomalies. What happens when highly competent leaders are perceived as corrupt but still win elections? The synthesis completely ignores clientelism and patronage, which dominate local politics in many regions. To claim that corruption is always a disqualifier contradicts empirical realities in emergent democracies. The authors must introduce nuance here and address paradoxes in voter behavior.

Discussion and Conclusion-The discussion reiterates the results rather than elevate the discourse. The assertion that leadership qualification is publicly constructed is valid but not novel. The authors suggest that institutions should assess participatory capacity. How can institutions operationalize this metric? The conclusion provides a neat summary but fails to offer actionable policy frameworks. The recommendations for future research are far too broad. The authors should specify exact methodological designs for future scholars to adopt.

Tables and Figures-Table 1 and Table 2 are clear. However, Figure 2 attempts to show causal mechanisms but appears overly simplistic. The arrows suggest unidirectional relationships, whereas public perception and leader behavior operate in a bidirectional loop. The visual model needs a complete redesign to reflect system dynamics accurately.

Please indicate your recommendation by checking the appropriate box below.

<u>Decision</u>	
	Accept the manuscript for publication.
/	Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.
	Reject the manuscript, as it is not suitable for publication.



Reviewer 2

General Comments and Recommendations

My main concern on this manuscript is whether the evidence included in the review really measures leadership qualifications as claimed, and whether the process from hundreds of studies to the final 53 studies is sufficiently transparent. I believe the paper can still be improved considerably, but these issues should be resolved before the conclusions are treated as dependable.

Specific Comments and Recommendations

I am not fully convinced that the review is actually reviewing leadership qualifications. This is for me one of the most serious problem in the manuscript. The phenomenon is stated as citizen perception of leadership qualifications, but the inclusion criteria allow studies about trust, transparency, accountability, service delivery, participation, responsiveness, satisfaction, and other public evaluations. These are related concepts, but they are not automatically qualifications for leadership. A citizen being satisfied with a municipal service does not necessarily mean that the citizen is making an evaluation about who is fit to govern. Likewise, institutional trust may concern the government system rather than a particular leader. The authors need a much stricter conceptual rule separating direct evidence of perceived leadership qualification from studies that are only indirect proxies. Otherwise, the review may be synthesizing general good-governance literature and relabeling this as leadership fitness.

The paper says that five themes “emerged,” but these same five areas were already built into the framework before the synthesis. The conceptual framework already identifies integrity, trust/responsiveness, leadership perception, service delivery/accountability, and participation as the five dimensions, and the manuscript clearly says that these dimensions guided the coding. Later, almost the same five categories are reported as themes that “emerged” from the synthesis. This creates a circularity problem. If the categories were decided beforehand, they should be described as deductive themes, not as findings that emerged naturally from the literature. If an inductive component was also used, the authors need to explain how new or contradictory themes were allowed to appear. At present, the method almost guarantees that the review will find what its framework already assumes.

The second-stage reduction from 232 studies to only 53 core studies is not sufficiently transparent. This is a critical issue because the main conclusions depend only on these 53 studies. The manuscript states that studies were prioritized according to directness, governance relevance, empirical evidence, and leadership-related outcomes, but these criteria are too general to reproduce the decision. The flow shows 232 broadly included studies but only 53 retained for final thematic synthesis. What exact rule caused one study to enter the 53 while another one did not? This additional refinement can substantially influence the final themes and can unintentionally permit selective inclusion. I suggest the authors provide the operational criteria for this second screening and a transparent accounting of the studies removed at this stage.

The manuscript appears to rely heavily on one researcher for the most bias-sensitive parts of the review. The author contribution statement indicates that one author conducted the literature search, screening, data extraction, quality appraisal, formal analysis, and thematic synthesis, while the other authors mainly provided supervision or review. If this accurately reflects the actual process, this is a significant weakness. Study eligibility, quality judgement, and thematic coding all involve subjective decisions. The manuscript should clarify whether any portion was independently checked by another reviewer, how disagreements were resolved, and whether



agreement was assessed. If no independent verification was conducted, this limitation is much more important than currently acknowledged and should be addressed where possible.

Research Question 2 is not well matched with the literature search itself. One research question specifically asks which formal and informal qualifications citizens associate with fitness to govern. However, the reported search terms mainly concern citizen perception, leadership, governance, trust, integrity, accountability, transparency, participation, service delivery, responsiveness, and satisfaction. Terms directly targeting formal qualifications such as educational qualifications, credentials, prior political experience, administrative experience, professional expertise, candidate qualifications, or eligibility are not clearly represented in the search description. Nevertheless, the results make conclusions regarding education, experience, and technical expertise. The authors therefore need either a more targeted search for formal qualifications or should narrow this research question and the related conclusions.

Transferable evidence is used as an important inclusion principle, but the authors never explain what makes evidence transferable. The paper emphasizes local governance, yet it also includes national, comparative, and broader public-sector studies whenever their findings were considered transferable to leadership evaluation. Later, these different governance levels are synthesized together. This is not a small matter. Citizen judgement of a mayor, a national political leader, a government institution, and a frontline public-service agency may operate very differently. The authors should define beforehand what “transferable” means and preferably present the findings separately by governance level before combining them. Otherwise, the local-governance emphasis of the paper becomes difficult to defend.

There is an important numerical problem in the accounting of studies. Table 3 reports 232 broadly included studies. It then reports 53 core studies and 164 supplementary/contextual sources. However, 53 + 164 accounts for only 217 of the 232 studies, leaving 15 studies without a reported status. This may appear only numerical, but for a systematic review this is not a minor error because every included record must be traceable. Table 2 also combines “supplementary/contextual or excluded after refinement” into a single figure of 391, although these seem to represent different screening decisions. I strongly suggest the authors reconcile every number across the flow diagram and all tables before publication.

The reader cannot independently inspect the evidence supporting the five themes. The tables provide aggregated numbers and broad descriptions of the evidence clusters, but there is no study-level table for the 53 core studies showing, at minimum, country, governance level, sample, design, construct actually measured, whether evidence is direct or indirect, quality appraisal, assigned theme, and principal finding. Tables 3 and 4 are therefore informative only at summary level. Even Table 5 only gives aggregated quality categories. For a review making broad conclusions from heterogeneous evidence, a full study-characteristics/evidence matrix is necessary. Without it, the reader cannot determine whether the conclusions faithfully represent all 53 studies or mainly a smaller group repeatedly cited in the narrative.

Please indicate your recommendation by checking the appropriate box below.

<u>Decision</u>	
	Accept the manuscript for publication.
/	Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.
	Reject the manuscript, as it is not suitable for publication.



SECOND ROUND

<u>Decision</u>	<u>Editor</u>	<u>Reviewer 1</u>	<u>Reviewer 2</u>
Accept the manuscript for publication.	/	/	/
Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.			
Reject the manuscript, as it is not suitable for publication.			