Context Before Conclusions: A Diagnostic Framework for Contextual Validity in Policy Transfer

Victor Frimpong *

Posted Date: 27 May 2026

doi: 10.20944/preprints202605.1815.v1

Keywords: contextual validity; policy transfer; evidence-based policymaking; external validity; implementation science; digital governance; policy evaluation

 doi:10.20944/preprints202605.1815.v1

Article

Context Before Conclusions: A Diagnostic Framework for Contextual Validity in Policy Transfer

Victor Frimpong

SBS Swiss Business School, Kloten-Zurich, Switzerland; v.frimpong@research.sbs.edu

Abstract

Purpose: This paper explores why evidence-based policies that appear effective in one context frequently produce uneven outcomes, exclusion, or legitimacy challenges when transferred across policy and governance contexts. It introduces the Contextual Research Validity Index (CRVI) as a diagnostic framework for evaluating whether the contextual conditions necessary for policy validity remain aligned across settings. **Methodology:** The study develops a conceptual and diagnostic framework that assesses contextual validity across four dimensions: epistemic alignment, institutional fit, cultural resonance, and operational feasibility. The framework is illustrated through an interpretive analysis of India's Aadhaar digital identification system, drawing on secondary literature, policy reports, and institutional evidence. **Limitation:** The paper is conceptual and illustrative rather than predictive or causal. The CRVI scoring approach is heuristic and based on qualitative interpretation rather than statistical modelling or original empirical data collection. **Findings:** The analysis demonstrates that policy effectiveness cannot be separated from contextual conditions. The Aadhaar case shows that interventions regarded as technically successful may still generate exclusion, legitimacy disputes, and uneven outcomes when epistemic assumptions, institutional safeguards, cultural expectations, and operational realities are misaligned. **Practical Implications:** The CRVI provides policymakers and evaluators with a structured tool for assessing transfer readiness, contextual risk, and governance vulnerabilities before scaling or replicating interventions across settings. **Social Implications**: The framework supports more accountable and context-sensitive policymaking by helping reduce exclusion, governance failures, and legitimacy erosion in large-scale policy interventions, particularly in digital public infrastructure. **Originality/Value:** The paper contributes to the policy evaluation and governance literature by reframing policy failure as a problem of contextual misalignment rather than of insufficient evidence alone, and by operationalising contextual validity through a transferable diagnostic framework.

Keywords: contextual validity; policy transfer; evidence-based policymaking; external validity; implementation science; digital governance; policy evaluation

1. Introduction

Evidence-based policymaking is based on the idea that effective interventions in one context can be adapted or scaled to others. This approach has impacted global policies in welfare, education, public health, and digital governance over the past 20 years. Improved evaluation methods, like randomised trials and systematic reviews, have bolstered the use of solid evidence in policy decisions. However, a significant challenge remains: even well-supported policies can fail or lead to unintended consequences when applied in different contexts.

Interventions are often context-dependent, as shown by implementation science and external validity research. These studies indicate that the effectiveness of an intervention depends on specific contextual conditions and does not universally apply across settings (Mehrotra et al., 2019;

 doi:10.20944/preprints202605.1815.v1

Huebschmann et al., 2019). Transportability research regards generalisation as a causal inference problem, emphasising that evidence from one location cannot predict outcomes in another without accounting for contextual similarities (Mehrotra et al., 2019). Additionally, the focus on effectiveness in reviews and trials often neglects pragmatism and context, making it hard for practitioners to assess the applicability of results elsewhere (Huebschmann et al., 2019). In primary care, successful adaptation of interventions to local contexts is crucial for ongoing engagement and sustainability; poorly adapted transfers can reduce performance, even when the core intervention is evidence-based (Kwan et al., 2020). Furthermore, cross-national experiences with nursing home interventions, for example, demonstrate that successful implementation across countries often depends on cross-cultural adaptation, including attention to legal norms, language differences, and cultural sensitivities (Hockley et al., 2019).

Evaluation and implementation research increasingly acknowledges that intervention effects depend on context, yet context is often inadequately measured in validity assessments. This has led to calls for better reporting and evaluation of external validity and contextual factors in evidence reviews and trials (Huebschmann et al., 2019; Ziemann et al., 2019). External validity usually addresses generalisation concerns, whether study findings apply to a wider population or different conditions. This focus can overlook how an intervention's effects are influenced by varying institutional, organisational, or resource contexts across different adoption sites (Huebschmann et al., 2019; Andrade, 2023). As a result, contextual factors are often acknowledged but not effectively operationalised for systematic diagnosis in different implementation environments.

This is a major evaluative weakness, as highlighted by scholars who argue that pragmatism and context are not adequately assessed for decision-makers considering interventions in new settings (Huebschmann et al., 2019; Ziemann et al., 2019). Recent contributions focus on measurement, showing how to better understand context with validated tools. These include instruments to quantify organisational factors related to learning (Lyman & Thorum, 2022), scales for assessing perceived contextual fit and support for evidence-based practices (Coyle et al., 2022), and standardised tools for planning in low-resource public health emergencies (Demelash et al., 2025). Additionally, some work operationalises "environmental context" at the neighbourhood level using indices for violence, disorder, and other characteristics to inform place-based policymaking (Williams et al., 2020). All these insights indicate that without a focus on transportability and context measurement, policymakers lack the evaluative infrastructure needed to differentiate between interventions that work across settings and those dependent on specific local conditions (Huebschmann et al., 2019; Ziemann et al., 2019; Coyle et al., 2022; Williams et al., 2020).

