

A methodological framework for the global index of volunteer engagement (GIVE): Measuring volunteerism in official statistics

Statistical Journal of the IAOS

1–19

© The Author(s) 2026



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/18747655261486416

journals.sagepub.com/home/sji

Sage | IOS Press

Samuel Manda¹ , Nada Abdelatif², Harrid Nkhoma¹ , Kgothatso Mokutu¹,
Sollie Millard¹ and Tapiwa Kamuruko³

Abstract

Volunteerism drives sustainable development and social cohesion, yet official statistical systems fail to capture its full contribution. Existing metrics focus narrowly on participation rates and basic economic value, overlooking broader individual and societal benefits. This paper presents the methodological framework for the Global Index of Volunteer Engagement (GIVE), a composite index developed by the University of Pretoria, the United Nations Volunteers (UNV), and the International Labour Organisation (ILO). GIVE adopts a multidimensional approach across four equally weighted dimensions: Value to the Individual, Value to the Community, Value to the Economy, and the Enabling Environment. We outline the framework's conceptual foundation, data mapping, normalisation to a 0–100 scale, and its capacity to accommodate both survey and administrative data. To assess robustness, we conducted a Monte Carlo simulation study using 10 synthetic datasets representing heterogeneous country contexts and 500 replications, comparing nine specifications with different weighting and aggregation methods. Results indicate that country orderings remain robust across specifications. While arithmetic final aggregation generally minimised numerical error, with the GM+AM specification yielding the lowest mean absolute percentage error (5.22%), we recommend AM+GM aggregation, as it satisfies the commonly used 10% accuracy benchmark (MAPE = 9.64%) while retaining within-dimension indicator compensability and limiting cross-dimension substitutability. We discuss implications for National Statistical Offices, including integrating survey modules, aligning with SDG monitoring frameworks, and building capacity in lower-resource settings. Finally, we outline a Phase 2 roadmap (2027–2029) to roll out GIVE across 60–100 countries to establish a globally comparable evidence base.

Keywords

composite index, volunteer measurement, official statistics, data harmonisation, National Statistical Offices, SDG monitoring

Received: 22 June 2026; accepted: 24 August 2026

1 Introduction

Some of the most important contributions people make to society, especially volunteering and other community-sustaining actions, remain largely invisible in statistics and data systems. This presents a persistent paradox: society's most vital contributions often remain the least visible in official data – an insight that has markedly influenced how statisticians and researchers gauge social progress and development.¹ Across communities in the Global South and around the world, people dedicate their time, skills, and energy to helping others without expecting recognition or reward. From community-based conservation groups in Kenya, to volunteer firefighters in Australia, to youth organising neighbourhood clean-ups and immunisation campaigns across Latin America, these common acts of service – diverse in form yet integrated in purpose, form the backbone of societal resilience.¹ Volunteers respond to crises,

sustain public services, bolster social bonds, and build community endurance in both practical and meaningful ways. Yet despite its scale and significance, volunteerism remains under-researched and often absent from official statistical systems.^{1,2}

Thus, volunteerism is increasingly recognised as a strategic asset for sustainable development. The 2026 State of the World's Volunteerism Report (SWVR) estimates that 2.1 billion working-age individuals, with approximately one

¹Department of Statistics, University of Pretoria, South Africa

²Biostatistics Research Unit, South African Medical Research Council, Cape Town, South Africa

³Voluntary Advisory Services Section, United Nations Volunteers (UNV), Bonn, Germany

Corresponding author:

Samuel Manda, Department of Statistics, University of Pretoria, Hatfield, 0028, South Africa.

Email: samuel.manda@up.ac.za

in three adults globally, engage in volunteer work each month.² In Africa, the State of Volunteerism in Africa Report documents the highest regional participation rate at 58.5% of the working-age population, with youth (74%) and women (50.5%) constituting the majority of volunteer assignments.³ The recorded volunteer hours in Africa alone translate to an annual monetary value of USD 353 million; if the volunteers captured in that survey formed a nation, they would constitute the fourth-largest economy on the continent.³ Yet despite this scale, the full scope and impact of volunteer contributions remain poorly understood.

The measurement of volunteerism has a long history of collaborative international efforts to understand and quantify its contribution to individuals, communities, economies, and sustainable development. UNV's State of the World's Volunteerism Report 2011 provided an early global account of the contribution of volunteerism to well-being.⁴ This international agenda has continued to develop, with the United Nations Volunteers (UNV) and the International Labour Organisation (ILO) leading initiatives to create standardised tools for national data collection, including the ILO Manual on the Measurement of Volunteer Work and UNV's support for volunteer modules in national household surveys.^{5,6} This international agenda was further highlighted when the United Nations General Assembly proclaimed 2026 as the International Year of Volunteers for Sustainable Development (IVY2026) through Resolution A/RES/78/127.⁷ The 2026 State of the World's Volunteerism Report reflects the culmination of more than a decade of research, placing particular emphasis on the measurement of volunteerism and its contribution to sustainable development, and synthesising contributions from a global research consortium, including Northumbria University, the University of Pretoria, and the ILO, alongside input from volunteer practitioners, statisticians, and civil society organisations across all world regions (see, Chapter 1).² The report documents persistent evidence gaps, including fragmented data systems, inconsistent definitions, a bias toward formal volunteering, and the near-invisibility of informal, community-based, and digital volunteering (UNV, 2025, Chapters 2-3).²

One of the major challenges in developing a global framework for measuring volunteerism relates to the absence of universally harmonised definitions and the fragmented nature of available data systems. Existing surveys and statistical instruments differ substantially in how volunteer work is conceptualised, including variations in the treatment of formal versus informal volunteering, direct community support, digital volunteering, mutual aid, and unpaid civic participation.^{2,6,8} In addition, volunteer-related variables are collected across diverse platforms, including Labour Force Surveys, Time Use Surveys, the World Values Survey, Gallup World Poll, Afrobarometer, and organisation-specific surveys, each with different reference periods, recall durations, sampling designs, and levels

of disaggregation. These inconsistencies complicate cross-country comparability and often result in the underrepresentation of informal, community-based, and youth-led volunteer activities, particularly in lower-resource settings.^{2,8} As part of the GIVE development process, a comprehensive mapping of volunteer data sources was undertaken to identify the types of volunteer variables available, definitions used, data quality considerations, coverage gaps, and the extent to which existing datasets could support a harmonised global volunteer measurement framework.⁹

The framing of the Global Index of Volunteer Engagement (GIVE) emerged through an extensive and iterative consultative process involving the United Nations Volunteers (UNV), the International Labour Organization (ILO), academics, statisticians, National Statistical Offices, volunteer-involving organisations, and civil society stakeholders across multiple world regions. Early conceptual discussions on the need for a multidimensional global volunteer index began prior to 2025, informed by growing recognition of persistent gaps in volunteer statistics and increasing demand for evidence-based volunteer policymaking.⁸ A major milestone was the initial technical consultation convened by UNV and ILO in Bonn, Germany, in 2024, which brought together experts from Africa, Asia-Pacific, Latin America and the Caribbean, Europe, and the Arab States to deliberate on the conceptual dimensions, indicators, methodological considerations, and statistical feasibility of the index. This was followed by broader regional consultations and workshops aimed at ensuring that the framework reflected diverse volunteering realities, incorporated regional priorities and contextual forms of volunteering, assessed data availability, and strengthened global ownership of the index.^{2,9} The consultative process ultimately informed the identification of the four core dimensions of GIVE: Value to the Individual, Value to the Community, Value to the Economy, and the Enabling Environment Chapter 7, Section 7.3.²

Incorporating volunteer statistics into official statistical systems and national accounts is essential for accurately capturing the full contribution of volunteer activity to economic and social development. The experience of New South Wales, Australia, illustrates the scale of what can otherwise remain invisible: volunteering was estimated to generate \$205.3 billion in social and economic value through approximately 1.3 billion hours of volunteer activity.¹⁰ Excluding such contributions from conventional economic accounts can therefore result in their systematic underrecognition and obscure the economic and social value and impact of volunteering, limiting the evidence available to strengthen and enhance its contribution to society and the economy. Thus, without systematic inclusion in official statistics, the socio-economic contribution of billions of people remains invisible, leading to poorly targeted development, labour, and resource allocation policies.

National Statistical Offices (NSOs) need robust, comparable volunteer data to value unpaid work within satellite accounts, to identify gaps in public service delivery where mutual aid substitutes for state provision, to monitor SDG progress on poverty, health, education, and gender equality, to design evidence-based labour market and social protection policies, and to measure social capital and community resilience.^{6,11}

The Global Index of Volunteer Engagement (GIVE) directly addresses this need by providing a standardised, multidimensional framework for integrating volunteer statistics into routine official statistical production – working with NSOs, aligning with ILO standards, and using household survey instruments such as Labour Force Surveys and Time Use Surveys. The conceptual framework for GIVE was developed through extensive regional consultations with volunteer practitioners, statisticians, index developers from UN agencies, academics, and representatives from Africa, Asia-Pacific, Latin America and the Caribbean, Europe, and the Arab States Chapter 7.² These consultations confirmed strong support for a global measurement tool and clarified the four dimensions that form the core of the GIVE: Value to the Individual, Value to the Community, Value to the Economy, and the Enabling Environment Chapter 7, Section 7.3.²

A central aim of this paper is to introduce an innovative and timely standardised volunteerism measurement tool to national governments, stakeholders, and international partners. The GIVE index tool can be applied consistently across countries. To ensure cross-country comparability, NSOs align their methods with guidelines set by the ILO.⁵ Integration is primarily achieved through three instruments: Labour Force Survey (LFS) modules, which add standardised survey questions on time spent in unpaid, non-compulsory activities; Satellite Accounts, which integrate Non-Profit Institution data to explicitly value and detail the sectors benefiting from volunteer efforts;¹² and Time Use Surveys (TUS), which track daily time allocation to capture informal, unorganised, and direct volunteering between individuals.¹³ By embedding volunteer statistics within these established statistical instruments, GIVE enables the production of robust, comparable, and policy-relevant data – making volunteerism visible, valued, and actionable for evidence-based decision-making.

The remainder of this paper is organised as follows. Section 2 details the four dimensions of the GIVE framework. Section 3 presents the methodology: harmonisation, normalisation, weighting, and aggregation. Section 4 reports sensitivity analyses using synthetic datasets. Section 5 discusses implementation for NSOs and the role of key stakeholders. Section 6 concludes with alignment to SDG monitoring, the International Year of Volunteers 2026, and future directions.

