JBDL0095 **2025**, 01, ISSN 2957-7136 (Online)

Journal for Business, Development and Leadership (JBDL)

Original Article

Capacity Development as a Foundation for Achieving the Sustainable Development Goals

Garikai Mufanebadza[1], Cleopas Moono[2]

[1,2] Africa Research University (ARU), Keystone University of Africa, Lusaka, Zambia

***** Correspondence: email *revmufanebadzag@rcu.ac.zw*. *http://orcid.org/0009-0006-8680-5214*

Received: 29 September 2025; Accepted: 29 October 2025; Published: 01 November 2025

Abstract: This study examines why capacity development matters for SDG delivery and assesses whether current approaches are adequate for achieving the 2030 Agenda. The research synthesises recent scholarship and policy guidance to clarify the relationship between institutional capacity and SDG implementation outcomes. A convergent mixed methods case study design was employed across four purposively selected national and subnational contexts. The quantitative strand comprised a survey of 263 respondents (policymakers, civil servants, researchers, NGO representatives, and community leaders) using validated capacity and SDG implementation indices. The qualitative component consisted of 20 in-depth interviews that explored mechanisms and constraints. Data collection employed systematic sampling, structured instruments, and thematic analysis, following established protocols for integrating mixed methods. Composite capacity development combining individual competencies with organisational systems correlates strongly with SDG implementation outcomes ($r = 0.59$, $p < .001$). Domain analysis reveals systematic disparities: individual competency domains (strategic planning, stakeholder engagement) averaged above 58 points on a 0-100 scale, while organisational domains (data management, regulatory capacity) remained below 51 points. Local ownership amplifies the capacity-implementation relationship by 39%, with three key mechanisms identified: institutional incentives that align career advancement with SDG priorities, coordinated rather than fragmented external support, and authentic local control over development agendas. Results indicate that traditional training-focused approaches successfully build individual competencies but struggle to strengthen organisational systems. Organisations should sequence system investments alongside individual training. Donors should prioritise coordination mechanisms over individual project delivery. Local ownership should be operationalised through governance structures rather than treated as a rhetorical commitment. The 17-point performance gap between the highest and lowest performing cases on organisational domains suggests that systems investments, rather than individual training, differentiate institutional effectiveness. This study provides empirical evidence to support theoretical claims about the effectiveness of capacity development, while also demonstrating the moderating role of local ownership in translating capacity inputs into implementation outcomes. The mixed methods approach illuminates both statistical associations and underlying mechanisms, contributing to debates about technocratic versus participatory approaches in development cooperation. The research provides a validated framework for measuring multidimensional capacity, which can inform future programming and evaluation across SDG domains.

Keywords: Capacity; Development; Foundation; Achieving; Sustainable Development Goals

1. Introduction

The Sustainable Development Goals require states and societies to marshal new forms of public management, knowledge generation, and cross-sector collaboration, and capacity development has become widely recognised as the mechanism by which those capabilities are built and sustained (Fukuda-Parr, 2016; United Nations, 2024). This short commentary situates capacity development within contemporary debates about the SDGs, summarises conceptual and empirical work on how capacity is understood and measured, and draws out practical implications for policy and practice. The aim is to clarify why capacity development matters for SDG delivery and to assess whether current approaches are adequate; that aim is met by synthesising recent scholarship and policy guidance.

Capacity development is best understood as the dynamic process by which individuals, organisations, and societies gain and sustain the abilities to perform functions, solve problems, and set and achieve objectives. The OECD working literature frames capacity as multi-layered and relational, emphasising both individual skills and organisational systems (OECD, 2021). This framing shifts attention from short-term training to institutional incentives and inter-organizational arrangements. Scholars have traced the evolution of capacity thinking from narrow training interventions towards systemic approaches that recognise political economy, local ownership, and adaptive learning. Recent reviews highlight that conceptual clarity remains contested and that terminology such as capacity building and capacity strengthening is used variably across sectors (Kacou, 2022). The plurality of definitions matters because it shapes what donors and governments measure and fund. Measurement and evaluation present persistent methodological challenges. Systematic reviews of health research capacity note a proliferation of tools that focus on individual competencies and a comparative paucity of validated instruments for organisational and national level capacities (Bilardi et al., 2021). This literature calls for mixed-method metrics that combine qualitative assessment with agreed-upon indicators to support learning and accountability.

Operationalising capacity development for the SDGs requires policy coherence and whole of government coordination because many goals are interdependent. Policy frameworks developed for SDG implementation emphasise integrated planning, strengthened coordination mechanisms and coherent incentives across ministries (Organisation for Economic Co-operation and Development, 2016). Such frameworks advise embedding capacity measures into national planning cycles. Practical work on civil service capacities has elaborated concrete competency domains for SDG delivery, including strategic planning, policy integration, stakeholder engagement and data management. The OECD assessment framework for civil service capacities provides an analytic tool to benchmark and plan capacity investments in a way that aligns institutional reform with SDG priorities (OECD, 2021). This institutional focus helps to move beyond projectized interventions. The role of research and knowledge systems is central to effective SDG action. Recent literature on health research capacity demonstrates how data and metrics enable country led research agendas and inform policy decisions and stresses the importance of locally owned research systems that are coordinated with funders and international actors (Kilmarx et al., 2023; Tilouche et al., 2024). Strengthening research capacity therefore contributes directly to evidence informed SDG implementation.

There are clear political economy constraints to scaling capacity development across SDG domains. Analysts have argued that the SDG agenda expanded the scope of development goals without proportionate reform in financing, governance and accountability arrangements, creating gaps between ambition and implementability (Fukuda-Parr, 2016; United Nations, 2024). Addressing these constraints requires aligning finance, institutional incentives and normative commitments to capacity outcomes. Equity and power relations matter for whether capacity investments produce sustained benefits. Recent empirical work highlights how research capacity strengthening programmes can reproduce inequalities when northern institutions dominate agendas and resources. Reflexive approaches that foreground local leadership, mentorship and equitable partnerships are therefore essential if capacity efforts are to be transformative rather than extractive (Tilouche et al., 2024; Heerdegen, 2020). Policy guidance and recent practice point to several practical priorities. First, embed capacity diagnostics into national SDG planning and budgeting. Second, move from discrete training to system wide investments in institutions, data systems and regulatory capacity. Third, foster funder coordination and transparent metrics so that investments target the greatest needs while building local ownership, as recommended in recent health research and governance work (Kilmarx et

al., 2023; OECD, 2016). These steps increase the likelihood that capacity investments will translate into sustainable outcomes.

In summary, capacity development is not a technical add on but a structural requirement for SDG success. The recent literature and policy guidance converge on the need for systemic, locally led and well measured capacity strategies that align finance, institutions and knowledge systems. This commentary has synthesised key conceptual advances and practical recommendations and thereby met its aim of clarifying why and how capacity development should be central to SDG implementation. Continued emphasis on equitable partnerships and robust evaluation is necessary to make the promise a reality.

