Nothing about them without them: insights from six countries on involving "beneficiaries" in co-design, monitoring, and evaluation of research-to-impact projects

Zenebe Uraguchi*

Bern University of Applied Sciences, School of Agricultural, Forestry, and Food Sciences (HAFL), Zollikofen, Switzerland

This study examines six participatory development cases situated in diverse institutional and sectoral contexts across Albania, Bangladesh, Ethiopia, Kosovo, the Philippines, and Tanzania. The study explores how systemic transformation emerges through co-design, adaptive learning, and institutional alignment,moving beyond linear input-output models of research impact. Using a developmental systems lens, the study treats impact not as a final result, but as a dynamic outcome shaped by the interplay of behavior, norms, and structures over time. Over 1,800 semi-structured interviews were triangulated with project documents, observational data, and binary logistic regression models to examine the influence of 10 participatory design features on sustained developmental outcomes. A key finding is that participation, particularly in monitoring, adaptive learning, and early framing, is not merely procedural but a systemic driver of institutional legitimacy, stakeholder trust, and long-term uptake. Countries with embedded participatory mechanisms, such as Ethiopia and Albania, showed deeper policy integration and structural change, while fragmented governance contexts, such as Tanzania and Kosovo, saw limited institutional embedding despite localized behavioral shifts. Crucially, the study argues that how research is done—who frames it, who participates in it, and how it adapts—is as consequential as what it seeks to achieve. Methodological integration of qualitative sensemaking and quantitative modeling offers practical insights into navigating complexity in research-to-impact pathways. Rather than serving as a report on six distinct cases, this article positions them as illustrations of a broader paradigm shift: from static, technocratic models to dynamic, participatory systems approaches. It offers both theoretical grounding and actionable guidance for researchers, implementers, and policymakers seeking to align evaluation and design with the realities of complex social systems.

KEYWORDS

participatory research, systemic change, research-to-impact pathways, institutional embedding, sensemaking analysis, developmental evaluation, evolutionary economics

Introduction

In the evolving field of development finance and cooperation, the relevance of Pyle's (1984) work, "Life After Project," remains crucial, particularly in understanding the inclusion, sustainability, and scalability of development outcomes and the lessons that can be learned from both successes and failures. Pyle critiqued the tendency of pilot projects to generate initial enthusiasm by showcasing small-scale successes, only for their impact to diminish or disappear altogether once the projects end. This highlights the fundamental challenge of transitioning from pilot projects to sustainable, scalable initiatives that continue to generate impact long after the formal conclusion of the project.

This paper is primarily a research-based analysis informed by a developmental systems lens, a conceptual framework that emphasizes the complex, adaptive, and evolving nature of change processes across time and scales. Unlike conventional linear models of research dissemination, this study recognizes that the translation of research into societal impact is shaped by a constellation of dynamic interactions across actors, institutions, and contexts. In doing so, the study bridges a critical gap in current research by offering both empirical evaluation and theoretical reflection, with a particular focus on participatory development (Tacchi et al., 2010; Cousins and Earl, 2004; Chouinard and Milley, 2018).

While the cases are based in six countries, the study does not approach them as representative "country cases." Instead, each case is treated as a distinct contextual system, comprising sectoral, institutional, and political dimensions, through which systemic impact is examined. This allows the analysis to avoid essentialist assumptions and focus on situated dynamics. Having this in mind, what distinguishes this paper is its dual aim: to empirically examine the factors shaping research-to-impact pathways, and to explore how participatory development both influences and is influenced by dynamic systems. Unlike many existing studies that assess impact retrospectively through structured models or treat participation as procedural, this paper positions participation as a systemic and epistemological force. For instance, Belcher and Hughes (2021) emphasize outputs within logic-based frameworks, overlooking adaptive complexity. Wailzer and Soyer (2022) co-develop participatory models but frame them within bounded theories of change. Faure et al. (2020) apply participatory assessments but focus on validation, not emergence. Even reflexive approaches like Blundo-Canto and Ferré (2022) stop short of theorizing participation as a driver of systemic change. Likewise, while Douthwaite et al. (2023) address long causal chains, their model lacks a fully integrated developmental systems perspective. This paper fills that gap by treating participation as a co-evolving mechanism in complex systems of change.

This study positions itself by theorizing participatory development as a co-evolutionary process within developmental systems. It provides a forward-looking model in which participation and system transformation are mutually constitutive, enabling more adaptive, inclusive, and contextually responsive research-to-impact pathways. The study also critiques conventional research-impact frameworks and advocates for approaches informed by evolutionary economics and complex systems thinking (Nishibe, 2006; Helbing and Kirman, 2013), which are more suitable for capturing the dynamics of systemic change.

Success in development, this study argues, should be redefined to include structural transformations and long-term impacts rather than focusing solely on short-term metrics. Through a comparative analysis of projects in the six countries, this study highlights how context-specific factors influence development outcomes. By emphasizing stakeholder-centered approaches, it advocates for policies and practices that prioritize inclusivity, sustainability, and scalability. The findings aim to provide actionable guidance for researchers and practitioners to design and implement development projects that place disadvantaged groups at the core of defining success, ensuring their voices shape impactful, long-term outcomes.

The principle "Nothing About Them Without Them" serves as a guiding philosophy for effective development work, emphasizing the importance of involving primary stakeholders, often labeled as "beneficiaries," in every phase of a project. From co-design to monitoring, evaluation, and learning, their involvement ensures that development initiatives genuinely address their needs and aspirations. This principle aligns with a broader focus on social justice, human rights, and participatory development (Gupta et al., 2015; Gupta and Pouw, 2017; Kyamusugulwa, 2013). The United Nations Sustainable Development Goals (SDGs) further underline this by adopting the "Leave No One Behind" framework, which commits to prioritizing disadvantaged groups in development efforts (Weber, 2018).

At the heart of impactful development interventions is the identification and understanding of disadvantaged and excluded groups. These groups are often heterogeneous, including women, ethnic minorities, internal migrants, individuals with disabilities, and those living in remote areas. Structural inequalities, as noted by Bailey et al. (2017), limit access to essential resources, opportunities, and capabilities, leading to exclusion. This calls for development interventions to identify and address these root causes of exclusion during the development initiatives' analysis or research phase. As Stockless and Brière (2024) emphasize, this requires asking the right questions during the design phase, supported by research, to understand the systemic inequalities that prolong exclusion.

Recent research shows that inclusiveness in development is essential for translating research into impactful, long-term outcomes (Grin et al., 2010; Bradshaw et al., 2017). It extends beyond economic growth and income equality, focusing on empowering disadvantaged groups to actively participate in and benefit from development processes. Studies demonstrate that interventions grounded in inclusive research design, such as social and communication skills training and personalized support, significantly improve the social integration of individuals with disabilities in low- and middle-income countries (Saran et al., 2020). Research also confirms that enhancing individual agency through participatory approaches leads to better mental health outcomes and improved capacity for navigating social and economic opportunities (Shankar et al., 2019). This means that projects that integrate such evidence-driven strategies during the design and implementation phases achieve more sustainable and meaningful impacts, reducing the risk of perpetuating structural inequalities.

This study further highlights that effective development projects require collaboration across multiple stakeholders, emphasizing the importance of engaging those with power

and resources alongside disadvantaged groups. Community-Based Inclusive Development (CBID) models illustrate how research-driven approaches enhance inclusiveness by involving disadvantaged populations in project planning and implementation (Varughese et al., 2025). This collaboration creates synergies between research insights and on-the-ground interventions, ensuring that projects are contextually responsive and adaptable.

However, many development projects fail to adequately incorporate insights from post-project evaluations (ex-post) into the design of new initiatives (ex-ante). As Garbero et al. (2023) argue, "inclusion by design" must be intentional and proactive, identifying barriers and opportunities for disadvantaged groups from the outset. This approach resonates with Sen's (1999) capabilities framework, which asserts that meaningful development occurs when interventions enhance the wellbeing, freedom, and agency of primary stakeholders.

Ensuring inclusivity in development requires multiple research-informed pathways to create systemic and transformative impact. One pathway involves leveraging opportunities within specific sectors. For example, in Albania, the information, communication, and technology (ICT) sector was identified for its potential to employ disadvantaged groups, including women, Roma communities, and disabled youth. This reflects labor market segmentation theory, which suggests that certain sectors offer more equitable access when inclusive strategies are implemented (Grimshaw et al., 2017). Another pathway focuses on geographical targeting, addressing spatial inequalities by implementing interventions tailored to regions with high concentrations of disadvantaged populations. In Kosovo, a job portal accessible in the Serbian language targeted Serbian communities, demonstrating strategies that address location-based barriers to opportunity (Bhaumik et al., 2011).

Inclusive development also requires engagement with both formal and informal institutions that shape power relations. Projects challenge institutional barriers by involving public institutions, civil society organizations, and community leaders. This aligns with Gaventa's (2021) "power cube" framework, which emphasizes addressing visible, hidden, and invisible dimensions of power to enhance systemic inclusion. However, a key challenge for projects promoting inclusivity is balancing the pressure to deliver short-term results with the need for long-term systemic change by understanding key constraints during the research and design process. Donors and policymakers often prioritize immediate, measurable outcomes, such as job creation or increased investment, over deeper, structural transformation (Stanley and Connolly, 2023). This pressure can hinder projects from addressing the root causes of exclusion, resulting in only incremental improvements.

This means that achieving systemic change requires sustained efforts to reform how institutions operate and how disadvantaged groups access opportunities. These transformations often take time and do not yield immediate, quantifiable results, making it difficult to align with traditional evaluation models that emphasize linear cause-effect relationships and short-term outputs (Gutheil, 2021). To navigate these tensions, research suggests taking up adaptive management approaches that allow projects to respond to evolving conditions. Agile monitoring and evaluation systems track real-time changes and support adaptive

interventions in complex environments (Synowiec et al., 2023). Despite these recommendations, many development evaluation models remain ill-equipped to capture non-linear, feedback-driven processes, creating a bias against systemic change initiatives (Reilly-King et al., 2024).

Materials and methods

This study explores how participatory development research can facilitate inclusive, sustainable, and scalable impact by driving systemic change within complex development contexts. Rather than treating impact as a linear outcome of knowledge dissemination, the research frames it as an emergent property of dynamic interactions between actors, institutions, and environments. It employs a developmental systems lens that highlights the importance of feedback loops, adaptive learning, and institutional evolution over time (Capra and Luisi, 2014; Senge, 2006). This lens enables a focus on how change unfolds across interdependent domains—behavioral, structural, and normative—rather than isolating discrete interventions.

However, systems thinking alone is often critiqued for its abstract, depoliticized treatment of power. It may obscure the influence of historical inequality, institutional inertia, and elite capture, which frequently undermine participatory intent and distort pathways to impact (Cooke and Kothari, 2001; Gaventa, 2006). To address these blind spots, the study integrates a Participatory Action Research (PAR) methodology that foregrounds the lived experiences, agency, and knowledge of stakeholders, particularly those traditionally excluded from decision-making. PAR offers a practical and ethical commitment to epistemic justice, ensuring that affected communities are not just consulted but actively shape the research process, from agenda-setting to interpretation and validation. This methodological synergy between systems thinking and participatory action enables a more reflexive, power-aware analysis of how research contributes to transformative development.

PAR: justification and implementation

This study employed PAR as a guiding methodological orientation to ensure that research processes were inclusive, contextually grounded, and responsive to the lived realities of local actors. PAR was used across all six countries as a framework for engaging stakeholders not only as informants but as co-designers and validators of the research process. This approach was particularly relevant in contexts where participatory development projects were already underway, and where community actors had experiential knowledge that could meaningfully inform both the framing and interpretation of research-to-impact pathways (Reason and Bradbury, 2008; Kindon et al., 2007; Pain and Kesby, 2007).