2 The Four Dimensions of GIVE

The GIVE framework captures the multidimensionality of volunteering across four equally weighted dimensions: Value to the Individual, Value to the Community, Value to the Economy, and the Enabling Environment. These four dimensions were identified and refined through extensive consultation with volunteer participants and researchers and form the initial conceptual structure of the GIVE Index (SWVR, 2026). Together, they reflect the principal ways in which volunteer engagement generates value, encompassing benefits to individuals and communities, economic contributions, and the broader conditions that enable and sustain volunteering.

2.1 Value to the Individual and Community

The Value to the Individual dimension captures the personal benefits that volunteers derive from their engagement, comprising six indicators: skills acquisition, life satisfaction, sense of purpose, social interactions, mental/physical well-being, and improved employment prospects.^{14–19} Complementarily, the Value to the Community dimension captures the social and collective benefits generated through volunteering, evaluated across five indicators: trust, resilience, mutual support, social cohesion, and feeling safe and secure.^{14,20–22}

2.2 Value to the Economy

The *Value to the Economy* dimension captures the economic value generated through volunteer work. Volunteerism contributes to economies not only through the provision of unpaid labour but also through the opportunity costs borne by volunteers and the direct financial resources they contribute to volunteer activities. Existing approaches to valuing volunteer work have primarily focused on estimating the replacement cost of volunteer labour at the aggregate level.^{6,23} While these approaches provide valuable estimates of the macroeconomic contribution of volunteering, they provide limited insight into the economic commitment made by individual volunteers, including the opportunity costs associated with time devoted to volunteering. The GIVE framework therefore adopts a bottom-up approach in which economic indicators are measured at the individual level and subsequently aggregated to produce district, national, regional and global estimates.

To reflect the multidimensional nature of economic contribution, five illustrative indicators are proposed:

1. Average monthly volunteer hours per volunteer, measuring the intensity of volunteer engagement through the amount of time voluntarily contributed.
2. Relative Economic Sacrifice (RES), measuring the proportion of an individual's available economic resources voluntarily devoted to volunteering.

3. Estimated replacement value of volunteer time, representing the market value of volunteer labour calculated by applying an appropriate replacement wage rate to reported volunteer hours.
4. Proportion of volunteers reporting unreimbursed out-of-pocket expenses, capturing direct financial contributions such as transport, communication, equipment, childcare, accommodation, or meals incurred while volunteering.
5. Economic Commitment Index (ECI), measuring the extent to which volunteers invest both time and financial resources in volunteer activities.

The Relative Economic Sacrifice (RES) is introduced as a novel indicator within the GIVE framework. For volunteer i , RES is defined as

$$RES_i = \frac{w_i h_i + c_i}{Y_i}, \quad (1)$$

where h_i denotes the number of volunteer hours contributed during the reference period, w_i denotes the estimated hourly opportunity cost of time, c_i denotes unreimbursed out-of-pocket volunteering expenses, and Y_i denotes the individual's disposable income or equivalent economic resources over the same period. The numerator therefore represents the total economic resources voluntarily contributed by an individual, while the denominator standardises this contribution relative to the individual's economic capacity.

2.3 Enabling Environment

The Enabling Environment dimension captures the structural, policy, and institutional context that supports volunteering in a country. Unlike the three individual-level dimensions, which are mainly derived from volunteer surveys, the Enabling Environment is a macro-level construct computed from country-level administrative data.²⁵ Based on the GIVE conceptual framework,¹⁴ this dimension comprises seven binary indicators, each indicating the presence or absence of a specific enabling factor.

3 Methodology for Composite Index Construction

3.1 Data Sources and Mapping

The construction of GIVE was informed by a systematic mapping of available data sources on volunteering and its associated outcomes. The objective of the mapping was to identify sources that could provide information relevant to the four GIVE dimensions while also assessing their geographic coverage, definitions, measurement approaches, and suitability for cross-country comparison. The sources considered included global surveys such as the World Values Survey, Afrobarometer, European Social Survey,

and Gallup World Poll; global indices such as the Charities Aid Foundation World Giving Index; regional sources including the Latin American Public Opinion Project and IFRC surveys; national instruments such as Time Use Surveys, Labour Force Surveys with volunteer modules, and country-specific household surveys; and complementary sources such as the UNV Global Volunteer Survey Chapter 7, Section 7.5.1.²

The mapping exercise identified substantial heterogeneity in how volunteering is defined, measured, and reported across these sources. Three challenges were particularly important for the development of GIVE. First, definitions of volunteering differ across surveys and statistical systems, affecting the comparability of participation and engagement measures. Second, existing data systems tend to be fragmented and often focus on formal, organisation-based volunteering, with less systematic coverage of informal and community-based activities. Third, important coverage gaps remain for emerging forms of volunteering, including digital and youth-led volunteering, as well as for disaggregated information by gender, age, disability, and geographic location Chapter 3.² These limitations motivated the harmonisation procedures described below and informed the selection and interpretation of indicators included in the GIVE framework.

3.2 Harmonisation Protocol

Because the identified data sources differ in definitions, reference periods, measurement instruments, and data quality, harmonisation is required before information from different sources can be used within a common index framework. GIVE therefore applies a four-step harmonisation protocol designed to improve comparability while maintaining a transparent record of the assumptions and transformations applied to each source.

In Step 1, relevant variables are identified and extracted, including measures of volunteering participation, frequency, duration, type of activity, motivation, and associated outcomes. Step 2 aligns the definitions and concepts across sources by mapping variables to the United Nations General Assembly definition of volunteering and the ILO statistical definition of volunteer work established by the 19th International Conference of Labour Statisticians (19th ICLS). Definitions that cannot be directly reconciled are documented in a harmonisation codebook rather than being implicitly treated as equivalent.^{5,6}

Step 3 addresses differences in reference periods. Where volunteering rates are reported over different observation periods, we convert them to four-week equivalents using adjustment coefficients derived from average volunteer rates across the relevant reference periods Appendix B.² This provides a common temporal basis for comparison while retaining information from the original data sources.

Finally, Step 4 assesses each dataset's quality and consistency by considering coverage, methodology, periodicity, and internal consistency. Data points are excluded where direct volunteering rates exceed corresponding total rates or where pronounced discontinuities suggest methodological changes rather than genuine changes in volunteering. The harmonisation process therefore serves not only to make the data comparable, but also to provide a transparent basis for documenting data limitations and determining which observations are sufficiently reliable for inclusion in the GIVE database. The resulting harmonisation codebook will be published as a public good alongside the GIVE database.

3.3 Normalisation

As the GIVE index combines indicators measured on different scales and, in some cases, with different units of measurement, the indicators must first be transformed to a common scale before they can be combined. Normalisation ensures that differences in the numerical scale of individual indicators do not inadvertently determine their contribution to the composite index. This is particularly important for a multidimensional index such as GIVE, where indicators representing individual, community, economic, and enabling-environment outcomes may have substantially different ranges and units.

Several approaches can place indicators on a common scale. Common methods include Z-score standardisation, which centres and scales observations by their mean and standard deviation, and Min-Max normalisation, which transforms observations to a specified interval. The choice of normalisation method is therefore an important component of composite-index construction because it can influence the relative contribution of individual indicators and, consequently, the resulting index values.²⁶

For GIVE, we use a 0–100 normalisation, whereby each indicator is transformed to a common scale ranging from 0 to 100. This scale is particularly appropriate for GIVE because it provides an intuitive interpretation of the resulting scores and facilitates comparison across indicators and dimensions. A score of 0 represents the lower end of the defined indicator scale, while 100 represents the upper end. Importantly, the common scale allows indicators with different original units to contribute to the composite index without one indicator dominating solely because of its numerical magnitude.

The use of Min-Max normalisation, however, requires some caution because the resulting values are sensitive to the observed minimum and maximum. In particular, an extreme observation can substantially affect the transformed values of the remaining observations. For example, if 99 observations lie between 0 and 40 while one observation has a value of 100, a conventional Min-Max transformation would place the 99 observations within a relatively compressed portion of the 0–100 scale. Thus,

although Min-Max normalisation provides a convenient common scale, its sensitivity to extreme values needs to be considered when constructing a composite indicator.²⁶

3.4 Weighting

As the GIVE dimensions are constructed from multiple indicators, a weighting and within-dimension aggregation scheme is required to determine how the indicators contribute to each dimension score. GIVE comprises four dimensions, indexed by $d = 1, \dots, 4$, with potentially different numbers of indicators contributing to each dimension. In the baseline specification, indicators receive equal weights within their respective dimension. Thus, for dimension d containing k_d indicators, each indicator receives weight

$$w_{dj} = \frac{1}{k_d}.$$

In the simulations described in Section 3.5 below, we use $k_1 = 6$ indicators for Value to the Individual, $k_2 = 5$ for Value to the Community, $k_3 = 6$ for Value to the Economy, and $k_4 = 7$ for the Enabling Environment. Equal weighting provides a transparent, reproducible approach and avoids imposing empirically unsupported assumptions that one indicator should matter more than another.

In practice, however, indicators within a dimension may not have equal substantive importance. Social, psychological, economic, or other theoretical considerations may suggest that some indicators are more important than others in representing the underlying dimension. In such cases, the weighting scheme should reflect the differential importance assigned to the indicators, with weights determined based on an explicit conceptual framework, empirical evidence, or a combination of both. Thus, equal weighting provides a transparent baseline rather than assuming all indicators are equally important. This approach is also consistent with established practice in constructing multidimensional indices, while allowing alternative data-driven weighting schemes to be considered as sensitivity analyses.^{26,27}

In the current GIVE specification, equal weighting serves as the baseline because the available evidence does not provide a sufficiently strong basis for assigning differential substantive weights to individual indicators. In addition to equal weighting, the equally weighted indicators within each dimension can be combined using either the arithmetic mean (AM) or geometric mean (GM). For normalised indicator values $z_{d1}, \dots, z_{dk_d}$ within dimension d , the two alternatives are

$$AM_d = \frac{1}{k_d} \sum_{j=1}^{k_d} z_{dj}, \quad GM_d = \left(\prod_{j=1}^{k_d} z_{dj} \right)^{1/k_d}.$$

The arithmetic mean permits full compensation among indicators within a dimension, so that a higher value for one

indicator can offset a lower value for another. The geometric mean provides a less compensatory alternative because the dimension score is more sensitive to low indicator values and therefore places greater emphasis on balanced performance. These formulations allow examination of how sensitive dimension scores are to alternative assumptions about combining indicators.

To assess how much the construction of GIVE depends on these choices, the simulation considers four within-dimension approaches: arithmetic-mean weighting (AM), geometric-mean weighting (GM), entropy weighting, and principal component analysis (PCA). AM and GM provide transparent equal-weight formulations, whereas entropy weighting and PCA derive weights from the statistical properties of the reference data.