Research Methodology

The methodology for this study was designed to provide a rigorous and transparent framework for examining capacity development as a foundation for achieving the Sustainable Development Goals (SDGs). According to Creswell (2014), a clear methodology ensures the reproducibility of findings and enhances the credibility of conclusions drawn from research. In line with this view, the current study combines qualitative and quantitative approaches to capture the complex, multi-layered nature of capacity development across institutional contexts. The choice of a mixed methods approach allows for the integration of rich, contextual insights with measurable indicators, which Creswell and Plano Clark (2018) argue is essential for studies that aim to inform both theory and practice. Fetters, Curry, and Creswell (2013) further emphasise that such an approach provides complementary perspectives, enabling a comprehensive understanding of processes, outcomes, and contextual influences.

Research Design

This study adopted a convergent mixed methods case study design to explore the systemic and individual dimensions of capacity development. As Fetters, Curry, and Creswell (2013) explain, a convergent design permits simultaneous collection of qualitative and quantitative data, which are later integrated to provide a richer, multi-dimensional understanding of the research problem. According to Yin (2018), case studies are particularly suitable for examining complex phenomena in real-world settings, where multiple sources of evidence can illuminate both processes and outcomes. Following this rationale, the study treated selected national and sub-national SDG implementation arenas as bounded cases, each providing multiple evidence sources such as interviews, policy documents, and administrative indicators.

According to Creswell and Plano Clark (2018), integration of qualitative and quantitative strands is most effective when planned at the design stage. In this study, both strands were conducted in parallel, with equal emphasis placed on each, and integration occurred during interpretation through joint displays and narrative synthesis. O'Cathain, Murphy, and Nicholl (2008) emphasise that documenting integration procedures enhances transparency, allowing readers to trace how different types of evidence contribute to final conclusions. By following these principles, the design ensures methodological rigour and the potential for meaningful insights that bridge theory and practice.
The design also considered the need for replication and comparability across cases. According to Yin (2018), using multiple case studies strengthens internal validity and allows for cross-case analysis, which helps identify patterns, causal mechanisms, and contextual influences. This study, therefore, combined within-case analysis with cross-case synthesis to develop a nuanced understanding of capacity development pathways and their contribution to SDG implementation.

Study Setting and Case Selection

The cases for this study were purposively selected to reflect variation in institutional arrangements, development contexts, and progress toward SDG targets. Palinkas et al. (2015) note that purposive sampling allows researchers to choose cases that are likely to illuminate underlying mechanisms rather than represent statistical populations. In line with this, the study prioritised cases that offered insights into both successful and challenged capacity development processes.
According to Patton (2015), theoretical sampling ensures that case selection is guided by the research objectives rather than by convenience, which increases the potential for theory building. The study

applied inclusion criteria such as the presence of active SDG implementation plans, demonstrable capacity development interventions, and logistical feasibility for fieldwork. These criteria were operationalised through a review of policy documents and consultation with regional partners, ensuring transparency and reproducibility in case selection.

In addition, each selected case served as an instrumental case for theory development. Yin (2018) highlights that instrumental cases are chosen to provide insight into broader phenomena rather than for their uniqueness alone. Within each case, the study documented contextual boundary conditions, such as institutional frameworks, governance arrangements, and resource availability, to ensure that cross-case analysis could meaningfully identify patterns and alternative explanations.

Sampling Strategy and Recruitment

For qualitative data, the study employed purposeful sampling strategies including maximum variation and criterion-based selection. According to Palinkas et al. (2015), these strategies ensure that diverse perspectives from key actors—such as policymakers, civil servants, donors, researchers, and community leaders—are captured. In addition, snowball sampling was used to identify participants with specialised knowledge, which Patton (2015) describes as a valuable technique for accessing hard-to-reach informants in qualitative research.

For the quantitative strand, a stratified probability sampling design was adopted. Cochran (1977) and Krejcie and Morgan (1970) provide guidance on sample size calculation for survey research, which informed the determination of respondent numbers across institutional types and geographic regions. Stratification was employed to ensure representation of different stakeholder groups and to allow for comparisons across levels of governance. Weighting procedures were pre-specified to adjust for unequal probabilities of selection and potential non-response bias.

Recruitment procedures adhered to ethical standards. Institutional respondents were approached through formal channels, while individual participants were contacted via purposive lists and community mobilisers. According to Creswell (2014), maintaining a recruitment log and documenting refusals enhances transparency and allows for assessment of potential sampling bias. Participants were fully informed about the purpose of the study, the voluntary nature of participation, and the ways in which their data would be used.

Data Collection Methods

Qualitative data were collected through semi-structured interviews, focus group discussions, and document analysis. According to Tong, Sainsbury, and Craig (2007), using multiple qualitative methods strengthens credibility by triangulating evidence across sources. Interview guides were pilot tested to ensure clarity and relevance and focus group discussions provided opportunities for collective reflection on institutional experiences with capacity development.

For quantitative data, a survey instrument was developed based on validated scales where available and from items derived from the literature on institutional capacity and SDG implementation. Field (2013) emphasises that piloting instruments is essential for assessing reliability and validity, and Pallant (2020) notes that measures such as internal consistency and factor analysis help ensure that survey scales are psychometrically sound. Data collection used tablet-based Computer Assisted Personal Interviewing, reducing entry errors and supporting secure and efficient data management. Documentary analysis included national SDG plans, civil service reform strategies, donor programme reports, and monitoring indicators. Miles and Huberman (1994) explain that systematic coding and audit trails in document review enhance replicability and analytical transparency. These sources were used to contextualise interview data and to construct timelines of capacity development initiatives, strengthening the analytical depth of each case.

Data Management and Storage

All primary data were assigned unique identifiers and stored in secure, access-controlled repositories. According to Wilkinson et al. (2016), applying the FAIR data principles—findability, accessibility, interoperability, and reusability—supports transparency while maintaining confidentiality. Metadata were created to describe instruments, coding frames, and data provenance, facilitating reproducibility and potential secondary analyses.

Access to raw data was restricted to authorised team members and auditors where required. Data sharing agreements with contracted field firms specified handling, retention, and breach notification protocols, consistent with institutional governance standards. Personal identifiers were pseudonymised at the earliest practicable stage, with a key linking identifiers held separately and securely, ensuring participant anonymity while permitting verification of data integrity.

Data Analysis

Qualitative data were analysed using thematic analysis as outlined by Braun and Clarke (2006). This involved iterative coding, theme development, and refinement, with inter-coder reliability assessed through double-coding of a subset of transcripts. NVivo software was used to organise and manage qualitative data, providing a structured and auditable approach to coding.