Co-design workshops were held at the start of the research process to identify priority themes and refine tools in collaboration with local actors, such as civil society groups in Kosovo, farmer cooperatives in Tanzania, and frontline service providers in

Ethiopia. These engagements contributed to the adaptation of interview instruments to local cultural and institutional contexts and ensured that key concepts (e.g., "impact," "inclusion," "systemic change") were aligned with stakeholders' lived experience. Additionally, facilitators were trained and supported to assist with interviews, contextual translation, and post-interview discussions, particularly in linguistically or politically sensitive settings.

PAR also supported iterative data validation through structured feedback sessions, where preliminary themes were shared with community participants for comment, correction, or expansion. For example, in Bangladesh, stakeholder feedback led to the refinement of one category from "technical training" to "confidence-building spaces," reflecting participants' emphasis on emotional security as a precondition for participation. These exchanges helped ensure that interpretive accuracy was grounded in local narratives, not only the researcher's assumptions.

However, the use of PAR also presented practical and epistemological limitations. While participatory processes were widely welcomed in principle, time constraints, political sensitivities, the COVID pandemic, and power asymmetries within communities often limited the depth of engagement. In some settings, gatekeeping by local elites or organizational partners influenced who could participate in workshops or interviews, potentially reinforcing existing hierarchies, an issue noted in broader PAR literature as a recurring challenge (Cooke and Kothari, 2001; Cornwall, 2008). Furthermore, participation was uneven across sites. In Ethiopia and the Philippines, conflict and COVID-19-related disruptions meant that participatory activities had to be adapted for remote formats, which limited the inclusion of the most vulnerable populations.

There were also methodological trade-offs. While PAR enabled responsiveness and co-production, it introduced variability across contexts in how research was shaped and carried out. This posed challenges for comparability and consistency in data collection. In some contexts, researchers had to balance collaborative ideals with the need for analytical clarity and timeline management, highlighting the tension between deliberative depth and operational feasibility (Cahill, 2007; Pain and Francis, 2003). In addition, while many participants expressed appreciation for their involvement, not all had the capacity or interest to engage in reflexive processes, raising ethical questions about participation fatigue and over-asking in settings where time and resources were limited. Despite these constraints, PAR added significant value by enhancing contextual relevance, stakeholder legitimacy, and trust. It helped surface hidden narratives. This was true especially around power, legitimacy, and emotional safety, which might have been missed in conventional methods.

Country selection: rationale and limitations

It is important to emphasize that the term "country" is used here as a geographic anchor, not as an explanatory category. The study focuses on case-based systemic configurations, not national attributes, thereby avoiding reductionist comparisons. Based on this understanding, the six countries of Albania, Bangladesh, Ethiopia, Kosovo, the Philippines, and Tanzania were selected

through purposive sampling to reflect institutional, geographic, and sectoral variation in research-to-impact dynamics. The selection was not designed for representativeness in a statistical sense, but rather to enable analytical generalization by examining contrasting systems where participatory development projects had been implemented with varying levels of institutional maturity and civic engagement (Flyvbjerg, 2006; Yin, 2017).

The selection was guided by five interlinked criteria. First, each country had an established participatory research initiative active during or before the study period, providing a foundation for stakeholder engagement and retrospective inquiry. All six projects explicitly stated that they used a systems approach, often presented as "Market Systems Development" (MSD). For example, Bangladesh featured gender-responsive agricultural programming focused on medicinal herbs for landless women, Ethiopia emphasized women's participation in income-generating sectors such as agriculture and hospitality, and the Philippines supported livelihood transitions for farmers post-disaster. Tanzania explored horticulture as a pathway to income and employment for youth and women, while Albania and Kosovo focused on youth education and employment.

Second, the study prioritized sectoral diversity to capture how inclusive development unfolds across different systems. The selected cases span agriculture, education, labor markets, and local governance. These were sectors marked by both institutional change and political contestation. These domains allowed analysis of varying institutional dynamics and power relations, from education-to-employment transitions in the Balkans to gendered agricultural innovation in South Asia and East Africa, and disaster-related reforms in Southeast Asia.

Third, institutional diversity was key. Countries ranged from relatively stable bureaucracies (e.g., Bangladesh, Albania) to more fragile or post-conflict settings (e.g., Ethiopia, Kosovo), allowing the study to observe how context-specific political and historical legacies shaped participatory processes and impact pathways (Thelen, 2004; Gaventa, 2006). Fourth, cultural and geographic variation was deliberately built into capture differences in local governance traditions, social hierarchies, and knowledge practices across South Asia, Sub-Saharan Africa, and Southeast Europe. This heterogeneity was valuable for analyzing how localized meanings of participation, legitimacy, and research relevance were co-constructed and challenged. Finally, all selected countries had established research partnerships or trusted local collaborators, which were essential for ensuring ethical engagement, cultural translation, and logistical feasibility (Clark et al., 2021).

Despite this strategic design, the approach carries important limitations. First, the purposive and context-driven selection means that findings cannot be generalized statistically to all development contexts; rather, they offer theory-building insights through comparative exploration of diverse cases (Yin, 2017). Second, access to some populations was constrained. In Ethiopia and parts of the Philippines, armed conflict and COVID-19 restrictions disrupted field activities, limited travel to remote regions, and restricted the participation of vulnerable groups, such as displaced persons, informal workers, and minority communities (Hickey and Mohan, 2004).

While local facilitators conducted interviews in participants' preferred languages and helped contextualize meanings, the process of transcription and translation inevitably led to some loss of nuance, particularly in emotionally charged or culturally specific terms. Even with back-translation checks, certain expressions of legitimacy, trust, or power could not be fully captured in English without interpretive framing.

Cross-country comparison also posed analytical challenges. Although a shared framework was applied across all cases, the degree of methodological adaptation required in each country, due to institutional realities, language, and political sensitivities, meant that comparisons had to remain interpretive and flexible. This shows the importance of reflexivity and humility in comparative research, particularly in global South contexts where uniformity may be both unrealistic and undesirable (Cornwall, 2008).

Data collection: semi-structured interviews and triangulation

The core qualitative data collection method was semi-structured interviews, chosen for their ability to strike a balance between comparability across sites and contextual depth. This approach allowed the research to pursue key themes: project participation, institutional shifts, adaptive learning, and developmental outcomes. The study tried, as much as possible, to remain responsive to localized meanings and participant narratives. Interviews typically lasted between 45 and 90 min, conducted in the native languages of participants by trained local facilitators and/or bilingual researchers, which significantly enhanced both rapport and data quality (Barriball and While, 1994; Ahlin, 2019).

Participants were briefed on the purpose of the research and informed of their rights, including confidentiality, voluntary participation, and the right to withdraw. Interviews were audio-recorded, transcribed verbatim, and anonymized to ensure ethical data handling and minimize identification risks. To further enhance data reliability and mitigate researcher bias, triangulation was employed through three complementary strategies:

Document reviews

Data collectors examined project reports, monitoring and evaluation documents, training manuals, and institutional frameworks relevant to each case. This helped situate participant accounts within broader institutional processes and timelines and allowed researchers to validate or contrast interview findings against the official record (Kallio et al., 2016).

Project visits

Data collectors participated in site visits to project locations, where they observed the physical and organizational settings in which interventions were implemented. These visits facilitated real-time conversations with stakeholders, exposure to operational dynamics, and a deeper understanding of implementation realities, especially useful in evaluating participation, accessibility, and visible outcomes (Phillippi and Lauderdale, 2018).

Field observations

Structured and unstructured non-participant observation was conducted in community meetings, training sessions, market settings, and project demonstration sites. These observations helped capture non-verbal cues, gendered interactions, group hierarchies, and informal dynamics often missed in interviews. They also served as a critical tool to verify the coherence between stated narratives and observed practices (Gubrium and Holstein, 1998). Where permitted, transcripts were returned to participants for member checking, a process that increased trust and allowed respondents to review or clarify statements, enhancing both accuracy and reflexivity. However, logistical barriers such as limited connectivity, literacy constraints, and time availability made this step uneven across sites.

While semi-structured interviews provided rich and adaptable insights, challenges were evident. Comparability across contexts varied depending on how participants interpreted or re-framed questions. Interviewers had to balance following the guide with allowing space for emergent, locally significant themes. In some cases, power dynamics, especially in hierarchical or gendered contexts, subtly shaped what was voiced and what remained unspoken. The triangulated approach not only deepened validity but also clarified gaps between policy and practice, revealed discrepancies between official reports and community narratives, and strengthened the interpretive robustness of the qualitative analysis (Roulston and Choi, 2018; Cachia and Millward, 2011). To ensure relevance, inclusivity, and contextual sensitivity, the study employed a two-tiered sampling strategy, combining purposive and snowball sampling methods. These approaches are widely recognized in qualitative research for their ability to access information-rich cases and marginalized voices within complex social systems (Patton, 2011; Noy, 2008).

The primary method was purposive sampling, through which participants were deliberately selected based on their direct engagement in the observed projects. These included local project implementers, government partners, facilitators, and community-based "beneficiaries". The aim was to capture perspectives from both "formal actors" (e.g., government or NGO staff) and "experiential actors" (e.g., youth participants, women entrepreneurs, or smallholder farmers) who were integral to the implementation or impact of interventions. Care was taken to ensure balance across gender, institutional roles, age, and geographic location, thereby increasing representational diversity and enhancing the credibility of findings (Palinkas et al., 2015). Particular attention was given to including women and youth in traditionally underrepresented sectors such as agricultural R&D and vocational education.

In more remote, fragile, or socially embedded contexts, such as ethnic minority groups, women in rural governance, or informal sector workers, snowball sampling was applied to supplement participant lists. Initial informants recommended others whose insights were vital for understanding institutional dynamics, exclusions, and informal networks within local systems. This method is especially effective in reaching "hidden" or "hard-to-reach" populations and generating insider knowledge (Noy, 2008; Sadler et al., 2010). While snowball sampling enriched the diversity of perspectives, it also presented risks of network bias, wherein dominant or vocal actors might overrepresent particular

viewpoints. To counteract this, the study sought referrals across multiple entry points and community gatekeepers to ensure triangulated access.

Local facilitators were indispensable in both phases of sampling. Their presence helped establish community trust, facilitated linguistic translation of technical concepts, and ensured that culturally sensitive engagement protocols were followed. In many cases, they also identified non-verbal cues or informal hierarchies that helped the research team navigate power asymmetries. These relationships were especially important in communities where formal authority structures were weak or contested. Moreover, facilitators helped bridge logistical and sociocultural barriers, including low literacy levels, limited digital connectivity, and local dialect variations. These challenges are commonly cited in participatory research in low-income settings (Creswell and Poth, 2016; Guest et al., 2006).

Analytical strategy: integrating quantitative and qualitative methods

By integrating binary logistic regression and sensemaking-based qualitative analysis, the study sought to capture both measurable patterns of impact and the subjective meanings that shaped stakeholder engagement and outcomes. This approach aligned with the study's PAR ethos, which emphasized collaboration, contextual sensitivity, and mutual learning. It is important to clarify that while the regression model helps identify statistical associations, it is not the centerpiece of this study. Its purpose is to scaffold the more central methodological innovation: integrating participatory sensemaking with quantitative trends to advance complexity-informed evaluation.

The quantitative component used binary logistic regression to model the likelihood of specific development outcomes, such as behavioral adoption, institutional engagement, or sustained use of tools, based on a range of independent variables, including gender, age, education, institutional role, and level of participation. The regression models enabled the identification of statistically significant associations while controlling for confounding factors, offering a structured view of how participation patterns varied across demographic and institutional contexts. For instance, the models demonstrated that individuals who engaged in co-design or monitoring platforms were more likely to report continued use of research-derived practices. These findings echoed evidence from other development evaluations (McEvoy et al., 2016; Rinaldi, 2020), affirming that structured stakeholder engagement can influence long-term adoption of new behaviors or norms.