The limitations of integrating technology into social policy are evident in large-scale reforms such as Aadhaar, India's national digital ID program. While Aadhaar was praised for enhancing welfare targeting, improving service-delivery efficiency, reducing fraud, and expanding access to public services through biometric verification, its implementation revealed significant challenges. Research highlights issues such as exclusion errors, authentication failures, uneven administrative capacity, and legitimacy concerns, especially for those reliant on last-mile service delivery. These problems have prompted legal and institutional scrutiny, thereby limiting Aadhaar's applicability. This raises a crucial question: how can something seem successful in evaluation while facing substantial practical issues?

This paper contends that contradictions in evidence-based models are not just due to poor implementation or political resistance. Instead, they highlight a more fundamental problem: the assumption of contextual validity without proper assessment. These models often evaluate effectiveness without verifying that the necessary epistemic assumptions, institutional structures, cultural expectations, and operational conditions for success are in place. When these alignments are lacking, evidence may be persuasive in theory but ineffective in practice.

Many evaluators still prioritise causal estimates over context when assessing policy effectiveness, even though they recognise that interventions depend heavily on surrounding conditions. As a result, policies supported by strong evidence in one setting are frequently transferred

 doi:10.20944/preprints202605.1815.v1

elsewhere without verifying whether the original enabling conditions remain intact (Cartwright & Hardie, 2012).

This paper uses the Contextual Research Validity Index (CRVI) (Frimpong, 2026) to evaluate the cross-context effectiveness of policies. The CRVI assesses contextual coherence across four dimensions: epistemic alignment, institutional fit, cultural resonance, and operational feasibility, helping analysts pinpoint where evidentiary claims lose credibility during policy transfer or scaling. Rather than replacing current evaluation methods, the CRVI enhances them by clarifying and assessing contextual validity.

The paper does not provide a new empirical evaluation of Aadhaar or determine its overall success or failure. Instead, Aadhaar serves as a clear case study illustrating how transfer-condition failures arise and why traditional evaluation frameworks often miss them. This analysis supports a broader argument about policy transfer and governance: understanding why evidence fails to apply across contexts requires a systematic approach to diagnosing contextual fit rather than focusing solely on effectiveness.

2. Policy Evaluation and the Limits of Context-Blind Transfer

Contemporary policy evaluation focuses on causal identification and generalisability. Methods such as randomised controlled trials, quasi-experimental designs, and systematic reviews are central to evidence-based policymaking, helping estimate average treatment effects and link outcomes to interventions (Angrist & Pischke, 2009; Banerjee et al., 2020). While these approaches enhance internal validity and reduce uncertainty about intervention effects, they also promote a narrow view of validity that ties methodological rigour to credibility, often overlooking contextual factors.

Within this paradigm, transferability is typically evaluated through external validity, which assesses whether findings from one population can be generalised to others. While important, this approach has limitations; it primarily focuses on population traits and treatment differences rather than on the institutional, cultural, and operational factors that influence an intervention's success. As a result, an intervention may meet external validity standards yet remain unsuitable for different environments (Cartwright & Hardie, 2012).

Realist evaluation emphasises that outcomes arise from mechanisms within specific contexts, as summarised in context–mechanism–outcome (CMO) configurations (Pawson & Tilley, 1997). This approach shifts focus from universal causal claims to conditional explanations of “what works, for whom, and under what circumstances.” Similarly, implementation science and process evaluation emphasise how organisational capacity, frontline discretion, and delivery systems influence policy outcomes (Mayne, 2012; Peters et al., 2013).

Despite advancements, the context of evaluation approaches is often described rather than systematically diagnosed, limiting policymakers' ability to assess whether interventions remain applicable across implementation settings (Huebschmann et al., 2019). Evaluators typically document contextual factors after the fact and lack a structured method for assessing whether an intervention's assumptions align with a specific setting before or during implementation. As a result, context is acknowledged but not systematically evaluated, limiting policymakers' ability to anticipate failures and to distinguish between interventions that are universally effective and those that depend on specific conditions.

The focus on "best practices" and evidence hierarchies by international organisations can overlook the importance of context. While evidence-based interventions are promoted for their effectiveness, critical scholars argue that such an approach prioritises methodology over contextual relevance (Peck & Theodore, 2015; Rodrik, 2008). Failures are frequently attributed to poor implementation rather than a misalignment between intervention assumptions and the new context.

There's a key paradox in evidence-based policymaking: while context is acknowledged as crucial, it is rarely considered in evaluation frameworks. Evaluations can demonstrate if an intervention works under specific conditions but often don't explain why results may fail, reverse, or

 doi:10.20944/preprints202605.1815.v1

lead to unintended consequences when implemented elsewhere. This is especially important in complex governance reforms, like digital identification or welfare targeting, where success depends on institutional capacity, social legitimacy, and operational reliability.

Although realist and implementation-oriented approaches provide detailed insights into how context influences outcomes, they fall short of offering a verifiable pre-transfer evaluation of whether the contextual factors that support policy validity are expected to remain consistent across different environments. Consequently, while context is recognised, it is not systematically evaluated as a factor affecting the transferability of evidence. This oversight can lead to misattributing failures to implementation issues rather than acknowledging a fundamental misalignment in context.

Despite advances in external validity research, realist evaluation, implementation science, and transportability analysis, current methods lack a structured approach to assessing whether the contextual conditions for policy validity align across settings. Context is often mentioned but not systematically evaluated as a criterion. This gap is particularly significant in policy transfer analysis of governance reforms and digital public infrastructure, where effectiveness depends on institutional, cultural, and operational factors. The Contextual Research Validity Index (CRVI) framework is proposed as a structured tool for evaluating contextual coherence across different settings, enhancing existing evaluation methods and addressing a significant gap in policy transfer analysis.