Entropy weighting. The entropy method is the most representative objective weighting tool. This method determines the relative importance of the attributes and how they are related to the evaluation or final result. In the standard Entropy Weight Method (EWM) used in multi-criteria decision-making, we explicitly reward indicators that exhibit high variation (differentiation), because they provide the most useful information for distinguishing between respondents.²⁸ For normalised indicator values z_{ij} , define the empirical probabilities $p_{ij} = z_{ij} / \sum_i z_{ij}$. The normalised entropy is

$$e_j = -\frac{1}{\ln(n)} \sum_{i=1}^n p_{ij} \ln(p_{ij}),$$

with $0 \ln 0 = 0$. The degree of diversification is the complement of entropy, $d_j = 1 - e_j$, and the resulting entropy weight is

$$w_j^{\text{Ent}} = \frac{d_j}{\sum_{k=1}^K d_k}.$$

As shown above, entropy weighting assigns greater weight to indicators that provide more discriminatory information in the observed indicator values, based purely on their statistical distribution rather than on their conceptual or normative significance.^{28–31} Indicators with high top-tier probabilities tend to have responses heavily concentrated in the upper Likert categories, showing less variation within the country. Such indicators provide less information for differentiating observations and consequently tend to receive lower entropy weights. Conversely, indicators with more dispersed responses provide greater discriminatory information and receive higher weights. Thus, an indicator with highly diverse responses within a country may exhibit greater entropy-related discriminatory information and command a larger weight, whereas an indicator with highly concentrated responses will provide less information for differentiation and receive a smaller weight. This feature distinguishes statistical differentiation from substantive

importance: an indicator judged to be universally important may receive a relatively low entropy-based weight if its observed responses are highly concentrated, even though its conceptual or policy relevance remains substantial.

In our simulation, we specify the top-tier probability separately for each indicator and country. Once the country-specific top-tier probability is set, it determines the response distribution for that indicator within the country and, consequently, the entropy-based weight assigned to that indicator for that country. The resulting weights are therefore country-dependent: an indicator may receive a relatively high weight in one country, where its response distribution provides greater discriminatory information, but a lower weight in another country, where responses are more concentrated or otherwise provide less information for differentiation. Thus, under entropy weighting, an indicator's contribution to the index is not fixed across countries but depends on the empirical distribution generated for each country. This purely data-driven characteristic underscores why equal arithmetic weighting remains an important baseline in GIVE, ensuring that the conceptual framework determines the relative importance of dimensions and indicators rather than allowing observed statistical variation to determine their weights. In the present simulation study, entropy weighting therefore serves as a statistical alternative for sensitivity analysis, allowing us to assess how country-specific differences in indicator distributions affect index values and the stability of country rankings.

Principal component analysis (PCA). : Principal component analysis provides a different data-driven approach. PCA transforms correlated indicators into a smaller set of orthogonal components that successively explain variation in the original indicators. In this application, PCA serves as a weighting procedure, using the first principal component to derive indicator weights because it explains the largest share of variation among the indicators.^{26,32,33} If ℓ_j denotes the loading of indicator j on the first principal component, the weight is

$$w_j^{\text{PCA}} = \frac{\ell_j^2}{\sum_{k=1}^K \ell_k^2}.$$

Squaring the loadings produces non-negative weights, which are then normalised to sum to one. The resulting PCA weights therefore reflect each indicator's relative contribution to the dominant common pattern of variation in the reference data.

Clarification on the role of entropy and PCA in the GIVE framework: Unlike the arithmetic and geometric mean approaches, which differ mainly in their compensatory properties at the aggregation stage, entropy and PCA are weighting methods that determine indicator importance based on the data's statistical properties. Entropy rewards

indicators with high variation, while PCA rewards indicators that contribute most to the dominant shared pattern of variation. Both methods are used exclusively as *weighting* techniques within their respective dimensions; they do not, by themselves, define how indicators are aggregated. In both cases, the derived weights, whether w_j^{Ent} or w_j^{PCA} , are applied to the normalised indicator values using a weighted arithmetic sum to produce the dimension score:

$$D_d = \sum_{j=1}^{k_d} w_j z_{dj}, \quad \text{where } \sum_{j=1}^{k_d} w_j = 1.$$

Thus, entropy and PCA determine which indicators contribute more or less to the dimension score, but they do not replace the subsequent aggregation step. Once these weights are applied, the resulting dimension scores can be combined using either the fully compensatory arithmetic mean or the less compensatory geometric mean in the final aggregation stage. Therefore, the choice between AM and GM determines the degree of compensation, irrespective of whether equal, entropy, or PCA weights are used within the dimensions. The final combination of the four dimension scores into the overall GIVE index is always governed by the second part of the specification (AM or GM), as described in the following section.

3.5 Aggregation

To clarify the notation used throughout this paper and in the simulation study, each method is denoted as $\mathbf{X} + \mathbf{Y}$, where:

- **The first part (X)** specifies the *within-dimension* procedure used to combine the individual indicators into each of the four dimension scores.
 - **AM:** Indicators within the dimension receive equal weights ($1/k$) and are combined using the arithmetic mean.
 - **GM:** Indicators within the dimension receive equal weights ($1/k$) and are combined using the geometric mean.
 - **Entropy:** Weights are derived empirically using the Entropy Weight Method (rewarding indicators that exhibit greater variation across respondents), and indicators are combined using a weighted arithmetic sum.
 - **PCA:** Weights are derived empirically using Principal Component Analysis (rewarding indicators that contribute most to the dominant pattern of shared variation), and indicators are combined using a weighted arithmetic sum.
- **The second part (Y)** specifies the *final aggregation* rule used to combine the four dimension scores (Individual, Community, Economic, and Enabling Environment) into the final GIVE index.

- **AM:** The four dimension scores are combined using the arithmetic mean, which is fully compensatory.
- **GM:** The four dimension scores are combined using the geometric mean, which is less compensatory and penalises unbalanced performance across dimensions.

For example, **AM+GM** denotes equal weighting and arithmetic aggregation within each dimension, followed by geometric aggregation across the four dimensions to produce the final GIVE score. This is the primary specification recommended in this paper. Similarly, **Entropy+AM** denotes entropy-derived weights applied within each dimension, followed by arithmetic aggregation across dimensions. We use this notation consistently throughout the simulation results and sensitivity analyses.

As the four GIVE dimension scores are constructed separately, a final aggregation rule is required to obtain a single overall measure of volunteer engagement. For the final aggregation stage, the four separate dimension scores ($D1, \dots, D4$) are combined into a single overall measure of volunteer engagement using either an arithmetic mean (AM) or a geometric mean (GM) formulation, applying the same mathematical definitions established in Section 3.4. The choice between these final aggregation rules matters significantly because it determines how much performance in one dimension can compensate for lower performance in another. In particular, the arithmetic mean and geometric mean represent different assumptions about the relationship among the four dimensions and therefore provide useful alternative formulations for evaluating the robustness of the GIVE index.^{26,34–37} In particular, the arithmetic mean is fully compensatory: a sufficiently high score in one dimension can offset a corresponding deficit in another. The geometric mean is less compensatory because the overall score is more sensitive to low-performing dimensions, thereby placing greater emphasis on balanced performance across dimensions. Geometric aggregation does not eliminate substitution altogether, but it limits the extent to which exceptionally high performance in one dimension can compensate for poor performance in another.

This distinction is substantively relevant to GIVE because the four dimensions, namely Value to the Individual, Value to the Community, Value to the Economy, and the Enabling Environment, capture complementary aspects of volunteering. A country may perform strongly in one dimension while having substantially weaker outcomes in another. An aggregation rule that permits unrestricted compensation could therefore mask important weaknesses in particular dimensions. Conversely, an overly restrictive approach could make the overall index excessively sensitive to a single low score. Considering both final aggregation alternatives therefore allows the simulation to assess the

implications of these assumptions rather than treating one aggregation rule as inherently correct.

The choice is also informed by established international composite indices. The modern Human Development Index (HDI) replaced arithmetic aggregation with geometric aggregation in 2010, specifically to reduce substitutability across its dimensions.^{38,39} In contrast, the Sustainable Development Goals (SDG) Index uses equal weights and arithmetic means to construct goal-level scores and to aggregate the 17 goals, providing an important precedent for an AM+AM structure.⁴⁰ The Global Innovation Index (GII) similarly relies primarily on weighted arithmetic aggregation to construct its sub-pillars and pillars and to combine its Input and Output sub-indices; its statistical audit also examines the robustness of the arithmetic-average assumption against alternative aggregation approaches.⁴¹ The Multidimensional Poverty Index (MPI) provides a complementary precedent through explicit indicator weights and the Alkire–Foster identification of joint deprivation, although its mathematical construction differs from that of GIVE.^{27,42} These examples demonstrate that both arithmetic and geometric aggregation are established approaches to multidimensional index construction, with the appropriate choice depending on the substantive meaning assigned to the dimensions.

To assess the sensitivity of GIVE to these alternative methodological choices, the simulation combines the two final aggregation alternatives with the four within-dimension approaches: AM, GM, entropy weighting, and principal component analysis (PCA). This yields eight hierarchical specifications. A ninth flat, equal-weight specification is included as a benchmark, in which all underlying indicators contribute directly to the overall index rather than first being organised into the four GIVE dimensions. The resulting nine specifications allow the numerical performance and country rankings of the GIVE index to be examined across a broad range of reasonable methodological choices.

The simulation results subsequently inform the selection of the primary specification. In particular, the choice of AM+GM is based not on the assumption that geometric aggregation is universally superior, but on its strong empirical performance, stable country rankings, conceptual transparency, and ability to limit compensability across the four complementary GIVE dimensions. AM+AM remains the principal additive sensitivity benchmark, while GM+AM, GM+GM, entropy, PCA, and the flat specification are retained to assess methodological sensitivity.

Figure 1 summarises the four GIVE dimensions, the illustrative indicators associated with each dimension, and the data sources used to measure them.

3.6 Methodology and Simulation Design

The simulation study assessed how well alternative GIVE weighting and aggregation specifications recover known country-specific population values under realistic sampling variation and between-country heterogeneity. The final simulation comprised 10 countries and 500 Monte Carlo replications. In each replication, the Individual and Economic dimensions were generated for the same 500 volunteers, while the Community dimension was generated from a separate sample of 2,000 community members. The Enabling Environment was generated at the country level using seven binary indicators. This design reflects GIVE's intended structure while allowing us to evaluate competing index-construction rules under controlled conditions.