However, logistic regression, while useful for identifying associations, could not explain the underlying mechanisms or contextual conditions behind these patterns. It was therefore complemented by a robust qualitative analysis using sensemaking. Sensemaking, as developed by Weick (1995) and extended by Maitlis and Christianson (2014), focuses on how individuals and groups interpret change in situations of complexity or uncertainty. It is not merely about reporting facts, but about how stakeholders frame, justify, and emotionally respond to development processes. In this study, sensemaking enabled researchers to interpret not just

whether change occurred, but how it was understood, what it meant to different actors, and why it was embraced or resisted.

Qualitative data, primarily from semi-structured interviews with over 1,800 participants, were analyzed using NVivo 12. The coding process was both deductive, based on the study's conceptual framework (shifts in behaviors, structures, and norms), and inductive, allowing new themes to emerge from the data itself. This dual approach enabled consistent thematic structuring across countries while also capturing local nuances, contested narratives, and emergent insights. For example, in the Philippines, research outputs were viewed by local officials not only as technical tools but as instruments of political legitimacy. In Ethiopia, women described project engagement as a process of redefining their public identities, not just as economic actors but as legitimate leaders within their communities.

This combined method provided powerful triangulation. Where regression models revealed patterns (e.g., platform participation correlated with behavior change), qualitative sensemaking helped explain the relational and institutional mechanisms behind those patterns, such as trust, peer learning, or symbolic recognition. In cases where quantitative data showed no significant effect, qualitative findings often exposed invisible barriers like elite capture, local skepticism, or governance failures that suppressed impact. This kind of integrative analysis enhanced both the validity and interpretive richness of the study, providing a layered understanding of how research engagement leads to, or fails to produce, systemic transformation.

Nevertheless, the integration of these methods was not without its challenges. One of the primary difficulties was the temporal and epistemological disconnect between quantitative and qualitative data. Regression analysis required standardized variables and clear causal relationships, while sensemaking relied on narrative depth and contextual specificity. Aligning these two modes of evidence required careful iterative work. The study revisited codes, refined models, and negotiated tensions between generalization and nuance. Another challenge involved data comparability across countries. While regression relied on consistent metrics, qualitative interviews varied in length, tone, and cultural framing, making synthesis across settings complex. Furthermore, software constraints occasionally made cross-case NVivo analysis cumbersome, particularly when working with multilingual datasets and team-based coding protocols.

Despite these challenges, the participatory design of the research process helped mitigate several risks. Adhering to the principle of "Nothing About Them Without Them," the study included participants in co-analysis sessions, transcript reviews, and validation workshops. These steps ensured that stakeholder voices remained central in the interpretation of both quantitative results and qualitative insights. This reflexive engagement also helped ensure the cultural and ethical appropriateness of findings, while enhancing their policy relevance and stakeholder legitimacy.

Binary regression model

The detailed regression modeling is offered not as a demonstration of technical rigor alone, but to show how traditional

statistical tools can be repositioned within a participatory and systems-oriented research logic. This reframing is central to methodological shifts needed in dynamic evaluation settings. To provide a rigorous, transparent, and actionable understanding of the conditions under which development research achieves systemic impact, the study used five steps of model specification (Equation 1), probability estimation (Equation 2), interpretation (Equation 3), marginal insights (Equation 4), and marginal effects and policy insights. Beyond confirming that participation matters, the model demonstrated that structured stakeholder engagement, adaptive learning, and gender-responsiveness are not simply ideals; they are also empirical drivers of success.

Given the binary nature of the outcome, where $Y = 1$ indicates a successful research-to-impact transition and $Y = 0$ indicates failure, the logistic regression model is structured to estimate the log odds of success as a function of key project design and stakeholder engagement variables. The general form of the model is expressed as:

$$Logit(P(Y = 1)) = log\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right) = \beta 0 + \\ + \sum_{i=1}^{n} \beta i X i + \epsilon \tag{1}$$

In this model, $P(Y=1)$ denotes the probability of achieving a successful development outcome, where Y represents the dependent variable, indicating whether the desired outcome was realized. The model incorporates:

- β_0: the intercept, representing the baseline log odds of success when all independent variables are zero.
- $\beta_1, \beta_2, \ldots, \beta_n$: coefficients for independent variables $X_1, X_2, \ldots, X_n$, which correspond to various factors associated with the 10 variables— from stakeholder involvement in project to co-creation of research questions, research relevance in design, valuing local knowledge, adaptive learning, resource allocation, collaborative networks, gender-sensitive indicators, stakeholder participation in M&E, and feedback mechanisms.
- ϵ: the error term, accounting for unexplained variability in the model.

The probability function derived from Equation 1 is used to calculate individual probabilities of success as:

$$P(Y = 1|X) = \frac{e^{\beta_0 + X_i \sum_{i=1}^{k} \beta i X i}}{1 + e^{\beta_0 + X_i \sum_{i=1}^{k} \beta i X i}} \tag{2}$$

Equation 2 allows the study to model how different combinations of engagement and design features influence the predicted likelihood of success. For the estimation and interpretation, the coefficients β_i were estimated using maximum likelihood estimation (MLE), which identifies the values of β that maximize the probability of observing the actual outcomes in the data. For interpretation, odds ratios are derived from the estimated coefficients:

$$Odds\ Ratoii = e^{\beta_i} \tag{3}$$

Equation 3 provides an intuitive measure: an odds ratio >1 indicates a positive association between X_i and project success, while a value below 1 suggests a negative relationship.

To assess the reliability of the binary logistic regression results, two key diagnostic tests were employed: the Hosmer–Lemeshow test and McFadden's pseudo-R^2. The Hosmer–Lemeshow goodness-of-fit test evaluates whether the model's predicted probabilities align with actual outcomes. A non-significant result suggests good model fit. In this study, the test produced a p-value of 0.43, indicating no significant deviation between predicted and observed values, thus confirming a satisfactory fit.

The McFadden's pseudo-R^2 was used to assess the model's explanatory power. With a value of 0.23, the model falls within the range typically deemed acceptable for social science models dealing with complex, behaviorally driven outcomes (0.2–0.4). This suggests that the selected predictors meaningfully account for variation in whether projects achieved sustained, inclusive impact. These diagnostics affirm that the regression estimates derived from Equations 1–3 are both statistically valid and substantively relevant.

While odds ratios provide useful information on relative likelihoods, they are often less intuitive for policy application. Hence, the study computed Average Marginal Effects (AMEs) to determine how much the probability of success changes with a one-unit increase in each X_i, holding other variables constant:

$$AMEi = \frac{1}{N} = \sum_{j=0}^{N} \frac{\partial P(Y_j = 1)}{\partial X_{ij}} \tag{4}$$

This marginal effect function (Equation 4) was particularly valuable in translating technical results into actionable recommendations. For example, stakeholder participation in co-design activities increased the probability of success by 15%−18% across multiple countries. In practice, this means that enhancing co-design practices can significantly improve the developmental outcomes of research investments.

Results

As presented in Tables 1, 2, the quantitative method shows a converging empirical narrative: participatory and adaptive features embedded in project design are not peripheral, but they are essential predictors of whether development interventions produce inclusive and sustained impact. As shown in Table 3, the consistency of these patterns across diverse countries and sectors enhances the generalizability of the study's conclusions and substantiates the claim that systemic change is not merely about delivering outputs but about shaping institutional relationships, trust, learning processes, and collective legitimacy. This reinforces a central argument of the study: how research is done—who frames it, who participates in it, and how it adapts—is as consequential as what it seeks to achieve.

The logistic regression findings presented in Table 4 show empirical evidence for the central role of participatory design features in predicting the likelihood of achieving inclusive, sustained, and scalable development outcomes. Across the six countries, key predictors such as stakeholder participation in monitoring and evaluation (X_2), adaptive learning mechanisms

TABLE 1 Sample size, participants, and type of projects covered.

Country	Sample size	Type of participants	Year(s) of data collection	Type of project
Albania	275	Youth trainees, ICT instructors, and vocational program administrators	2021–2023	ICT employment and inclusive skills training
Bangladesh	350	Female farmers, NGO facilitators, and local agricultural officers	2022–2023	Access to land and gender-responsive agriculture
Ethiopia	325	Women farmers, agricultural entrepreneurs, and local development planners	2019–2023	Livelihood innovation and women's participation in agriculture and services
Kosovo	275	Local entrepreneurs, municipal officials, and CSO representatives	2021–2023	Inclusion in local economic development, such as skills and education
Philippines	300	Women farmers, disaster recovery personnel, and education officials	2020–2022	Post-disaster recovery and adaptive skills in relevant sector initiatives
Tanzania	275	Youth farmers, women in horticulture, and agricultural extension agents	2022–2023	Agricultural R&D and inclusion in market access

TABLE 2 Regression model variables.

Variable code	Name	Type	Concept captured
$X1$	Stakeholder involvement in framing	Binary	Early co-creation and ownership
X_2	Participation in M&E	Binary	Reflexivity, mutual accountability
$X3$	Integration of local knowledge	Binary	Contextual alignment and epistemic inclusion
$X4$	Gender-sensitive indicators	Binary	Gender equity and inclusion
$X5$	Adaptive learning mechanisms	Binary/ordinal	Flexibility, responsiveness
$X6$	Resource adequacy	Ordinal	Implementation capacity and responsiveness
$X7$	Collaborative networks	Binary/ordinal	Partnerships, institutional bridges
$X8$	Policy/institutional alignment	Binary	System integration and policy uptake
$X9$	Feedback accessibility	Binary	Knowledge, usability, and transparency
$X10$	Stakeholder trust	Binary/ordinal	Legitimacy and relational quality

TABLE 3 Model diagnostics summary.

Diagnostic test	Purpose	Observed result
Hosmer–Lemeshow Test	Evaluates the goodness-of-fit by comparing predicted probabilities with observed outcomes across deciles of risk. A non-significant p-value suggests that the model fits the data well	$p = 0.43$, indicating no significant discrepancy—thus, the model is a good fit
McFadden's pseudo-R^2	Measures the explanatory power of the logistic regression model. Values between 0.2 and 0.4 are typically considered acceptable in social sciences	$R^2 = 0.23$, suggesting moderate explanatory power and adequate model fit

emerged as a consistently significant driver of success. Projects with iterative feedback loops, reflexive adjustment mechanisms, and stakeholder-responsive learning were significantly more likely to sustain impact. These findings align with growing evidence in development research that highlights the importance of adaptive governance and learning-based implementation in achieving systemic change in complex contexts. Rather than following rigid plans, successful initiatives rely on iterative problem-solving and embedded learning structures. This supports approaches such as Problem-Driven Iterative Adaptation (PDIA), which emphasize local ownership, feedback loops, and institutional flexibility (Andrews et al., 2017; Ramalingam et al., 2014; Valters et al., 2016). The observed significance of co-design, monitoring engagement, and adaptive learning in this study reflects these principles and reinforces the need for research to act as a platform for continuous system learning.

Figure 1 shows the odds ratios for five key predictors of inclusive and sustained development impact across six countries. These predictors—stakeholder framing ($X1$), M&E participation (X_2), adaptive learning ($X5$), policy alignment ($X8$), and stakeholder trust ($X10$)—consistently show odds ratios above 1, with most ranging between 1.5 and 2.1, confirming their strong positive association with project success. This reinforces the conclusion that participatory and adaptive elements are not

($X5$), and policy alignment ($X8$) consistently demonstrated strong and statistically significant associations with project success. These variables often reached the 1% level of significance, with odds ratios ranging from 1.61 to 2.10, suggesting that these features do not just correlate with impact; they function as enabling conditions for systemic change.

As a case in point, policy alignment ($X8$), significant across all six models, highlights the necessity of embedding project goals within existing institutional frameworks to secure legitimacy, coherence, and continuity. Similarly, adaptive learning ($X5$)

TABLE 4 Logistic regression results by country.