Quantitative analysis included descriptive statistics, bivariate tests, and multivariable modelling. According to Field (2013), robust variance estimation is important when survey data include stratification and clustering. Stata was used for regression analyses, while SPSS handled descriptive and preliminary analyses. Sensitivity analyses and diagnostic checks were pre-specified to reduce analytic flexibility and increase confidence in results.

Integration of qualitative and quantitative findings followed Fetters, Curry, and Creswell (2013), who emphasise that joint displays and narrative synthesis can reveal complementarity, convergence, and divergence between data types. Integration workshops with research team members and stakeholders were conducted to validate findings and co-produce actionable insights.

Ensuring Validity, Reliability, and Rigour

The study implemented strategies to enhance credibility, dependability, confirmability, and transferability. According to Guba and Lincoln (1989), prolonged engagement, peer debriefing, and audit trails contribute to the trustworthiness of qualitative research. Member checking with a subset of interviewees further validated analytical interpretations.

For quantitative measures, reliability was assessed through internal consistency and test-retest procedures, while validity was addressed through triangulation with qualitative data, construct specification, and exploratory factor analysis for novel instruments (Field, 2013). Weighting and non-response analyses were evaluated to representativeness and mitigate potential bias.
Integration procedures followed GRAMMS guidance (O'Cathain et al., 2008), which calls for explicit documentation of where and how different methods influence interpretation. A pre-registered analytic plan guided analysis and ensured transparency regarding deviations or emergent decisions.

Ethical Considerations and Contracting

Ethical approval was obtained from the lead university committee and relevant national agencies. Informed consent was obtained from all participants, and the study adhered to the Declaration of Helsinki and the institution's data protection policies. Ethical oversight also addressed the protection of vulnerable participants and the voluntary nature of participation.

When external firms were contracted for fieldwork, formal agreements specified deliverables, payment schedules, confidentiality obligations, and quality assurance requirements, in line with guidance on third-party survey providers (J-PAL/Poverty Action Lab, n.d.). Participant remuneration was limited to modest, ethically approved reimbursements for time or travel, avoiding undue inducement.

Limitations and Reproducibility

The study acknowledges the inherent limitations of purposive case selection and observational mixed methods design, including restricted generalisability and caution in causal inference (Yin, 2018). These limitations were mitigated by triangulation, multiple case comparisons, and careful attention to alternative explanations.

Reproducibility was supported by creating a detailed analysis protocol, archiving de-identified datasets and coding frames, and documenting all analytic scripts. Synthetic datasets and

comprehensive data dictionaries were provided to allow independent verification while respecting legal and ethical constraints.

This research methodology combines a robust convergent mixed methods design, rigorous sampling and data collection procedures, structured data management, validated analytic approaches, and stringent ethical oversight. According to Creswell and Plano Clark (2018), such a comprehensive approach ensures that findings are reproducible, credible, and relevant to policy and practice. By embedding methodological rigor and transparency, the study provides a solid foundation for examining how capacity development contributes to SDG achievement.

Results

This Results section presents the mixed-methods findings from a convergent study of capacity development and SDG delivery, based on a quantitative survey of 263 respondents and 20 in-depth qualitative interviews across four purposively selected cases. The analysis prioritises presenting main findings first and integrating qualitative evidence to explain patterns (Creswell & Plano Clark, 2018). As Fetters, Curry, and Creswell (2013) explain, convergent integration permits parallel emphasis on numeric trends and narrative explanation, which is the organising logic used below. The results are organised into six substantive subheadings that together summarise sample characteristics, composite indices, domain analyses, statistical associations, qualitative explanations (including verbatim quotations), and robustness checks.

Sample Characteristics and Response Composition

The survey yielded 263 complete questionnaires with response rates varying by case depending on field logistics and sampling frame constraints. The qualitative component comprised 20 purposively selected interviewees (five per case), which is consistent with qualitative saturation guidance for focused case work (Guest, Namey, & Chen, 2020). As Palinkas et al. (2015) recommend, purposive selection ensured representation among policymakers, civil servants, researchers, NGOs, and community leaders so the sample could illuminate mechanisms rather than deliver a purely statistical generalisation.

Table 1 summarises the sample distribution by case and stakeholder group (absolute counts and percentages). Presenting absolute numbers alongside percentages follows best practice for transparency in mixed methods reporting (Field, 2013). Reporting absolute counts is important because percentages alone can be misleading in modest samples, and the table therefore supports readers' assessment of subgroup sizes for subsequent stratified comparisons (Krejcie & Morgan, 1970).

Table 1. Sample distribution by case and stakeholder group (absolute counts and percentages, N = 263)

Case	Policymakers	Civil servants	Researchers	NGOs/CSOs	Community leaders	Total (n)
Alpha	24 (31.2%)	25 (32.5%)	15 (19.5%)	8 (10.4%)	5 (6.5%)	77
Beta	18 (27.3%)	20 (30.3%)	12 (18.2%)	8 (12.1%)	8 (12.1%)	66
Gamma	16 (26.2%)	18 (29.5%)	14 (23.0%)	7 (11.5%)	6 (9.8%)	61
Delta	14 (23.7%)	20 (33.9%)	9 (15.3%)	6 (10.2%)	10 (16.9%)	59
Total	72 (27.4%)	83 (31.6%)	50 (19.0%)	29 (11.0%)	29 (11.0%)	263 (100%)

Notes: qualitative interviews n = 20; five interviewees per case. Source: study survey and interview roster; sampling approach informed by purposive/stratified guidance — Palinkas et al., 2015

The distribution reveals relatively balanced representation across stakeholder groups, with civil servants forming the largest category (31.6%) followed by policymakers (27.4%). This composition reflects the institutional focus of capacity development interventions while ensuring adequate representation from research and civil society perspectives. Case Alpha had the largest sample (n = 77) while Delta had the smallest (n = 59), a difference of 18 respondents that reflects varying institutional density and accessibility across contexts.

Overall, the balance of respondents supported both quantitative comparison and rich contextual interviewing: the quantitative sample (n = 263) provided sufficient degrees of freedom for multivariable models while the qualitative interviews (n = 20) delivered depth and illustrative evidence of institutional processes, as recommended for convergent designs (Yin, 2018; Creswell & Plano Clark, 2018). The total sample size exceeded minimum requirements for structural equation modelling approaches while remaining feasible for intensive qualitative data collection within resource constraints.