3. Research Method

This paper employs a conceptual and diagnostic research design to explore how contextual validity affects the transferability of evidence-based policies across settings. Instead of conducting causal evaluations or original empirical tests, it introduces the Contextual Research Validity Index (CRVI) as a framework for assessing how well policies align with different contexts.

The analysis relies on secondary sources, including academic literature, policy reports, institutional documents, and legal materials on India's Aadhaar digital identification system. Aadhaar was chosen because it exemplifies both successful large-scale implementation and ongoing debates around exclusion, legitimacy, institutional capacity, and operational reliability. This makes it ideal for exploring the challenges evidence-supported interventions face in diverse contexts.

The CRVI framework assesses validity conditions through four dimensions: epistemic alignment, institutional fit, cultural resonance, and operational feasibility. Scores are assigned based on a qualitative review of the evidence and literature, with a focus on diagnostic insights rather than predictive measurements. The goal is to enhance transparency in contextual assessments and highlight discrepancies between policy assumptions and real-world implementation.

This paper serves as a conceptual-methodological contribution with case applications, rather than a formal empirical evaluation of Aadhaar.

4. Contextual Research Validity Index (CRVI) as a Diagnostic Framework

This paper uses the CRVI to analyse why evidence-based policies often fail when applied in different contexts. The CRVI complements traditional evaluation methods by making contextual validity clear and measurable, rather than implicit. It does not replace causal evaluation; rather, it focuses on whether the necessary conditions for an intervention’s credibility are maintained when policies are transferred, scaled, or adapted to new settings.

4.1. Conceptual Rationale

The CRVI framework emphasises that the success of research and policy depends on relational factors rather than on methods alone. While current evaluation frameworks effectively assess whether interventions work under certain conditions (Pawson & Tilley, 1997; Cartwright & Hardie, 2012), they often fail to address whether the underlying assumptions, institutional structures, cultural norms, and operational capacities align across different contexts (Frimpong, 2026). This misalignment

 doi:10.20944/preprints202605.1815.v1

can lead to policies that appear evidence-based but fail, distort, or face challenges when applied in new settings.

To address this issue, the CRVI framework treats contextual coherence as a vital aspect of validity. It defines context not just as background information but as specific conditions that influence how evidence is understood, implemented, and maintained. This approach allows evaluators and policymakers to move beyond simple success-or-failure evaluations and understand where and why misalignments arise.

4.2. Core Dimensions of the CRVI

The CRVI evaluates validity conditions through four key dimensions:

- **Epistemic Alignment:** This dimension assesses how well an intervention's underlying knowledge aligns with the understanding of problems, evidence, and solutions in a specific context, including how information is produced and expectations about behaviour.
- **Institutional Fit:** This refers to how well an intervention aligns with existing organisational structures, governance, and accountability mechanisms. Even well-designed interventions can fail if institutional responsibilities are unclear or incentives are misaligned.
- **Cultural Resonance:** This dimension evaluates whether an intervention is viewed as legitimate and fair within the local social norms and values. Policies that clash with community expectations may face resistance or cause unintended exclusion.
- **Operational Feasibility:** This assesses whether the necessary resources, infrastructure, and logistical support for implementation are available. Many operational challenges become apparent only when initiatives are scaled up, impacting policy success.

These dimensions highlight that effective policies require sound design and strong alignment with the local context.

4.3. Scoring Logic and Interpretive Use

The CRVI dimensions can be assessed on a scale (e.g., *1–5*) to create a composite profile of contextual validity. However, the CRVI does not provide precise measurements or causal estimates. Instead, scores serve as diagnostic indicators, revealing strengths, vulnerabilities, and trade-offs. A moderate or low composite score does not necessarily indicate an ineffective intervention; rather, it suggests that effectiveness depends on specific contextual factors.

The CRVI functions as a diagnostic and anticipatory tool. It can assess transfer readiness before implementation, explain failures afterwards, or compare performance across different settings. By highlighting contextual coherence, the framework clarifies contradictions often obscured by traditional evaluation metrics, such as when overall performance improves while issues such as exclusion or legitimacy persist.

4.4. Positioning Relative to Existing Evaluation Approaches

The CRVI recognises the importance of rigorous causal evaluation but emphasises aspects often overlooked by traditional methods. While conventional evaluations focus on whether an intervention achieves desired outcomes under specific conditions, the CRVI assesses if the necessary epistemic assumptions, institutional structures, cultural expectations, and operational capacities are aligned across different contexts.

This framework complements existing evaluation approaches by linking evidence claims to their contextual relevance, enhancing policy transfer analysis and governance assessment. It is especially valuable for interventions that are transferred, scaled, or adapted to different institutional and sociocultural environments.

 doi:10.20944/preprints202605.1815.v1

The subsequent section uses the CRVI framework to examine India's Aadhaar digital identification system, highlighting how contextual validity risks can arise across dimensions, even when the system is considered evidence-based and successful in practice.

Figure 1 summarises the CRVI framework, showing how evidence-based interventions can lose relevance when applied in different contexts. It highlights how epistemic, institutional, cultural, and operational dimensions can diagnose risks of misalignment, exclusion, instability, and loss of legitimacy.