Predictor	Albania (OR, CI, Sig)	Bangladesh (OR, CI, Sig)	Ethiopia (OR, CI, Sig)	Kosovo (OR, CI, Sig)	Philippines (OR, CI, Sig)	Tanzania (OR, CI, Sig)
X_1: stakeholder framing	1.55 [1.01–2.37]*	1.44 [0.97–2.13]	1.65 [1.12–2.43]*	1.62 [1.05–2.51]*	1.50 [1.00–2.24]*	1.48 [1.01–2.19]*
X_2: M&E participation	1.69 [1.11–2.58]*	1.61 [1.08–2.42]*	1.82 [1.21–2.74]**	1.77 [1.15–2.72]*	1.71 [1.15–2.55]*	1.56 [1.06–2.30]*
X_3: local knowledge	1.49 [1.00–2.21]*	1.52 [1.01–2.31]*	1.47 [0.98–2.21]	1.41 [0.94–2.10]	1.45 [0.98–2.17]	1.42 [0.95–2.11]
X_4: gender indicators	1.40 [0.91–2.15]	1.38 [0.93–2.07]	1.34 [0.89–2.03]	1.36 [0.88–2.10]	1.32 [0.87–2.01]	1.30 [0.88–1.91]
X_5: adaptive learning	1.83 [1.22–2.75]**	1.92 [1.30–2.85]**	1.79 [1.15–2.78]*	1.85 [1.21–2.81]**	1.80 [1.23–2.65]**	1.78 [1.19–2.66]*
X_6: resource adequacy	1.08 [0.76–1.53]	1.06 [0.74–1.53]	1.10 [0.78–1.56]	1.14 [0.79–1.66]	1.18 [0.81–1.71]	1.12 [0.79–1.59]
X_7: networks	1.29 [0.91–1.84]	1.43 [1.01–2.03]*	1.25 [0.89–1.76]	1.32 [0.90–1.94]	1.30 [0.89–1.90]	1.36 [1.01–1.95]*
X_8: policy alignment	1.97 [1.29–3.01]**	1.87 [1.21–2.91]**	2.10 [1.35–3.29]**	2.02 [1.31–3.11]**	1.95 [1.24–3.07]**	1.90 [1.26–2.87]**
X_9: feedback access	1.32 [0.92–1.89]	1.34 [0.89–2.02]	1.58 [1.02–2.43]*	1.36 [0.95–1.97]	1.40 [0.94–2.09]	1.29 [0.91–1.84]
X_{10}: stakeholder trust	1.85 [1.21–2.82]**	1.76 [1.18–2.64]*	1.90 [1.22–2.96]**	1.89 [1.23–2.91]**	1.83 [1.15–2.89]**	1.81 [1.20–2.71]*

Significance levels for odds ratios are indicated as follows: *$p < 0.05$ (statistically significant at the 5% level). **$p < 0.01$ (statistically significant at the 1% level).
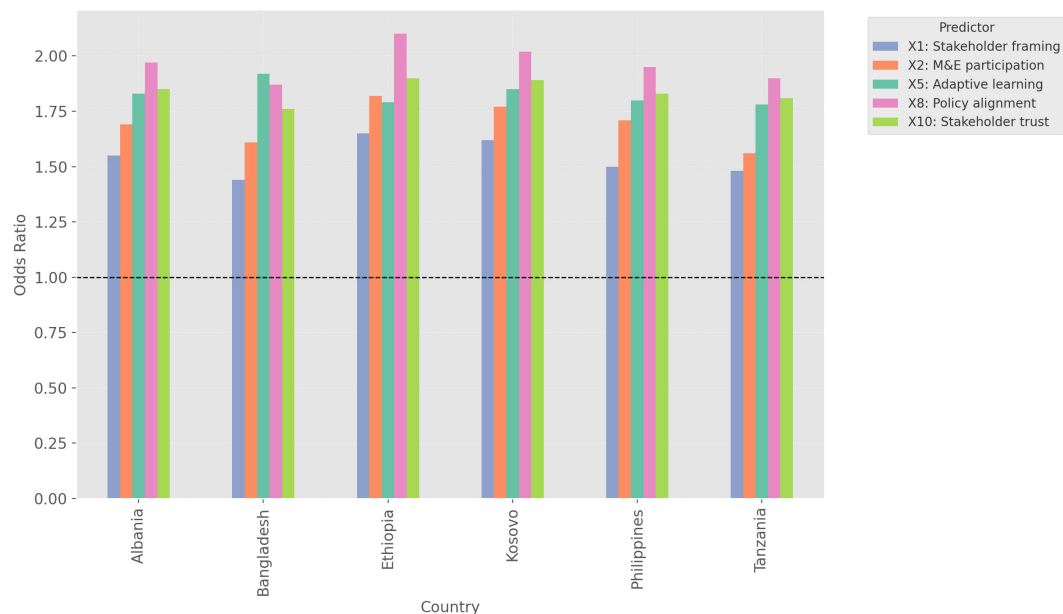OR, odds ratio; CI, confidence interval; Sig, significance level.

FIGURE 1
Odds ratios of key predictors across countries.

auxiliary features but are core enabling conditions for systemic development impact.

Although not all variables were significant in every country, several others, particularly co-design of research framing (X_1) and stakeholder trust (X_{10}), achieved significance in at least four country models, reinforcing the centrality of epistemic inclusion and relational legitimacy. In contrast, resource adequacy (X_6) and gender-sensitive indicators (X_4) were less consistently significant. While their theoretical relevance remains intact, the statistical inconsistency suggests that their influence may be contingent upon institutional culture, project scale, or the depth of participatory practice, rather than their presence alone.

The robustness of the models is supported by the diagnostics presented in Table 5. All models demonstrated Hosmer–Lemeshow p-values above 0.38, indicating that the predicted probabilities were well-calibrated to observed outcomes. The McFadden's pseudo-R^2 values, ranging from 0.21 to 0.25, are considered satisfactory within applied policy and social science research, especially when modeling complex behavioral, institutional, or relational processes. Importantly, Variance Inflation Factors (VIFs) remained below the accepted threshold of 2.5 across all models, confirming the absence of multicollinearity and enhancing the interpretability of individual predictors. Sample sizes per country (between 275 and 350) further ensured adequate statistical power.

TABLE 5 Model diagnostics summary.

Model	McFadden's R^2	Max VIF	Hosmer-Lemeshow p	N
Albania	0.23	2.2	0.42	275
Bangladesh	0.21	2.2	0.39	350
Ethiopia	0.23	2.4	0.43	300
Kosovo	0.24	2.1	0.38	275
Philippines	0.25	2.3	0.41	300
Tanzania	0.22	2.3	0.4	275

TABLE 6 Descriptive statistics for independent variables.

Variable	Mean	SD	Min	Max
$X1$: stakeholder framing	0.64	0.48	0	1
X_2: M&E participation	0.58	0.49	0	1
$X3$: local knowledge	0.61	0.49	0	1
$X4$: gender indicators	0.53	0.5	0	1
$X5$: adaptive learning	0.66	0.47	0	1
$X6$: resource adequacy	3.45	0.91	1	5
$X7$: networks	0.6	0.49	0	1
$X8$: policy alignment	0.59	0.49	0	1
$X9$: feedback access	0.56	0.5	0	1
$X10$: stakeholder trust	0.62	0.49	0	1

Complementing these findings, Table 6 presents descriptive statistics that contextualize the coding and variable distribution across the study sample. Binary variables such as $X1$, $X2$, and $X3$ had mean values between 0.53 and 0.66, indicating that a majority of respondents across sites reported exposure to these participatory or inclusion-oriented interventions. This distribution also ensured adequate variation for regression modeling. The lone ordinal variable, resource adequacy ($X6$), had a mean score of 3.45 on a 5-point scale, with a standard deviation of 0.91, reflecting moderate variation in perceived implementation quality. These descriptive statistics affirm that the predictors are not overly skewed or clustered, and they align well with assumptions of logistic regression modeling.

Sensemaking and qualitative analysis across six countries

The qualitative component of this study was grounded, as mentioned in the previous section, in a sensemaking approach, following the work of Weick (1995) and Maitlis and Christianson (2014). This approach focuses not only on what changed in development processes but on how those changes were interpreted, contested, or embraced by individuals and groups. Rather than treating participants as passive informants, sensemaking positioned them as agents navigating uncertainty, institutional complexity, and shifting power dynamics. The goal was to understand how stakeholders constructed meaning around participation, knowledge use, and system transformation.

The qualitative findings show a fundamental insight: how stakeholders make sense of participation significantly shapes its outcomes. Whether interpreted as empowerment, legitimacy, strategy, or tokenism, participation takes on contextual meaning that either facilitates or constrains systemic change. The study's sensemaking approach—by foregrounding emotional, symbolic, and institutional interpretations—offers critical explanatory power that complements statistical generalizability. This reinforces a central argument in adaptive development literature: that development is not just about changing structures, but about shifting the meanings and relationships that sustain them (Andrews et al., 2017; Ramalingam et al., 2014).

A total of 1,800 semi-structured interviews were conducted across six countries—Albania, Bangladesh, Ethiopia, Kosovo, the Philippines, and Tanzania. These interviews involved local beneficiaries, civil society representatives, implementation staff, and government actors, selected through purposive sampling to capture diverse perspectives. Interviews were transcribed and analyzed using NVivo 12, following a dual-phase coding strategy.

First, deductive coding was applied using the study's conceptual framework, focusing on three interdependent dimensions of systemic change: behavioral shifts, structural transformations, and normative reorientations (Ramalingam et al., 2014). This ensured theoretical consistency across cases. Second, inductive coding was performed using in-vivo and emergent thematic approaches, allowing locally constructed meanings and unexpected insights to surface organically. This included participant metaphors, affective framings, and culturally embedded expressions.

Cross-case patterns were examined using NVivo's matrix coding tools, enabling both comparative thematic synthesis and attention to context-specific deviations. Inter-coder agreement was tested through iterative refinement of the codebook to strengthen reliability and interpretive alignment (Valters et al., 2016).

Analysis revealed that the interpretive dynamics of participation varied significantly across settings but also coalesced around a set of recurring themes. In Albania, youth engagement in ICT-based platforms was described not merely as technical capacity-building but as a route to visibility and institutional legitimacy, what one participant called "being seen by the system." This framing resonated with Weick's (1995) conception of sensemaking as identity-relevant enactment. In Bangladesh, women engaged in gender-responsive agriculture acknowledged new agency, but often spoke of navigating institutional silences—spaces where formal recognition coexisted with persistent exclusions from authority structures.

In Ethiopia, cooperative and governance participation among women was framed as both symbolic empowerment and practical inclusion, aligning with Maitlis and Christianson's (2014) idea of "emotion-infused meaning-making." In Kosovo, economic development projects were appreciated, but stakeholders emphasized that civic recognition and interethnic trust carried more transformative value than material outputs. In the Philippines, participatory research tools were often interpreted as political instruments, useful for enhancing legitimacy in

interactions with government institutions, what Ramalingam et al. (2014) refer to as "adaptive brokerage." In Tanzania, youth and women narrated their involvement in horticultural interventions as a means of repositioning within labor and social hierarchies, shifting institutional roles, and challenging generational norms.

To do a deeper analysis of the findings, as shown in Figure 2, the theme of symbolic empowerment referred to how individuals perceived their participation as conferring status, legitimacy, or recognition, rather than simply transferring resources or knowledge. Symbolic empowerment played a critical role in normative shifts. It helped shift the social scripts around who was seen as capable, legitimate, or visible in institutional settings. This reinforced the idea that development outcomes are mediated by recognition and identity, not only resource flows.

Symbolic empowerment was most pronounced in Albania, where young participants framed their involvement in ICT-based platforms as a means to be "seen" by institutions and the labor market. This reflects a context where institutional visibility is scarce, and formal inclusion mechanisms are limited. Participation was thus performative, and it reshaped participants' identities as legitimate social actors. In Ethiopia, especially among women in agricultural cooperatives and local governance, participation was experienced as public affirmation. Many described their increased involvement not just in terms of income or output, but as a way of transforming how others—and they themselves—understood their role in society.