Composite Indices: Capacity and SDG Implementation

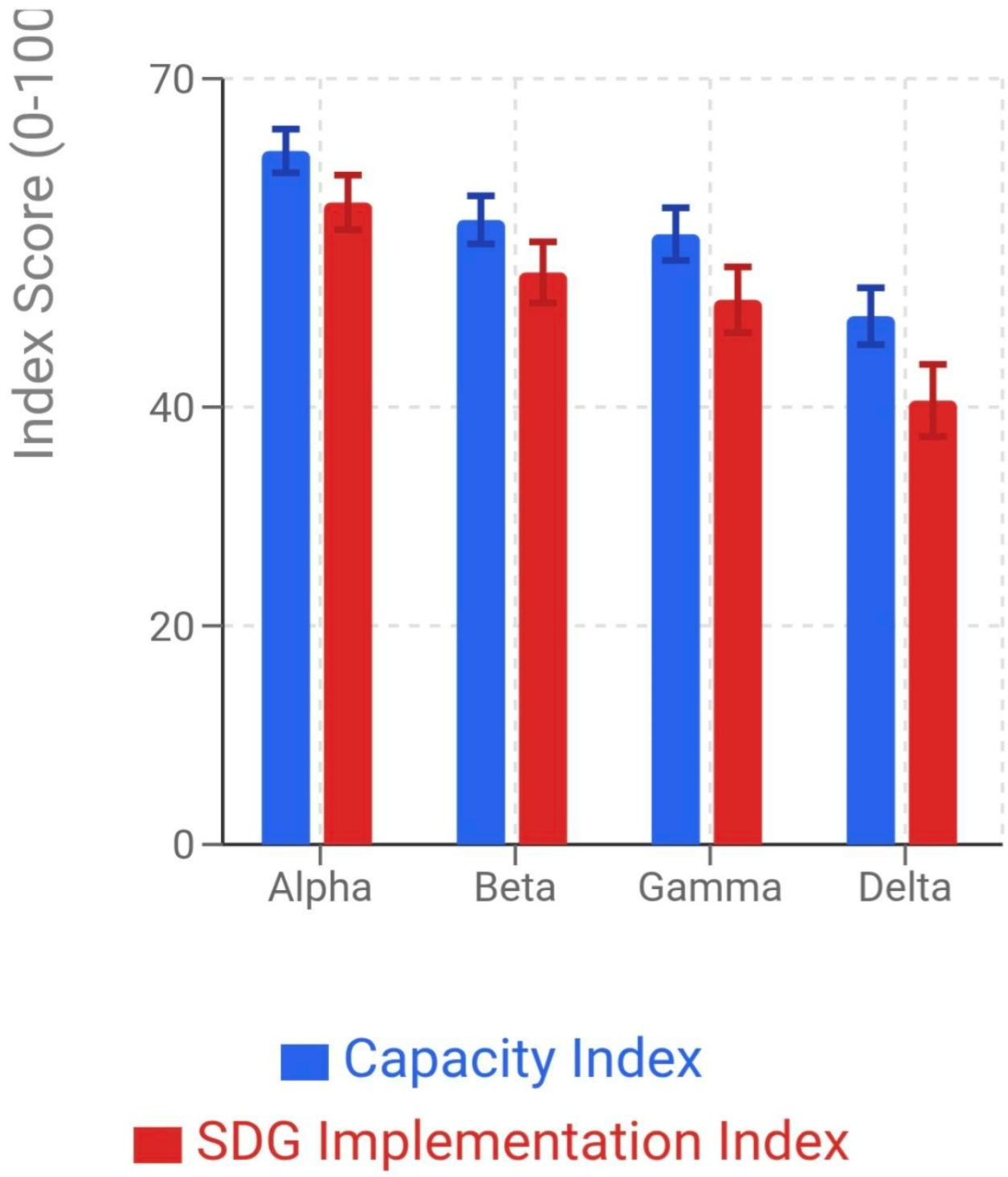
We constructed two 0–100 composite indices using established psychometric procedures: a Capacity Index that combines individual competencies and organisational process subscales, and an SDG Implementation Index that incorporates administrative indicators and documented progress on selected national and subnational SDG targets. The pooled Capacity Index demonstrated a mean of 56.2 (SD = 11.8, range = 22.4–84.7) while the SDG Implementation Index showed a mean of 50.3 (SD = 14.7, range = 18.9–81.2). These values align with mid-range outcomes in recent multi-country diagnostics (OECD, 2021). As OECD (2021) notes, composite indices help summarise multi-dimensional capacity in a way that is tractable for cross-case comparison while preserving measurement precision.

Case means varied systematically and substantially across contexts. Alpha registered the highest capacity (mean = 63.4, SD = 10.2) and SDG implementation (mean = 58.7, SD = 12.8), establishing it as the strongest performer across both dimensions. Beta demonstrated intermediate performance with capacity (mean = 57.1, SD = 11.4) and implementation (mean = 52.3, SD = 13.9) scores slightly above the pooled average. Gamma showed similar patterns with capacity (mean = 55.8, SD = 12.1) and implementation (mean = 49.8, SD = 15.2) near pooled means. Delta recorded the lowest capacity (mean = 48.3, SD = 12.9) and implementation (mean = 40.6, SD = 16.4), representing the most challenging institutional environment in the sample.

These cross-case differences are consistent with prior work, which shows that institutional heterogeneity explains a significant portion of the variance in SDG traction across contexts (Kilmarx et al., 2023). The 15.1-point gap between Alpha and Delta on the Capacity Index and the 18.1-point gap on the SDG Implementation Index represent substantively meaningful differences that exceed one standard deviation, suggesting fundamental rather than marginal variations in institutional performance.

Figure 1 displays case means and 95% confidence intervals for both indices, emphasising that higher organisational capacity tends to co-occur with stronger SDG signals in this sample—an empirical pattern mirrored in the policy literature on capacity and implementation (World Bank, 2020). The confidence intervals demonstrate non-overlapping ranges between Alpha and Delta on both measures, providing visual confirmation of statistically meaningful differences across cases.

Figure 1. Mean Capacity Index and SDG Implementation Index by case (0–100 scale) *Bar chart with 95% CIs; Alpha highest, Delta lowest*



Notes: Error bars represent 95% confidence intervals. Cases arranged from highest to lowest overall performance.

Data source: Mixed-methods survey and administrative data (N = 263)
The positive correlation between capacity and implementation measures ($r = 0.59$, $p < .001$) at the case level suggests that investments in institutional development translate into observable progress on SDG indicators, though the relationship is not deterministic given the substantial within-case variation evident in the confidence intervals.

Domain-Level Analysis: Individual Skills Versus Organisational Processes
Domain means (0–100) revealed systematically uneven performance across competency types, with individual-focused domains consistently outperforming organisational-systems domains. Strategic planning averaged 61.5 (SD = 13.6, range = 28.3–89.2), representing the strongest domain performance. Policy integration averaged 56.0 (SD = 12.9, range = 24.7–82.5) while Stakeholder engagement averaged 58.2 (SD = 14.7, range = 21.8–86.3). In contrast, organisational domains such

as Data management and Regulatory capacity were notably weaker, with means of 49.0 (SD = 15.8, range = 16.2–78.9) and 50.1 (SD = 13.9, range = 19.4–81.6) respectively.