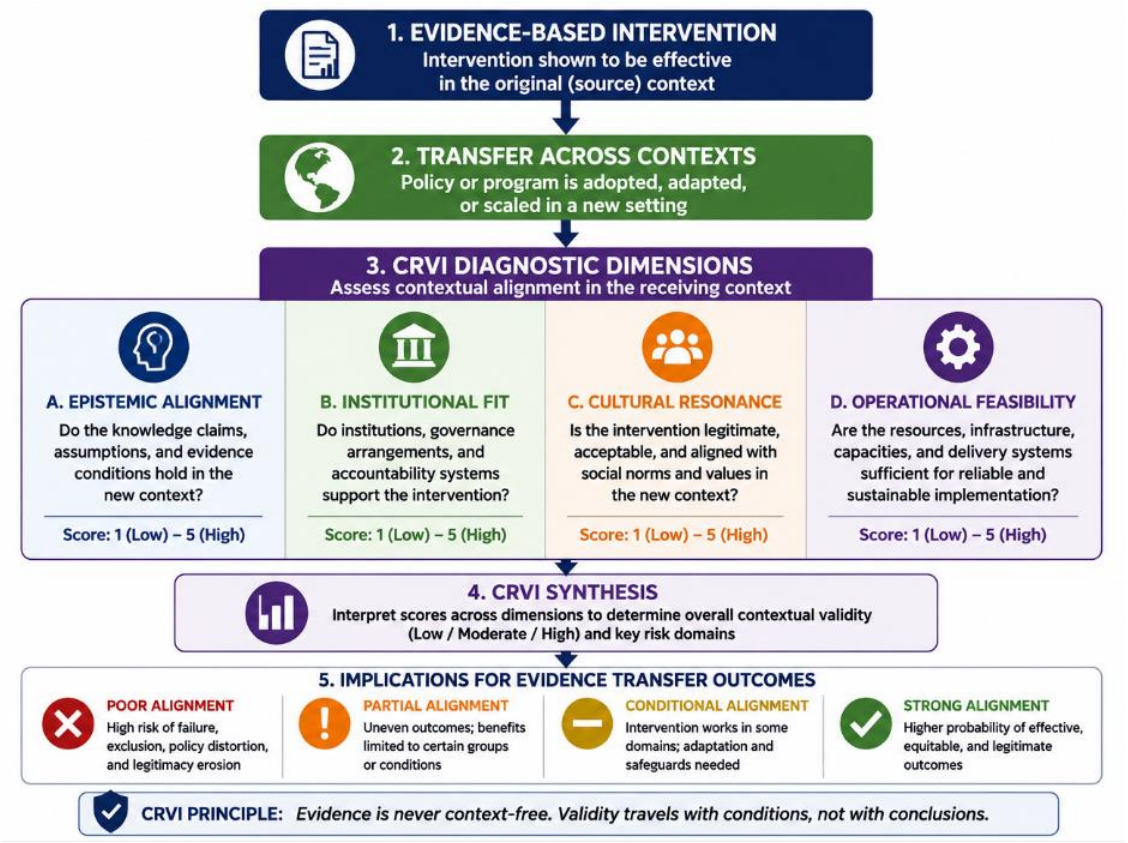

Figure 1. CRVI Framework for Diagnosing Evidence Transfer Across Contexts. Source: Developed by the Author, 2026.

5. Diagnostic Example of Contextual Validity in Policy Transfer: India's Aadhaar Case

India's Aadhaar system is a comprehensive biometric digital identification framework that assigns residents a unique identity number linked to biometric and demographic information. Established to enhance the efficiency, inclusivity, and reliability of welfare provision and public services, Aadhaar has evolved into one of the world's largest digital identity frameworks and is often cited as a paradigm for digital public infrastructure and policy adaptation. Nonetheless, its implementation has faced ongoing issues related to exclusion errors, data precision, institutional protections, and legal challenges.

This blend of widespread adoption, strong evidence claims, and ongoing contextual debates makes Aadhaar an especially apt example of how the Contextual Research Validity Index (CRVI) can reveal why evidence-informed policies may lose validity when transferred, scaled up, or replicated across contexts.

The paper analyses the Aadhaar digital identification program in India. It demonstrates that a policy deemed evidence-based and effective can still result in exclusion, controversy, and inconsistent outcomes if contextual alignment is assumed without evaluation.

 doi:10.20944/preprints202605.1815.v1

The CRVI scoring system assesses contextual alignment between policy assumptions and implementation realities, using qualitative evaluations of relevant evidence and literature.

Scores range from *1 to 5*:

1: Severe misalignment; core assumptions are unsupported by the environment.

2: Substantial weaknesses that compromise reliability and consistency.

3: Partial alignment; the intervention can work under certain conditions but is vulnerable to changes.

4: Strong compatibility with only minor constraints.

5: High coherence; robust support and minimal friction exist.

The scores in **Table 1** are for illustrative and diagnostic purposes, not based on statistical analysis. They aim to enhance transparency in contextual assessment and demonstrate that validity conditions can vary across dimensions, even when an intervention appears successful based on overall evaluation metrics.

Table 1. CRVI Diagnostic Profile for Aadhaar as a Transferable Policy Model (Illustrative).

CRVI Dimension	***Indicative Score (1–5)***	**Diagnostic Rationale (based on secondary sources)**
Epistemic Alignment	*2.5*	The score shows partial alignment: Aadhaar's biometric logic works under normal conditions, but issues such as repeated failures, connectivity problems, and reliance on stable biometric capture reveal gaps between the system's assumptions and users' realities.
Institutional Fit	*3.0*	The score shows moderate, uneven alignment. Aadhaar benefits from strong infrastructure and admin integration, but effectiveness varies across agencies and regions due to differences in grievance handling, oversight, and implementation quality.
Cultural Resonance	*2.5*	The score shows contested legitimacy. Despite Aadhaar's widespread use, privacy issues, legal challenges, and debates over surveillance and exclusion show only partial acceptance.
Operational Feasibility	*3.5*	Aadhaar's higher score shows its nationwide deployment and integration capability. However, last-mile issues such as device reliability, connectivity, and frontline discretion hinder effective operations.
Composite CRVI (mean)	***2.88***	Indicates partial contextual validity: the model's effectiveness varies based on uneven enabling conditions that traditional "best practice" narratives often overlook.