Trust-building was another theme that captured how participatory processes create or reinforce relational legitimacy, especially in fractured or hierarchical institutional landscapes. Trust functions as an enabling condition for adaptive learning, institutional reform, and local ownership. It explains why M&E participation (X_2) and stakeholder trust ($X10$) emerged as statistically significant predictors in the regression models. Trust is not an outcome, but rather it is a mechanism. It featured prominently in Kosovo. In a context of ethnic fragmentation, trust was not assumed; it had to be built. Participants noted that local governance forums fostered mutual recognition across ethnic lines, giving participation a symbolic and reconciliatory dimension. It was also clear in Tanzania, in both youth and women's narratives, that trust was a prerequisite for risk-taking and engagement. Many described how repeated interactions with facilitators and peer groups helped them redefine institutional relationships and gain confidence in voicing needs.

Political capital use reflected the instrumental use of participatory tools by local actors to gain legitimacy, resources, or leverage within political hierarchies. This demonstrated how participatory mechanisms are not neutral—they are often repurposed by actors to navigate or challenge existing power structures. This aligns with adaptive development theories, which argue for recognizing the political economy of learning and feedback loops (Ramalingam et al., 2014). It was dominant in the Philippines. Participants, especially local officials and community leaders, referred to participatory research and data collection as strategic assets. Tools like community scorecards and monitoring reports were used to claim attention from provincial or national institutions. Participation was valued not just for internal empowerment but for its external signaling value.

Institutional navigation also described how participants understood their role as navigating, adapting to, or resisting institutional constraints, particularly in opaque or exclusionary systems. It highlighted that even well-designed participatory platforms can be reinterpreted based on local political and cultural norms. This reveals a tension between formal inclusivity and real-world power asymmetries, a key explanation for null effects in variables like resource adequacy ($X6$).
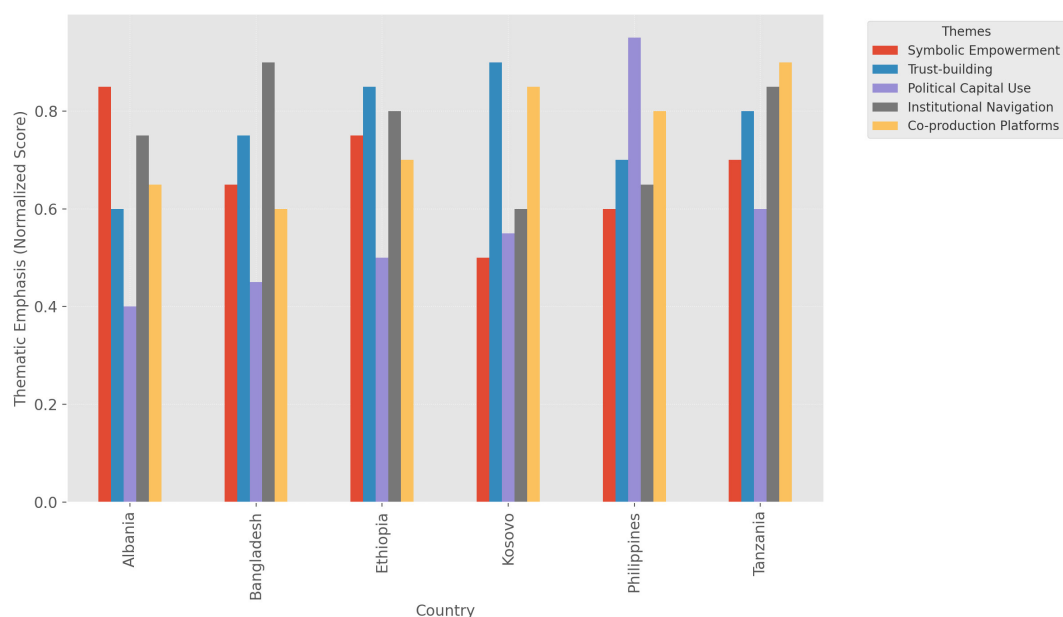

FIGURE 2
Cross-country emphasis on key qualitative themes.

It was especially important in Bangladesh, where women in agriculture projects framed their engagement not in terms of empowerment alone, but as calculated negotiation within patriarchal and bureaucratic spaces. They spoke of needing to "work around the rules" or engage selectively with officials. In Tanzania as well, youth participants explained how formal structures, such as cooperatives or producer groups, required constant adaptation to existing gatekeepers, local elites, or rigid administrative processes.

Lastly, the theme of co-production platforms captures instances where participants perceived themselves as active shapers of project design, decisions, or evaluation, moving beyond consultation into meaningful co-creation. In practice, this referred to co-production that reflected a high-trust, low-hierarchy engagement model. Its presence supported the quantitative finding that adaptive learning ($X5$) and stakeholder framing ($X1$) were significant predictors of success. This is the epistemic dimension of participation, where whose knowledge counts becomes a central axis of impact.

Co-production of platforms was most evident in Kosovo, in which participants emphasized that forums were not just ceremonial; they allowed for deliberative dialogue, joint priority-setting, and consensus-building. This created a sense of procedural justice and ownership. In Tanzania, co-production was visible in horticultural interventions where farmer-led innovation, youth-driven planning, and gender-sensitive adaptation were not only allowed but institutionalized.

Comparative discussion: pathways of systemic change

This article is best understood as a methodological contribution to the ongoing paradigm shift from linear, technocratic models of evaluation toward dynamic, participatory, and systems-based approaches. The six cases serve not as final proof points but as illustrations of how such a shift reconfigures research practice, insight generation, and institutional design.

Taking a step back from the individual findings, the most vital contribution of this article lies in its service to the broader methodological and conceptual reorientation currently underway in the field of development evaluation. Rather than offering six case-specific results, this study seeks to demonstrate what it means to operationalize complexity and systems thinking through participatory, context-responsive, and co-evolutionary research practice. The six cases serve not as bounded "country comparisons" but as living laboratories of this paradigm shift, each showing what such a transition demands in terms of design adaptations, epistemological commitments, and institutional recalibration.

The study provides concrete illustrations of how shifting from linear, input–output models to dynamic systems approaches transforms not only the framing of research questions but also the role of stakeholders, the nature of evidence, and the meaning of impact. Through its integration of PAR, sensemaking, and selective use of regression analysis, it advances methodological guidance on how to manage the tensions between rigor and relevance, standardization and flexibility, and institutional inertia and learning. The challenges and insights experienced, ranging from translation across epistemologies to issues of power asymmetry and symbolic legitimacy, are offered here as both challenges and opportunities for practitioners, funders, and policymakers attempting similar reconfigurations.

This study, thus, advocates for research and evaluation practices that are as adaptive and relational as the systems they seek to understand and influence. In doing so, it hopes to support those working at the frontlines of institutional innovation, especially where traditional regimes remain dominant but increasingly inadequate for achieving sustainable and inclusive development. The findings of this research point out systemic change through three lenses: behavioral, structural, and normative shifts. These dimensions represent distinct but interdependent levels through which research-led interventions influence inclusive and sustainable development outcomes. By triangulating regression findings, sensemaking data, and site-specific qualitative narratives, this study advances the literature by addressing key gaps: (i) a lack of granular comparative analysis of institutional change, (ii) insufficient empirical grounding of behavioral adaptations in participatory models, and (iii) limited theorization of normative transformation across diverse cultural contexts.

Behavioral shifts: co-design and adaptive learning

Behavioral shifts represent the most proximate and observable dimension of systemic change in development processes. These are changes in how individuals and groups engage, make decisions, and adapt their practices in response to new information, relationships, or tools. This study finds that such shifts were most clearly catalyzed through the mechanisms of co-design and adaptive learning approaches that allow stakeholders to move from passive participation to active ownership.

Across the six countries, Ethiopia and the Philippines are examples of the most substantive behavioral transformation. In Ethiopia, agricultural extension workers and women's cooperatives engaged in structured reflection sessions and peer-to-peer dialogues, facilitated by local researchers and NGOs. These sessions were not ancillary but built into the project's core delivery structure. Participants described refining strategies for livestock diversification and seasonal irrigation planning based on experiential feedback. These behaviors reflect the principles of double-loop learning, where actors not only adjust actions but reconsider underlying assumptions—a hallmark of adaptive systems (Argyris, 1976; Williams and Brown, 2018).

Similarly, in the Philippines, community members in disaster-prone areas, particularly in regions affected by typhoons and floods, utilized research outputs (e.g., flood risk maps, needs assessments) not just for awareness but as tools to negotiate with local authorities. Local leaders framed evidence as a means to justify their demands for budget reallocation and prioritized infrastructure. This practice exemplifies what the literature terms strategic appropriation of data, where behavioral agency is exercised in politically situated ways (Marcelo et al., 2016).

In contrast, the experience in Bangladesh and Tanzania reveals the limits of behavioral change when adaptive structures are underdeveloped. In Bangladesh, despite extensive NGO presence

and training programs, the absence of localized control over digital monitoring tools restricted community feedback. Participants often lacked the capacity—or perceived legitimacy—to propose changes midstream. This reflects critiques from Angeli and Montefusco (2020) that adaptive systems can fail when digital inclusion and decision-making are decoupled. Similarly, in Tanzania, although horticulture-based livelihood initiatives were implemented widely, extension officers described a top-down delivery style with limited space for participant-led iteration. Bureaucratic rigidity and fear of reprimand for deviation from protocol were cited as barriers, echoing Howard et al.'s (2021) observations on incentive misalignment within hierarchical development systems.

The regression analysis provides further empirical validation of these patterns. As shown in Table 4, the predictors *X5* (adaptive learning) and *X1* (co-design of framing) were statistically significant in at least four of the six countries, with odds ratios generally ranging from 1.50 to 1.92. This suggests a strong and consistent association between participatory behavioral practices and the likelihood of successful systemic outcomes. In countries like Ethiopia, Albania, and the Philippines, where participatory framing was operationalized early in the research cycle, the odds of a project achieving measurable institutional or normative impact increased markedly.

Importantly, these findings also advance the scholarly conversation by providing rare cross-country evidence for claims that are often asserted but seldom quantified. Much of the literature on co-design and learning (e.g., Chambers, 1997; Westley et al., 2017) emphasizes principles but lacks comparative empirical backing. This study addresses that gap by linking behavioral mechanisms directly to observed outcomes across varied governance and cultural contexts.

Moreover, the qualitative sensemaking process informed not just that behavioral change occurred, but how it was experienced and interpreted by participants. In Albania, young ICT trainees described digital platforms as "an escape from clientelism," signaling not only a new skillset but a shift in political consciousness. In Kosovo, community members cited improved confidence in engaging with municipal authorities after participating in framing workshops. These accounts reveal that co-design is not only a design method—it is a relational practice that builds cognitive, emotional, and political capacity.

Structural shifts: institutional alignment and collaborative networks

Structural shifts involve long-term transformations in how institutions function, how policies interact, and how organizations collaborate. They represent the "rules of the game" (North, 1990) that either constrain or enable systemic change. In the context of this study, structural change was operationalized through policy alignment (*X8*) and networked collaboration (*X7*), two variables that consistently influenced whether research processes translated into embedded, durable outcomes.

Among the six countries, Bangladesh and Albania demonstrated the strongest structural change outcomes.

In Bangladesh, the integration of project monitoring and evaluation (M&E) tools within district-level (*upazila*) planning cycles enhanced local ownership and responsiveness. Local administrators reported that evidence generated by the research teams helped structure resource allocation for land leasing governance and other services, contributing to long-term shifts in budgetary accountability. This aligns with the work of Andrews et al. (2017), who stress the importance of building "capability traps" into bureaucracies through iterative feedback systems.

In Albania (*rreth*), cross-sectoral ICT employment initiatives provided a model for institutional collaboration. The involvement of ministries, municipal offices, civil society organizations (CSOs), and youth groups fostered horizontal coordination that persisted beyond the lifespan of the intervention. Interviewees noted that government uptake of participatory digital tools (e.g., employment portals, certification tracking) changed not only administrative routines but also inter-ministerial communication practices. These findings align with Borrás and Edquist (2013), who argue that institutional coherence and horizontal policy coordination are key enablers of system innovation in development.