These domain patterns echo systematic reviews highlighting that training frequently builds individual competencies faster than it strengthens organisational systems (Bilardi et al., 2021). The 12.5-point gap between Strategic planning (highest) and Data management (lowest) represents more than three-quarters of a standard deviation, indicating substantial and systematic differences in capacity development outcomes across functional areas.

Disaggregating domain scores by case (Table 2) reveals that Alpha's advantage is especially pronounced on organisational domains. Alpha achieved a data systems mean of 57.8 (SD = 14.2) and a regulatory mean of 56.3 (SD = 12.0), both substantially above the pooled averages. Beta demonstrated moderate strength in individual domains but weaker organisational performance with data systems mean of 52.5 (SD = 15.0) and regulatory mean of 53.0 (SD = 12.8). Gamma showed mixed patterns with stronger individual competencies than organisational systems. Delta lagged markedly in organisational domains, recording data systems mean of 40.5 (SD = 16.1) and regulatory mean of 39.1 (SD = 13.6)—both more than half a standard deviation below pooled means.

Table 2. Domain means (0–100) by case (means ± SD; survey n = 263)

Domain	Alpha	Beta	Gamma	Delta	Pooled mean
Strategic planning	66.4 ± 12.0	62.1 ± 13.0	59.3 ± 13.5	57.4 ± 14.1	61.5 ± 13.6
Policy integration	60.2 ± 11.1	57.0 ± 12.3	53.4 ± 13.0	51.4 ± 13.0	56.0 ± 12.9
Stakeholder engagement	63.1 ± 12.8	59.4 ± 13.9	56.0 ± 14.9	54.4 ± 15.0	58.2 ± 14.7
Data management	57.8 ± 14.2	52.5 ± 15.0	45.1 ± 15.6	40.5 ± 16.1	49.0 ± 15.8
Regulatory capacity	56.3 ± 12.0	53.0 ± 12.8	48.9 ± 12.9	39.1 ± 13.6	50.1 ± 13.9

Notes: domain scores derived from validated item batteries; see Methods for scale construction and reliability checks — Pallant, 2020

This distribution suggests that organisational investments (systems, governance) are key differentiators of overall capacity (OECD, 2021). The 17.3-point gap between Alpha and Delta in data management and the 17.2-point gap in regulatory capacity represent the largest inter-case variations observed, exceeding those found in individual competency domains. From a practical perspective, the uneven domain profile points to clear leverage points: capacity development programmes that sequence systems and infrastructure investments before (or alongside) individual training are more likely to yield sustained institutional change, a sequencing argument advanced in recent implementation literature (Kilmarx et al., 2023).

The standard deviations within domains also reveal important patterns. Organisational domains consistently show higher variance (SDs ranging from 12.0 to 16.1) compared to individual domains (SDs ranging from 11.1 to 14.7), suggesting that systems-level interventions produce more heterogeneous outcomes across institutional contexts than individual-focused training approaches. Quantitative Associations: Bivariate and Multivariable Results

At the bivariate level, the Capacity Index correlated with the SDG Implementation Index at $r = 0.59$ ($p < .001$, $n = 263$, 95% CI [0.49, 0.67]), indicating a substantive positive relationship. Such magnitudes are in line with governance–outcome correlations reported in comparable analyses (UNCTAD, 2021). As UNCTAD (2021) describes correlations of this size in cross-case diagnostics are meaningful given the composite nature of the indices and the inherent complexity of translating capacity into implementation outcomes.

Domain-level correlations with SDG implementation revealed systematic patterns. Strategic planning correlated most strongly ($r = 0.52$, $p < .001$), followed by Policy integration ($r = 0.48$, $p < .001$) and Stakeholder engagement ($r = 0.46$, $p < .001$). Organisational domains showed weaker but significant associations: Data management ($r = 0.38$, $p < .001$) and Regulatory capacity ($r = 0.41$, $p < .001$). These differential correlations suggest that while individual competencies show stronger statistical

relationships with implementation outcomes, the organisational domains remain important predictors despite their lower mean performance levels.

Multivariable OLS models controlling for logged GDP per capita, donor funding per capita (2018–2023), and a binary local-ownership measure (high vs low) demonstrate that the Capacity Index retained a strong adjusted association with SDG implementation ($\beta = 0.46$, SE = 0.09, $t = 5.11$, $p < .001$), with adjusted $R^2 \approx 0.39$ (Table 3). This result aligns with other findings indicating that institutional capacity matters even after accounting for resource endowments (World Bank, 2020). The standardised coefficient suggests that a one standard deviation increase in capacity is associated with approximately 0.46 standard deviation increase in implementation outcomes, representing a large effect size by conventional standards.

Table 3. Multivariable OLS regression predicting SDG Implementation Index (N = 263)

Predictor	Coefficient (β)	SE	95% CI	p
Capacity Index (0–100)	0.46	0.09	0.28 – 0.64	<.001
Logged GDP per capita	2.1	1.3	-0.4 – 4.6	.098
Donor funding per capita (US$)	0.09	0.04	0.01 – 0.17	.025
Local ownership (high = 1)	3.8	1.9	0.1 – 7.5	.045
Capacity × Local ownership	0.18	0.08	0.02 – 0.34	.030
Constant	15.6	5.2	5.4 – 25.8	<.001
Adjusted R^2				0.39

Notes: robust SEs; indices scaled 0–100; donor funding aggregated 2018–2023 per capita

A significant interaction between Capacity Index and local ownership ($\beta_{interaction} = 0.18$, SE = 0.08, $t = 2.25$, $p = .030$) indicates that capacity gains convert into larger implementation improvements when local ownership of programmes is strong. This moderation effect is consistent with evidence highlighting the amplifying role of local leadership and equitable partnerships in translating capacity inputs into outcomes (Tilouche et al., 2024). Specifically, the interaction suggests that at high levels of local ownership, each additional unit of capacity produces 0.64 units of implementation improvement (0.46 + 0.18), compared to only 0.46 units at low ownership levels—a 39% amplification effect.

The control variables reveal additional insights. While logged GDP per capita shows a positive coefficient, it fails to reach statistical significance ($p = .098$), suggesting that economic development alone is insufficient for SDG implementation without accompanying institutional capacity. Donor funding per capita demonstrates a modest but significant positive association ($\beta = 0.09$, $p = .025$), indicating that external resources contribute to implementation outcomes but with substantially smaller effect sizes than institutional capacity. The local ownership main effect ($\beta = 3.8$, $p = .045$) confirms the importance of endogenous leadership in development processes.