Source: Developed by the Author, 2026.

The CRVI (Table 1) is a structured diagnostic tool, not a predictive or causal metric. Scores for each dimension are assigned based on qualitative judgment, using documentary evidence, institutional analysis, and contextual knowledge. Low scores (*1–2*) reflect significant misalignment between policy assumptions and context; mid-range scores *(3)* show partial alignment; and high scores (*4–5*) indicate strong alignment with minor issues. Composite scores are used to profile

 doi:10.20944/preprints202605.1815.v1

contextual risk rather than to provide exact thresholds or rankings. The goal of scoring is to ensure transparency and comparability in evaluation, rather than numerical precision.

5.1. Epistemic Alignment: Assumptions about Identity and Access

Aadhaar's approach to biometric authentication aims to reliably establish identity at service points, enabling accurate targeting and reducing waste. However, research shows that authentication failures caused by biometric mismatches, connectivity issues, or device malfunctions can result in the denial of welfare benefits for those without alternative identification or effective recourse (Khera, 2017; Hundal, 2020).

This reveals a gap between the system's assumptions and users' realities. As a result, the overall effectiveness score is ***2.5***, indicating a mismatch between design intentions and actual access experiences.

5.2. Institutional Fit: Administrative Capacity and Accountability

Aadhaar relies on a complex network involving central authorities, state governments, local agencies, and service points. Its success hinges on strong administrative capabilities to handle exceptions, address grievances, and coordinate across these levels. Evidence shows that when these capabilities are lacking, Aadhaar can increase procedural burdens and exclusions rather than improve service delivery (Khera, 2017). Additionally, overall "savings" metrics may mix reduced leakage with non-payments due to exclusion, a nuance often lost in traditional evaluations.

The CRVI score of ***3.0*** underscores this issue: while the framework functions under certain administrative conditions, its effectiveness declines when accountability and grievance mechanisms are weak.

5.3. Cultural Resonance: Legitimacy, Trust, and Contestation

Aadhaar's development highlights the need for cultural acceptance and legitimacy in digital identification. For many users, it connects with their existing practices, trust in government, and privacy concerns. These issues led to legal challenges and to the Supreme Court of India imposing restrictions on the use of Aadhaar in 2018.

From a CRVI perspective, this demonstrates that sociopolitical acceptance is essential for effectiveness. The indicative score of **2.5** reflects the extent to which contested legitimacy influences both adoption and long-term sustainability.

5.4. Operational Feasibility: Scaling and Last-Mile Conditions

Aadhaar demonstrates strong operational feasibility at scale, as evidenced by its nationwide enrolment and integration with various services. However, its effectiveness declines significantly in the last mile. Issues such as connectivity gaps, unreliable devices, biometric errors, and frontline discretion particularly affect vulnerable populations in rural and resource-limited areas (Dixon, 2017; Hundal, 2020).

The CRVI score of ***3.5*** highlights this issue: while large-scale deployment is possible, consistent functionality depends on uneven infrastructural and logistical conditions.

5.5. Interpreting the Composite CRVI Profile

Aadhaar's composite CRVI score of ***2.88*** highlights that evidence supporting an intervention's technical efficacy may not always translate well. This score does not imply that Aadhaar is ineffective or non-transferable; rather, it indicates that its success depends on epistemic, institutional, cultural, and operational conditions that vary across contexts. Traditional evaluation methods often overlook these dependencies, whereas the CRVI framework makes contextual validity explicit, enabling analysts and policymakers to identify which interventions are consistently effective and which depend on specific environments.

 doi:10.20944/preprints202605.1815.v1

Cross-Dimensional Interactions and Compounding Risk

The CRVI dimensions are distinct, yet the Aadhaar case shows that issues in one area can worsen problems in others. Operational failures such as biometric errors, poor connectivity, and insufficient frontline capacity not only reduce system efficiency but also erode cultural legitimacy by fostering perceptions of exclusion. When grievance and redress mechanisms are weak, issues such as false negatives and data errors lead to actual exclusion without recourse. Thus, the challenges faced by Aadhaar stem from a combination of contextual failures rather than isolated technical or administrative issues. CRVI highlights how these risks accumulate across dimensions rather than cancelling each other out.

6. What the CRVI Reveals That Evidence-Based Models Miss

Using the CRVI with Aadhaar offers valuable insights that go beyond traditional evaluation methods, which mainly focus on effectiveness, efficiency, and coverage. While these methods are important for identifying causal effects, they often overlook how stable the conditions for success are across different environments. The CRVI enhances our understanding by revealing how contextual misalignments affect policy transfer, legitimacy, operational consistency, and governance outcomes. This allows for more informed and adaptable strategies.

6.1. From Outcome Validation to Assumption Diagnosis

Traditional evaluation frameworks focus on measurable outcomes of interventions, often highlighting enrolment rates, transaction volumes, and cost savings. The CRVI takes a different approach by examining the assumptions behind these outcomes. It assesses whether the necessary conditions for reliable identification, access, and beneficiary interaction are maintained across various implementation settings. This analysis helps clarify why interventions that appear effective at scale can still lead to exclusion and uneven access in practice (Khera, 2017; Hundal, 2020).

6.2. From Best Practice to Conditional Transferability

Evidence-based policymaking often highlights successful interventions as "best practices" that can be easily transferred. However, the Aadhaar case shows that policy transfer depends on specific conditions rather than being universally applicable. Effective interventions rely on institutional safeguards, reliable infrastructure, administrative coordination, and sociocultural acceptance, which may not be present in every setting. The CRVI reframes policy diffusion as a problem of contextual alignment, urging policymakers to assess whether the essential conditions for successful transfer are in place before scaling interventions.

6.3. Revealing Legitimacy as an Evaluative Blind Spot

Standard evaluation frameworks often overlook the importance of legitimacy, trust, and sociopolitical acceptance in ensuring policy sustainability. The Aadhaar case highlights how issues such as legal challenges, privacy concerns, and feelings of exclusion can affect public acceptance, even when the system performs well technically. The CRVI addresses this by prioritising cultural resonance and legitimacy as key factors in evaluation. This approach broadens policy assessment by incorporating social acceptance and institutional trust alongside technical performance (Pawson & Tilley, 1997).

6.4. Distinguishing Scalability from Operational Robustness

Large-scale implementation is often seen as a sign of policy success. However, Aadhaar shows that scalability does not equal operational effectiveness. While the system was widely deployed, it faced ongoing last-mile issues, including connectivity problems, biometric errors, and variability in frontline execution, which undermined reliability (Dixon, 2017; Hundal, 2020). The CRVI highlights

 doi:10.20944/preprints202605.1815.v1

the difference between the ability to expand interventions and the capacity to maintain equitable and reliable operations in challenging real-world conditions.

6.5. Reframing Policy Failure as Contextual Misalignment

The CRVI redefines policy failure as a result of contextual misalignment, rather than just poor implementation or lack of evidence. Traditional evaluation models assume that evidence remains valid across different settings, but the Aadhaar case shows that interventions can lose reliability and legitimacy when contextual conditions vary. By highlighting these conditions, the CRVI helps identify where evidence portability fails and why policy effectiveness can be inconsistent across different environments.

Comparative Interpretive Value of CRVI

The mixed outcomes of the Aadhaar system often lead to the conclusion that it is partially effective, primarily due to implementation issues, capacity gaps, or inconsistent compliance. This perspective maintains that the intervention itself is valid but blames execution for its shortcomings. However, CRVI offers a different view, suggesting that failures stem from deep-rooted misalignments across dimensions such as knowledge, institutions, culture, and operations. Instead of seeing problems like exclusion or breakdowns as mere implementation flaws, CRVI recognises them as indicators that the necessary conditions for the policy's success are lacking. This reinterpretation does not dismiss the causal evidence; rather, it highlights when and why evidence-backed policies fail outside their original supportive contexts.

The CRVI shifts the focus of policy evaluation by positioning contextual alignment as essential for policy validity, rather than treating it as a secondary factor. While traditional evaluation methods often view context as a mere background element, the CRVI emphasises that contextual coherence is crucial for maintaining the credibility of evidence across different settings. This approach moves beyond simply asking if interventions are effective to examining whether the conditions needed for evidence to be relevant in new contexts are preserved during policy transfer.

7. Implications for Policy Evaluation and Governance

The CRVI analysis highlights important implications for the production, evaluation, and governance of evidence in policy systems. These insights are relevant not only to the Aadhaar case but also to wider challenges in evidence-based policymaking, especially when policies are transferred, scaled, or standardised across different contexts.

7.1. Rethinking Evaluation Beyond "What Works"

The findings indicate the need to move beyond evaluation frameworks that focus solely on outcome validation. While it's important to know if an intervention works in specific conditions, this alone isn't enough for effective policy transfer. The CRVI emphasises the need to pair causal evaluations with contextual diagnostics to ensure the conditions for effectiveness exist elsewhere. Evaluators should shift their focus from just attribution to assessing both effectiveness and contextual validity. Including these diagnostics earlier in the evaluation process—especially during design and scaling—can help prevent the misuse of reliable evidence.

7.2. From Best Practice to Best Fit in Policy Design

The analysis highlights the shortcomings of "best practice" narratives in policy diffusion. Often, successful policies are presented as universally applicable, with adaptation overlooked. The CRVI emphasises that adaptation should be central, making contextual fit a key evaluation criterion. Policymakers should adopt a best-fit approach, assessing interventions not only for effectiveness but also for compatibility with local knowledge, institutional structures, cultural norms, and operational capacities. This approach promotes more selective, conditional, and context-aware policy transfer.

 doi:10.20944/preprints202605.1815.v1

7.3. Strengthening Governance Through Contextual Audits

The CRVI impacts governance and accountability. When contextual validity is assumed without assessment, policy failures are often misattributed to implementation issues or non-compliance. This can hide responsibility and damage trust in institutions and evidence-based methods. The CRVI provides a framework for contextual audits to clarify accountability when outcomes don't meet expectations. By making assumptions about context explicit, governance systems can better identify design flaws, contextual challenges, and execution failures, leading to more transparent and accountable decision-making.

7.4. Anticipating Legitimacy and Sustainability Risks

Addressing cultural resonance and legitimacy is crucial for the sustainability of policies. Even if interventions excel in technical or administrative measures, they may encounter resistance or legal challenges if they conflict with existing norms and values. The CRVI urges policymakers to consider legitimacy alongside operational feasibility to mitigate these risks. This is especially important for technology-driven governance reforms, in which technical efficiency may exceed social acceptance. By prioritising legitimacy, the CRVI fosters more lasting and socially relevant policy design.

7.5. Integrating the CRVI into Evaluation and Policy Cycles

The CRVI can be seamlessly integrated into current evaluation and governance processes without replacing established methods. It serves as a diagnostic tool for assessing transfer readiness, interpreting uneven outcomes, and comparing the performance of similar interventions across different settings. By doing so, the framework strengthens evidence-based approaches, aligning them with the realities of policy implementation and governance.

Overall, these insights indicate that better policy outcomes depend not only on improved evidence but also on stronger alignment between that evidence and the specific context. The CRVI operationalises contextual validity, providing policymakers and evaluators with a practical way to ensure this alignment and minimise the risks of context-blind policy transfer.

7.6. Operational Instruments for Contextual Validity

7.6.1. CRVI Pre-Transfer Audit

CRVI can be used as a pre-transfer audit, added to evaluation reports, for policies intended for replication, scaling, or international use. This audit would not reassess causal impact but will document essential assumptions about reliability, safeguards, cultural acceptance, and operational capacity in the target context. By highlighting these contextual factors in advance, the audit enables a shift from retrospective outcome assessment to proactive risk diagnosis.

Illustrative Ex Ante Use of CRVI

To apply CRVI ex ante, imagine a policymaker wanting to transfer a biometric digital identification system from a well-connected, centralised country to a low-connectivity area with fragmented services and limited grievance mechanisms. While biometric matching is technically reliable, operational challenges such as inconsistent electricity supply, poor network coverage, and insufficient frontline capacity would result in low scores. Additionally, weak legal safeguards, the absence of appeal mechanisms, and accountability issues may hinder institutional alignment. Cultural legitimacy could also be questionable in areas with low trust in government data practices. The CRVI profile would not dismiss the intervention but would highlight significant contextual risks, signalling the need for design adjustments, careful sequencing, or additional investments before transferring evidence from the original context.

This shows that the framework can be utilised prior to policy transfer, rather than merely serving to analyse failures after implementation.

 doi:10.20944/preprints202605.1815.v1

7.6.2. Contextual Risk Register

CRVI should be integrated into a contextual risk register that connects to governance and accountability systems. Instead of viewing contextual factors as mere background, the register should document identified misalignments in CRVI dimensions, assign responsibility for addressing them, and monitor changes in contextual risks throughout implementation. This method links contextual validity with standard risk management practices and clarifies if failures arise from design assumptions, contextual limitations, or execution issues.

7.6.3. Legitimacy Stress-Test for Digital Public Infrastructure

CRVI advocates for legitimacy stress-tests during the implementation of digital public infrastructure and data governance systems. These tests would evaluate not only technical performance but also access to redress, error correction, and public trust under real-world conditions. By prioritising legitimacy as a key factor in validation rather than a result of success, this approach addresses a common oversight in digital policy interventions.

The CRVI reframes contextual validity as a primary factor in evaluation instead of a secondary one. Traditional evaluation methods often see context as merely background or a moderating influence on outcomes. In contrast, the CRVI views contextual alignment as essential to policy validity. This approach shifts the focus of policy evaluation from merely assessing whether interventions are effective to ensuring that the necessary conditions for credible evidence are maintained across different environments.

8. Conclusions

This paper tackles a key issue in evidence-based policymaking: why interventions with strong evidence in one context often fail or face challenges when applied elsewhere. Although modern evaluation frameworks recognise the importance of context, they lack systematic tools to assess whether the conditions that support policy effectiveness in one setting will apply in another. Consequently, policy failures are often mistakenly attributed to poor implementation or lack of effort rather than to fundamental contextual mismatches.

To address this, the paper introduces the Contextual Research Validity Index, a framework that views validity as relational rather than an inherent quality of evidence. CRVI evaluates contextual alignment across four dimensions: epistemic, institutional, cultural, and operational. This enables a clearer assessment of evidence transferability and helps explain failures after they occur. An illustrative application to India's Aadhaar system reveals how strong evidence claims can lead to exclusion and disputes when contextual conditions are inconsistent or when failures compound across dimensions.

The analysis emphasises that policy success cannot be judged solely by effectiveness in the original context. When policies are transferred, their validity relies on governance capacity, legitimacy, institutional safeguards, and operational realities that may not be captured by standard evaluation metrics. CRVI complements causal evaluation by clarifying when evidence loses its intended function outside its original context and by reframing policy failure as an issue of contextual validity rather than merely an implementation issue.

Several limitations exist. The application of CRVI here is illustrative and based on secondary sources rather than original data. The scoring method is heuristic and aims to facilitate diagnostic reasoning rather than precise measurement. Future research could enhance the framework by developing standardised scoring rubrics, assessing inter-rater reliability, and applying CRVI across cases to identify patterns of contextual risk. Additionally, exploring how CRVI interacts with formal models of transportability and decision-theoretic approaches to policy transfer would be beneficial.

Despite its limitations, this paper advances a more context-sensitive understanding of evidence portability and policy evaluation. Rather than assuming that evidence retains its validity once transferred across settings, the CRVI emphasises that policy effectiveness depends on the continued

 doi:10.20944/preprints202605.1815.v1

alignment of epistemic assumptions, institutional arrangements, cultural legitimacy, and operational capacity. By making these dependencies explicit, the framework provides policymakers and evaluators with a structured way to diagnose contextual risk before implementation failures emerge.

In an era of rapid policy diffusion, digital governance expansion, and global replication of "best practices," responsible evidence-based policymaking requires more than identifying what works. It also requires understanding where, for whom, and under what conditions policy validity can survive transfer across contexts. The CRVI contributes to this shift by repositioning contextual alignment from a secondary implementation concern to a central condition of evidence credibility itself.

References

1. Andrade, C. (2023). Real World Studies: What They Are and What They Are Not. Indian Journal of Psychological Medicine, 45(5), 537–538. https://doi.org/10.1177/02537176231188563
2. Angrist, J. D., & Pischke, J. S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400829828
3. Banerjee, A., Duflo, E., & Kremer, M. (2020). The Influence of Randomized Controlled Trials on Development Economics Research and on Development Policy. The MIT Press EBooks, 439–487. https://doi.org/10.7551/mitpress/11130.003.0015
4. Cartwright, N., & Hardie, J. (2012). Evidence Based Policy: A Practical Guide to Doing it Better. Oxford University Press. http://ukcatalogue.oup.com/product/9780199841622.do
5. Coyle, K., April Idalski Carcone, Seyram Butame, Meardith Pooler-Burgess, Chang, J., & Naar, S. (2022). Adapting the self-assessment of contextual fit scale for implementation of evidence-based practices in adolescent HIV settings. Implementation Science Communications, 3(1). https://doi.org/10.1186/s43058-022-00349-4
6. Demelash, S., Ayano, G., Girma, M., Zuylen, P. V., Aragaw, F. M., Darebo, T. D., … & Spigt, M. (2025). Development and validation of an assessment tool for public health emergency management program. Global Health Research and Policy, 10(1). https://doi.org/10.1186/s41256-025-00423-z
7. Dixon P. (2017). A Failure to "Do No Harm" -- India's Aadhaar biometric ID program and its inability to protect privacy in relation to measures in Europe and the U.S. Health and technology, 7(4), 539–567. https://doi.org/10.1007/s12553-017-0202-6
8. Frimpong, V. (2026). The Challenge of Context-Free Validity: Introducing the Contextual Research Validity Index Framework for Situated Legitimacy under Socioeconomic Challenges. SocioEconomic Challenges, 10(1), 42–49. https://doi.org/10.61093/sec.10(1).42-49.2026
9. Hockley, J., Froggatt, K., Van den Block, L., Onwuteaka-Philipsen, B., Kylänen, M., Szczerbińska, K., Gambassi, G., Pautex, S., & Payne, S. A. (2019). A framework for cross-cultural development and implementation of complex interventions to improve palliative care in nursing homes: the PACE steps to success programme. BMC Health Services Research, 19(1). https://doi.org/10.1186/s12913-019-4587-y
10. Huebschmann, A. G., Leavitt, I. M., & Glasgow, R. E. (2019). Making Health Research Matter: A Call to Increase Attention to External Validity. Annual Review of Public Health, 40(1), 45–63. https://doi.org/10.1146/annurev-publhealth-040218-043945
11. Hundal, H. S., & Chaudhuri, B. (2020). Digital Identity and Exclusion in Welfare. Proceedings of the 2020 International Conference on Information and Communication Technologies and Development. https://doi.org/10.1145/3392561.3397583
12. Khera, R. (2017). Impact of Aadhaar on Welfare Programmes. Economic and Political Weekly, 52(50), 61–70. http://www.jstor.org/stable/45132600
13. Kwan, B. M., Rementer, J., Ritchie, N. D., Nederveld, A. L., Phimphasone-Brady, P., Sajatovic, M., Nease, D. E., & Waxmonsky, J. A. (2020). Adapting Diabetes Shared Medical Appointments to Fit Context for Practice-Based Research (PBR). The Journal of the American Board of Family Medicine, 33(5), 716–727. https://doi.org/10.3122/jabfm.2020.05.200049
14. Mayne, J. (2012). Contribution analysis: Coming of age? Evaluation, 18(3), 270–280. https://doi.org/10.1177/1356389012451663

 doi:10.20944/preprints202605.1815.v1

15. Mehrotra, M. L., Petersen, M. L., & Geng, E. H. (2019). Understanding HIV Program Effects: A Structural Approach to Context Using the Transportability Framework. JAIDS Journal of Acquired Immune Deficiency Syndromes, 82(3), S199–S205. https://doi.org/10.1097/qai.0000000000002202
16. Pawson, R., & Tilley, N. (1997). Realistic evaluation. Sage Publications, Inc.
17. Peck, J., & Theodore, N. (2015). Fast Policy. https://doi.org/10.5749/minnesota/9780816677306.001.0001
18. Peters, D. H., Tran, N. T., & Adam, T. (2013). Implementation research in health: a practical guide. World Health Organisation.
19. Rodrik, D. (2008). Second-Best Institutions. American Economic Review, 98(2), 100–104. https://doi.org/10.1257/aer.98.2.100
20. Supreme Court of India. (2018). Justice K.S.Puttaswamy(Retd) vs Union Of India on 26 September, 2018. Indiankanoon.org. https://indiankanoon.org/doc/127517806/
21. Williams, S. C., Milam, A. J., Furr-Holden, D. M., & Salgado, C. C. (2020). Measurement of Social Processes at the Neighborhood Level in Baltimore City. American Journal of Community Psychology, 66(1-2), 53–64. https://doi.org/10.1002/ajcp.12424
22. World Bank (2022), Annual Reports | Identification for Development. Worldbank.org. https://id4d.worldbank.org/annual-reports
23. Ziemann, A., Brown, L., Sadler, E., Ocloo, J., Boaz, A., & Sandall, J. (2019). Influence of external contextual factors on the implementation of health and social care interventions into practice within or across countries—a protocol for a "best fit" framework synthesis. Systematic Reviews, 8(1). https://doi.org/10.1186/s13643-019-1180-8