Kosovo also showed emerging forms of structural transformation, particularly in the economic development and rural inclusion sectors. Here, CSO–government collaboration focused on integrating local business priorities into municipal planning. While not yet institutionalized into national policy, this cooperation marks an important transition from donor-driven programming to embedded, local governance-led planning processes. These types of "meso-level coalitions" are frequently cited as precursors to broader structural change (Geels and Schot, 2007).

By contrast, the Philippines showed the challenges of institutional lock-in. Despite strong technical outputs from disaster resilience research, fragmented mandates between national agencies and local government units (LGUs), combined with frequent staff turnover, have weakened coordination. Several respondents described stalled initiatives due to unclear ownership or conflicting regulations. This reflects Arthur's (1989) theory of increasing returns, where path-dependent systems reinforce suboptimal equilibria, and North's (1990) claim that institutions tend to persist even when ineffective, due to sunk costs and political resistance to reform.

The regression results substantiate these qualitative findings. Policy alignment (*X8*) was a statistically significant predictor of impact in five of the six countries, with odds ratios between 1.87 and 2.10 (see Table 4). This variable was particularly influential in Albania, Ethiopia, and Bangladesh—countries where institutional relationships between research implementers and public administrators were strong from the outset. Moreover, collaborative networks (*X7*) showed positive associations in countries like Kosovo and Bangladesh, where inter-organizational relationships were either institutionalized or leveraged through formal coordination platforms.

Importantly, diagnostics such as Variance Inflation Factors (VIFs) remained below 2.5 for all variables, indicating that structural indicators operated independently of behavioral ones. This is conceptually important: structural shifts do not simply emerge from changed behavior—they must be deliberately

cultivated, politically negotiated, and technically supported. The findings thus support the position of Westley et al. (2011), who argue that multi-level transitions require both enabling structures and active institutional entrepreneurs to overcome lock-in and inertia.

Additionally, this study contributes to the literature on absorptive capacity (Cohen and Levinthal, 1990) by showing that systems with pre-existing channels for feedback, such as Bangladesh's local government planning committees or Albania's ICT hubs, were more capable of translating research into embedded institutional processes. Conversely, systems with fragmented authority or poor retention (e.g., the Philippines) failed to absorb even high-quality research due to organizational instability.

Normative shifts: gender equity and social recognition

Normative change refers to the transformation of socially shared values, perceptions of legitimacy, and expectations about roles and behavior. In development practice, these shifts are essential but complex, as they touch on how people interpret identity, power, and recognition. As theorized by North (1990) and refined through sensemaking frameworks (Weick, 1995; Maitlis and Christianson, 2014), norms are not merely regulatory. They are deeply emotional and relational, maintained through symbolic practices and community consensus. This study adopts this conceptual lens to examine how development interventions influenced gender roles, social visibility, and symbolic legitimacy across six diverse contexts.

Findings from this study reveal that normative change was the most uneven of the three domains assessed (in addition to behavior and structures). It is more difficult to quantify and more sensitive to context than behavioral or structural shifts. Nevertheless, powerful shifts did emerge in Ethiopia and Bangladesh, where gender-responsive interventions were explicitly designed to reshape social hierarchies. In Ethiopia, for example, women participating in agricultural cooperatives reported that their roles evolved from supportive laborers to knowledge holders, individuals whose decisions carried weight within their communities (Uraguchi, 2010). This transition was not only about increased participation; it was also about symbolic recognition, or what Johnson et al. (2016) describe as emotional legitimacy, which is the moment when a person's contributions are publicly acknowledged as valid and valued.

Similarly, in Bangladesh, projects targeting land governance and farming systems made deliberate efforts to include women in training, monitoring, and planning processes. As women gained skills and voice in these forums, they were increasingly accepted as decision-makers. This reconfiguration of gender roles was particularly meaningful in rural areas, where patriarchal structures often constrained female agency. The data show that this visibility of women being seen and heard in public forums was itself a form of transformation, confirming the work of Tavenner and Crane (2019) as well as Carter et al. (2014) on the role of public acknowledgment (e.g., bilateral donors) in norm change with what evidence and outcomes.

By contrast, in Tanzania and Kosovo, attempts to shift gender or inclusion norms encountered resistance, and normative gains remained marginal. In these settings, local stakeholders framed development interventions, especially those focused on women's leadership or minority inclusion, as externally imposed. In Kosovo, the entrenchment of political polarization and ethnic tensions created barriers to trust, while in Tanzania, traditional leaders questioned the legitimacy of changing gender roles, citing cultural and religious norms. These findings align with Duran (2019) and Lončar (2016), who argue that normative interventions that lack cultural resonance or local ownership can provoke backlash, rather than transformation.

The results of the logistic regression offer partial empirical support for these patterns. Gender-sensitive indicators ($X4$) showed statistically significant associations with positive project outcomes in Ethiopia and Bangladesh ($\beta = 0.47$ and $\beta = 0.46$, respectively), but not in Tanzania or Kosovo. Yet this statistical variability does not diminish the relevance of the shifts observed. As sensemaking interviews confirmed, even in countries where gender-sensitive indicators ($X4$) lacked statistical significance, symbolic actions, such as women publicly moderating meetings or young people questioning hierarchy, represented critical steps toward social realignment.

From a theoretical perspective, these findings make three important contributions. First, they extend sensemaking theory by showing that normative change is not only about framing new meanings, but also about earning recognition within emotionally charged social orders. Second, the study challenges the adequacy of universal indicators for norm change, highlighting that transformation is often context-dependent, non-linear, and mediated through symbolic performances. Third, it reinforces the idea that normative legitimacy emerges relationally, from trust, endorsement, and repeated social interactions, not solely from access or inclusion.

From a practical standpoint, the study emphasizes that symbolic infrastructure is as important as institutional architecture. In other words, projects must not only provide roles for underrepresented actors; they must build spaces where these actors can be seen, heard, and affirmed. This includes designing participatory processes that make emotional and social shifts visible (Cornwall, 2008), such as storytelling platforms, public forums, and peer-to-peer exchanges. Moreover, interventions need to be culturally embedded, drawing on local idioms of legitimacy and aligning with community-held values to avoid perceptions of imposition or political manipulation.

Conclusion: practical implications for research-to-impact pathways

More than a set of findings, this article contributes to the methodological transformation of research-to-impact studies. By centering participatory co-production and systemic sensemaking, it provides a prototype for how developmental evaluation can align with complexity thinking and epistemic justice. The concept of "Nothing About Them Without Them" emphasizes the need for inclusive, participatory research-to-impact pathways. Yet, doubts about the effectiveness of research continue to grow, largely due

to a failure to consistently link research with context-specific, sustainable improvements (Platts-Fowler and Robinson, 2016). This disconnect is further worsened by austerity in development finance, which pressures projects to deliver immediate, measurable results at the expense of long-term systemic change (Major, 2014; Done et al., 2011). Without robust evidence of impact, trust in research is weakened, particularly among disadvantaged groups who feel excluded from design and decision-making processes.

This study provides empirical and conceptual contributions to understanding how participatory research translates into systemic development outcomes. Drawing from over 1,800 stakeholder interviews, triangulated with observational data, document analysis, and logistic regression models, it offers a dynamic, systems-based understanding of how change unfolds across behavioral, structural, and normative domains. In doing so, it moves beyond static, linear notions of research utilization and advances a co-evolutionary model of research-to-impact grounded in participatory development science.

Participation as a systemic driver, not an output

This study contributes to closing a critical gap in the research-to-impact literature by evidencing how embedded participation functions not as a peripheral activity but as a systemic driver of change across behavioral, structural, and normative domains. Regression analysis revealed that co-design ($X1$), monitoring engagement (X_2), and adaptive learning ($X5$) were statistically significant predictors of project success in at least four of the six countries, with odds ratios ranging from 1.41 to 1.87. These results were reinforced through qualitative sensemaking, which showed that participation shaped how stakeholders interpreted their roles, with women in Ethiopia describing themselves as "knowledge brokers" and youth in the Philippines reframing their identity as "planners" rather than "volunteers." These experiences resonate with Weick's (1995) and Maitlis and Christianson's (2014) work on sensemaking, which highlights the importance of participation in generating shared understanding and emotional legitimacy.

Yet, a significant gap remains in how participation is conceptualized and measured in mainstream development practice. Often reduced to workshop attendance or consultation records (Chambers, 1997; Belcher and Hughes, 2021), participation is rarely evaluated for its influence on institutional memory, decision-making, or legitimacy formation. While participatory rural appraisal (PRA) and stakeholder matrices remain common tools, they frequently lack mechanisms to capture how power, trust, or epistemic authority are renegotiated through participatory processes (Gaventa and Cornwall, 2008).

Future research must advance this agenda by developing metrics that assess participatory influence over time, such as the degree to which local actors shape monitoring frameworks, reinterpret evidence, or contest dominant narratives. Longitudinal approaches and mixed methods, particularly those attentive to narrative, emotion, and symbolic acts, are essential to understanding how participation becomes institutionalized.

Practically, this implies a reorientation of funding and project design toward long-term facilitation support, community feedback loops, and safe spaces for dissent and co-learning. Donors and evaluators must shift from counting participants to analyzing whose voice shapes direction, policy, and meaning. This way, it is possible to adopt an approach that treats participation as a living infrastructure for system-wide learning and inclusive governance.

Linear to co-evolutionary pathways

This study offers a key theoretical intervention by challenging the dominant linear models of research impact that assume a unidirectional flow from research to policy to practice. Instead, the empirical findings suggest that research-to-impact processes are best understood as co-evolutionary, marked by feedback loops, contested meanings, and iterative adaptation. In particular, the study demonstrates how impact often arises not from direct policy uptake but through localized reinterpretation, informal experimentation, and relational learning among actors across sectors and levels.

For instance, in Albania and the Philippines, stakeholders used research outputs not simply as tools for implementation, but as instruments to renegotiate roles, coordinate fragmented mandates, or legitimize new governance practices. These behaviors indicate that impact was shaped by contextual interactions, not simply by evidence quality or dissemination efforts. Such patterns are consistent with developmental systems theory, which views complex social change as emergent and relational rather than mechanistic or input-output driven (Capra and Luisi, 2014). The frequent finding that policy alignment ($X8$) was statistically significant, yet functioned differently across countries, underscores that impact is relationally situated and path-dependent (North, 1990).

This co-evolutionary framing is especially relevant in politically fluid or resource-constrained environments, where centralized uptake is limited, and change relies on distributed agency. In Bangladesh, for example, community-level actors integrated research into district (*upazila*) planning cycles despite the absence of formal mandates. In contrast, in Kosovo, political fragmentation undermined institutional memory, demonstrating how structural discontinuities can disrupt impact even when local ownership exists. These cases support calls from complexity scholars for adaptive impact pathways that evolve with shifting actor networks, institutional constraints, and opportunity structures (Westley et al., 2011; Andrews et al., 2017).

However, existing impact assessment frameworks, such as logic models and results chains, still reflect a linear epistemology. They prioritize attribution over contribution, uptake over iteration, and technical evidence over relational learning. As a result, they often fail to detect the more nuanced, indirect, or symbolic pathways through which research shapes discourse, networks, and behavior over time (Belcher and Hughes, 2021). This study helps fill that gap by evidencing how outcomes often emerge through recursive interaction, boundary-spanning roles, and dynamic sensemaking processes.

Future theoretical work should focus on developing models that integrate causal complexity, multi-actor feedback, and adaptive timing. This includes combining quantitative tools (e.g., contribution analysis, Bayesian models) with qualitative methods like process tracing and longitudinal sensemaking. Donors, meanwhile, must support evaluation frameworks that capture indirect, iterative, and symbolic forms of change, not just linear indicators of adoption or output delivery.