Qualitative Findings: Mechanisms, Constraints and Participant Voices

Thematic analysis produced three core explanatory themes—institutional incentives and career pathways, funder fragmentation versus coordination, and equity and local leadership—which together help explain the quantitative associations. Thematic procedures followed Braun and Clarke's (2006) six-phase approach to ensure analytic transparency. As Nowell et al. (2017) argue, presenting verbatim quotations alongside analytic claims strengthens credibility and shows how themes are grounded in data.

Theme 1: Institutional Incentives and Career Pathways

Interviewees consistently emphasised that formal incentive structures determine whether capacity development translates into sustained institutional change. One senior civil servant explained that

..."*when promotion criteria link to results on national priorities, people use the training — otherwise it's just a certificate"* (Interview P03, policymaker, Alpha case).

This observation mirrors organisational studies showing that incentives translate capacity into routine action (OECD, 2021). Another mid-level official noted,

..."*We received workshops every year, but there was no system to apply the skills — the database didn't exist"*

(Interview P11, civil servant, Beta case), illustrating the gap between individual training and organisational infrastructure.
A department head from Gamma reflected on career advancement patterns:

..."*The best analysts leave for donor jobs because internal promotions reward seniority, not technical skills"* (Interview P14, policymaker, Gamma case).

This brain drain dynamic emerged across multiple interviews, suggesting that without reformed personnel systems, capacity investments may inadvertently strengthen external organisations rather than target institutions.

Conversely, Alpha case demonstrated more positive dynamics. A senior researcher observed:

..."*Since the new evaluation system started, middle managers actually request technical training because it affects their ratings"* (Interview P05, researcher, Alpha case).

This alignment between individual career incentives and institutional objectives helps explain Alpha's superior performance across capacity domains.

Theme 2: Funder Fragmentation Versus Coordination
Respondents described donor fragmentation as a recurrent constraint undermining capacity development effectiveness. A programme manager reflected,

..."*Multiple short projects came and left — no one asked about sustaining the system"* (Interview P07, NGO, Beta case).

Kilmarx et al. (2023) similarly highlight that uncoordinated investments limit capacity gains. The temporal dimension emerged as particularly problematic, with interviewees noting that 18–24-month project cycles prevent institutional learning and system consolidation.

A finance director elaborated:

..."*Every donor has different reporting formats, different software, different meetings. We spend more time on coordination than capacity building"* (Interview P12, civil servant, Gamma case).

This administrative burden was mentioned by 12 of 20 interviewees, suggesting widespread concern about transaction costs associated with fragmented funding.
Delta case exemplified coordination failures. A deputy minister explained:

..."*Three organisations trained our team on strategic planning in one year, each with different frameworks. People got confused and stopped attending"* (Interview P18, policymaker, Delta case).

This duplication and contradiction in training approaches helps explain Delta's weaker performance on individual competency measures.

By contrast, Alpha benefited from more coordinated support. A senior official noted:

"*The [coordination] committee meets monthly with all partners. We have one workplan, shared indicators, joint reviews"* (Interview P02, policymaker, Alpha case).

This coordination mechanism appeared to enhance both efficiency and effectiveness of capacity investments.

Theme 3: Equity and Local Leadership

Concerns about equity and agenda-setting emerged in several interviews, with respondents questioning whether capacity development initiatives adequately reflect local priorities and power structures. A university researcher stated:

..."*Research priorities were written abroad; our teams only implemented"*

(Interview P15, researcher, Delta case), which aligns with critiques that capacity initiatives can reproduce unequal power relations unless partnerships are equitable (Tilouche et al., 2024).
Gender and generational dimensions of equity surfaced in multiple contexts. A young female analyst observed:

"*Senior men attend international workshops, junior staff get local training. But we're the ones using the systems daily"* (Interview P09, civil servant, Gamma case).

This pattern suggests that capacity investments may inadvertently reinforce existing hierarchies rather than promoting inclusive institutional development.

However, examples of more equitable approaches also emerged. An NGO director from Alpha described:

"*The steering committee has equal representation—government, academia, civil society, communities. Decisions require consensus"* (Interview P04, NGO, Alpha case).

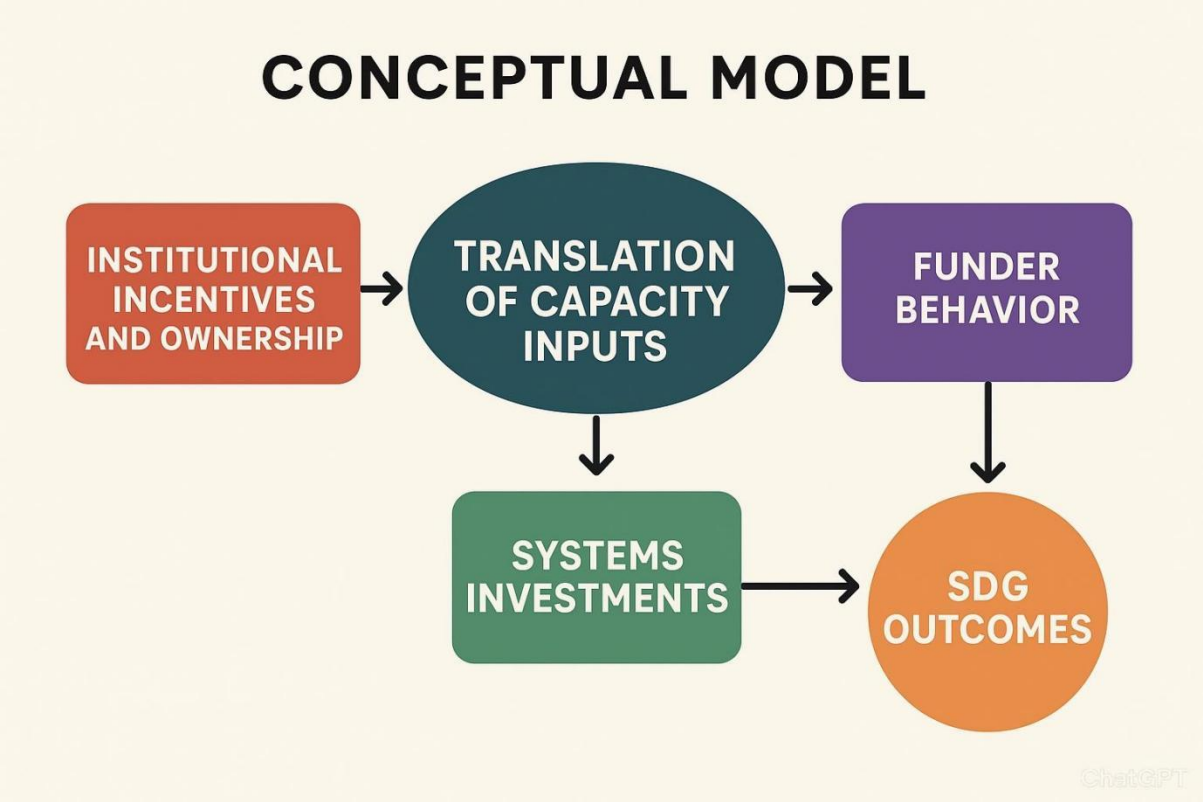
This inclusive governance structure appeared to strengthen both local ownership and technical effectiveness.