3.6.1 Empirically motivated global benchmarks for ordinal indicators. For each of the six Individual and five Community ordinal indicators, we specified a global top-tier probability

$$d_j = P(Y_j \geq 4),$$

where responses of 4 or 5 represent a positive response. We assigned the individual indicators to empirically motivated performance tiers. Sense of purpose and mental/physical well-being were assigned the highest global benchmarks (0.94 and 0.92, respectively), social interactions and skills acquisition were assigned intermediate values (0.83 and 0.80), while life satisfaction and improved employment prospects were assigned lower values (0.78 and 0.65). These values reflect the asymmetric and generally positive distribution of reported volunteer outcomes rather than a uniform distribution. In particular, the NCVO *Time Well Spent* study reports high overall satisfaction among volunteers and substantial reported benefits to wellbeing, supporting high but indicator-specific positive-response benchmarks.⁴³

For the Community dimension, mutual support and social cohesion were assigned the highest benchmarks (0.85 and 0.82), feeling safe and secure an intermediate value (0.72), and trust and resilience lower values (0.68 and 0.64). These values were intended as plausible empirical benchmarks rather than universal population parameters. The tiering reflects the expectation that immediate interpersonal and collective experiences may respond more directly to community activity, whereas institutional trust and resilience are slower-moving and more context-dependent constructs. This is consistent with the broader social-capital literature on networks, reciprocity, trust and community cohesion.^{44,45}

For each country c , a country-specific top-tier prevalence was then generated as

$$d_{jc} \sim \text{Uniform} \left[\max(0.10, d_j - 0.20), \min(0.90, d_j + 0.20) \right], \quad c = 1, \dots, 10.$$

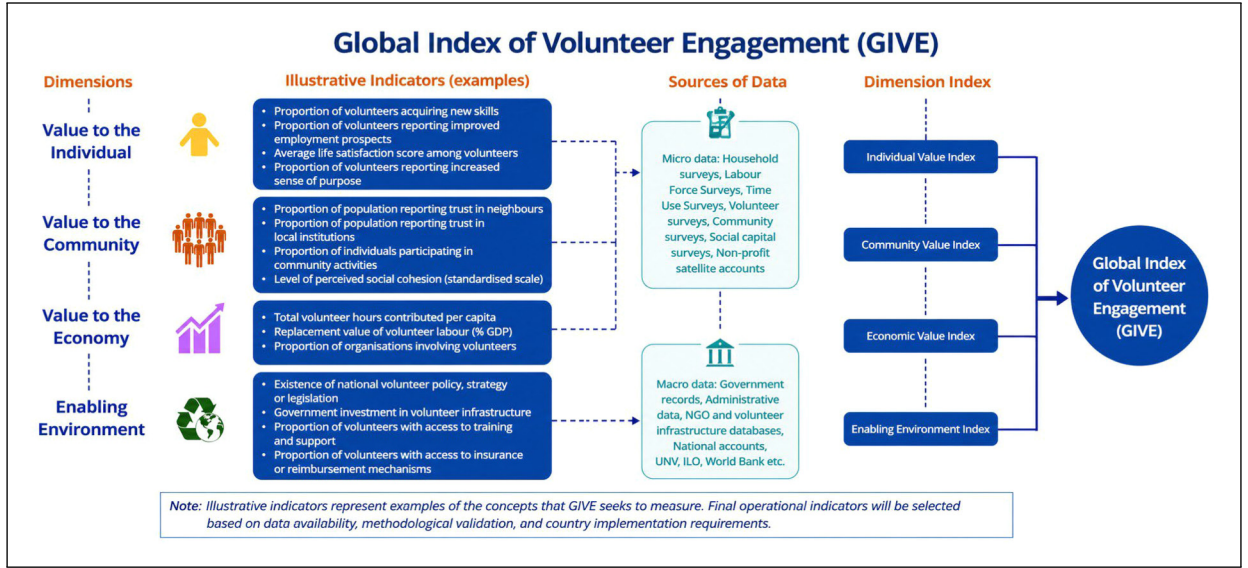


Figure 1. Global Index of Volunteer Engagement (GIVE): Dimensions, Illustrative Indicators, and Data Sources. The illustrative indicators represent examples of the concepts that GIVE seeks to measure. Final operational indicators will be selected based on data availability, methodological validation, and country implementation requirements.

Table 1. Global top-tier probabilities used in the simulation.

Dimension	Indicator	$d_j = P(Y \geq 4)$
Individual	Sense of purpose	0.94
Individual	Mental/physical well-being	0.92
Individual	Social interactions	0.83
Individual	Skills acquisition	0.80
Individual	Life satisfaction	0.78
Individual	Improved employment prospects	0.65
Community	Mutual support	0.85
Community	Social cohesion	0.82
Community	Feeling safe and secure	0.72
Community	Trust	0.68
Community	Resilience	0.64

Each resulting d_{jc} was generated once and held fixed across all 500 replications. Thus, the global value defines the empirical centre of the simulation, while each country has its own underlying population prevalence. The corresponding ordinal probabilities were

$$P(Y_{ijc} = r) = \frac{1 - d_{jc}}{3}, \quad r = 1, 2, 3,$$

$$P(Y_{ijc} = r) = \frac{d_{jc}}{2}, \quad r = 4, 5.$$

The cumulative probabilities determine country- and indicator-specific latent-normal cut-points. Consequently, repeated Monte Carlo samples represent repeated samples from the same underlying country populations rather than repeated regeneration of the population parameters Table 1.

3.6.2 Correlation structure for the Individual and Community dimensions. Because the indicators within the Individual and Community dimensions are conceptually related, the simulation allows for positive dependence among indicators rather than treating them as independent outcomes. The dependence structure is introduced at the latent continuous level. For each country c , a country-specific equicorrelation parameter was generated once and then held fixed across the Monte Carlo replications. For the six Individual indicators, the correlation parameter was drawn as

$$\rho_{I,c} \sim \text{Uniform}(0.3, 0.8),$$

whereas for the five Community indicators,

$$\rho_{C,c} \sim \text{Uniform}(0, 0.5).$$

Thus, countries differ in the strength of dependence between indicators, while their underlying correlation structure remains fixed across replications. The ranges represent plausible heterogeneous dependence structures rather than fixed empirical correlation coefficients.

For each country and replication, the latent indicator vector was generated jointly from a multivariate normal distribution with zero means and the corresponding country-specific correlation matrix. For the Individual dimension, for example,

$$\mathbf{Z}_{cr} = (Z_{1cr}, \dots, Z_{6cr})^T \sim N_6(\mathbf{0}, \mathbf{\Sigma}_{I,c}),$$

where $\mathbf{\Sigma}_{I,c}$ has ones on the diagonal and $\rho_{I,c}$ in all off-diagonal elements. An analogous construction was used

for the five Community indicators using $\Sigma_{C,c}$. This corresponds directly to the simulation implementation, in which the latent variables are generated jointly using `mvrnorm()` with the specified correlation matrix.⁴⁶

The latent variables were subsequently transformed into the observed five-category Likert responses. Importantly, the category cut-points were determined from the indicator-specific probability distributions rather than being fixed across replications. For indicator j in country c and replication r , let

$$\mathbf{p}_{jcr} = (p_{jcr1}, \dots, p_{jcr5})$$

denote the replication-specific probabilities of the five Likert categories. The corresponding cumulative probabilities define the latent normal cut-points

$$b_{jcr,k} = \Phi^{-1} \left(\sum_{\ell=1}^k p_{jcr\ell} \right), \quad k = 1, \dots, 4,$$

where $\Phi^{-1}(\cdot)$ is the inverse standard normal distribution function. The observed Likert response is then obtained by assigning Z_{jcr} to the corresponding interval defined by these cut-points.

This construction separates two sources of variation in the simulation. The country-specific correlation parameter and the underlying country-level probability structure define the population setting, whereas the replication-specific latent probabilities and the resulting volunteer-level draws generate Monte Carlo sampling variation. Consequently, each replication produces a different sample of volunteers while retaining the same underlying country population truth. The correlation is therefore not merely specified as a parameter of the simulation; it is transmitted through the jointly generated latent variables into the observed indicators.

For the Individual dimension, the resulting six indicators were observed on a five-point Likert scale and subsequently transformed to the common 0–100 scale using

$$X_{jcr}^* = 100 \left(\frac{X_{jcr} - 1}{5 - 1} \right).$$

The same latent-correlation construction was used for the five Community indicators, with a separate sample of 2,000 community members per country and replication. The resulting country-level indicator means were then used in the subsequent weighting and aggregation procedures.

3.6.3 Economic dimension. We generated the economic dimension at the individual level using the same 500 volunteers who contributed the Individual data. We retained four indicators: monthly volunteer hours, individual hourly rate, replacement value of volunteer time, and Relative Economic Sacrifice (RES). Monthly hours were generated from a right-skewed Gamma distribution with shape 2 and scale

3, with a latent normal component introduced to generate dependence with the other economic measures. Country-specific hourly rates were generated once for each of the 10 countries from a Gamma distribution with mean USD 22 per hour and then bounded to the range USD 15–25 per hour. For each country, the 500 volunteers subsequently received individual hourly rates drawn uniformly around that country-specific rate, subject to the same USD 15–25 bounds. Replacement value was calculated as

$$r_{ic} = h_{ic} w_{ic},$$

where h_{ic} is volunteer hours and w_{ic} is the individual hourly rate. RES was generated as an hours-dependent measure, averaging about 5% and bounded between 0.5% and 20%. We normalised the resulting economic indicators to a common 0–100 scale.

This construction explicitly distinguishes the country-level economic context from individual-level variation. The country rate is a fixed population characteristic across replications, while individual rates and hours vary between volunteers and between Monte Carlo samples. The economic indicators are therefore not generated as independent variables: replacement value is directly determined by hours and rate, while RES is positively related to hours. This induces the expected dependence among economic indicators and provides a more realistic basis for comparing weighting methods.

3.6.4 Simulation of Enabling Environment Indicators. The Enabling Environment dimension captures the structural, policy, and institutional context that supports volunteering within a country. Unlike the three individual-level dimensions—Value to the Individual, Value to the Community, and Value to the Economy—which are derived from volunteer surveys, the Enabling Environment is a macro-level construct based on country-level institutional and administrative conditions. Based on the GIVE conceptual framework,¹⁴ this dimension comprises seven binary indicators, each representing the presence or absence of a specific enabling factor.

Table 2 presents the seven Enabling Environment indicators and their estimated global proportions, derived from the UNV Knowledge Portal and relevant literature.⁴⁷ We used these global proportions as benchmark probabilities for the simulation, allowing the simulated indicators to reflect the prevalence of enabling conditions observed in the available empirical sources.

For each country i and indicator j , the binary enabling indicator was generated independently from a Bernoulli distribution using the corresponding global probability p_j :

$$E_{ij} \sim \text{Bernoulli}(p_j), \quad j = 1, \dots, 7,$$

where $E_{ij} = 1$ indicates the presence of the enabling condition and $E_{ij} = 0$ its absence. The same global benchmark

Table 2. Enabling environment indicators and estimated global proportions.

Indicator	Global Proportion	Source / Rationale
National Volunteer Policy	0.32	62 out of 194 countries have established national volunteer policies ⁴⁷
National Volunteer Legislation	0.43	83 out of 194 countries have enacted volunteer laws ⁴⁷
National Volunteer Strategy	0.34	At least 68 countries have stand-alone national strategies ⁴⁷
Government Budget Allocation	0.20	Countries with dedicated budget lines for volunteer infrastructure
Volunteer Access to Training	0.78	Surveys indicate 76.6–80.1% of volunteers have access to training ⁴⁸
Volunteer Access to Insurance	0.40	IFRC data shows varying coverage; global average approximately 40%
Volunteer Access to Reimbursement	0.53	Surveys indicate 53–55% of volunteers receive reimbursement ⁴⁸

probability was used for each country for a given indicator, while the Bernoulli realisation was allowed to vary across countries. Consequently, countries could have different realised configurations of the seven enabling indicators, despite sharing the same underlying global benchmark probabilities.

The Enabling Environment score for country i was then calculated as the proportion of the seven indicators realised as present, multiplied by 100:

$$\text{Enabling}_i = \frac{1}{7} \sum_{j=1}^7 E_{ij} \times 100.$$

Because the indicators are binary, the resulting country-level score takes values in increments of 100/7, ranging from 0 to 100. This construction reflects the interpretation of the Enabling Environment as a set of institutional conditions that may or may not be present in a country, rather than as a continuous individual-level outcome.

3.6.5 Country-level dimension scores and population true parameters. For the Individual and Community dimensions, the country-level estimate in each replication is the simple mean across the 500 volunteers or 2,000 community members, respectively. In a future official survey implementation, sampling weights would estimate population quantities; the present simulation intentionally uses equal-weight samples so the comparison isolates the effects of weighting and aggregation at the index-construction stage. For the Economic dimension, the same 500 volunteers provide all four economic indicators, which are aggregated at the country level in the same manner.

The true population-level GIVE values are determined analytically from the specified data-generating distributions and remain fixed across replications. Importantly, each weighting and aggregation specification has its own corresponding true population-level GIVE value. Thus, the accuracy of AM+AM, AM+GM, GM+AM, GM+GM, Entropy+AM, Entropy+GM, PCA+AM and PCA+GM is assessed against the population value generated under that same specification, rather than against a single arbitrary benchmark. The Flat specification has its own equal-weighted population target across all underlying indicators.

For arithmetic aggregation of the four dimensions,

$$\text{GIVE}_c^{\text{true,AM}} = \frac{1}{4} \sum_{d=1}^4 D_{dc}^{\text{true}},$$

whereas for geometric aggregation,

$$\text{GIVE}_c^{\text{true,GM}} = \left(\prod_{d=1}^4 D_{dc}^{\text{true}} \right)^{1/4}.$$

This method-specific truth construction is essential because it separates estimation error from differences caused simply by changing the definition of the target index.

3.6.6 Replication-specific latent probabilities and sampling variation. For each country i and indicator j , a country-specific latent probability p_{ij} is first generated from the specified top-tier probability distribution. This probability is treated as fixed across the Monte Carlo replications and defines the underlying country-level truth for indicator j . The corresponding true value of each dimension and, ultimately, the true GIVE index are therefore determined from these fixed probabilities.

To introduce sampling variation across Monte Carlo replications without repeatedly using essentially the same volunteer sample, we introduce a replication-specific latent probability, denoted by p_{ijr} , for country i , indicator j , and replication r . For each replication, p_{ijr} is generated from a distribution centred on the fixed country-specific probability p_{ij} . Thus,

$$p_{ijr} \sim F(\cdot | p_{ij}), \quad r = 1, \dots, R,$$

where $R = 500$ is the number of Monte Carlo replications. The replication-specific probability p_{ijr} determines the category probabilities used to generate the volunteer-level observations in replication r . Consequently, each replication produces a new set of 500 simulated volunteers, while the samples remain centred on the same underlying country-specific probability p_{ij} .

For example, where the five response categories are defined using the top-tier probability, the category probabilities for replication r are constructed from p_{ijr} rather than

directly from p_{ij} . The resulting volunteer responses therefore vary across replications because the latent cut-point varies slightly from one replication to another. This provides sampling variation while preserving a common underlying country-specific truths. This distinction separates two sources of variation in the simulation: the fixed country-specific latent probability p_{ij} defines the data-generating truth, whereas p_{ijr} introduces replication-specific sampling variation around that truth. The revised procedure therefore avoids repeatedly resampling essentially the same volunteer observations while maintaining a common population parameter against which the performance of the alternative GIVE specifications can be assessed.

3.6.7 Monte Carlo performance measures. Following the simulation strategy described in the preceding sections, we assessed the performance of the different weighting and aggregation methods using several well-known Monte Carlo performance measures.⁴⁹ Let $\hat{G}_{crm}$ denote the GIVE estimate for country c , replication r , and method m , and let G_{crm}^{true} denote the corresponding fixed population truth. As explained in Section 3.6.6, the population truth is fixed across replications, while each replication draws a new sample using a perturbed latent cut-point. The country-specific Monte Carlo bias is defined as

$$\text{Bias}_{cm} = \frac{1}{R} \sum_{r=1}^R (\hat{G}_{crm} - G_{crm}^{\text{true}}), \quad R = 500.$$

As described in the previous section, the population truth is fixed across replications, the bias could equivalently be written as

$$\text{Bias}_{cm} = \left(\frac{1}{R} \sum_{r=1}^R \hat{G}_{crm} \right) - G_{cm}^{\text{true}}.$$

In addition to bias, we evaluated model performance using two other metrics: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), which were calculated as follows:

$$\text{MAE}_{cm} = \frac{1}{R} \sum_{r=1}^R |\hat{G}_{crm} - G_{cm}^{\text{true}}|$$

and

$$\text{MAPE}_{cm} = \frac{100}{R} \sum_{r=1}^R \left| \frac{\hat{G}_{crm} - G_{cm}^{\text{true}}}{G_{cm}^{\text{true}}} \right|.$$

We report the overall summaries, which are obtained by averaging the corresponding performance measures across the 10 countries for each of the four GIVE dimensions and the overall GIVE score. Importantly, p_{ijr} is used only to generate the observed data within replication r . It is not used as the population truth for evaluating the performance of

the competing GIVE specifications. For every replication, the estimated GIVE value is compared with the fixed true value obtained from p_{ij} . Thus, bias, mean absolute error (MAE), mean squared error (MSE), and mean absolute percentage error (MAPE) measure each specification's ability to recover the same underlying country-level truth across independently generated samples.

3.6.8 Results of the Monte Carlo Simulations. The first step in interpreting the Monte Carlo experiment was to establish the population values against which the competing GIVE specifications were evaluated. To obtain sufficiently accurate approximations to these population quantities, we generated a large reference population of $N_{\text{TRUE}} = 100,000$ observations for each country and used these populations to calculate the population-level GIVE values under all nine weighting-aggregation specifications. This procedure was particularly important for the entropy- and PCA-based weighting schemes, for which the population weights cannot be obtained directly from the prescribed dimension means. We then treated the resulting population values as fixed truths when evaluating the 500 Monte Carlo samples.

Table 3 reports the country-specific parameters used to generate the simulation data, together with the resulting population dimension scores. Table 4 reports, for each weighting-aggregation specification, the corresponding population truth followed by the mean GIVE estimate obtained across the 500 Monte Carlo replications. Thus, these tables evaluate each estimator against the population target defined by the same weighting and aggregation procedure, rather than comparing all estimators against a single common GIVE value.

The population values in Table 3 demonstrate substantial heterogeneity across countries. Under arithmetic weighting, the true Individual dimension ranges from 69.44 to 74.50, the Community dimension from 64.58 to 72.87, and the Economic dimension from 25.63 to 31.04. The country-specific Enabling Environment scores range from 36.60 to 46.06. The correlation parameters also vary considerably across countries, with Individual correlations ranging from 0.363 to 0.697 and Community correlations from 0.019 to 0.490. The economic hourly rates range from USD 15.00 to USD 25.00. These values confirm that the simulation incorporates meaningful between-country heterogeneity while keeping the underlying population-generating mechanisms fixed over the repeated Monte Carlo samples.

An important feature of the simulation concerns the Enabling Environment dimension. Unlike the Individual, Community and Economic dimensions, the enabling indicators are specified as country-level indicators rather than characteristics generated separately for individual observations. Consequently, their values are not affected by the individual-level sampling variation or by the weighting procedures applied within the Individual, Community and Economic dimensions. The Enabling Environment score

Table 3. Country-specific population parameters and dimension scores.

Country	Rate	ρ_I	ρ_C	Individual	Community	Economic	Enabling
Country 1	25.00	0.533	0.112	72.81	72.63	31.04	46.06
Country 2	16.56	0.430	0.101	72.69	71.22	27.12	36.60
Country 3	21.99	0.511	0.169	71.07	68.92	30.21	40.44
Country 4	21.01	0.397	0.332	70.42	66.78	28.75	43.00
Country 5	15.72	0.403	0.070	70.94	64.58	28.62	39.92
Country 6	15.00	0.697	0.019	72.40	68.64	26.76	45.77
Country 7	25.00	0.363	0.343	74.50	70.60	28.95	42.75
Country 8	22.06	0.482	0.363	69.44	72.87	25.63	43.78
Country 9	20.91	0.636	0.490	73.60	69.61	30.35	36.60
Country 10	19.37	0.681	0.196	72.79	65.78	29.20	43.88

Note: The table reports the realised country-specific hourly economic rate, the Individual and Community latent-scale correlations used to generate the Monte Carlo populations, and the resulting population dimension scores. The Enabling Environment score is a country-level quantity and is therefore fixed across Monte Carlo replications and is not re-estimated from individual-level observations.

is therefore fixed for a given country across Monte Carlo replications and is invariant to whether arithmetic, geometric, entropy or PCA weighting is used for the other dimensions. This invariance is a consequence of the hierarchical structure of the simulated data and should not be interpreted as evidence that the Enabling Environment is intrinsically insensitive to weighting in the construction of GIVE. Rather, its country-level indicators enter the composite index directly as a fixed dimension score.

Table 4 demonstrates why it is necessary to define method-specific population truths. The population GIVE value changes systematically across specifications because the weighting and aggregation rules define different composite measures. For example, in Country 1, the population value is 55.64 under AM+AM, 52.44 under AM+GM, 68.45 under GM+AM and 67.00 under GM+GM. The corresponding mean estimates across the 500 replications are 55.61, 51.81, 68.55 and 66.29, respectively. The entropy- and PCA-based specifications similarly have their own population targets. Differences between specifications therefore reflect two distinct sources: differences in the estimands defined by the weighting and aggregation rules, and sampling error in estimating those estimands. The bias, MAE, RMSE and MAPE calculations consequently compare each estimator with its own method-specific population truth.

Overall, the mean Monte Carlo estimates in Table 4 are generally close to their corresponding population values for the arithmetic-, geometric- and entropy-based specifications, although the magnitude of discrepancy varies across methods and countries. The large reference population provides a substantially more reliable benchmark for the entropy- and PCA-based methods than the earlier simulations in which population truths were available only for selected specifications.

The nine specifications produced generally small errors, although the magnitude of error depended on both the weighting and final aggregation rule. Table 5 summarises the overall bias, MAE, RMSE, and MAPE across the 10 countries. GM+AM produced the lowest MAPE (5.22%)

and near-zero bias (0.004), indicating highly accurate recovery of the population GIVE under that specification. AM+AM achieved the smallest MAE (3.36) and RMSE (4.22), with a small negative bias (−0.15). Entropy+AM performed similarly (MAE = 3.38, RMSE = 4.24, MAPE = 6.79%). The Flat benchmark had larger errors (MAE = 4.27, RMSE = 5.37, MAPE = 7.92%).

Among the geometric-aggregation specifications, AM+GM (our primary specification) yielded a MAPE of 9.64%, MAE of 4.75, and RMSE of 8.01, with a negative bias of −1.51. This bias reflects the systematic underestimation inherent in geometric aggregation when any dimension is low. GM+GM showed a similar MAPE (9.66%) but higher MAE and RMSE (6.05 and 10.17, respectively). Entropy+GM also had a similar MAPE (9.74%) but with lower MAE and RMSE (4.34 and 7.27) than AM+GM. The PCA-based methods had the largest errors: PCA+AM had a positive bias of 3.45, while PCA+GM performed worst in terms of MAPE (12.23%), though its RMSE (7.92) was lower than several geometric specifications.

As presented in Table 5, all non-PCA specifications achieved a MAPE below the 10% threshold for highly accurate parameter recovery.⁵⁰ This indicates that sampling variation introduces negligible estimation error when reconstructing the true population parameters under these configurations. Consequently, the choice between the AM+GM and AM+AM specifications does not materially affect the framework's numerical reliability. The primary AM+GM specification is therefore preferred on robust theoretical grounds, as it balances the core principles of the composite index framework by retaining full indicator compensability within dimensions while limiting cross-dimension substitutability to align with global index standards, and it achieves a statistically rigorous level of estimation accuracy (MAPE = 9.64%)

Taken together, these results indicate that the principal GIVE specification, AM+GM, provides reasonably accurate recovery of its population target, although it is less

Table 4. Method-specific true GIVE value / mean estimated GIVE over 500 replications.

Country	AM+AM	AM+GM	GM+AM	GM+GM	Entropy+AM	Entropy+GM	PCA+AM	PCA+GM	Flat
Country 1	55.64 / 55.61	52.44 / 51.81	68.45 / 68.55	67.00 / 66.29	52.43 / 52.39	47.45 / 46.99	49.49 / 52.10	45.76 / 47.02	56.67 / 56.65
Country 2	51.91 / 51.76	47.61 / 46.04	63.21 / 63.19	60.59 / 58.69	48.02 / 47.82	42.51 / 41.15	44.63 / 48.83	41.56 / 43.73	52.59 / 52.40
Country 3	52.66 / 52.76	49.46 / 47.92	65.07 / 65.35	63.14 / 61.33	49.58 / 49.64	44.81 / 43.49	47.20 / 49.89	43.69 / 44.10	53.41 / 53.55
Country 4	52.24 / 52.39	49.10 / 48.25	63.70 / 64.08	62.34 / 61.46	49.20 / 49.31	44.43 / 43.74	45.79 / 51.03	42.86 / 46.89	53.29 / 53.49
Country 5	51.01 / 51.05	47.83 / 46.45	63.42 / 63.59	61.60 / 59.93	47.83 / 47.81	43.23 / 41.98	45.16 / 50.62	42.28 / 46.28	51.93 / 51.99
Country 6	53.39 / 52.80	49.67 / 47.52	63.46 / 63.04	62.29 / 59.74	50.65 / 49.99	45.05 / 43.11	48.31 / 50.84	44.96 / 45.34	54.77 / 54.03
Country 7	54.20 / 53.78	50.51 / 48.90	65.31 / 65.02	63.68 / 61.75	51.48 / 51.03	45.96 / 44.56	49.82 / 51.73	45.73 / 46.34	55.23 / 54.71
Country 8	52.93 / 52.99	48.82 / 48.06	63.27 / 63.48	61.67 / 60.83	49.77 / 49.77	43.93 / 43.26	47.00 / 50.28	43.92 / 45.49	54.09 / 54.18
Country 9	52.54 / 52.17	48.84 / 46.18	64.63 / 64.42	62.02 / 58.76	49.27 / 48.85	44.20 / 41.79	45.23 / 48.96	41.92 / 42.25	53.06 / 52.59
Country 10	52.91 / 52.59	49.77 / 47.87	64.73 / 64.58	63.35 / 61.09	49.92 / 49.54	45.13 / 43.44	47.25 / 50.12	43.98 / 44.78	54.07 / 53.67

Note: The table reports, in each cell, the method-specific population truth followed by the mean estimated GIVE across the 500 Monte Carlo replications. We obtained population truths for all nine specifications using a large reference population of $N_{\text{TRUE}} = 100,000$. Thus, "55.64 / 55.61" for AM+AM in Country 1 denotes a population value of 55.64 and an average estimated value of 55.61.

Table 5. Monte carlo performance of the nine GIVE specifications.

Method	Bias	MAE	RMSE	MAPE (%)
GM+AM	0.004	3.365	4.227	5.22
AM+AM	-0.153	3.360	4.225	6.35
Entropy+AM	-0.200	3.378	4.243	6.79
Flat	-0.185	4.266	5.369	7.92
AM+GM	-1.505	4.752	8.010	9.64
GM+GM	-1.782	6.048	10.174	9.66
Entropy+GM	-1.321	4.342	7.267	9.74
PCA+AM	3.453	4.599	5.774	9.84
PCA+GM	1.555	5.313	7.919	12.23

Note: Bias is the average country-specific bias across the 10 countries. MAE, RMSE and MAPE are similarly averaged. All error measures are on the 0–100 GIVE scale except MAPE, which is expressed as a percentage.

numerically accurate than the corresponding arithmetic-aggregation alternatives. The results also demonstrate that the apparent differences in accuracy between specifications should not be interpreted as evidence that one method is estimating the same quantity better than another. Rather, each weighting–aggregation combination defines a distinct population-level composite index. The simulation therefore evaluates both the numerical recovery of these distinct targets and the robustness of the resulting country comparisons.

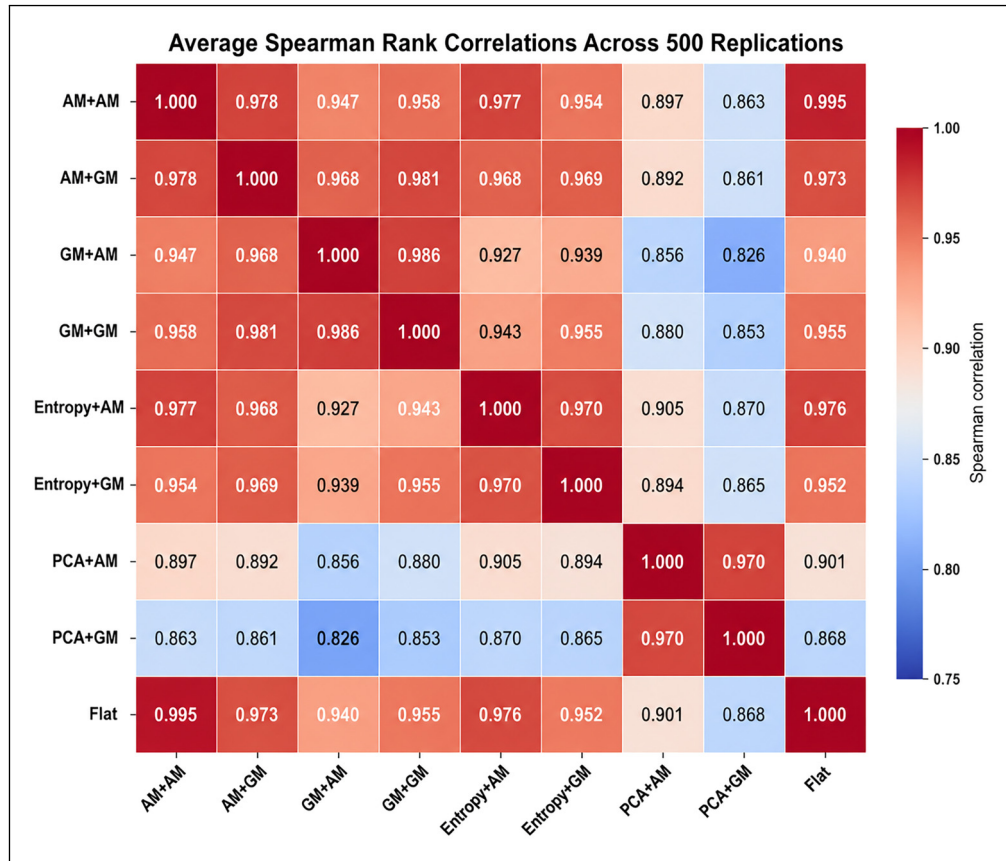
3.6.9 Stability of country rankings. A second criterion for evaluating the sensitivity of GIVE is whether alternative weighting and aggregation procedures materially change the ranking of countries. For each Monte Carlo replication, countries were ranked separately under each of the nine specifications, and pairwise Spearman rank correlations were calculated. These correlations were then averaged across the 500 replications.

The results in Table 6 show very high rank agreement among the non-PCA specifications. The AM+AM and AM+GM specifications have an average Spearman correlation of 0.978, indicating that the two approaches produce almost identical country rankings despite differences in their absolute GIVE scores. Similarly, the correlations among the AM, GM, entropy and Flat specifications are generally above 0.92, with most exceeding 0.94. The two PCA specifications are also highly correlated (approximately 0.970), although their correlations with the other specifications are somewhat lower, generally ranging from approximately 0.83 to 0.91.

Figure 2 provides a visual summary of these rank correlations. The predominance of high correlations among the non-PCA specifications indicates that the ordering of countries is relatively insensitive to the choice of weighting and aggregation procedure. This is particularly important for the intended policy use of GIVE, where the purpose is not only to produce an absolute index value but also to facilitate meaningful comparison across countries.

Table 6. Average spearman rank correlations across the 500 monte carlo replications.

	AM+AM	AM+GM	GM+AM	GM+GM	Entropy+AM	Entropy+GM	PCA+AM	PCA+GM	Flat
AM+AM	1.000	0.978	0.947	0.958	0.977	0.954	0.897	0.863	0.995
AM+GM	0.978	1.000	0.968	0.981	0.968	0.969	0.892	0.861	0.973
GM+AM	0.947	0.968	1.000	0.986	0.927	0.939	0.856	0.826	0.940
GM+GM	0.958	0.981	0.986	1.000	0.943	0.955	0.880	0.853	0.955
Entropy+AM	0.977	0.968	0.927	0.943	1.000	0.970	0.905	0.870	0.976
Entropy+GM	0.954	0.969	0.939	0.955	0.970	1.000	0.894	0.865	0.952
PCA+AM	0.897	0.892	0.856	0.880	0.905	0.894	1.000	0.970	0.901
PCA+GM	0.863	0.861	0.826	0.853	0.870	0.865	0.970	1.000	0.868
Flat	0.995	0.973	0.940	0.955	0.976	0.952	0.901	0.868	1.000

**Figure 2.** Spearman rank correlations between GIVE specifications across the 500 Monte Carlo replications. Higher values indicate greater agreement in the ordering of the 10 countries.

The high rank stability suggests that methodological choices affect the magnitude and scale of GIVE more strongly than they affect the broad ordering of countries. In particular, the very high correlation between AM+AM and AM+GM indicates that replacing arithmetic aggregation with geometric aggregation changes the degree of compensability between dimensions without substantially altering the comparative ordering. This provides reassurance that the substantive conclusions drawn from GIVE are not driven primarily by the particular weighting or aggregation rule adopted.

These findings distinguish two dimensions of robustness. Numerical accuracy concerns how closely an estimated GIVE score recovers the population target defined by its own weighting and aggregation rule, whereas rank stability concerns whether the relative ordering of countries is preserved when alternative specifications are used. The Monte Carlo experiment indicates that the latter is particularly strong across the non-PCA specifications. Thus, although the choice of weighting and aggregation procedure can materially change the index's numerical scale, it

has a much smaller effect on the comparative ranking of countries.

3.6.10 Comparison with established global composite indices.

The methodological choices considered for GIVE are also informed by established approaches to the construction of international composite indices. The Human Development Index (HDI) provides an important precedent for geometric aggregation: since 2010, the HDI has used the geometric mean to combine its dimensions, specifically to reduce the extent to which high achievement in one dimension can compensate for low achievement in another.^{38,39} In contrast, the Sustainable Development Goals (SDG) Index provides a strong precedent for arithmetic aggregation, using equal weights and arithmetic means in constructing goal-level scores and in aggregating the 17 goals.⁴⁰ The Global Innovation Index (GII) similarly relies primarily on weighted arithmetic aggregation in constructing its sub-pillars and pillars and in combining its Input and Output sub-indices, while its statistical audit examines the robustness of the arithmetic-average assumption against alternative aggregation approaches.⁴¹ The Multidimensional Poverty Index (MPI) provides a further relevant precedent for multidimensional measurement through explicit indicator weights and the Alkire–Foster identification of joint deprivation, although its construction differs mathematically from GIVE.^{27,42} Taken together, these indices demonstrate that both arithmetic and geometric aggregation are established approaches in international composite-index construction, with the choice reflecting the substantive relationship among the dimensions being measured. These precedents do not imply that one rule is universally correct for GIVE; rather, they provide methodological context for evaluating the alternative specifications considered in the simulation.

3.6.11 Compensatory and substitution mechanisms.

The distinction between compensatory and substitution mechanisms provides an additional conceptual basis for the aggregation choice. Arithmetic aggregation is fully compensatory: a sufficiently high score in one dimension can offset an equally sized deficit in another. Geometric aggregation is less compensatory because the product of the dimension scores is more strongly affected by a low-performing dimension. In this sense, the geometric mean limits substitution rather than eliminating it. For GIVE, this distinction matters because the Individual, Community, Economic, and Enabling Environment dimensions capture complementary aspects of volunteer engagement. A country with very strong individual benefits but weak community or enabling conditions should not necessarily obtain the same overall assessment as a country performing more evenly across all four dimensions. The AM+GM formulation therefore combines transparent equal weighting within dimensions with a final aggregation rule that places greater emphasis on balanced performance across dimensions.

3.6.12 Rationale for AM+GM as the primary GIVE specification.

The simulation results, together with principles from the international composite-index literature, motivate the selection of AM+GM as the primary GIVE specification. Equal arithmetic weighting within dimensions is transparent and avoids assigning greater normative importance to an indicator solely because it exhibits greater statistical variation. At the final aggregation level, geometric aggregation substantially reduces compensability across dimensions while retaining a continuous composite measure.

The simulation also demonstrates that selecting AM+GM does not materially alter the substantive ranking of countries relative to the AM+AM benchmark. The average Spearman correlation between AM+AM and AM+GM was 0.978, indicating very strong agreement in the ordering of countries. Although AM+AM achieved slightly lower numerical error than AM+GM, the differences in absolute performance should be considered alongside the substantive properties of the aggregation rules. In particular, AM+GM provides a deliberate trade-off between numerical accuracy and limited compensability: a relatively low score in one GIVE dimension cannot be completely offset by high scores in the remaining dimensions. The objective is therefore not to claim universal statistical superiority for AM+GM, but to select a specification that balances estimation performance, interpretability, normative transparency, and limited compensability.

Accordingly, we recommend AM+GM as the primary GIVE specification, with AM+AM retained as the principal additive sensitivity benchmark and GM+AM, GM+GM, entropy, PCA, and Flat specifications retained for methodological sensitivity analysis. This recommendation reflects both the substantive rationale for limiting compensability across the GIVE dimensions and simulation evidence that the principal specifications reasonably recover their respective population targets and, importantly, produce highly consistent country rankings. The strategy makes explicit that the final GIVE index is a normative composite measure rather than a uniquely determined statistical quantity, while demonstrating empirically that the principal substantive conclusions are robust to reasonable alternative construction choices.^{26,34}

3.7 Summary of Simulation Results

Overall, the Monte Carlo study supports the robustness of the GIVE framework, particularly for the comparative ranking of countries. The simulation incorporated empirically motivated global benchmarks, country-specific heterogeneity, correlated ordinal indicators, country-specific economic rates, and fixed population parameters across repeated samples. Across 500 replications and 10 countries, the alternative specifications generally showed modest bias and moderate absolute error, although numerical performance varied by weighting and aggregation method. Arithmetic

final aggregation generally produced lower numerical error than geometric final aggregation, while the GM+AM specification achieved the lowest MAPE. The non-PCA specifications produced particularly high agreement in country rankings, whereas the PCA-based specifications generated somewhat more distinct rankings. These findings support using AM+GM as the primary specification on substantive and methodological grounds, while retaining the alternative specifications as sensitivity analyses.

4 Implementation for Official Statistics

4.1 Phase 2 Roadmap (2027–2029)

Building on the validated framework, Phase 2 will operationalise GIVE across an initial cohort of 40–60 countries through five interlinked steps: (1) data compilation and quality assessment (MOUs with NSOs, quality report); (2) index computation (GIVE scores for 40–60 countries, disaggregated results); (3) validation and stakeholder review (regional workshops, refined methodology); (4) global reporting and dissemination (technical report, country profiles, dashboards); and (5) institutionalisation and sustainability (GIVE Technical Task Force, capacity-building) Chapter 7, Section 7.8.²

4.2 Role of Key Stakeholders (NSOs, Academia, CSOs, and INGOs)

National Statistical Offices (NSOs) are essential partners for GIVE implementation. Specific contributions include: integrating volunteer modules into existing household surveys (Labour Force Surveys, Time Use Surveys); applying harmonisation protocols to ensure cross-country comparability; validating country-level data and providing contextual interpretation; and building capacity for ongoing volunteer measurement within routine statistical production. For lower-resource settings, Phase 2 includes targeted capacity-building on survey design, data harmonisation, and index construction.

Academic institutions have a critical role in moving GIVE from a conceptual framework to a fully operationalised global index. Universities can serve as hubs for pilot testing, methodological refinement, and capacity building – particularly in lower-resource settings where national statistical systems are still developing. By conducting targeted surveys, training local researchers, and publishing harmonised code and data tools, academia can help translate the GIVE framework into practical, replicable measurement tools that countries can adopt and adapt.

Civil society organisations (CSOs) and international non-governmental organisations (INGOs) are equally essential partners in this journey. They work directly with volunteers and communities on the ground, possessing deep experiential knowledge of what works, what data is

collectable, and what measurement approaches are meaningful. CSOs and INGOs can strengthen the connection between national governments and UNV by providing evidence of practice – real-world examples that inform indicator design, validate measurement approaches, and demonstrate the value of integrating volunteer statistics into official data systems. Their role is to bridge community realities and national statistical processes, ensuring that GIVE remains grounded in volunteers' lived experiences.

5 Conclusion

Our study presents extensive Monte Carlo simulations across 10 representative country settings and 500 replications to evaluate nine specifications for constructing the first-ever Global Index of Volunteer Engagement (GIVE). The simulation analysis provides empirical evidence of stable country rankings and acceptable numerical accuracy under heterogeneous simulated country-level conditions. Although arithmetic final aggregation yielded the lowest numerical error, the recommended arithmetic-geometric (AM+GM) specification achieved a mean absolute percentage error (MAPE) of 9.64%, below the 10% accuracy threshold. These findings demonstrate that aggregation choices should be guided by substantive conceptual alignment rather than numerical error alone.

Based on these empirical results and drawing on precedents from established global composite indices such as the Human Development Index (HDI) and the Global Multidimensional Poverty Index (MPI), we recommend a two-stage AM+GM aggregation strategy.^{27,38,39,41} After normalising indicators to a 0–100 scale, indicators within the three individual-level dimensions, namely Value to the Individual, Value to the Community, and Value to the Economy, are combined using simple arithmetic means, while the resulting three dimension scores are combined with the country-level Enabling Environment score using a geometric mean. This structure provides transparent and equal representation of indicators within dimensions while limiting cross-dimension compensability and placing greater emphasis on balanced performance. It is also consistent with established approaches to global composite-index construction, including those used in the Sustainable Development Goals Index⁴⁰ and the Global Innovation Index,⁴¹ while allowing AM+AM and other specifications to serve as sensitivity benchmarks.

The proposed GIVE framework provides a practical, flexible, and scalable foundation for National Statistical Offices to integrate volunteerism modules into household surveys, labour-force surveys, and official administrative records. Its flexible architecture permits iterative refinement as additional data become available and cross-country implementation provides new evidence on indicator comparability and measurement performance. Looking ahead,

the next phase of the GIVE initiative will focus on international field testing and progressive expansion to additional countries. Guided by the Phase 2 roadmap for 2027-2029, this expansion will focus on refining data sources, assessing cross-country measurement performance, and establishing GIVE as a globally comparable statistical framework for measuring volunteer engagement.

Acknowledgements

The authors gratefully acknowledge the United Nations Volunteers (UNV) for their partnership and financial support, the International Labour Organization (ILO) for its methodological guidance and support, and the United Nations Development Programme (UNDP) for its contributions throughout the development of the Global Index of Volunteer Engagement (GIVE). We also extend our sincere gratitude to the many civil society organisations and volunteers who contributed data, perspectives, and insights to the development of the framework.

We particularly acknowledge the valuable contributions and guidance of Vladimir Ganta, Senior Statistician, International Labour Organization (ILO); and Emiliya Asadova, Regional Policy Specialist and Team Lead on Evidence, United Nations Volunteers (UNV) programme; Rafael Diez Medina, Chief Statistician and Director, Department of Statistics, ILO; Yanchun Zhang, Chief Statistician, United Nations Development Programme (UNDP); Emiliya Asadova, Regional Policy Specialist and Team Lead on Evidence, United Nations Volunteers (UNV) programme; Jacob Mwathi Mati, Associate Professor, University of the Witwatersrand, South Africa; Alok Rath, Global Head of Knowledge, Evidence and Learning, VSO International, New Delhi, India; Jin-Kyung Jung, Professor, Department of Public Administration, Kwangwoon University, Seoul, South Korea and Desiree Manamela, and Chief Director: Labour Statistics, Statistics South Africa, Pretoria, South Africa. Their expertise, constructive advice, and engagement were valuable in shaping the conceptual and methodological development of GIVE.

We further acknowledge the following individuals for their technical advice, inputs, and support during the development of GIVE: Omar Agbangba, Francisco Javier Ampuero, Juan Pablo Gordillo, Njeri Wabiri, Ngianga-Bakwin Kandala, Bayarmaa Luntan, Christopher Millora, Jessika Naidoo, Bogna Kietli?ska-Radwa?ska, Nujoud Serhan, Vera Chrobok, Louise Chamberlain, Enkhmandakh Ishdorj, Andrey Pogrebnyak, Anastasiya Krasnoshchoka, Narendra Mishra, and Jacqueline Butcher García-Colín, Director, Centro de Investigación y Estudios sobre Sociedad Civil (CIESC), Mexico.

ORCID iDs

Samuel Manda  <https://orcid.org/0000-0002-9672-3312>

Harrid Nkhoma  <https://orcid.org/0009-0004-7981-0563>

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily represent the views of UNV, ILO, the University of Pretoria, or any of the affiliated institutions.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Manda S. *The Paradox of Invisible Contributions*. United Nations Volunteers, 2026.
2. United Nations Volunteers (UNV). *2026 State of the World's Volunteerism Report: Volunteerism and its measurements*. Bonn: UNV, 2025.
3. African Union Commission (AUC). *State of Volunteerism in Africa Report*. Addis Ababa: Women, Gender and Youth Directorate, 2025.
4. United Nations Volunteers. *State of the World's Volunteerism Report 2011: Universal Values for Global Well-being*. UNV, 2011.
5. International Labour Organization (ILO). *Resolution concerning statistics of work, employment and labour underutilization (19th ICLS)*. Geneva: ILO, 2018.
6. International Labour Organization (ILO). *Manual on the measurement of volunteer work*. Geneva: ILO, 2021.
7. United Nations Volunteers (UNV). *UN General Assembly proclaims 2026 as the International Year of Volunteers for Sustainable Development*. (Resolution A/RES/78/127), 2023.
8. Russell B. *Conceptualising volunteerism and global volunteer measurement: challenges, definitions and emerging frameworks*. Paper presented at the Twelfth international conference of the international society for third sector research (ISTR), Stockholm, 2016.
9. Manda S. *Global Volunteer Index Project: An Overview*. Presentation delivered at the Regional Technical Consultation Workshop, Crowne Plaza Hotel, Santo Domingo, Dominican Republic, 2025.
10. NSW Government. *2025 NSW State of Volunteering Report*. The Centre for Volunteering, 2025.
11. United Nations. *Transforming our world: The 2030 Agenda for Sustainable Development*. (Resolution A/RES/70/1). New York: UN, 2015.
12. United Nations. *Handbook on non-profit institutions in the system of national accounts*. UN, 2003.
13. United Nations Statistics Division. *Guidelines on Time Use Surveys*. New York: UN, 2017.
14. United Nations Volunteers. *State of the World's Volunteerism Report 2026: Volunteerism and Its Measurement*. Bonn: UNV, 2026.
15. Corporation for National and Community Service. *Volunteering as a Pathway to Employment*. Washington, DC: CNCS, 2013.

16. Deloitte. *2016 Deloitte Impact Survey: Building Leadership Skills Through Volunteerism*. New York: Deloitte, 2016.
17. Independent Sector. *The Value of Volunteer Time*. Independent Sector, 2025.
18. Jenkinson CE, Dickens AP, Jones K, et al. Is volunteering a public health intervention? A systematic review and meta-analysis of the health and survival of volunteers. *BMC Public Health* 2013; 13: 773.
19. Morrow-Howell N, Hinterlong J, Rozario PA, et al. Effects of volunteering on the well-being of older adults. *J Gerontol: Ser B* 2003; 58: S137–S145.
20. Abrams D, Lalot F and Hogg MA. *Community, belonging and wellbeing: The impact of volunteering on social cohesion*. British Academy, 2023.
21. Davies B, Abrams D, Horsham Z, et al. The causal relationship between volunteering and social cohesion: A large scale analysis of secondary longitudinal data. *Soc Indic Res* 2024; 171: 809–825.
22. Horsham Z, Abrams D, Davies B, et al. Social cohesion and volunteering: Correlates, causes, and challenges. *Trans Issues Psychol Sci* 2024; 10: 51–68.
23. Salamon LM, Sokolowski SW and Haddock MA. Measuring the economic value of volunteer work globally: Concepts, estimates, and a roadmap to the future. *Annals of Public and Cooperative Economics* 2011; 82: 217–252.
24. Wallrodt S and Thieme L. The opportunity costs of volunteering: evidence from Germany. *Nonprofit Volunt Sect Q* 2022; 52: 1129–1157.
25. Sidel M. *Law and policies affecting volunteerism since 2001*. University of Wisconsin Law School Legal Studies Research Paper No. 1676, 2009.
26. Organization for Economic Co-operation and Development. *Handbook on Constructing Composite Indicators: Methodology and User Guide*, 2008. Paris, 2008.
27. Oxford Poverty and Human Development Initiative (OPHI) & United Nations Development Programme (UNDP). *Global Multidimensional Poverty Index 2023: Unstacking global poverty*. OPHI/UNDP, 2023.
28. Zhu Y, Tian D and Yan F. Effectiveness of entropy weight method in decision-making. *Mathemat Problems Eng* 2020; 2020: 1–5.
29. Chen P. *Effectiveness of Entropy Weight Method in Decision-Making*. Wiley Online Library, 2020.
30. Karagiannis R and Karagiannis G. Constructing Composite Indicators with Shannon Entropy: The Case of Human Development Index. *Working Paper*, 2019.
31. Shannon CE. A mathematical theory of communication. *Bell System Technical Journal* 1948; 27: 379–423.
32. Jolliffe IT. *Principal Component Analysis*. 2nd ed. Springer, 2002.
33. Vyas S and Kumaranayake L. Constructing socio-economic status indices: How to use principal components analysis. *Health Policy Plan* 2006; 21: 459–468.
34. Decancq K and Lugo MA. Weights in multidimensional indices of wellbeing: An overview. *Econom Rev* 2013; 32: 7–34.
35. Mazziotta M and Pareto A. On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Soc Indic Res* 2015; 127: 983–1003.
36. Mazziotta M and Pareto A. Mazziotta-Pareto Index in Composite Indicator Construction. In: F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research*, 2022. Springer.
37. De Muro P, Mazziotta M and Pareto A. Composite indices of development and poverty: An application to MDGs. *Soc Indic Res* 2011; 104: 1–18.
38. United Nations Development Programme (UNDP). *Human Development Report 2021/2022: Technical Notes*. United Nations Development Programme, 2022.
39. Herrero C, Mart'inez R and Villar A. *Improving the measurement of human development*. United Nations Development Programme, 2010.
40. Sustainable Development Solutions Network (SDSN). *Sustainable Development Report 2026: Methodology*. SDSN, 2026.
41. World Intellectual Property Organization (WIPO) and Joint Research Centre (JRC). (2025). *Global Innovation Index 2025: JRC Statistical Audit*. Geneva: WIPO.
42. Alkire S and Foster J. Counting and multidimensional poverty measurement. *J Public Econ* 2011; 95: 476–487.
43. National Council for Voluntary Organisations (NCVO). *Time Well Spent 2023: A National Survey on the Volunteer Experience*. NCVO, 2023.
44. Helliwell JF and Putnam RD. The social context of well-being. *Philosop Trans R Soc London Ser B: Biol Sci* 2004; 359: 1435–1446.
45. Putnam RD. *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, 2000.
46. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2026.
47. United Nations Volunteers. *Knowledge Portal: National Volunteer Policies and Legislation*. UNV, 2025.
48. Volunteering Australia. *National Survey of Volunteers*. Volunteering Australia, 2010.
49. Chai T and Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. *Geosci Model Dev* 2014; 7: 1247–1250.
50. Lewis CD. *Industrial and Business Forecasting Methods*. Butterworths, 1982.