Institutionalization and sustainability

A key insight emerging from this study is that behavioral change, while essential, is insufficient on its own to sustain long-term development outcomes. Rather, sustainable impact depends on whether and how research becomes embedded in institutional structures through policy alignment, inter-organizational coordination, and systems of local ownership. Regression analysis confirmed that policy alignment ($X8$) was one of the most robust predictors of impact, with statistically significant odds ratios across five of the six countries studied. These quantitative findings were reinforced through qualitative interviews, where stakeholders in Bangladesh (*upazila*) and Albania (*rreth*) described how research outputs were integrated into district planning processes, sectoral programming, and inter-agency collaboration.

In Bangladesh, for instance, participatory monitoring data were used to update local policies on leasing *khas* land—government-owned land that is not currently under private ownership—refine irrigation schedules and coordinate agricultural calendars to better support the cultivation of medicinal herbs intended for pharmaceutical markets. Similarly, in Albania, ICT-based training platforms were institutionalized via joint initiatives between youth organizations and municipal authorities (*autoritete komunale*). These cases show that impact materializes when research is designed to interface with existing governance routines, rather than operating in isolation. Such insights align with evolutionary institutional theory, which views systems change as a balance between innovation and institutional inertia (Nelson and Winter, 1982).

By contrast, the Philippines and Tanzania illustrate the constraints of institutional path dependency (North, 1990). In both settings, even when behavioral readiness was evident—e.g., youth engagement, frontline innovation—the absence of enabling structures, political continuity, or policy coherence constrained the translation of research into formal change. Fragmented mandates, high staff turnover, and a lack of cross-sector coordination impeded uptake and rendered adaptive learning less effective. This reinforces the view that institutional embedding is not a downstream outcome but a design requirement, especially in volatile or resource-constrained systems (Andrews et al., 2017). Despite these patterns, current impact frameworks often overlook institutional diagnostics in favor of knowledge dissemination indicators. Most evaluation models focus on whether findings were shared, but not on whether the receiving institutions had the absorptive capacity, political incentives, or feedback systems to act on them. This creates a critical gap in research design

and policy engagement strategies, especially in low-governance or high-fragmentation contexts.

To address this, future research should systematically integrate assessments of institutional readiness, absorptive capacity, and policy coherence into research-to-impact planning. This includes identifying entry points in bureaucratic processes, leveraging intermediary organizations, and designing outputs that align with legal mandates, budget cycles, and staff routines. It also calls for deeper inquiry into the informal governance systems, such as local coalitions, networks, and social capital, that mediate institutional uptake but are rarely captured in formal models.

From a policy perspective, donors and implementing agencies should prioritize incentives for cross-sector collaboration, support institutional brokerage roles, and invest in technical assistance that strengthens systems integration. Importantly, sustainability must be defined not by the persistence of activities but by the degree to which research-derived insights become institutional memory and operational norm.

Refining the evidence base

Despite its robust design and comparative approach, this study acknowledges several limitations that point to key areas for future research. First, the issue of temporality remains a persistent challenge. As the study relied on retrospective data collection, including participant recall and ex-post documentation, it was difficult to fully establish causal sequencing between research interventions and observed systemic changes. While triangulation helped mitigate some of this uncertainty, the absence of real-time tracking limited the ability to isolate how change emerged across different stages of the project life cycle. This gap echoes longstanding critiques in impact evaluation literature about the difficulty of capturing dynamic, unfolding processes in cross-sectional designs (Rogers, 2008; Belcher and Hughes, 2021).

Second, although binary logistic regression proved effective in highlighting statistically significant associations, for example, between policy alignment ($X8$) and institutional uptake, its explanatory power was constrained in capturing relational, affective, and processual mechanisms. Variables such as trust-building, legitimacy negotiation, and symbolic recognition, while central to normative change, remained analytically underrepresented in the quantitative models. This is consistent with broader critiques of positivist approaches in complex development contexts, where linear causality often obscures the messy, negotiated realities of social transformation (Patton, 2011; Westhorp, 2014).

Moreover, the qualitative dimension, though methodologically rigorous, encountered practical difficulties related to cross-country coding and standardization. The combined use of deductive (framework-guided) and inductive (emergent) coding in NVivo allowed flexibility, but also introduced inconsistencies in thematic depth across cases. Multilingual contexts added further complexity, where digital access, translation accuracy, and interviewer positionality influenced both data quality and participant voice. These factors are a good reminder of the

need for more reflexive and equity-conscious qualitative designs, particularly in comparative research (Janssen et al., 2022).

Looking forward, there is a clear need to invest in real-time, longitudinal, and adaptive evaluation frameworks that can better capture the evolution of change processes. Tools such as process tracing (Collier, 2008), developmental evaluation (Patton, 2011), and realist evaluation (Pawson and Tilley, 1997) offer promising avenues. These approaches are well-suited to unpack the interaction of context, mechanism, and outcome over time, especially in settings where interventions evolve in response to shifting political and institutional conditions. Moreover, mixed-method designs that intentionally integrate emotional, symbolic, and institutional dimensions can enrich our understanding of how impact is interpreted, internalized, and sustained.

Lastly, future research should consider the politics of research itself—whose voices define impact, whose data is privileged, and how reflexivity is institutionalized in evaluation systems. Addressing these questions will help advance an evidence base that is not only more rigorous but also more socially attuned and methodologically just. This study thus positions itself in service of a broader reconfiguration, where evaluation becomes a reflexive, participatory, and systemic learning process, rather than a retrospective scorecard.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

ZU: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahlin, E. M. (2019). Semi-structured interviews with expert practitioners: their validity and significant contribution to translational research. *Qual. Rep.* 24, 581–593. doi: 10.4135/9781526466037

Andrews, M., Pritchett, L., and Woolcock, M. (2017). *Building State Capability: Evidence, Analysis, Action.* Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198747482.001.0001

Angeli, F., and Montefusco, A. (2020). Sensemaking and learning during the COVID-19 pandemic: a complex adaptive systems perspective on policy decision-making. *World Dev.* 136:105106. doi: 10.1016/j.worlddev.2020.1 05106

Argyris, C. (1976). Single-loop and double-loop models in research on decision making. *Adm. Sci. Q.* 21, 363–375. doi: 10.2307/2391848

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Econ. J.* 99, 116–131.

Bailey, Z. D., Krieger, N., Agénor, M., Graves, J., Linos, N., and Bassett, M. T. (2017). Structural racism and health inequities in the USA: evidence and interventions. *Lancet* 389, 1453–1463. doi: 10.1016/S0140-6736(17)30569-X

Barriball, K. L., and While, A. (1994). Collecting data using a semi-structured interview: a discussion paper. *J. Adv. Nurs.* 19, 328–335. doi: 10.1111/j.1365-2648.1994.tb01088.x

Belcher, B., and Hughes, K. (2021). Understanding and evaluating the impact of integrated problem-oriented research programmes: concepts and considerations. *Res. Eval.* 30, 154–164. doi: 10.1093/reseval/rvaa024

Bhaumik, S. K., Gang, I. N., and Yun, M. S. (2011). "Gender and ethnicity in fragile states: the case of post-conflict Kosovo," in *Fragile States: Causes, Costs, and Responses*, eds. W. Naudé, A. U. Santos-Paulino, and M. McGillivray (Oxford: Oxford University Press), 133–147. doi: 10.1093/acprof:oso/9780199693153.003.0007

Blundo-Canto, G., and Ferré, M. (2022). How can the past inform and benefit research-for-development interventions? The value of reflexive and participatory approaches. *Anthropol. Dév.* 56, 107–133. Available online at: https://journals.openedition.org/anthropodev/1808

Borrás, S., and Edquist, C. (2013). "The governance of change in socio-technical and innovation systems: three pillars for a conceptual framework," in *The Governance of Socio-Technical Systems: Explaining Change*, eds. S. Borrás and J. Edler (Cheltenham: Edward Elgar), 135–152. doi: 10.4337/9781784710194.00011

Bradshaw, J., Tomlinson, M., Pantazis, C., Pemberton, S., Bramley, G., Manley, D., et al. (2017). *Poverty and Social Exclusion in the UK: Volume 2-the Dimensions of Disadvantage*. Bristol: Policy Press.

Cachia, M., and Millward, L. J. (2011). The telephone medium and semi-structured interviews: a complementary fit. *Qual. Res. Organ. Manag.* 6, 265–277. doi: 10.1108/17465641111188420

Cahill, C. (2007). The personal is political: developing new subjectivities through participatory action research. *Gend. Place Cult.* 14, 267–292. doi: 10.1080/09663690701324904

Capra, F., and Luisi, P. L. (2014). *The Systems View of Life: A Unifying Vision*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511895555

Carter, J., Byrne, S., Schrader, K., Kabir, H., Uraguchi, Z. B., Pandit, B., et al. (2014). Learning about women's empowerment in the context of development projects: do the figures tell us enough? *Gend. Dev.* 22, 327–349. doi: 10.1080/13552074.2014.920986

Chambers, R. (1997). *Whose Reality Counts? Putting the First Last*. London: Intermediate Technology Publications. doi: 10.3362/9781780440453

Chouinard, J. A., and Milley, P. (2018). Uncovering the mysteries of inclusion: empirical and methodological possibilities in participatory evaluation in an international context. *Eval. Program Plann.* 68, 90–98. doi: 10.1016/j.evalprogplan.2017.12.001

Clark, L., Higdon, G. L., Thompson, S., Tofaris, E., and Rose, P. (2021). *Celebrating the Impact of the Raising Learning Outcomes in Education Systems Programme*. Brighton: Institute of Development Studies. doi: 10.35648/20.500.12413/11781/ii370

Cohen, W. M., and Levinthal, D. A. (1990). Absorptive capacity: a new perspective on learning and innovation. *Adm. Sci. Q.* 35, 128–152. doi: 10.2307/2393553

Collier, P. (2008). *The Bottom Billion: Why the Poorest Countries are Failing and What can be Done About it*. Oxford: Oxford University Press.

Cooke, B., and Kothari, U. (eds.) (2001). *Participation: The New Tyranny?* London: Zed Books.

Cornwall, A. (2008). Unpacking 'participation': models, meanings and practices. *Community Dev. J.* 43, 269–283. doi: 10.1093/cdj/bsn010

Cousins, J. B., and Earl, L. M. (2004). *Classroom Assessment for Student Learning: Doing It Right – Using It Well*. Alexandria, VA: ASCD.

Creswell, J. W., and Poth, C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. London: Sage Publications.

Done, A., Voss, C., and Rytter, N. G. (2011). Best practice interventions: short-term impact and long-term outcomes. *J. Oper. Manag.* 29, 500–513. doi: 10.1016/j.jom.2010.11.007

Douthwaite, B., Proietti, C., Polar, V., and Thiele, G. (2023). Outcome trajectory evaluation (OTE): an approach to tackle research-for-development's long-causal-chain problem. *Am. J. Eval.* 44, 263–280.

Duran, E. (2019). *Healing the Soul Wound: Trauma-informed counseling for Indigenous Communities*. New York, NY: Teachers College Press.