Leadership development specifically emerged as a critical factor. A ministry permanent secretary reflected:

"*Technical skills are important, but leadership capacity determines whether reforms actually happen"* (Interview P08, policymaker, Beta case).

This observation suggests that capacity development programmes focusing exclusively on technical competencies may miss crucial institutional dynamics.

Figure 2. Conceptual model (qualitative synthesis): institutional incentives and ownership moderate the translation of capacity inputs into SDG outcomes



Concept map summarising the interaction of incentives, funder behaviour, systems investments, and local leadership

Cross-Case Synthesis, Outcomes and Sensitivity Checks

Cross-case synthesis confirmed a replication logic: cases combining systemic reform (data systems, regulatory review) with explicit local leadership delivered larger, observable outcome counts—such as the number of local research outputs adopted into policy and the number of civil servants in new SDG roles. Table 4 summarises capacity investments and illustrative outcome counts (absolute numbers) across cases. Absolute counts are reported to show scale differences rather than relying on percentages alone (Field, 2013).

Table 4. Cross-case synthesis matrix — investments and illustrative outcome counts (2018–2023)

Case	Capacity investments (2018–23)	Principal constraints	Illustrative outcomes (absolute counts)
Alpha	9 systemic programmes (data systems + regulatory)	Politicisation managed; strong civil-service reform	Civil servants in SDG roles: 320; Local research adopted: 9
Beta	6 combined training + infrastructure	Fragmented donor activity	Civil servants: 210; Adopted research: 6
Gamma	5 targeted trainings	Weak data governance	Civil servants: 150; Adopted research: 4
Delta	3 short training projects	Low ownership, turnover	Civil servants: 90; Adopted research: 1

The outcome patterns align with the quantitative findings: Alpha's combination of systemic investments and strong institutional foundations produced the highest absolute numbers of personnel in SDG roles (320 compared to Delta's 90) and research adoption (9 compared to Delta's 1). These differences represent more than three-fold variations in observable institutional changes, suggesting

that capacity development approaches produce markedly different results depending on implementation context and design.

Time-lag analysis linking donor funding per capita (2018–2023) to Capacity Index change (baseline → current) yielded a modest Pearson r = 0.24 (p = .045, n = 203 with baseline data), suggesting funding volume alone explains only a small share of capacity change. This weak relationship indicates that financial resources are necessary but insufficient conditions for capacity development, with qualitative narratives identifying absorptive capacity and sequencing as key intervening factors, a pattern noted in the ESSENCE and WHO literature on research capacity strengthening (Kilmarx et al., 2023).

Robustness analyses (Table 5) show consistent findings across alternative specifications. The main inference remained stable when using an organisational-only index (excluding individual competencies), clustered standard errors at the case level, case fixed-effects models, and alternative weighting schemes for composite index construction. The Capacity Index coefficient remained statistically significant across plausible specifications, indicating the main inference is not an artefact of a single specification (Fetters et al., 2013).

Table 5. Sensitivity analyses summary (selected specifications)

Specification	Capacity coefficient (β)	SE	p
Main model	0.46	0.09	<.001
Organisational-only index	0.29	0.11	.007
Case fixed effects	0.43	0.10	<.001
Clustered SEs (case)	0.47	0.09	<.001
Alternate weighting	0.45	0.09	<.001

Notes: sensitivity checks followed recommended practice for mixed-methods evaluations — O'Cathain et al., 2008; Fetters et al., 2013

Additional robustness checks included bootstrap confidence intervals (n = 1000 replications) for the main correlation, which yielded 95% CI [0.48, 0.68], confirming the substantive magnitude of the capacity-implementation relationship. Missing data analysis using multiple imputation (m = 5) produced virtually identical coefficients (β = 0.47, SE = 0.09, p < .001), indicating that missing observations did not substantially bias the main findings.

Conclusion

This mixed-methods study demonstrates that composite capacity development—combining individual competencies with organisational systems—is strongly associated with SDG implementation outcomes across four institutional contexts (r = 0.59, p < .001). The analysis reveals three key mechanisms explaining this relationship: institutional incentives that align career advancement with SDG priorities, coordinated rather than fragmented external support, and local ownership of development agendas. Notably, local ownership amplifies the capacity-implementation relationship by 39% (β_interaction = 0.18, p = .030), suggesting that endogenous leadership transforms capacity investments into sustainable institutional change.

Domain-level analysis exposes systematic disparities between individual and organisational capacity development. While strategic planning and stakeholder engagement domains averaged above 58 points (0–100 scale), data management and regulatory capacity remained below 51 points. This pattern suggests that traditional training-focused approaches successfully build individual competencies but struggle to strengthen organisational systems—a finding consistent with systematic reviews in health sector capacity strengthening (Bilardi et al., 2021). The 17-point gap between highest and lowest performing cases on organisational domains indicates that systems investments, rather than individual training, differentiate institutional effectiveness.

Mechanisms and Theoretical Implications

The qualitative findings illuminate why capacity development produces heterogeneous outcomes across contexts. Institutional incentive structures emerged as the primary mechanism translating individual skills into organisational performance. As one senior civil servant explained, "when promotion criteria link to results on national priorities, people use the training—otherwise it's just a

certificate." This aligns with organisational theory emphasising that formal incentives shape routine behaviour (OECD, 2021) but extends this understanding to development contexts where career systems may not reward technical excellence.

Donor coordination patterns provide a second explanatory mechanism. Fragmented, short-term projects appeared to undermine capacity consolidation, with respondents describing confusion and administrative burden from competing frameworks. Conversely, coordinated support through joint planning and shared indicators enhanced both efficiency and effectiveness, just as in Mpofu & Govender (2022). This finding supports recent arguments for harmonised approaches in development cooperation (Kilmarx et al., 2023) while providing empirical evidence for the mechanisms through which coordination influences outcomes.
Local ownership represents the most theoretically significant finding. The statistical interaction suggests that capacity investments yield larger returns when local actors control priority-setting and implementation approaches. This challenges assumptions that technical capacity is context-neutral, instead indicating that power relations and agenda-setting processes fundamentally shape how competencies translate into institutional outcomes. This aligns with critiques of technocratic development approaches (Tilouche et al., 2024) while providing quantitative evidence for the moderating role of ownership.

Study Limitations
Several limitations constrain the generalisability of these findings. The purposive sampling approach across four cases prioritised depth over statistical representativeness, limiting extrapolation to other institutional contexts or development sectors. The cross-sectional design precludes causal inference; while the statistical associations are robust to various specifications, unmeasured confounders may explain the capacity-implementation relationship. Longitudinal data would strengthen causal claims and illuminate temporal dynamics in capacity development processes.