Faure, G., Blundo-Canto, G., Devaux-Spatarakis, A., Le Guerroué, J. L., Mathé, S., Temple, L., et al. (2020). A participatory method to assess the contribution of agricultural research to societal changes in developing countries. *Res. Eval.* 29, 158–169. doi: 10.1093/reseval/rvz036

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qual. Inq.* 12, 219–245. doi: 10.1177/1077800405284363

Garbero, A., Sakos, G., and Cerulli, G. (2023). Towards data-driven project design: providing optimal treatment rules for development projects. *Socio-Econ. Plann. Sci.* 89:101618. doi: 10.1016/j.seps.2023.101618

Gaventa, J. (2006). Finding the spaces for change: a power analysis. *IDS Bull.* 37, 23–33. doi: 10.1111/j.1759-5436.2006.tb00320.x

Gaventa, J. (2021). "Linking the prepositions: using power analysis to inform strategies for social action," in *Essays on Evolutions in the Study of Political Power*, ed. G. M. Gallarotti (London: Routledge), 109–130. doi: 10.4324/9781003200673-8

Gaventa, J., and Cornwall, A. (2008). "Power and knowledge," in *The Sage Handbook of Action Research: Participative Inquiry and Practice, 2nd Edn*, eds. P. Reason, and Bradbury (Thousand Oaks, CA: Sage), 172–189. doi: 10.4135/9781848607934.n17

Geels, F. W., and Schot, J. (2007). Typology of sociotechnical transition pathways. *Res. Policy* 36, 399–417. doi: 10.1016/j.respol.2007.01.003

Grimshaw, D., Fagan, C., and Hebson, G. and Tavora, I. (2017). "A new labour market segmentation approach for analysing inequalities: introduction and overview," in *Making Work More Equal* (Manchester: Manchester University Press, 1–32. doi: 10.7765/9781526125972.00007

Grin, J., Rotmans, J., and Schot, J. (2010). *Transitions to Sustainable Development: New Directions in the Study of Long Term Transformative Change*. London: Routledge. doi: 10.4324/9780203856598

Gubrium, J. F., and Holstein, J. A. (1998). Narrative practice and the coherence of personal stories. *Sociol. Q.* 39, 163–187. doi: 10.1111/j.1533-8525.1998.tb02354.x

Guest, G., Bunce, A., and Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods* 18, 59–82. doi: 10.1177/1525822X05279903

Gupta, A., Pouw, N., and Ros-Tonen, M. (2015). Towards an elaborated theory of inclusive development. *Eur. J. Dev. Res.* 27, 541–559. doi: 10.1057/ejdr.2015.30

Gupta, J., and Pouw, N. (2017). Towards a trans-disciplinary conceptualization of inclusive development. *Curr. Opin. Environ. Sustain.* 24, 96–103. doi: 10.1016/j.cosust.2017.03.004

Gutheil, L. (2021). Adaptive project management for the civil society sector: towards an academic research agenda. *Int. Dev. Plann. Rev.* 43, 393–418. doi: 10.3828/idpr.2020.17

Helbing, D., and Kirman, A. (2013). Rethinking economics using complexity theory. *Real-World Econ. Rev.* 23—52. Available online at: https://papers.ssrn.com/abstract=2292370; http://www.paecon.net/PAEReview/issue64/HelbingKirman64.pdf

Hickey, S., and Mohan, G. (Eds.) (2004). *Participation: From Tyranny to Transformation? Exploring New Approaches to Participation in Development*. London: Zed Books.

Howard, N., Thomas, P. R., Boydell, V., and Squires, F. (2021). Power, bureaucracy and the 'incentive trap': how development programming can distort the agency of frontline workers. *World Dev.* 140:105283. doi: 10.1016/j.worlddev.2020.105283

Janssen, M. J., Bergek, A., and Wesseling, J. H. (2022). Evaluating systemic innovation and transition programmes: towards a culture of learning. *PLOS Sustain. Transform.* 1:e0000008. doi: 10.1371/journal.pstr.0000008

Johnson, C., Hegtvedt, K. A., Khanna, N., and Scheuerman, H. L. (2016). Legitimacy processes and emotional responses to injustice. *Soc. Psychol. Q.* 79, 95–114. doi: 10.1177/0190272516645317

Kallio, H., Pietilä, A. M., Johnson, M., and Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *J. Adv. Nurs.* 72, 2954–2965. doi: 10.1111/jan.13031

Kindon, S., Pain, R., and Kesby, M. (eds.) (2007). *Participatory Action Research Approaches and Methods: Connecting People, Participation and Place*. London: Routledge. doi: 10.4324/9780203933671

Kyamusugulwa, P. M. (2013). Participatory development and reconstruction: a literature review. *Third World Q.* 34, 1265–1278. doi: 10.1080/01436597.2013.824653

Lončar, J. (2016). State-building and local resistance in Kosovo: minority exclusion through inclusive legislation. *Communist Post-Communist Stud.* 49, 279–290. doi: 10.1016/j.postcomstud.2016.06.004

Maitlis, S., and Christianson, M. (2014). Sensemaking in organizations: taking stock and moving forward. *Acad. Manage. Ann.* 8, 57–125. doi: 10.5465/19416520.2014.873177

Major, C. (2014). *Teaching Online: A Guide to Theory, Research, and Practice*. Baltimore, MD: Johns Hopkins University Press.

Marcelo, D., Mandri-Perrott, C., House, S., and Schwartz, J. (2016). "Prioritizing infrastructure investment: a framework for government decision making," in *World Bank Policy Research Working Paper*, 7674. doi: 10.1596/1813-9450-7674

McEvoy, P., Brady, M., and Munck, R. (2016). Capacity development through international projects: a complex adaptive systems perspective. *Int. J. Manag. Proj. Bus.* 9, 528–545. doi: 10.1108/IJMPB-08-2015-0072

Nelson, R. R., and Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press of Harvard University Press.

Nishibe, M. (2006). Innovation, institutions, and evolution of economic systems. *Evol. Inst. Econ. Rev.* 3, 29–57.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511808678

Noy, C. (2008). Sampling knowledge: the hermeneutics of snowball sampling in qualitative research. *Int. J. Soc. Res. Methodol.* 11, 327–344. doi: 10.1080/13645570701401305

Pain, R., and Francis, P. (2003). Reflections on participatory research. *Area* 35, 46–54. doi: 10.1111/1475-4762.00109

Pain, R., and Kesby, M. (2007). "Geographies of impact: power, participation and potential," in *Participatory Action Research Approaches and Methods: Connecting People, Participation and Place*, eds. S. Kindon, R. Pain, and M. Kesby (London: Routledge), 184–190.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., Hoagwood, K., et al. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Adm. Policy Ment. Health Ment. Health Serv. Res.* 42, 533–544. doi: 10.1007/s10488-013-0528-y

Patton, M. Q. (2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use.* New York, NY: Guilford Press.

Pawson, R., and Tilley, N. (1997). "An introduction to scientific realist evaluation," in *Evaluation for the 21st Century: A Handbook*, eds. E. Chelimsky, and W. R. Shadish (London: SAGE), 405–418. doi: 10.4135/9781483348896.n29

Phillippi, J., and Lauderdale, J. (2018). A guide to field notes for qualitative research: context and conversation. *Qual. Health Res.* 28, 381–388. doi: 10.1177/1049732317697102

Platts-Fowler, D., and Robinson, D. (2016). Community resilience: a policy tool for local government? *Local Gov. Stud.* 42, 762–784. doi: 10.1080/03003930.2016.1186653

Pyle, D. (1984). *Life After Project. A Multi-Dimensional Analysis of Implementing Social Development Programs at the Community Level.* Boston, MA: John Snow, Inc.

Ramalingam, B., Laric, M., and Primrose, J. (2014). *From Best Practice to Best Fit: Understanding and Navigating Wicked Problems in International Development.* London: Overseas Development Institute (ODI).

Reason, P., and Bradbury, H. (eds.) (2008). *The SAGE Handbook of Action Research: Participative Inquiry and Practice, 2nd Edn.* London: SAGE Publications. doi: 10.4135/9781848607934

Reilly-King, F., Duggan, C., and Wilner, A. (2024). Foresight and futures thinking for international development co-operation: promises and pitfalls. *Dev. Policy Rev.* 42:e12790. doi: 10.1111/dpr.12790

Rinaldi, L. (2020). "Integrated thinking for stakeholder engagement: a processing model for judgments and choice in situations of cognitive complexity," in *The Routledge Handbook of Integrated Reporting*, eds. C. Villiers, P.-C. K. Hsiao, and W. Maroun (London: Routledge), 280–292. doi: 10.4324/9780429279621-19

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14, 29–48. doi: 10.1177/1356389007084674

Roulston, K., and Choi, M. (2018). "Qualitative interviews," in *The SAGE Handbook of Qualitative Data Collection*, ed. U. Flick (London: SAGE Publications), 233–249. doi: 10.4135/9781526416070.n15

Sadler, G. R., Lee, H. C., Lim, R. S., and Fullerton, J. (2010). Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy. *Nurs. Health Sci.* 12, 369–374. doi: 10.1111/j.1442-2018.2010.00541.x

Saran, A., White, H., and Kuper, H. (2020). Evidence and gap map of studies assessing the effectiveness of interventions for people with disabilities in low-and middle-income countries. *Campbell Syst. Rev.* 16:e1070. doi: 10.1002/cl2.1070

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

Senge, P. M. (2006). *The Fifth Discipline: The Art and Practice of the Learning Organization.* Revised Edn. New York, NY: Doubleday.

Shankar, A., Sundar, S., and Smith, G. (2019). Agency-based empowerment interventions: efforts to enhance decision-making and action in health and development. *J. Behav. Health Serv. Res.* 46, 164–176. doi: 10.1007/s11414-018-9592-0

Stanley, A., and Connolly, L. (2023). Implications of time on donor behavior and processes in relation to localization. *Negot. J.* 39, 455–480. doi: 10.1111/nejo.12446

Stockless, A., and Brière, S. (2024). How to encourage inclusion in a qualitative research project using a design-based research methodology. *Int. J. Qual. Methods* 23, 16094069241227852. doi: 10.1177/16094069241227852

Synowiec, C., Fletcher, E., Heinkel, L., and Salisbury, T. (2023). Getting rigor right: a framework for methodological choice in adaptive monitoring and evaluation. *Global Health Sci. Pract.* 11(Supplement 2), e2200243. doi: 10.9745/GHSP-D-22-00243

Tacchi, J., Slater, D., and Hearn, G. (2010). *Ethnographic Action Research: A User's Handbook.* New Delhi: UNESCO

Tavenner, K., and Crane, T. A. (2019). Beyond "women and youth": applying intersectionality in agricultural research for development. *Outlook Agric.* 48, 316–325. doi: 10.1177/0030727019884334

Thelen, K. (2004). *How Institutions Evolve: The Political Economy of Skills in Germany, Britain, the United States, and Japan.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511790997

Uraguchi, Z. B. (2010). Food price hikes, food security, and gender equality: assessing the roles and vulnerability of women in households of Bangladesh and Ethiopia. *Gend. Dev.* 18, 491–501. doi: 10.1080/13552074.2010.521992

Valters, C., Cummings, C., and Nixon, H. (2016). *Putting Learning at the Centre: Adaptive Development Programming in Practice.* London: Overseas Development Institute (ODI).

Varughese, S., Gale, L., Annuncia, S. F., Deore, Y., Grills, N., Soji, F., et al. (2025). The impact of a competency-based community-based inclusive development (CBID) training program in India. *Disabil. CBR Incl. Dev.* 36, 6–25.

Wailzer, M., and Soyer, L. (2022). Co-developing an impact model for evaluating the societal impact of participatory research approaches. *Fteval J. Res. Technol. Policy Eval.* 53, 111−118. doi: 10.22163/fteval.2022.547

Weber, H. (2018). "Politics of 'leaving no one behind': contesting the 2030 Sustainable Development Goals agenda," in *The Politics of Destination in the 2030 Sustainable Development Goals* (Routledge), 64–79.

Weick, K. E. (1995). *Sensemaking in Organizations.* Thousand Oaks, CA: SAGE Publications.

Westhorp, G. (2014). *Realist Impact Evaluation: An Introduction.* Methods Lab, London: Overseas Development Institute (ODI).

Westley, F., McGowan, K., and Tjörnbo, O. (2017). *The Evolution of Social Innovation: Building Resilience through Transitions.* Cheltenham: Edward Elgar Publishing. doi: 10.4337/9781786431158

Westley, F., Zimmerman, B., and Patton, M. Q. (2011). *Getting to Maybe: How the World is Changed.* Toronto, ON: Vintage Canada.

Williams, B. K., and Brown, E. D. (2018). Double-loop learning in adaptive management: the need, the challenge, and the opportunity. *Environ. Manage.* 62, 995–1006. doi: 10.1007/s00267-018-1107-5

Yin, R. K. (2017). *Case Study Research and Applications: Design and Methods*, 6th Edn. Thousand Oaks, CA: Sage Publications.