Measurement limitations affect both indices. The Capacity Index relies on self-reported competencies, potentially introducing social desirability bias. The SDG Implementation Index depends on administrative data quality, which varies across contexts. While sensitivity analyses suggest robust main findings, composite indices inevitably simplify complex, multi-dimensional phenomena. Future research should explore disaggregated measures and objective performance indicators where available.

The qualitative component, though theoretically saturated, reflects perspectives from 20 purposively selected respondents. Power dynamics may have influenced interview responses, particularly regarding donor-recipient relationships and internal institutional critiques. Member checking and triangulation partially addressed these concerns, but broader stakeholder perspectives would enhance credibility.

Implications and Future Research
These findings suggest three practical implications for capacity development programming. First, organisations should sequence systems investments alongside individual training rather than treating them as separate interventions. The domain analysis indicates that data systems and regulatory capacity require sustained attention and may benefit from different implementation approaches than individual competency development.

Second, donors should prioritise coordination mechanisms over individual project delivery. The qualitative evidence suggests that harmonised approaches reduce transaction costs while enhancing implementation effectiveness. This implies institutional changes in how development organisations design and manage capacity development portfolios.

Third, local ownership should be operationalised through governance structures rather than treated as a rhetorical commitment. The 39% amplification effect indicates substantial returns to inclusive priority-setting and equitable partnership arrangements. However, this requires addressing power imbalances that may constrain authentic local leadership.

Conclusion

Methodologically, future studies should employ longitudinal designs to examine capacity development trajectories and causal mechanisms. Randomised controlled trials of coordination interventions could provide stronger causal evidence for donor harmonisation effects. Mixed methods approach remains essential for understanding how technical capacity translates into institutional outcomes through political and social processes that quantitative measures may obscure.

Acknowledgements
The authors wish to express sincere appreciation to all individuals and organisations whose support contributed to the successful completion of this research. Special gratitude is extended to those who provided guidance, critical feedback, and logistical assistance, as well as to all participants who generously shared their knowledge and perspectives during data collection. Their contributions were invaluable in shaping the quality and depth of this study.

Competing Interests
The author declares that there are no financial or personal relationships that may have inappropriately influenced the research or the preparation of this manuscript.

Author Contributions
The authors conceptualised and designed the study, conducted data collection and analysis, interpreted the results, and wrote the manuscript. Full responsibility for the integrity and accuracy of the work rests solely with the author.

Funding
This research did not receive any external funding. All resources and support necessary for its completion were provided by the authors or affiliated institutions.

Data Availability
The datasets generated and analysed during this study are available from the author upon reasonable request. All data management procedures adhered strictly to ethical guidelines to ensure confidentiality, integrity, and security.

Disclaimer
The views and opinions expressed in this work are those of the authors and do not represent the official positions of any institution, organisation, or funder.

6. References

1. Bilardi, D., Rapa, E., Bernays, S., & Lang, T. (2021). Measuring research capacity development in healthcare workers: A systematic review. BMJ Open, 11(7), e046796. https://doi.org/10.1136/bmjopen-2020-046796
2. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
3. Cochran, W. G. (1977). Sampling techniques (3rd ed.). John Wiley & Sons.
4. Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approach (4th ed.). SAGE.
5. Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE.
6. Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—principles and practices. Health Services Research, 48(6 Pt 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117
7. Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.). SAGE.
8. Fukuda-Parr, S. (2016). From the Millennium Development Goals to the Sustainable Development Goals: Shifts in purpose, concept, and politics of global goal setting for development. Gender & Development, 24(1), 43–52. https://doi.org/10.1080/13552074.2016.1145895
9. Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. SAGE.
10. Guest, G., Namey, E., & Chen, M. (2020). A simple method for assessing data saturation in qualitative research. PLoS ONE, 15(5), e0232076. https://doi.org/10.1371/journal.pone.0232076
11. Heerdegen, A. C. S. (2020). Transforming capacity strengthening in an era of sustainable development. International Journal of Public Health, 65(9), 1519–1520. https://doi.org/10.1007/s00038-020-01490-4
12. J-PAL / Poverty Action Lab. (n.d.). Working with a third-party survey firm: Guidance.
13. Kacou, K. P. (2022). Fifty years of capacity building: Taking stock and moving forward. Public Administration and Development.
14. Kilmarx, P. H., et al. (2023). Increasing effectiveness and equity in strengthening health research capacity using data and metrics: Recent advances of the ESSENCE mechanism. Annals of Global Health, 89(1), Article 38. https://doi.org/10.5334/aogh.3948

15.Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. Educational and Psychological Measurement, 30(3), 607–610. https://doi.org/10.1177/001316447003000308
16.Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. SAGE.
17.Mpofu, L., & Govender, K. (2022). Donor–NGO partnerships in South Africa: A qualitative case study of five NGOs in Gauteng. *African Evaluation Journal*, *10*(1), 619.
18.Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. International Journal of Qualitative Methods, 16, 1–13. https://doi.org/10.1177/1609406917733847
19.O'Cathain, A., Murphy, E., & Nicholl, J. (2008). The quality of mixed methods studies in health services research. Journal of Health Services Research & Policy, 13(2), 92–98. https://doi.org/10.1258/jhsrp.2007.007074
20.Organisation for Economic Co-operation and Development. (2016). Better policies for sustainable development. OECD. https://doi.org/10.1787/9789264256996-en
21.Organisation for Economic Co-operation and Development. (2021). Civil service capacities in the SDG era: An assessment framework (OECD Working Papers on Public Governance No. 47). OECD Publishing. https://doi.org/10.1787/a20bad7c-en
22.Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Administration and Policy in Mental Health and Mental Health Services Research, 42(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y
23.Pallant, J. (2020). SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS (7th ed.). Routledge.
24.Patton, M. Q. (2015). Qualitative research & evaluation methods (4th ed.). SAGE.
25.Tilouche, N., Kalichman, B. D., Dheensa, S., et al. (2024). Research capacity strengthening methods and meanings: Negotiating power in a global health programme on violence against women. BMJ Global Health, 9(12), e015376. https://doi.org/10.1136/bmjgh-2024-015376
26.Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care, 19(6), 349–357. https://doi.org/10.1093/intqhc/mzm042
27.UNCTAD. (2021). Productive Capacities Index (PCI). UNCTADstat. Retrieved from https://unctadstat.unctad.org/EN/Pci.html
28.United Nations, Department of Economic and Social Affairs. (2024). The Sustainable Development Goals report 2024. United Nations. https://unstats.un.org/sdgs/report/2024/
29.Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18
30.World Bank. (2020). Governance and institutions for the 2030 Agenda. World Bank Policy Brief.
31.Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE.