sustainability

MDPI

Article

Comparative Sustainability Under the 5P Framework: Construction and Evaluation of a Composite Index for Colombia, Mexico, and Economies and Territories of the Middle East and Its Regional Environment, 2020–2023

Alejandro Acevedo Amorocho [1,*], José Gerardo De la Vega Meneses [2], Henry Fernández Pinto [3], Jesús David Franco Rangel [1], Jenny Johana Ramírez Acuña [1] and Zamir Rojas Pulido [1]

[1] Research Group for Accounting Development, Universidad Santo Tomás, Bucaramanga 680001, Colombia; jesusdavid.franco@ustabuca.edu.co (J.D.F.R.); jenny.ramirez01@ustabuca.edu.co (J.J.R.A.); zamir.rojas@ustabuca.edu.co (Z.R.P.)

[2] Business School Research Group, Universidad Popular Autónoma del Estado de Puebla, Puebla 72017, Mexico; josegerardo.delavega@upaep.mx

[3] Virtual Education and Information and Communication Technologies Research Group, E-Innovare, Unidades Tecnológicas de Santander, Bucaramanga 680001, Colombia; hfernandezp@correo.uts.edu.co

* Correspondence: alejandro.acevedo@ustabuca.edu.co

Abstract

This article constructs and evaluates a composite sustainability index aligned with the 5P framework (People, Planet, Prosperity, Peace, and Partnerships) to compare Colombia and Mexico with 14 countries and territories from the Middle East and its regional environment, using information available for 2020–2023. The design is quantitative, documentary, and comparative and integrates eight international indicators. The main specification is estimated for 16 comparable cases after verifying coverage, harmonizing directionality, applying a fixed bilateral p5–p95 treatment of extreme values, and normalizing the variables on a scale from 0 to 100. Uniform weighting by dimension constitutes the baseline scenario, while an expanded 32-case reference database is used for exploratory KMO and Bartlett diagnostics and to estimate alternative indicator weights through principal component analysis. The countries with the best aggregate performance are Colombia, Israel, Mexico, and Kuwait, although their results are not homogeneous across the evaluated dimensions. Spearman and Kendall coefficients, deterministic weighting scenarios, a direct indicator-level PCA composite, z-score standardization, geometric-mean aggregation, leave-one-dimension-out tests, reference-base re-estimation, constrained weights, and Monte Carlo simulations indicate high overall concordance under moderate perturbations but also reveal sensitivity in some individual positions when assumptions are relaxed. Comparison with the SDG Index yields moderate convergence (Spearman $\rho = 0.503$; Kendall $\tau = 0.367$; Pearson $r = 0.554$), with wide confidence intervals due to the small sample. Under these conditions, the index is useful for identifying performance groups, characterizing national profiles, and detecting gaps between countries, but its results should be interpreted as an analytical approximation rather than as a definitive sustainability hierarchy.

Keywords: composite index; 5P framework; sustainability indicators; sustainable development; international comparison

Check for updates

Academic Editor: Jianming Cai

Received: 19 September 2026
Revised: 20 October 2026
Accepted: 21 October 2026
Published: 24 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Sustainable development as set out in the 2030 Agenda proposes an integrated vision of sustainable development, articulated through five axes that encompass people, planet, prosperity, peace, and partnerships [1]. The strength of this framework lies in recognizing interdependencies, whereby progress in both the social and economic spheres is conditioned by the protection of environmental systems, the development of institutional capacities, and collaboration among actors [1].

Given the particular characteristics of this integrative framework, a weakness can be identified in terms of evaluation due to the current structure of unidimensional indicators, which limits the structural reading of results and intersectoral trade-offs.

According to the methodological consensus, the development of composite indices enhances analytical capacity and the interpretation of results; however, to achieve this, their construction must be based on transparent constructs that explicitly incorporate uncertainty assessment and subsequently subject the results to robustness and sensitivity checks [2–7].

It is important to note that the use of composite indices to construct rankings entails a series of both technical and methodological challenges, because the results are sensitive to the selection, normalization, weighting, and aggregation criteria employed [2,7].

Even small changes in Min–Max reference limits can alter relative positions, which makes sensitivity analysis important when a composite indicator is used for ranking [8]. The 5P Index is specified as a formative composite because its dimensions represent distinct components of sustainability that jointly define the construct; high inter-item correlations are therefore neither expected nor required for indicator retention. Accordingly, the quality of the measurement framework is assessed through conceptual coherence, methodological traceability, and the stability of results under alternative methodological choices rather than through a single internal-consistency criterion [2,3].

Given the advances in international measurement systems and metrics, it can be observed that the empirical operationalization of the 5P framework still lacks comparable indicators grounded in explicit uncertainty principles. This deficit becomes more evident when comparing regions such as Latin America and the Middle East, where productive, environmental, and geopolitical structures are often represented through aggregate global indicators. In this context, the study proposes a formative measurement architecture that methodologically separates index construction (design, weighting, and aggregation of dimensions) from robustness analysis (sensitivity, stability, and uncertainty tests), thereby differentiating the two analytical levels.

Under this approach, two analytical expectations are examined rather than causal hypotheses. H1 expects the general ordering of the 5P Index to remain highly concordant under plausible alternative specifications. H2 expects countries with similar aggregate scores to display heterogeneous dimensional profiles and, consequently, performance blocks to be more stable than exact rank positions. H1 is assessed through Spearman and Kendall rank correlations, maximum rank shifts, alternative weighting and normalization scenarios, and leave-one-dimension-out tests. H2 is assessed by comparing dimensional profiles and the persistence of block membership across these scenarios. Neither expectation is interpreted as causal or confirmatory; both guide the robustness analysis and the comparative interpretation.

Based on the preceding premises, the general objective of the research is to construct and evaluate a composite sustainability index aligned with the 5P framework that allows comparison of the relative position of Colombia and Mexico with economies and territories of the Middle East and its regional environment over the 2020–2023 reference window. Specifically, the study addresses three questions: (i) How are countries ordered when People, Planet, Prosperity, Peace, and Partnerships are integrated into a common measure?

(ii) Which dimensions and compensation patterns explain interregional and intraregional gaps? (iii) How robust is the ordering to variations in normalization, the normalization reference base, weighting, aggregation, and omission of dimensions? The p5–p95 winsorization rule is held fixed in the reported robustness scenarios; sensitivity to alternative winsorization cut points is therefore treated as an explicit limitation rather than as a tested source of uncertainty.

The article makes three contributions. First, it provides a replicable methodological framework for constructing a composite index aligned with the 5P framework and with international guidance for comparative composite indicators and rankings [2,7]. Second, it offers comparative evidence for Colombia and Mexico relative to countries and territories in the Middle East and its regional environment, with interpretation at both aggregate and dimensional levels. Third, it explicitly evaluates ranking robustness under alternative methodological specifications, thereby reducing the risk of overinterpreting a single ranking configuration [3,8].

To address the issues outlined above, the article is organized into six sections. Following this introduction, Section 2 develops the theoretical framework; the methodological design is addressed in Section 3; Section 4 presents the aggregate, dimensional, and robustness results; Section 5 discusses the findings and their strategic implications; and Section 6 presents the conclusions, limitations, and future lines of work.

2. Theoretical Framework

2.1. Sustainable Development and the 5P Approach

Sustainable development has become established as a normative and analytical framework that seeks to integrate social, environmental, economic, and institutional objectives simultaneously into public and private decision-making processes. Within the structure of the 2030 Agenda, this integration is explicitly reflected in the 5P approach (People, Planet, Prosperity, Peace, and Partnerships), which states that sustainable progress requires coordinated efforts in human well-being, prosperity, environmental protection, institutional strength, and cooperation among actors [1]. From this perspective, sustainability is presented as a multidimensional phenomenon in which improvements in one area depend on enabling conditions and factors inherent in others, such as institutional stability and the development of partnerships to implement policies, or environmental resilience to ensure well-being and productivity over time [1].

In this context, the 5P approach is particularly valuable for comparative research because it provides a conceptual framework that integrates a cohesive scheme from which national realities with heterogeneous trajectories and diverse structural conditions can be addressed. Its intrinsic strength, however, also constitutes an operational weakness: making the 5Ps comparable is not a simple exercise of translating categories into numbers; rather, it is a process involving choices, including which indicators to prioritize and how to formalize disparate time series, how to standardize what is not originally standardized, and how to choose an aggregation criterion that does not distort the meaning of each dimension. In addition, the process must leave a sufficient methodological trail so that an external reader can follow it and independently assess the soundness of the comparisons drawn [9].

2.2. Composite Indices for Measuring Multidimensional Phenomena

In applied research, composite indices are used to condense multidimensional realities into a manageable figure. The apparent simplicity of an index conceals a methodical process structured around a network of technical choices and normative assumptions that are often not reported with the transparency and detail they require [9]. A relevant point is that anyone seeking to structure this type of tool must pass through four critical stages,

each with its corresponding methodological rigor that will shape the result: selecting the indicators or alpha factors that will enter the structure, standardizing their scales to make them measurable, assigning weights that establish their relative importance, and defining the aggregation rule that combines them into a single value [9–11].

In addition, sustainability is a matter of greater complexity because countries rarely exhibit a balanced profile: the evidence indicates that while some may stand out in a particular area, they may show imbalances in other dimensions. Given this complexity, linear aggregation (a widely used approach) allows strengths in one area to compensate for weaknesses in other areas, potentially producing a synthetic result that is more favorable than the overall profile would warrant [12]. For this reason, experts recommend interpreting global indices together with the disaggregated profile and providing a detailed discussion of the degree of substitutability assumed by the aggregation procedure [9,12].

It is also important to clarify that even after taking the above precautions, the soundness of the index remains an open issue. This is supported by the findings of several studies showing that the nominal weights assigned to an indicator do not always correspond to the influence it ultimately exerts on the final result. It is therefore necessary not to rely on appearances alone, but to complement the analysis with estimates of effective contribution and to subject the index to a varied set of tests that modify the weights in order to assess ranking robustness [3,11].

2.3. Robustness, Sensitivity, and Ranking Stability

When a composite index is used to generate rankings, the methodological demands increase because relative positions may change under reasonable choices concerning normalization, weighting, aggregation, or the treatment of extreme observations. For this reason, the literature recommends evaluating ranking robustness as a central quality criterion: a ranking is more informative when its main ordering patterns persist across plausible specifications [4]. In practice, robustness can be examined by comparing alternative rankings generated through systematic methodological changes and quantifying concordance with rank-correlation coefficients such as Spearman's ρ and Kendall's τ [5,6].

Empirical evidence shows that ranking behavior is not always stable, a dynamic that may arise from the interaction among weights, aggregation, and data structure. Experts therefore recommend incorporating alternative weighting scenarios, comparing normalization methods such as min–max versus z-scores, and examining the effect of extreme values [8,10]. It has also been noted that seemingly technical decisions, such as the thresholds set in min–max scaling, can produce fragile rankings, reinforcing the need to report sensitivity as part of methodological transparency [8,13]. Taken together, these recommendations justify that a comparative sustainability study should not be limited to reporting the aggregate ranking; it should go further by presenting a disaggregated profile by dimension and demonstrating how the ordering responds under alternative conditions.

2.4. Global Sustainability Indices as External References

In the field of sustainability, various comparative frameworks have been developed, and tools such as the SDG Index have become a key reference for measuring progress toward the 2030 Agenda. These systems serve as an external benchmark for comparing results, identifying coincidences, or analyzing why two metrics differ according to their coverage or weighting [14]. However, the methodological literature calls for these data to be examined in sufficient detail, because differences between indices do not necessarily indicate real progress; rather, they may reflect disparities in the conceptual framework, aggregation technique, or data quality itself [2,9].

In comparative studies involving heterogeneous regions such as Latin America and the Middle East, these external benchmarks help determine whether the ordering obtained from the index is consistent with global assessments, while also allowing analysis of the influence of structural variables. This dynamic also opens discussion of underlying structural factors, such as political instability, institutional gaps, and data scarcity, which jointly condition territorial realities and the way these dimensions are measured [14].

In light of the considerations described above, the 5P index incorporates an explicit conceptual design together with robustness controls, in contrast with global frameworks, thereby supporting interpretation with methodological transparency.

2.5. International Indicators and Multidimensional Measurement

Assessing sustainable development requires metric tools capable of capturing its multiple dimensions. Accordingly, in recent decades international indicators have become established for measuring human well-being, environmental performance, economic growth, resource security (water–energy–food), democratic quality, peace, and intersectoral cooperation; however, observing each indicator individually provides only a fragmented view [15,16]. In this context, expert guidance favors the integration of metrics that make it possible to identify patterns, detect imbalances, and establish correlations among dimensions, thereby reducing biases and enabling more robust comparisons [17,18], which is particularly important in regions with contrasting realities such as Latin America and the Middle East.

Within this framework, economic growth, commonly estimated through changes in real GDP, becomes a relevant factor for sustainable development because its influence extends to variables such as employment, income, investment, and the state's fiscal revenue capacity to provide public goods and services [19,20]. On the other hand, its analytical scope is limited when used as an isolated indicator and must be complemented by indicators of distribution, well-being, environmental sustainability, and institutional stability [21]. This multidimensional framework must also incorporate the resource-security dimension, because demographic pressures, accelerated urbanization, and changes in consumption patterns intensify the use of natural resources, reinforcing the water–energy–food nexus as an indispensable analytical framework for assessing systemic resilience and resource interdependence [18,22,23].

In addition, the Human Development Index (HDI) synthesizes the dimensions of health, education, and real income, making it possible to identify differences in human development among countries with similar income levels and revealing well-being asymmetries that are not adequately explained by per capita income. Nevertheless, it has limitations in explicitly capturing inequality, multidimensional poverty, or social security [24,25].

Complementarily, the Happy Planet Index (HPI) links subjective well-being, life expectancy, and ecological footprint, making it possible to estimate how efficiently national systems transform natural capital into human development outcomes [15,16,26].

The Green Growth Index, in turn, quantifies performance in resource efficiency, preservation of natural capital, green economic opportunities, and social inclusion, providing guidance for strengthening sustainability policies [27–29].

With regard to the institutional dimension, the Democracy Matrix assesses democratic quality through sociopolitical dimensions and institutions, providing empirical evidence to identify strengths and weaknesses in democratic systems [30–32]. Regarding collaborative governance, the Concordia Partnership Index assesses the ecosystem most conducive to multi-actor governance by analyzing the viability of public–private partnerships, particularly in terms of shared-risk mitigation and the degree to which states and intersectoral actors can leverage their capacities [33–35]. Finally, the Global Peace Index provides a

methodology for measuring the Peace dimension that goes beyond the absence of armed conflict, incorporating dimensions of security and stability and thereby facilitating international comparisons and trend analysis [36]. In this context, the convergence of these instruments provides a rationale for developing an integrated and multidimensional approach capable of addressing the complexity of sustainable development.

3. Materials and Methods

The methodology is organized into two complementary components that directly correspond to the scope announced in the title. The first comprises the construction of the 5P Composite Index: conceptual delimitation, indicator selection and screening, coverage verification, directionality adjustment, normalization, subindex calculation, weighting, and aggregation. The second comprises analytical evaluation of the resulting index through exploratory covariance diagnostics, sensitivity to scaling and the normalization reference base, alternative weighting and aggregation rules, ordinal stability, leave-one-dimension-out analysis, weight-uncertainty simulation, and external convergence. This separation follows the traceability and sensitivity principles recommended for composite indicators and prevents exploratory diagnostics from being interpreted as confirmatory validation [3,9,37].

3.1. Study Design and Unit of Analysis

The study adopts a quantitative, documentary, and comparative design aimed at constructing and applying a composite index aligned with the 5P framework (People, Planet, Prosperity, Peace, and Partnerships) to assess relative performance across countries. The unit of analysis is the country, and the level of observation corresponds to aggregate indicators obtained from international sources. Consistent with methodological guidance for composite indicators, the construction process explicitly documents indicator selection, directionality, normalization, weighting, aggregation, uncertainty, and sensitivity choices [3,9,37].

3.2. Data, Sample, and Reference Window

The study sample comprises 16 units of analysis (15 countries and one territory): Colombia, Mexico, Israel, Kuwait, the United Arab Emirates, Oman, Türkiye, Saudi Arabia, Jordan, Qatar, Pakistan, Egypt, Bahrain, Iran, Palestine, and Iraq. Colombia and Mexico serve as Latin American reference cases, while the remaining units represent economies and territories in the Middle East and its regional environment.

The study follows a contrasting-case design to examine how different configurations of human well-being, environmental performance, economic dynamism, peace, and cooperation translate into an integrated synthetic metric.

The expression regional environment reflects the variability in international geographical classifications of Türkiye and Pakistan, as well as the status attributed to Palestine as a sovereign territory in statistical repositories. Criteria of completeness, consistency, and interinstitutional comparability were also used to delimit the sample for the eight constituent indicators over the 2020–2023 observation window.

The baseline ordering was estimated for the observations retained in the validated main sample. The study therefore adopts a harmonized cross-sectional design, using the latest comparable observation available for each indicator within the stated analysis window.

Sample selection was organized in two phases. First, temporal and cross-sectional coverage was assessed and only cases with complete information for the eight indicators were retained for the applied ranking, thereby avoiding statistical imputation. Second, an expanded reference database of 32 countries satisfying the same indicator-coverage requirement was assembled. This expanded database is not used to rank additional

countries or to alter the composition of the 16-case applied sample; it is used only to provide a broader covariance basis for exploratory KMO and Bartlett diagnostics and for deriving alternative PCA weights.

Separating the two analytical samples is a deliberate design decision. The 16-case sample preserves the geographic and comparative scope of the applied ranking, whereas the expanded 32-case reference database provides a broader basis for estimating preprocessing parameters and exploratory covariance diagnostics. The expanded database therefore supports sensitivity and exploratory weighting exercises without redefining the substantive country ranking.

Consequently, analyses based on the expanded database (n = 32) are used only as exploratory contrasts to the baseline ranking. This design has important limitations: the covariance structure of the expanded sample may differ from that of the 16 applied cases; the subject-to-variable ratio is four observations per indicator, which remains below conventional recommendations for confirmatory factor analysis; and the empirical PCA weights should neither be given causal meaning nor be treated as substitutes for the normative baseline specification.

Reproducibility is addressed at two levels. For the 16 ranked cases, the baseline index and the direct indicator-level PCA sensitivity scenario can be reconstructed from the source definitions, transformation rules, equations, dimensional aggregation procedure, the PCA weights reported later in the alternative-weighting analysis, and the processed indicator-level matrix reproduced in Appendix A. Replication of the expanded 32-case preprocessing and exploratory covariance stage additionally requires the underlying calculation workbook and its reference-base audit records. Because the complete 32-case roster and exclusion ledger are not reproduced in the article or deposited in a persistent public repository at this revision stage, the manuscript does not claim full independent reproducibility of that auxiliary stage; this boundary is stated explicitly in Section 5 and the Data Availability Statement.

3.3. Construction Process of the 5P Composite Index

The 5P Index was developed through a seven-stage procedure consistent with the construction sequence recommended for composite indicators [37]. The first stage conceptually aligned the indicators with the five dimensions and defined their directionality. This was followed by coverage and temporal checks, assessment of potential redundancy, treatment of extreme values, normalization to a common metric, calculation of dimensional subindices, and definition of the weighting and aggregation structure. The index-evaluation analyses were conducted subsequently and are not operations required for construction of the baseline score.

3.3.1. Coding, Screening, and Conceptual Alignment

Based on the methodological sequence described above, the indicators were coded according to their 5P dimension, operational definition, source, temporal rule, directionality, coverage, redundancy risk, and inclusion status. Table 1 reflects this traceability at the point where the decision is made to retain, reassign, or exclude each variable.

Table 1. Coding and Screening Matrix of Candidate Indicators for the 5P Index (2020–2023 Window).

5P Dimension	Candidate Indicator	Operational Definition	Source	Time Window (Criterion)	Directionality	Expected Coverage *	Risk of Redundancy/Dominance	Decision (F1–F4)	Justification (Summary)
People	Human Development Index (HDI)	Synthetic measure of human development (health, education, and income)	UNDP	Latest data within 2020–2023	↑ better	High	Low	Retain (F1–F2)	Theoretical support and international comparability; proxy for human well-being.
People	Happy Planet Index (HPI)	Well-being and ecological efficiency (well-being vs. footprint)	HPI	Latest data within 2020–2023	↑ better	Medium	Medium	Retain (F1–F2), review (F3)	Complements People; requires sensitivity analysis due to dispersion/dominance.
Planet	Green Growth Index (GGI)	Synthetic measure of green growth	GGI	Latest data within 2020–2023	↑ better	Medium–High	Medium	Retain (F1–F2)	Captures environmental-productive performance; assess sensitivity due to variance.
Planet	WEF Nexus Index	Security/resilience of the water–energy–food nexus	WEF/Nexus	Latest data within 2020–2023	↑ better	Medium	Medium	Reassign to Planet (F1)	Aligned with resource security; improves Planet coherence.
Prosperity	GDP growth (average 2021–2022)	Proxy for economic dynamism (short/medium term)	World Bank/IMF/UN	Average 2021–2022 (within 2020–2023)	↑ better	High	Low–Medium	Include (F1–F2)	Retained because of availability and cross-sectional comparability; its sensitivity is controlled in the robustness analysis.
Peace	Global Peace Index (GPI)	Peace/security (invert scale if applicable)	IEP (GPI)	Latest data within 2020–2023	↑ better **	Medium	Low	Retain (F1–F2)	Relevant to Peace; ensure comparable directionality.
Peace	Quality of Democracy	Democratic quality (institutionality)	International index	Latest data within 2020–2023	↑ better	Medium	Medium	Retain (F1–F2)	Captures institutionality; review redundancy if additional governance measures are added.
Partnerships	Concordia Partnership Index	Partnerships and cooperation	Concordia	Latest data within 2020–2023	↑ better	Medium	Medium	Retain (F1–F2)	Direct alignment with Partnerships; verify coverage.
Prosperity (not included)	GDP per capita (PPP)/GNI per capita (PPP)	Structural material prosperity (level)	World Bank/IMF/UN	Latest data within 2020–2023	↑ better	Medium	Medium	Exclude (F1–F3)	Excluded in this version because of missing data and measurement heterogeneity; retained as a future extension.

Note. * Expected coverage indicates anticipated cross-country data availability within the 2020–2023 window. ** For the Global Peace Index, the original lower-is-better scale is inverted before normalization. The symbol ↑ indicates that higher values correspond to better performance after any required directionality adjustment.

To strengthen content validity, a dimension-level congruence review was conducted to verify that the selected indicators adequately represented the construct and that their directionality was homogeneous [9]. For the People dimension, the selection focused on documented well-being and human-development indicators (HDI), to which the HPI was added as a sustainability-oriented measure of well-being. It is important to note that the HPI may amplify cross-country differences because of its dispersion, which justifies including sensitivity analysis to assess its impact [8]. In the Planet dimension, environmental-performance and resource-resilience metrics (Green Growth Index and WEF Nexus Index) were included, because the water–energy–food nexus represents enabling conditions for environmental sustainability and resource security [9].

In the Prosperity dimension, average GDP growth for 2021–2022 was used as a proxy for short-term economic dynamism because it provided homogeneous coverage across the sample. This variable does not represent structural prosperity and may partly capture the rebound following the 2020 contraction. For that reason, it is interpreted separately from income-level measures and its influence is examined through weighting sensitivity and leave-one-dimension-out tests. The indicator should therefore be read as a cyclical prosperity signal within the selected window, not as evidence of sustained improvements in living standards. Future extensions should incorporate GDP per capita, employment, productivity, or national income when consistent coverage is available for all cases.

Because composite metrics were included in the structural development of the index, there is a potential risk of implicit collinearity.

Indicators such as the HDI, HPI, Green Growth Index, and WEF Nexus Index intrinsically contain subdimensions in the socio-environmental and economic fields, while the GPI and democratic-quality measures share attributes related to governance and cooperation capacities.

Given the evident interdependence among the variables, the classification for the 5P index was resolved by identifying the dominant conceptual component of each metric while acknowledging the covariance inherent among the selected alpha variables. For reasons of temporal coverage and international comparability, the Prosperity and Partnerships dimensions were each synthesized using a single indicator, which ensures parsimony and sample consistency while narrowing the descriptive scope of both domains.

These methodological limitations are addressed through weighting simulations, together with unidimensional analyses and leave-one-out validation. For this reason, the resulting instrument is a formative index intended to condense the complexity of development rather than a reflective scale composed of interchangeable variables.

Table 1 presents the structure of the matrix in which the indicators considered for constructing the 5P index are coded and screened.

Table 1 summarizes the screening sequence. The first step assessed conceptual alignment with the 5P framework; the second verified coverage and temporal consistency within the 2020–2023 window; the third examined redundancy, dominance, directionality, and potential sensitivity to extreme observations; and the fourth established the robustness checks used after index construction. Indicators with sufficient and comparable information were retained, while inverse-scale variables were directionally recoded before normalization. Given the multidimensional nature of the construct, indicator retention was guided by theoretical relevance, data comparability, and subsequent robustness analysis rather than by a single statistical criterion [3,7–9]. Source: Authors' compilation based on the international indicator sources listed in Table 1.

Table 1 shows that indicator selection balances data availability, conceptual coverage, scale directionality, and the risk of undue influence from extreme observations. Although this procedure reduces discretion in model specification, partial conceptual overlap among

some source indices remains possible. The aggregate result is therefore interpreted together with omission tests, alternative weighting schemes, and the other sensitivity analyses described below.

3.3.2. Operational Map of the 5P Composite Index

Based on the screening reported in Table 1, the 5P Index is operationalized through dimension-level subindices and a global index. Table 2 summarizes the indicators included, the temporal criterion applied, and the adjustments required to maintain a consistent interpretation across scales (for example, when an original indicator is interpreted in the opposite direction).

Table 2. Operational Map of the 5P Index: Dimensions, Final Indicators, and Measurement Rules (2020–2023).

5P Dimension	Final Indicator	Temporal Rule	Directionality	Transformation/Operational Observation
People	HDI	Latest available data within 2020–2023	↑ better	No additional transformation (subsequent normalization).
People	HPI	Latest available data within 2020–2023	↑ better	Review sensitivity due to dispersion; min–max normalization (baseline).
Planet	Green Growth Index	Latest available data within 2020–2023	↑ better	Min–max normalization (baseline).
Planet	WEF Nexus Index	Latest available data within 2020–2023	↑ better	Placed in Planet for conceptual coherence; min–max normalization (baseline).
Prosperity	GDP growth (average 2021–2022)	Average 2021–2022 (within 2020–2023)	↑ better	Used because of comparative availability; ranking sensitivity to its weight is reported.
Peace	Global Peace Index	Latest available data within 2020–2023	↑ better	If the original scale is inverse, invert before normalization.
Peace	Quality of Democracy	Latest available data within 2020–2023	↑ better	No additional transformation (subsequent normalization).
Partnerships	Concordia Partnership Index	Latest available data within 2020–2023	↑ better	Verify coverage; min–max normalization (baseline).

Note. Baseline normalization uses min–max scaling with parameters estimated from the expanded reference base. Ranking robustness is assessed through alternative weighting rules, z-score standardization, reduced-compensability aggregation, leave-one-dimension-out analysis, constrained weights, and re-estimation of the min–max reference base using the 16 ranked cases. The bilateral p5–p95 winsorization rule is held fixed across these scenarios and is therefore not presented as a separately tested source of uncertainty. Source: Authors' compilation from the indicator providers listed in the table. The symbol ↑ indicates that higher values correspond to better performance after any required directionality adjustment.

Table 2 shows that the dimensions are not composed of the same number of indicators: People, Planet, and Peace each comprise two measures, whereas Prosperity and Partnerships depend on a single indicator.

For the Partnerships pillar, the Concordia Partnership Index was retained because it captures conditions for public–private coordination and multi-actor collaboration. At the same time, it does not fully cover other partnership mechanisms such as trade integration, technical cooperation, or international development finance. This restriction improves

operational transparency but narrows the empirical coverage of the pillar and reinforces the need to interpret the index as a formative composite. The influence of Partnerships is therefore examined through dimension-omission tests and alternative weighting scenarios.

3.3.3. Directionality Adjustment and Normalization

Because the indicators come from heterogeneous scales, the baseline specification transforms them to a common scale from 0 to 100. Equation (1) is applied to indicators with positive directionality.

$$x'_{i,j} = 100\left(\frac{x_{i,j} - min(x_j)}{max(x_j) - min(x_j)}\right) \quad (1)$$

where $x'_{i,j}$ is the normalized value of indicator j for country i, expressed on a scale from 0 to 100; $x_{i,j}$ is the original observed value of indicator j for country i; $min(x_j)$ and $max(x_j)$ are, respectively, the minimum and maximum values of indicator j across the countries in the sample, after the conservative treatment of extreme values has been applied. The subscript i identifies the country and the subscript j identifies the indicator. Equation (1) is applied to variables with positive directionality, for which a higher original value represents better performance: the minimum is transformed to 0, the maximum to 100, and intermediate values retain their proportional position within the observed range.

When the original scale interprets a lower value as better performance, the inverse transformation in Equation (2) is used.

$$x'_{i,j} = 100\left(\frac{max(x_j) - x_{i,j}}{max(x_j) - min(x_j)}\right) \quad (2)$$

where $x'_{i,j}$ is the normalized value of indicator j for country i, expressed on a scale from 0 to 100; $x_{i,j}$ represents the original observed value of indicator j for country i; $max(x_j)$ corresponds to the maximum value of indicator j in the sample; and $min(x_j)$ corresponds to the minimum value of the same indicator. The subscript i identifies the country and the subscript j identifies the indicator. Equation (2) is applied to variables with inverse directionality, that is, variables for which a lower original value represents better performance. With this transformation, the lowest observed value receives 100 and the highest receives 0. In this study it was applied to the Global Peace Index (GPI), whose original scale assigns better peace conditions to lower values; after inversion, a higher value of $x'_{i,j}$ represents better relative performance in the Peace dimension, consistent with the other components of the 5P Index.

3.3.4. Data Coverage and Comparability Criterion

Data coverage was verified before calculating the subindices and the global index. No statistical imputations or weighting readjustments due to partial availability were applied; only countries with sufficient, consistent, and temporally compatible information for all eight indicators were retained. Cases without minimum coverage were excluded from the main ranking. The values used come from the international sources identified for each indicator and were subjected to a common rule for directionality adjustment, treatment of extremes, and normalization. Consequently, all countries included in the ranking were evaluated using the same structure of eight indicators and five dimensions, preventing relative position from depending on different information sets across cases.

3.3.5. Construction of Subindices, Weighting, and Global Aggregation

The baseline specification uses uniform weighting because of its transparency. First, each dimensional subindex is calculated using Equation (3).

$$I_{p,i} = \frac{1}{k_p} \sum_{j \in p} x'_{i,j} \quad (3)$$

where $I_{p,i}$ is the subindex obtained by country i in dimension p of the 5P framework; k_p is the number of indicators assigned to that dimension; $x'_{i,j}$ is the normalized value for country i on indicator j; and the condition $j \in p$ indicates that the sum includes only the indicators belonging to dimension p. The index j runs over the indicators in the dimension. Equation (3) calculates each subindex as a simple arithmetic mean, so all indicators included within the same dimension receive the same internal weight. The result remains on the 0-to-100 scale.

The global index with equal dimension weights is expressed in Equation (4).

$$I_{5P,i} = \sum_{p=1}^{5} v_p I_{p,i} \quad (4)$$

where $I_{5P,i}$ is the global 5P Index score for country i; v_p is the weight assigned to dimension p; $I_{p,i}$ is the subindex of country i in that dimension; and the summation operator runs over the five dimensions, from $p = 1$ to $p = 5$. The weights must satisfy $v_p \geq 0$ and sum to one. Equation (4) aggregates People, Planet, Prosperity, Peace, and Partnerships into a single score from 0 to 100. Because the aggregation is linear and compensatory, high performance in one dimension may partially compensate for low performance in another; therefore, the global result is interpreted together with the dimensional profiles and sensitivity tests.

Under the uniform scenario, dimensional weights are defined using Equation (5).

$$v_p = \frac{1}{5}; \qquad \sum_{p=1}^{5} v_p = 1 \quad (5)$$

where v_p is the uniform weight assigned to dimension p and the value 5 represents the total number of dimensions in the framework. Consequently, $1/5 = 0.20$, so each dimension contributes 20% to the global index. Equation (5) defines the main weighting specification, chosen for its transparency and because it avoids assigning greater normative importance to any of the 5Ps. This assumption is subsequently contrasted with PCA-derived weights.

The baseline specification applies equal weights at the dimension level, not equal weights across the eight indicators. Each of the five dimensions receives 20% of the global score. Within People, Planet, and Peace, the two constituent indicators are averaged equally, so each has an effective baseline contribution of 10% to the final index; Prosperity and Partnerships contain one indicator each and therefore contribute 20% each. By contrast, PCA-derived weights are estimated at the indicator level and are used only as an alternative sensitivity specification.

This scheme is compensatory; therefore, its interpretation is accompanied by ranking-robustness tests to verify that the ordering does not depend on a single weighting or normalization structure [9,12].

The result of this process is a 5P Composite Index expressed on a scale from 0 to 100 for the 16 countries and territories in the main sample, accompanied by five dimensional subindices and a baseline ranking. The construction maintains a common information structure for all cases and makes explicit the decisions concerning directionality adjustment, normalization, weighting, and aggregation. Based on this specification, the following section evaluates the behavior of the index under alternative scenarios and complementary diagnostics.

3.4. Evaluation Strategy for the 5P Composite Index

Once the baseline index had been constructed, its evaluation was organized as an empirical protocol comprising five levels of evidence. The first, conceptual level examined content alignment across the eight indicators and five 5P dimensions. The second retained KMO, Bartlett, PCA, and Cronbach's alpha strictly as exploratory covariance diagnostics, not as validation of a reflective scale. The third and fourth levels examined methodological sensitivity to weighting, normalization, aggregation, the normalization reference base, and dimension omission, together with ordinal consistency through Spearman and Kendall coefficients. The fifth level used Monte Carlo simulation to quantify uncertainty attributable specifically to simultaneous variation in dimension weights. This protocol follows the OECD/JRC emphasis on transparent construction, uncertainty analysis, and sensitivity analysis for composite indicators [37].

3.4.1. Sensitivity, Ordinal Robustness, and Leave-One-Out Test

To verify whether variations in methodological design affect countries' relative positions [10,13,37], ranking robustness was examined through predefined scenarios that alter the weighting structure, scaling method, aggregation rule, normalization reference base, or inclusion of a 5P dimension. Ordinal changes are quantified with Spearman (ρ), Kendall (τ), and the maximum absolute shift. The omission exercise is implemented at the dimension level: each of the five 5P dimensions is removed once, the remaining four are reweighted equally, and the ranking is recomputed. A geometric-mean scenario reduces the degree of compensation allowed by the arithmetic mean, while a constrained-weight experiment evaluates 10,000 admissible vectors in which each dimension receives between 10% and 30% of the total weight. Finally, reference-base sensitivity is tested by re-estimating min–max limits within the 16 ranked cases while retaining the same processed inputs. These checks complement the indicator-level PCA specification and the Monte Carlo weight-uncertainty analysis [4,5,38,39].

Monotonic concordance between the baseline ranking and an alternative scenario is estimated with Spearman using Equation (6).

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{6}$$

where ρ is Spearman's rank-correlation coefficient; d_i is the difference between the position of country i in the baseline ranking and its position in the alternative ranking; d^2 represents the square of that difference; n is the total number of countries compared; and the summation includes all countries from $i = 1$ to $i = n$. Equation (6) quantifies overall ordinal concordance between two rankings: values close to 1 indicate a very similar ordering, values close to 0 indicate no monotonic association, and values close to -1 indicate an inverse ordering. The expression corresponds to the no-ties case; when ties occur, the coefficient is obtained from average ranks and their correlation.

As a contrast based on concordant and discordant pairs, Kendall is used according to Equation (7).

$$\tau = \frac{C - D}{n(n-1)/2} \tag{7}$$

where τ is Kendall's concordance coefficient; C is the number of concordant country pairs, that is, pairs that maintain the same relative order in both rankings; D is the number of discordant pairs, whose relative order changes; n is the number of countries; and $n(n-1)/2$ is the total number of possible pairs. Equation (7) evaluates ranking stability through pairwise comparisons. A value close to 1 indicates high preservation of pairwise ordering; a value close to 0 reflects low concordance; and a value close to -1 indicates

inversion of the ordering. The form presented corresponds to Kendall τ_a without correction for ties; if ties exist, the corresponding adjusted variant is used.

Sensitivity to the exclusion of each 5P dimension is summarized by the maximum rank change in Equation (8).

$$\Delta_d^{LOO} = max_i \left| r_i^{(-d)} - r_i^{(0)} \right| \quad (8)$$

where Δ_d^{LOO} represents the maximum absolute rank shift associated with the exclusion of dimension d; $r_i^{(-d)}$ denotes the position of country i in the ranking recalculated after removing dimension d; and $r_i^{(0)}$ denotes its position in the baseline ranking. The operator max_i identifies the largest absolute displacement observed across all countries. The procedure is implemented separately for each of the five 5P dimensions. In every iteration, one dimension is removed and the remaining four dimensions are reweighted equally before recalculating the composite index and the corresponding ranking. Thus, Equation (8) captures the maximum country-specific sensitivity to the omission of each dimension, while Spearman's ρ and Kendall's τ provide complementary measures of global ordinal stability.

3.4.2. Exploratory Covariance Diagnostics and Alternative Weighting

The formative specification of the 5P Index is established conceptually by the 5P framework: its dimensions are heterogeneous, non-interchangeable components that jointly define the composite rather than reflective manifestations of a single latent variable. KMO, Bartlett, PCA, and Cronbach's alpha are therefore retained only as exploratory covariance diagnostics. KMO and Bartlett assess whether the expanded covariance matrix is technically suitable for exploratory PCA; PCA provides a data-driven alternative weighting vector; and alpha describes within-set covariance patterns. None of these statistics validates, confirms, or causes the formative specification. In particular, the very high covariance in Planet ($\alpha = 0.990$) signals possible redundancy, whereas the negative alpha in Peace signals opposing covariance between its two components after directionality alignment; both findings would be problematic under a reflective interpretation but are treated here as diagnostics requiring transparent sensitivity analysis.

The operational sequence for the expanded reference database of 32 cases and eight indicators is summarized in Table 3. The sequence records the p5–p95 capping rule used before min–max scaling, the common reference-base role of the expanded database, and the exploratory covariance procedures applied after directionality alignment.

Table 3. Statistical Treatment Protocol and Exploratory Evaluation of the I5P–Score (2020–2023).

Stage	Procedure	Operational Criterion	Purpose	Evidence Reported
1	Directionality	Unify interpretation (↑ better); invert scale when applicable	Coherence and Comparability	List of inverted variables
2	Winsorization	p5–p95 (5% in each tail) by variable	Reduce outliers without eliminating cases	p5–p95 rule; four capped case-level values per indicator; exact cut points and case-level audit retained in the calculation workbook
3	Normalization (baseline)	Min–max on a 0–100 scale	Common scale for aggregation	Normalized variables

Table 3. *Cont.*

Stage	Procedure	Operational Criterion	Purpose	Evidence Reported
4	5P Scores	Simple averages by dimension	Identify drivers by pillar	5P subindices by country
5	Weighting (PCA)	PCA on directionally aligned and standardized variables	Alternative empirical weights	Explained variance and weights
6	Covariance diagnostics	Global alpha and subindex alpha for dimensions with two indicators	Complementary description; not reflective validation	Reported α and conceptual interpretation
7	Factorability	KMO and Bartlett	Adequacy for factor techniques	KMO; χ^2, gl, p
8	Ranking robustness	Scenarios (weights, normalization, aggregation, reference base) + leave-one-dimension-out	Ordering stability	ρ/τ, maximum rank changes, and reference-base sensitivity

Note. The sequence prioritizes traceability and ranking robustness; factor statistics are reported as complementary evidence. Source: Authors' methodological design. The symbol ↑ indicates that higher values correspond to better performance after directionality harmonization.

Table 3 distinguishes the construction operations from the complementary diagnostics used to evaluate the index. Directionality, winsorization, normalization, subindex calculation, and baseline aggregation produce the index itself; PCA, covariance diagnostics, and robustness tests are used only to assess how sensitive that result is to alternative analytical choices. Table 4 summarizes the main exploratory statistics and makes this functional separation explicit.

Table 4. Results of Statistical Treatment and Exploratory Diagnostics of the Instrument.

Component	Result/Statistic	Value	Observation
Reference sample	Expanded analytical base	n = 32	Used for winsorization limits, min–max reference parameters, KMO/Bartlett, and PCA.
Winsorization	Rule	p5–p95; 4 capped values per indicator	Applied by variable on the 32-case reference base before min–max scaling; the rule is fixed in the reported robustness scenarios and no countries are removed.
Normalization	Baseline method	Min–Max (0–100)	Reference limits estimated on n = 32 and then applied to the 16 ranked cases.
PCA	Explained variance (PC1)	68.8%	Exploratory PCA on directionally aligned and standardized variables.
PCA	Cumulative explained variance (PC1 + PC2)	84.9%	Used descriptively; no latent-factor validation is claimed.
PCA	Indicator weights (abs. PC1 loadings)	See Table 6	Indicator-level weights normalized to sum to 1 and applied directly to the eight normalized indicators in E3; not first collapsed to dimension weights.

Table 4. *Cont.*

Component	Result/Statistic	Value	Observation
Exploratory covariance	Global alpha (8 indicators)	0.469	Descriptive covariance diagnostic; not a formative-index validity threshold.
Exploratory covariance	Alpha People (HDI, HPI)	0.532	Moderate covariance between conceptually distinct measures.
Exploratory covariance	Alpha Planet (GGI, Nexus)	0.990	Very high covariance; signals possible redundancy and motivates sensitivity analysis.
Exploratory covariance	Alpha Peace (GPI, Democracy)	−0.492	Inverse covariance after GPI directionality adjustment; signals internal tension.
KMO/Bartlett	Global KMO	0.866	Technical adequacy for exploratory PCA; not construct validation.
KMO/Bartlett	Bartlett test	$\chi^2 = 359.75$; df = 28; $p < 0.001$	Rejects an identity correlation matrix but can coexist with multicollinearity.

Note. Cronbach's alpha is reported solely as an exploratory covariance diagnostic. The global value ($\alpha = 0.469$) and the People value ($\alpha = 0.532$) show that the corresponding indicator sets would not satisfy conventional internal-consistency expectations if they were treated as reflective measures; these values do not establish or validate a formative specification. The formative status of the 5P Index follows from its conceptual design, in which heterogeneous and non-interchangeable components jointly define the construct. The Planet coefficient ($\alpha = 0.990$) indicates possible redundancy between the Green Growth Index and the WEF Nexus Index. In Peace, the GPI was inverted before the calculation; the negative coefficient ($\alpha = -0.492$) indicates opposing covariance between the transformed peace/security measure and democracy rather than reliability. Alpha is not calculated for Prosperity or Partnerships because each contains a single indicator. Source: Authors' calculations.

The information presented in Table 4 is complementary to the index-evaluation process and should not be understood as conclusive validation evidence. Its interpretation must be integrated with content validity, traceability of the construction process, dimensional profiles, and ranking stability under different methodological assumptions.

To improve reproducibility, the manuscript reports the native scale, reference year, and transformation rule for each indicator before normalization. Winsorization is defined by a bilateral p5–p95 rule on the 32-case reference base. Under that rule, four case-level values were capped for each indicator in the preprocessing audit. The exact indicator-specific cut points and the expanded-base inclusion/exclusion ledger remain fields of the formula-driven calculation workbook rather than results that can be reconstructed from the 16-case ranking alone. Accordingly, the manuscript does not treat the expanded preprocessing stage as fully independently reproducible from the article by itself; this limitation is stated explicitly in the Discussion and Data Availability Statement. Table 5 documents the native measurement scale and temporal basis of each indicator, while Appendix A reproduces the processed indicator-level matrix used for the 16-case ranking and the direct PCA sensitivity scenario.

Table 5 clarifies an important temporal and measurement issue identified during review: the publication or edition year of a composite index does not necessarily coincide with the observation year of each underlying component. The study therefore treats 2020–2023 as an availability window for a harmonized cross-section. This design improves coverage but also introduces temporal heterogeneity, particularly because the window spans both the pandemic shock and the subsequent recovery. Results are consequently interpreted as a comparative snapshot rather than as a synchronized annual panel.

Table 5. Reference Year, Native Scale, and Pre-Normalization Treatment of the Eight Indicators.

Indicator	Reference/Edition Used	Native Unit/Scale	Temporal Basis and Transformation Before Normalization
Human Development Index (HDI)	2021	Index, 0–1	2021 country value; higher is better.
Happy Planet Index (HPI)	2020	Index score (provider scale)	2020 reference value; higher is better.
Green Growth Index (GGI)	2022	Index score (provider scale)	2022 edition; composite inputs may include lagged source data; higher is better.
GDP growth	Average 2021–2022	Annual growth rate (%)	Arithmetic average of 2021 and 2022; higher is better; interpreted as short-term dynamism.
WEF Nexus Index	2023	Index score (provider scale)	2023 edition using latest available component inputs; higher is better.
Global Peace Index (GPI)	2023	Index score; lower is better in source	2023 edition; inverted before normalization so higher transformed values indicate better peace performance.
Quality of Democracy	2020	Index score (provider scale)	2020 reference release; higher is better.
Concordia Partnership Index	2022	Index score (provider scale)	2022 edition; higher is better.

Note. The year shown is the reference or edition year used in the analytical database. For composite source indices, some component variables may originate from earlier observation years. GDP growth is the arithmetic average for 2021–2022. All indicators are directionally aligned before min–max normalization; the Global Peace Index is inverted because lower source values indicate better performance. Source: Authors' compilation based on UNDP, Happy Planet Index, Global Green Growth Institute, World Bank, WEF Nexus Index, Institute for Economics & Peace, Democracy Matrix, and Concordia.

The next step involved min–max normalization, which homogenized the scale of the indicators within an interval from 0 to 100. The baseline limits were estimated from the 32-case reference base and then applied to the 16 ranked cases. To isolate sensitivity to this reference-base decision, an additional scenario re-estimates the min–max limits within the 16 ranked cases while holding the directionally aligned, preprocessed indicator values fixed. This scenario changes only the normalization reference base; it does not alter the p5–p95 capping rule. Indicators whose source directionality was opposite to the common interpretation were recoded before scaling so that higher transformed values consistently denote better performance.

For the estimation of alternative weights, principal component analysis was implemented as an exploratory tool intended to characterize the interdependence among indicators. The first-component loadings are converted into non-negative normalized indicator weights, but the resulting vector is not interpreted as normative importance or as evidence of a latent 5P factor structure. In the E3 sensitivity scenario, these eight weights are applied directly to the eight normalized indicators. They are not first summed into dimension weights and then reapplied to equally averaged dimensional scores. This distinction preserves the indicator-level differences produced by PCA and keeps the empirical weighting exercise separate from the equal-dimension baseline.

The score of the first principal component is obtained using Equation (9).

$$PC_{1,i} = \sum_{j=1}^{m} a_{j1} z_{i,j} \quad (9)$$

where $PC_{1,i}$ is the score of the first principal component for country i; a_{j1} is the loading of indicator j on the first component; $z_{i,j}$ is the standardized value of indicator j for

country i, with mean zero and standard deviation one; m is the total number of indicators included in the PCA, which in this study equals 8; and the summation runs over $j = 1, \ldots, m$. Equation (9) linearly combines the standardized indicators to represent the covariance pattern that explains the largest proportion of variance in the expanded database of 32 countries. Its use is exploratory and is limited to generating an alternative weighting scenario; it does not validate latent factors or replace the main ranking of 16 countries.

The loadings are transformed into non-negative normalized weights using Equation (10).

$$w_j^{\mathrm{PCA}} = \frac{|a_{j1}|}{\sum_{h=1}^{m} |a_{h1}|} \tag{10}$$

where w_j^{PCA} is the alternative weight assigned to indicator j based on the PCA; a_{j1} is its loading on the first principal component; taking the absolute value of each loading neutralizes the sign to prevent negative weights; a_{h1} is the loading of generic indicator h on the same component; h runs over the m indicators; and the denominator normalizes the absolute values so that the sum of the weights equals 1. Equation (10) transforms the loading matrix into a comparable standardized weighting scheme, applied strictly as a robustness contrast to the original ranking, without assuming that the factor solution validates the latent structure of the 5P framework.

With the data screened, the next step is to evaluate the index and identify its sensitivity and stability when moving away from uniform weighting toward weights derived from the variability of the data themselves. The PCA weights are documented in Table 6.

Table 6. PC1 Loadings and Derived Weights (PCA).

Indicator	PC1 Loading	Weight (abs., Normalized)
HDI_2021	−0.201	0.076
HPI_2020	−0.413	0.157
GGI_2022	−0.413	0.157
GDPg_AVG22_21	0.349	0.133
Nexus_2023	−0.419	0.159
GPI_2023	0.380	0.144
Democracy_2020	−0.039	0.015
Concordia_2022	−0.417	0.158

Note. PC1 corresponds to the first principal component estimated from directionally aligned and standardized variables. The weights are derived from the absolute PC1 loadings and rescaled to sum to one. In E3, the eight weights are applied directly to the eight normalized indicators; implied dimension totals are descriptive only and are not used in the E3 calculation. Source: Authors' calculations from the expanded 32-case database.

Table 6 reports the PC1 loadings and the normalized absolute indicator weights. Before PCA, all variables were directionally aligned, including inversion of the Global Peace Index, and then standardized. The indicator weights are HDI 7.6%, HPI 15.7%, GGI 15.7%, GDP growth 13.3%, WEF Nexus 15.9%, GPI 14.4%, Democracy 1.5%, and Concordia 15.8%. For descriptive purposes only, summing these weights by 5P dimension yields 23.3% for People, 31.6% for Planet, 13.3% for Prosperity, 15.9% for Peace, and 15.8% for Partnerships; these dimension totals are not used to calculate E3. Operationally, E3 is obtained by multiplying each country's eight processed indicator values in Appendix A by the corresponding eight weights in Table 6 and summing the products. No subsequent within-dimension equal averaging is applied. This preserves the large indicator-level contrast between, for

example, Democracy (1.5%) and Concordia (15.8%) and directly addresses whether the PCA weighting itself changes the ranking.

When a dimension contains at least two indicators, alpha is calculated using Equation (11).

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \sigma_j^2}{\sigma_T^2}\right) \quad (11)$$

where α is Cronbach's alpha coefficient; k denotes the number of indicators included in the evaluated set; σ_j^2 is the variance of indicator j; σ_T^2 is the variance of the total score obtained by summing the k indicators; and the summation accumulates the individual variances from $j = 1$ to $j = k$. It is essential to indicate that Equation (11) is used only as an exploratory diagnostic of internal consistency. The 5P framework is conceived as a predominantly formative construct, so its dimensions represent complementary and not necessarily interchangeable facets.

Because the People, Planet, and Peace dimensions contain two indicators, their alpha coefficients are highly responsive to the association between each pair of variables, making automated cutoff rules inappropriate. Given this complexity, inclusion of each component is based on its theoretical role within the 5P structure and its effect on the final ordering. Accordingly, the interpretation process considers three criteria: thematic consistency, potential redundancy, and ranking sensitivity. This approach prevents a single parameter from determining the viability of a formative construct.

The global coefficient is $\alpha = 0.469$. If the eight indicators were interpreted as interchangeable manifestations of a single reflective construct, this value would indicate weak internal consistency; it does not, however, establish the formative nature of the index. The People coefficient ($\alpha = 0.532$) likewise indicates only modest covariance between HDI and HPI and is not interpreted as evidence of dimensional validity. The Planet coefficient ($\alpha = 0.990$) is sufficiently high to raise a redundancy concern between the Green Growth Index and the WEF Nexus Index. For Peace, after directionally inverting the Global Peace Index, $\alpha = -0.492$ indicates opposing covariance between the transformed peace/security measure and democracy in the expanded reference sample. Taken together, these statistics are used only to diagnose covariance patterns and potential redundancy. Retention of the indicators follows from the conceptual 5P architecture, coverage requirements, and the robustness analyses; alpha is not used as an inclusion or exclusion threshold. Alpha is not calculated for Prosperity and Partnerships because each dimension contains a single indicator.

Once an exploratory adequacy of the correlation matrix has been verified, graphical evidence is presented to facilitate interpretation of the components and alternative weights. These figures should be interpreted as descriptive support for the sensitivity analysis, not as confirmation of a definitive factor structure.

The Kaiser–Meyer–Olkin value (KMO = 0.866) and Bartlett's test of sphericity ($\chi^2 = 359.75$; df = 28; $p < 0.001$) show that the expanded correlation matrix is not an identity matrix and can be explored with PCA. However, the ratio of four observations per variable is small, and the strong covariance observed for the GGI–Nexus pair indicates potential multicollinearity. These diagnostics therefore support only the technical use of PCA as a sensitivity device; they do not validate latent 5P factors.

As an exploratory complement, the scree plot is used to observe the concentration of variance in the first components. The retained solution is used to describe covariance and support sensitivity analysis; it is not intended to establish a definitive latent dimensionality for the 5P framework.

To examine how variance is distributed across the principal components, Figure 1 presents the scree plot obtained from the standardized variables.

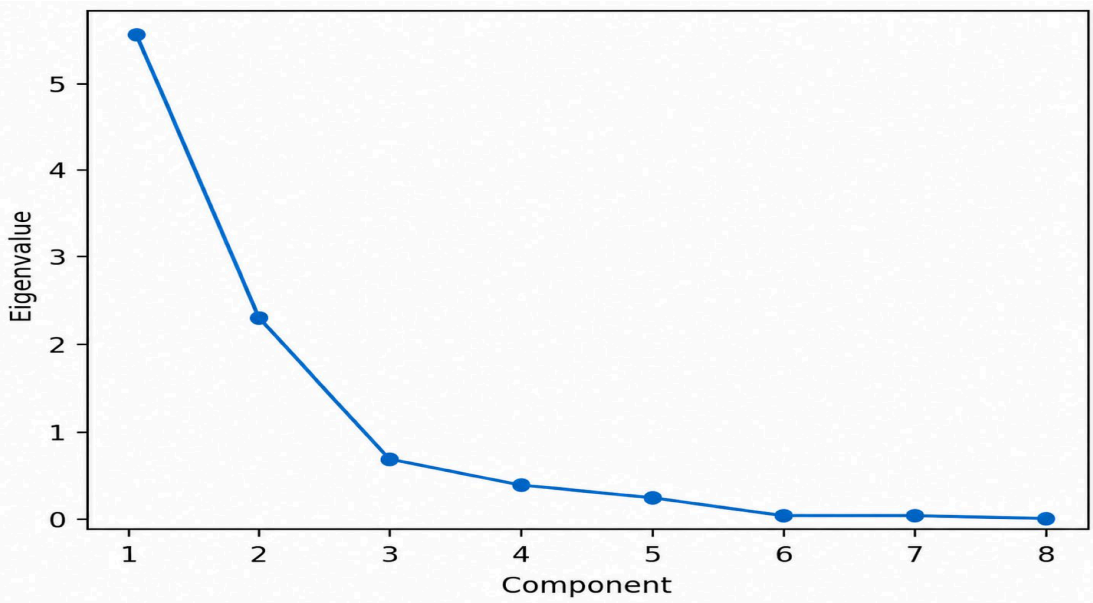


Figure 1. Scree Plot (PCA). Note. Eigenvalues associated with the principal components estimated from directionally aligned and standardized variables. Source: Authors' calculations.

Figure 1 shows a substantial concentration of variance in the first components and a marked decline in subsequent eigenvalues. This pattern supports the parsimonious use of PCA to generate alternative weights, although it does not by itself demonstrate that the 5Ps constitute latent factors or authorize a confirmatory interpretation of the structure.

Because the rotated three-factor solution was not required for weighting and was not sufficiently justified by the sample size, it was removed from the revised analysis. Figure 2 now focuses on the unrotated PC1–PC2 correlation circle, which is sufficient to visualize the dominant covariance pattern without implying a validated latent-factor structure.

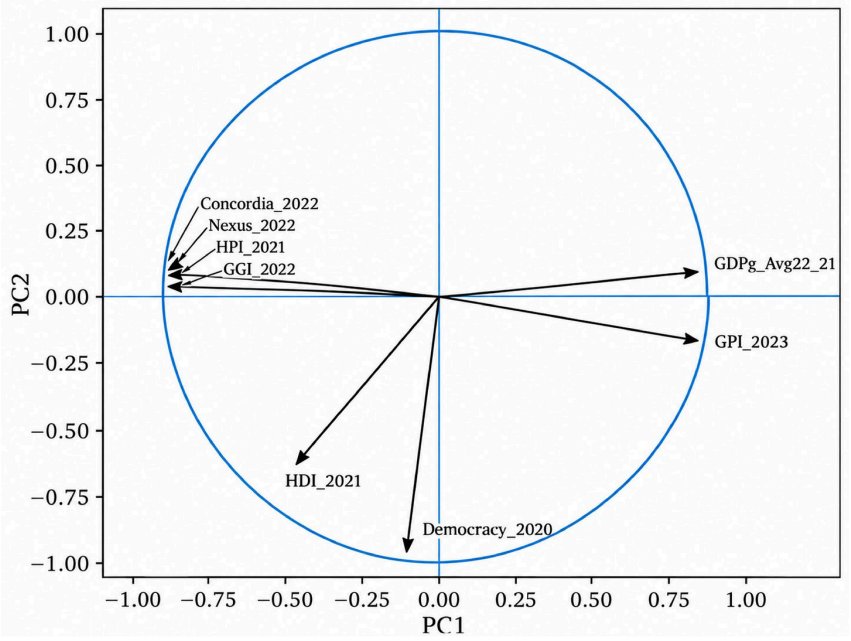


Figure 2. *Cont.*

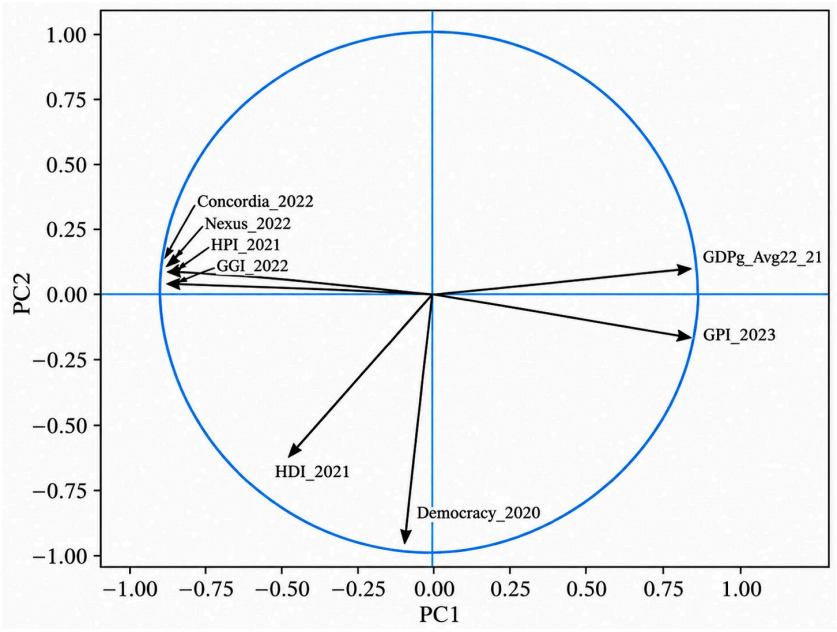


Figure 2. Correlation Circle (PCA). Note. Correlation circle in the PC1–PC2 plane. Vector length approximates the quality of representation of each indicator; direction reflects its association with the components. Source: Authors' calculations.

In Figure 2, the proximity and orientation of the vectors make it possible to identify positive associations, oppositions, and variables with low representation in the PC1–PC2 plane. Consequently, the graph should be interpreted as a covariance diagnostic useful for index sensitivity, not as definitive validation of the normative dimensions of the 5P framework.

Finally, Figure 3 reports the distribution of the weighted I5P–Score and indicates tertile cut-off points. This visualization supports a descriptive reading by groups (low, medium, and high) and makes it possible to assess whether the score distribution is concentrated within narrow ranges, which is relevant when interpreting small differences between adjacent ranking positions.

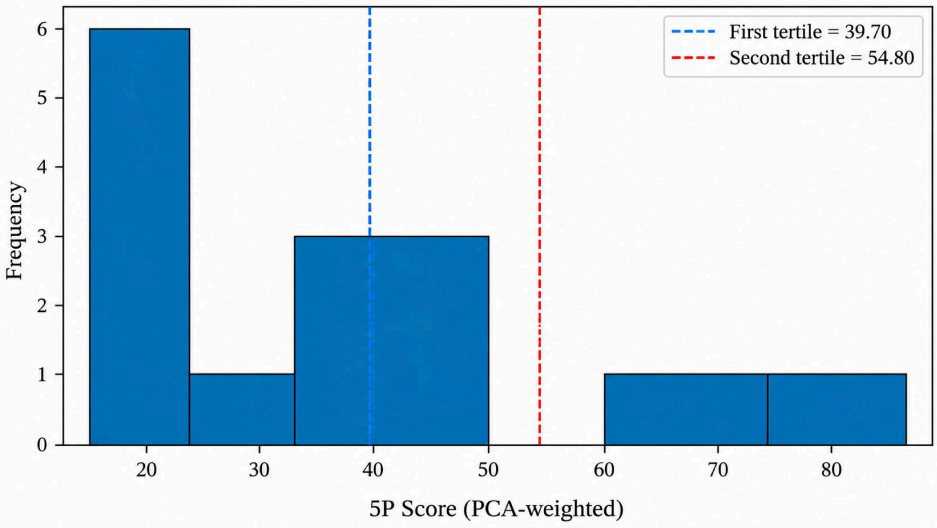


Figure 3. Distribution of the Weighted I5P–Score and Tertile Cut-off Points. Note. Distribution of the weighted I5P–Score (PCA weights) and tertile cut-off points for descriptive classification. Source: Authors' calculations. Dashed vertical lines indicate the first and second tertile cut-off points used for the descriptive classification.

Figure 3 shows that the scores are not uniformly distributed and that clusters exist in different performance segments. Tertile cut-offs facilitate descriptive classification, but the small distances around each threshold advise against rigid labels and support complementing the interpretation with uncertainty intervals and dimension-level profiles.

Finally, robustness was assessed by comparing the baseline ranking with normative weighting scenarios, the direct indicator-level PCA composite, alternative normalization and aggregation rules, leave-one-dimension-out tests, constrained dimension weights, and re-estimation of the normalization reference base. Rank correlations summarize global concordance, whereas maximum rank change identifies country-specific sensitivity. Consequently, a high correlation does not imply invariance: stability may be strong across performance blocks while exact positions still change. The fixed p5–p95 winsorization rule is not varied in these scenarios, and no leave-one-country-out result is interpreted as evidence of outlier-treatment sensitivity. Country identities associated with maximum shifts are reported only where they are directly determined by the recalculated scenario.

3.4.3. Monte Carlo Weight-Uncertainty Protocol

Deterministic sensitivity indicates how the ranking changes under predefined scenarios, but it does not describe the distribution of positions that may arise when dimension weights vary simultaneously. To address this issue, a Monte Carlo weighting simulation was performed on the five dimensional subindices of the 16 countries. A total of 10,000 iterations were carried out per scenario using a fixed random seed (20,261,009). In each iteration, a vector of five non-negative weights summing to one was generated, the aggregate score of each country was recalculated, and the resulting rank was recorded. Two complementary scenarios were considered: a symmetric Dirichlet distribution with $\alpha = 50$, which concentrates weights around equality, and another with $\alpha = 1$, which allows widely dispersed combinations and functions as a weighting stress test.

The weighting vector for iteration b is obtained using Equation (12).

$$w^{(b)} = (w_1^{(b)}, \ldots, w_5^{(b)}) \sim Dirichlet(\alpha, \ldots, \alpha),\ \sum_{d=1}^{5} w_d^{(b)} = 1, w_d^{(b)} \geq 0 \quad (12)$$

where $w^{(b)}$ is the weighting vector generated in iteration b; $w_d^{(b)}$ is the weight assigned to dimension d in that iteration; d runs over the five 5P dimensions, from 1 to 5; α is the common concentration parameter of the Dirichlet distribution. The weights are constrained to sum to one and each $w_d^{(b)}$ is non-negative. With $\alpha = 50$, the weights are concentrated around 0.20 and represent moderate perturbations of equal dimension weighting; with $\alpha = 1$, the distribution allows more dispersed combinations. Equation (12) therefore generates a complete alternative dimension-weighting structure in each iteration.

For each generated vector, the simulated score for country i is recalculated using Equation (13).

$$I_i^{(b)} = \sum_{d=1}^{5} w_d^{(b)}\, S_{i,d},\ i = 1, \ldots, 16 \quad (13)$$

where $I_i^{(b)}$ is the simulated aggregate score for country i in iteration b; $w_d^{(b)}$ is the random weight assigned to dimension d; $S_{i,d}$ is the dimensional subindex of country i for dimension d, expressed on the 0–100 scale; and the summation combines the five dimensions ($d = 1, \ldots, 5$). After calculating the score for the 16 countries, the values are ordered from highest to lowest and the position attained by each case is recorded. Equation (13) is used to obtain the empirical distribution of ranks, the median, interquartile range, Top-3 and Top-5 probabilities, and the best and worst positions observed.

The simulation complements, but does not replace, the deterministic tests. It is explicitly a weight-uncertainty analysis and does not propagate uncertainty in the source data, normalization limits, or indicator selection. Those dimensions of robustness are examined

separately through the z-score, geometric-mean, constrained-weight, and omission scenarios. The uncertainty tables presented in Section 4.4 therefore quantify the stability of positions under alternative dimension weights rather than total model uncertainty.

4. Results

The results are presented following the same analytical distinction used in the methodology. First, the product of the index-construction process is documented: final composition, scores, subindices, classification, and national profiles. The evaluation of its behavior is then presented through ordinal robustness, sensitivity, weighting uncertainty, and external comparison. This organization makes it possible to verify, on the one hand, how the 5P Composite Index is operationalized and, on the other, the extent to which its conclusions are maintained or change under alternative methodological decisions.

4.1. Results of the Construction of the 5P Composite Index

Construction culminated in a specification comprising eight indicators distributed across the five dimensions of the 5P framework, normalized on a common scale and aggregated using the uniform-weighting structure defined as the baseline scenario. This section presents the resulting index and does not yet constitute a robustness test; its purpose is to show the empirical product of the methodological decisions described in Section 3.3, including global scores, the ordering, and classification by performance blocks.

Figure 4 presents the aggregate 5P Index score (0–100) for the analyzed sample. The ranking places Colombia (83) first, followed by Israel (76) and Mexico (64). A second group comprises Kuwait (60), the United Arab Emirates (55), and Oman (54). Intermediate positions are occupied by Türkiye (51), Saudi Arabia (48), Jordan (46), and Qatar (42). Finally, the lowest scores correspond to Pakistan (38), Egypt (37), Bahrain (36), Iran (36), Palestine (34), and Iraq (32).

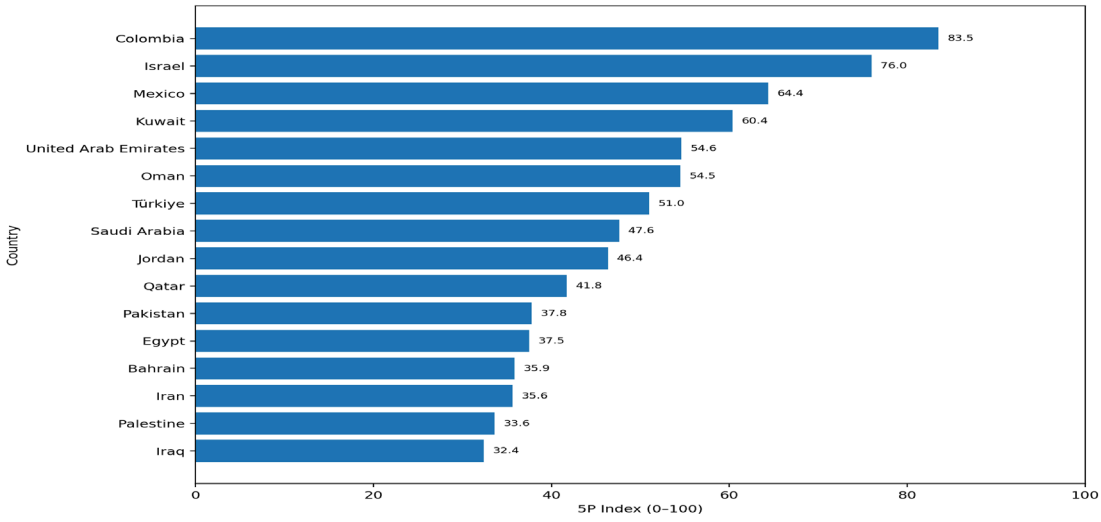


Figure 4. 5P Index by Country (Aggregate Score, 0–100 Scale). Note. Aggregate 5P Index scores (0–100) calculated from normalized indicators. Interpretation of the ranking is accompanied by dimension-level results and sensitivity/robustness tests [3,9].

For descriptive purposes and without overinterpreting small differences, the ranking can be read in three blocks:

- High block (≥60): Colombia, Israel, Mexico, and Kuwait.
- Medium block (42–55): United Arab Emirates, Oman, Türkiye, Saudi Arabia, Jordan, and Qatar.

- Low block (≤38): Pakistan, Egypt, Bahrain, Iran, Palestine, and Iraq.

This block-based reading is complemented by analysis by dimension (5P) to identify which components drive relative performance, as well as by ranking-robustness tests under alternative specifications, as recommended by methodological frameworks for comparative composite indices [3,7–9].

The overall comparison among the countries and territories analyzed is summarized in Figure 4, which presents the aggregate 5P Index score on a common scale from 0 to 100.

Figure 4 makes it possible to identify a top block led by Colombia and Israel, followed by Mexico and Kuwait, as well as intermediate and lagging groups. However, the exact ordinal position should not be interpreted as an absolute measure of sustainability: the result depends on the coverage of eight indicators, the temporal window, and the compensatory nature of the aggregation.

4.2. Dimensional Composition and Performance Profiles

Once the eight indicators had been normalized according to Equations (1) and (2), aggregation was performed in two stages. First, the People, Planet, and Peace subindices were calculated as simple averages of their two constituent indicators, whereas Prosperity and Partnerships retained the value of the single indicator available in each dimension. Second, the 5P Composite Index was obtained through uniform weighting of the five subindices. Table 7 presents the result of this aggregation and relates each country's dimensional profile to its global score, ordinal position, and classification by performance blocks.

Table 7. Dimensional Subindices, Global Score, Ordinal Position, and Block Classification of the 5P Composite Index.

Rank	Country	People	Planet	Prosperity	Peace	Partnerships	5P Index	Block
1	Colombia	82.00	90.00	84.00	70.00	91.50	83.50	High
2	Israel	90.00	82.00	75.00	45.00	88.00	76.00	High
3	Mexico	74.00	80.00	52.00	48.00	67.90	64.38	High
4	Kuwait	76.00	38.00	78.00	75.00	34.90	60.38	High
5	United Arab Emirates	82.00	45.00	68.00	58.00	20.10	54.62	Medium
6	Oman	72.00	50.00	62.00	61.00	27.50	54.50	Medium
7	Türkiye	70.00	55.00	45.00	42.00	43.00	51.00	Medium
8	Saudi Arabia	75.00	35.00	65.00	55.00	8.10	47.62	Medium
9	Jordan	65.00	48.00	35.00	52.00	31.90	46.38	Medium
10	Qatar	80.00	30.00	58.00	35.00	5.75	41.75	Medium
11	Pakistan	45.00	44.00	42.00	30.00	27.75	37.75	Low
12	Egypt	50.00	35.00	40.00	38.00	24.50	37.50	Low
13	Bahrain	72.00	32.00	45.00	25.00	5.40	35.88	Low
14	Iran	62.00	40.00	38.00	18.00	20.10	35.62	Low
15	Palestine	55.00	42.00	20.00	15.00	36.10	33.62	Low
16	Iraq	48.00	35.00	30.00	18.00	30.90	32.38	Low

Note. People, Planet, and Peace correspond to the simple average of their two constituent indicators; Prosperity and Partnerships are each represented by a single indicator. The 5P Index is the arithmetic mean of the five subindices. The block classification has a descriptive purpose: high (≥60), medium (40–59.99), and low (<40). Small differences between countries that are close in the ranking should be interpreted together with the sensitivity tests and not as conclusive substantive gaps. Source: Authors' calculations.

Table 7 shows that favorable overall concordance of the index can coexist with clearly differentiated dimensional profiles. Colombia is located in the top block because of the combination of favorable results in Planet and Partnerships, without depending exclusively on a single dimension. Israel maintains high aggregate performance because of its strengths in People and Planet, although Peace introduces a relevant compensation effect. Mexico benefits mainly from the environmental component, while Kuwait offsets its relatively lower performance in Planet through stronger results in Prosperity and Peace. Among countries in intermediate and lower positions, shortfalls are concentrated to varying degrees in Peace, Prosperity, or Partnerships. For this reason, classification by blocks provides a more consistent interpretation than an isolated reading of adjacent ordinal positions.

Methodologically, Tables 7 and 8 make it possible to reconstruct the complete sequence leading from indicator normalization to the calculation of subindices and the global score. This traceability is particularly relevant because the use of composite indices as inputs may generate conceptual overlap and compensation effects. Consequently, both tables should be interpreted as support for the construction and comparative evaluation of the 5P Index, and not as evidence of causal relationships between the indicators and the final position of the countries.

Table 8. Ranking Robustness: Scenarios and Stability (Spearman/Kendall).

Scenario	Description	Spearman (ρ) vs. Baseline	Kendall (τ) vs. Baseline	Maximum Rank Change
E0 (baseline)	Equal weights across the five 5P dimensions (20% each)	1.000	1.000	0
E1	Social–environmental emphasis: People 0.25, Planet 0.25, Prosperity 0.20, Peace 0.15, Partnerships 0.15	0.985	0.950	2 (Egypt; Bahrain)
E2	Institutional emphasis: Peace 0.25, Partnerships 0.25, People 0.20, Planet 0.15, Prosperity 0.15	0.988	0.933	1 (multiple cases)
E3	Indicator-level PCA composite: Table 6 weights applied directly to the eight normalized indicators	0.985	0.933	2 (Bahrain)
E4	Z-score standardization of the five dimension scores followed by equal aggregation	0.988	0.950	2 (Pakistan; Bahrain)
E5	Geometric mean of the five dimension scores (reduced compensability)	0.929	0.817	4 (Qatar)
E6	Leave-one-dimension-out (5 tests; remaining dimensions equally weighted)	0.953–0.982	0.850–0.917	2–4; maximum: Palestine when Prosperity is omitted
E7	10,000 admissible weight vectors with each dimension constrained to 10–30%	0.921–1.000 (median 0.991)	0.783–1.000 (median 0.950)	0–5
E8	Reference-base sensitivity: min–max limits re-estimated within the 16 ranked cases; preprocessed indicator values held fixed	0.988	0.950	2 (Pakistan; Bahrain)

Note. ρ denotes Spearman's rank correlation and τ Kendall's rank correlation, each computed against the equal-dimension baseline. Maximum rank change is the largest absolute difference in country position under the specified scenario. E3 applies the Table 6 PCA weights directly to the eight processed indicators in Appendix A. E6 summarizes five leave-one-dimension-out tests; E7 summarizes 10,000 admissible dimension-weight vectors with each weight constrained to 10–30%; and E8 re-estimates min–max limits using only the 16 ranked cases while holding the preprocessed inputs fixed. Source: Authors' recalculations from the dimensional scores in Table 7 and the indicator-level matrix in Appendix A.

On this basis, interpretation of the results is developed along two complementary dimensions: aggregate performance, which summarizes each country's relative position, and dimensional performance, which makes it possible to identify the strengths, shortfalls, and compensations underlying the global score.

Interpretation of the ranking is enriched by reviewing the relative contribution of each dimension (People, Planet, Prosperity, Peace, and Partnerships). In Colombia, the aggregate score is supported mainly by high values in Planet and Partnerships, accompanied by a Prosperity component that also contributes significantly. Overall, the result reflects favorable performance in more than one dimension rather than dependence on a single pillar.

In Israel, the high position is explained by strong performance in People and Partnerships, together with a high level in Planet. However, the Peace subindex is visibly lower, producing a high global result with an important internal contrast: strengths in well-being/institutionality and sustainability, but a relative shortfall in stability or security.

In Mexico, the component that most strongly drives the aggregate score is Planet. Even so, the final result is moderated by relatively lower values in Prosperity and Peace, suggesting robust environmental performance that is not reflected with the same intensity in structural prosperity or stability.

Kuwait illustrates the compensatory nature of the index: it stands out in Peace and Prosperity, while its performance in Planet is low. This shows that similar aggregate scores may arise from different internal combinations across dimensions.

In the lower aggregate-performance group, the most frequent pattern is the presence of shortfalls in Peace and/or Prosperity, with additional weaknesses in Partnerships in some cases. These results describe profiles with relative limitations in stability, cooperation/partnerships, or the economic base, which are ultimately reflected in the global index.

4.3. Sensitivity and Ordinal Robustness Assessment

The evaluation begins by examining whether the ordering obtained during construction remains similar when reasonable design decisions are modified. To this end, the baseline ranking is compared with alternative specifications using the Spearman and Kendall coefficients and is complemented by the maximum rank change and leave-one-out tests. Overall concordance and the sensitivity of individual countries are interpreted jointly, because a high correlation may coexist with relevant shifts in specific cases.

Because the ranking may vary with weighting, normalization, aggregation, reference-base definition, and omission of dimensions, the sensitivity analyses were recalculated from the processed indicator-level matrix and the dimensional scores. Table 8 reports the baseline, two normative dimension-weighting schemes, the direct indicator-level PCA composite (E3), z-score standardization, geometric-mean aggregation, leave-one-dimension-out tests, the constrained-weight experiment, and the 16-case reference-base normalization scenario (E8). All rank statistics are computed against the same equal-dimension baseline.

The stability of positions under methodological changes is summarized in Table 8. The table compares the baseline ranking with the main alternative scenarios using the Spearman and Kendall coefficients, together with the maximum absolute shift observed for a country.

Table 8 shows that the baseline and equal-dimension specification are identical by construction. The two normative weighting scenarios produce only small shifts. The direct indicator-level PCA scenario E3 retains high concordance with the baseline (Spearman ρ = 0.985; Kendall τ = 0.933), but it now preserves the indicator-level contrast in Table 6 and produces a maximum shift of two positions for Bahrain. The geometric mean and leave-one-dimension-out tests generate larger changes: Qatar records a four-position shift under

geometric aggregation, and Palestine records the largest leave-one-dimension-out shift (four positions when Prosperity is omitted). Re-estimating the min–max reference limits within the 16 ranked cases (E8) also preserves high global concordance ($\rho = 0.988$; $\tau = 0.950$), with maximum two-position shifts for Pakistan and Bahrain. These results support greater stability at the performance-block level than at the exact-position level, while showing that aggregation and dimension omission remain the more influential design choices.

Taken together, the tests show high overall concordance across specifications, but not complete invariance of positions. Some countries in the middle segment experience substantial shifts, particularly in leave-one-out exercises. Consequently, the result should be interpreted by performance blocks and dimensional profiles, avoiding claims that small ordinal differences represent substantive superiority.

A leave-one-country-out check was also considered. Because winsorization and min–max parameters are fixed from the 32-case reference database rather than re-estimated on the 16 ranked cases, removing one ranked country does not alter the transformed scores of the remaining cases; their relative ordering is therefore invariant under this specific deletion check. This is a mechanical consequence of the fixed-reference normalization design. It is reported only to clarify the deletion logic and is not used as evidence of sensitivity to the p5–p95 treatment or to the choice of normalization reference base; those issues are addressed, respectively, as an explicit limitation and through scenario E8.

4.4. Uncertainty Assessment Through Monte Carlo Simulation

In accordance with the protocol developed in Section 3.4.3, ranking uncertainty was evaluated using 10,000 iterations per scenario and five dimension weights generated from a symmetric Dirichlet distribution. Equations (12) and (13) were applied in each iteration. The $\alpha = 50$ scenario represents moderate perturbations around equal 20% dimension weights, whereas $\alpha = 1$ functions as a stress test by allowing much more unequal combinations. For each country, the median rank, interquartile range, probabilities of belonging to the Top-3 and Top-5, and best and worst positions are reported.

Before examining widely dispersed weights, ranking uncertainty was estimated under a scenario of nearly equal weights generated from a Dirichlet distribution with $\alpha = 50$. Table 9 presents, for each country, the median position, interquartile range, and probabilities of being in the Top-3 and Top-5, with the purpose of assessing ordering stability under moderate fluctuations around uniform weighting.

The results in Table 9 indicate that, when weights fluctuate moderately around equality, the top positions exhibit high stability and the interquartile ranges tend to be small. This behavior supports the baseline specification, although it does not eliminate uncertainty for countries located near the boundaries between performance blocks.

Subsequently, a more demanding scenario was evaluated in which dimension weights may vary considerably across iterations through a Dirichlet distribution with $\alpha = 1$. Table 10 reports the same statistics as Table 9 and identifies which countries retain relatively robust positions and which become more sensitive when the relative contribution of the five 5P dimensions changes substantially.

Table 10 shows that ranking variability increases when extreme weights are allowed. Colombia and Israel frequently remain in the top positions, whereas several countries in the middle segment display a wider range of positions. This result confirms that uncertainty is concentrated mainly in cases whose dimensional profiles depend more strongly on compensations between strengths and shortfalls.

Table 9. Ranking Uncertainty (α = 50; Nearly Equal Weights).

Country	Median	IQR	P(Top3)	P(Top5)	Best	Worst
Colombia	1	0	1.000	1.000	1	1
Israel	2	0	1.000	1.000	2	2
Mexico	3	0	0.988	1.000	3	4
Kuwait	4	0	0.012	1.000	3	4
United Arab Emirates	5	1	0.000	0.617	5	7
Oman	6	1	0.000	0.383	5	6
Türkiye	7	0	0.000	0.000	6	8
Saudi Arabia	8	0	0.000	0.000	7	9
Jordan	9	0	0.000	0.000	8	10
Qatar	10	0	0.000	0.000	9	12
Pakistan	11	1	0.000	0.000	10	14
Egypt	12	0	0.000	0.000	11	14
Bahrain	13	1	0.000	0.000	11	16
Iran	14	1	0.000	0.000	12	14
Palestine	15	0	0.000	0.000	13	16
Iraq	16	0	0.000	0.000	15	16

Note. The simulation with α = 50 represents moderate variation around equal dimension weights (0.20 each). Median indicates the typical rank; IQR measures rank dispersion. P(Top3) and P(Top5) are the proportions of 10,000 iterations in which a country ranks among the top 3 and top 5, respectively. Best and Worst report the extreme positions observed. Source: Authors' calculations from the five dimensional subindices in Table 7; seed = 20261009.

Table 10. Ranking Uncertainty (α = 1; Widely Varying Weights).

Country	Median	IQR	P(Top3)	P(Top5)	Best	Worst
Colombia	1	0	1.000	1.000	1	2
Israel	2	0	0.976	0.995	1	8
Mexico	3	1	0.597	0.863	3	8
Kuwait	4	1	0.389	0.918	1	10
United Arab Emirates	5	1	0.021	0.506	3	12
Oman	6	1	0.018	0.442	3	8
Türkiye	7	3	0.000	0.251	4	10
Saudi Arabia	8	2	0.000	0.022	5	14
Jordan	8	2	0.000	0.000	5	14
Qatar	10	3	0.000	0.002	4	16
Pakistan	12	3	0.000	0.000	8	16
Egypt	12	2	0.000	0.000	9	16
Bahrain	13	5	0.000	0.000	9	16
Iran	13	2	0.000	0.000	9	16
Palestine	15	4	0.000	0.000	6	16
Iraq	15	1	0.000	0.000	9	16

Note. A low median rank and a small IQR indicate positional stability, whereas a large IQR and a wide Best–Worst range indicate sensitivity to weighting. The α = 1 scenario is intentionally diffuse and should be interpreted as a stress test rather than as a plausible policy weighting scheme. Source: Authors' calculations from the five dimensional subindices in Table 7; 10,000 iterations; seed = 20261009.

The Monte Carlo results complement the deterministic sensitivity analysis by showing that positional stability is strongest at the top of the ranking and that dispersion increases as the admissible weight combinations become more unequal. This evidence reinforces the interpretation of the index in terms of performance blocks and dimensional profiles rather than as a rigid ordinal hierarchy.

4.5. Integrated Synthesis of Construction and Evaluation

A joint reading of the results makes it possible to separate two types of evidence. Construction evidence identifies each country's score, position, and dimensional profile under the baseline specification. Evaluation evidence determines which components of the result are relatively robust and which depend more strongly on weighting, normalization, or the inclusion of specific indicators. This distinction prevents the observed ranking from being equated with a definitive sustainability hierarchy.

In the aggregate 5P Index score, Colombia, Israel, Mexico, and Kuwait are located in the upper block ($\geq$60). However, the dimensional decomposition shows that this position does not arise from homogeneous performance but from different internal combinations of People, Planet, Prosperity, Peace, and Partnerships. In Colombia, the global result is explained mainly by high levels in Planet and Partnerships, accompanied by a Prosperity component that also contributes significantly. In Israel, the high position is sustained by strengths in People and Partnerships, although a notable shortfall in Peace creates an internally contrasting profile that must be considered when interpreting the ranking. Kuwait, in turn, stands out in Peace and Prosperity, while its performance in Planet is comparatively low; this pattern illustrates the compensatory nature of the index and shows why the aggregate score alone may conceal specific weaknesses.

The recalculated sensitivity results provide qualified support for H1. Rank concordance remains high under the principal deterministic scenarios, but exact positions are not invariant: the geometric-mean and dimension-omission scenarios generate larger shifts than the normative weighting scenarios. H2 is also supported only in relative terms. Performance blocks are generally more stable than exact ranks under moderate reweighting, yet omission of Partnerships produces several block changes. Accordingly, the evidence supports interpreting the index primarily through dimensional profiles and broad performance groups rather than as a rigid ordinal hierarchy.

4.6. External Evaluation Through Comparison with the SDG Index

As a check of external convergence, the aggregate score of the proposed index was compared with the SDG Index reported by the Sustainable Development Report (SDR). This comparison does not constitute criterion validation because the two instruments differ in indicator coverage, temporal structure, and aggregation. Its purpose is to assess whether the country ordering shows a meaningful association with a widely used external sustainability benchmark.

For the 16 comparable cases, Pearson correlation is r = 0.554 (p = 0.026; bootstrap 95% CI: 0.162 to 0.810), Spearman correlation is ρ = 0.503 (p = 0.047; bootstrap 95% CI: $-$0.004 to 0.817), and Kendall correlation is τ = 0.367 (p = 0.052; bootstrap 95% CI: 0.000 to 0.685). The point estimates indicate moderate convergence, but the wide intervals—especially for the rank-based coefficients—show that the external comparison is imprecise in a sample of 16 cases. Table 11 therefore serves as a country-level diagnostic of convergence and divergence rather than as a definitive validation test.

Table 11. Comparison with the SDG Index from the Sustainable Development Report.

Country	Proposed 5P Index (0–100)	SDG Index (SDR, Score)
Colombia	83.50	70.46
Israel	76.00	74.46
Mexico	64.38	70.80
Kuwait	60.38	63.28
United Arab Emirates	54.62	69.84
Oman	54.50	67.07
Türkiye	51.00	70.61
Saudi Arabia	47.62	65.19
Jordan	46.38	70.99
Qatar	41.75	65.09
Pakistan	37.75	56.97
Egypt	37.50	68.09
Bahrain	35.88	64.41
Iran	35.62	69.56
Palestine	33.62	65.61
Iraq	32.38	63.94

Note. Table 11 compares the 5P Composite Index score, constructed from eight indicators normalized on a scale from 0 to 100 with uniform weighting, with the SDG Index published by the Sustainable Development Report. The comparison is used as external evidence of partial convergence and does not imply conceptual or temporal equivalence between the two instruments. Cases without comparable information are not included in the correlation calculations. Source: Authors' calculations based on the proposed 5P Index and the Sustainable Development Report.

The country-level comparison confirms that convergence is uneven. Colombia scores 13.04 points higher on the proposed 5P Index than on the SDG Index, whereas Iran (−33.94), Palestine (−31.99), Iraq (−31.56), and Egypt (−30.59) show the largest negative gaps. These divergences are consistent with differences in coverage and weighting, particularly the stronger representation of Peace, Partnerships, and short-term Prosperity in the 5P architecture. Consequently, the external comparison supports partial convergence but also identifies cases in which the two measurement systems capture different aspects of sustainability.

5. Discussion

The study contributes along two complementary analytical lines. First, it constructs a transparent composite metric aligned with the 5P framework and documents the steps linking source indicators to dimensional and aggregate scores. Second, it evaluates how the resulting ordering changes under alternative weighting, normalization, aggregation, reference-base, and dimension-omission assumptions. The bilateral p5–p95 winsorization rule is held fixed in these analyses; therefore, the study does not claim robustness to alternative winsorization cut points. The contribution should consequently be understood not as the production of a single ranking, but as a framework that reports the ranking together with evidence on its stability, sensitivity, and interpretive limits, consistent with methodological guidance for composite indicators [2–6,9].

The results show that integrated sustainability does not follow a homogeneous pattern across countries. Colombia leads the baseline scenario because its aggregate score combines particularly strong results in Planet and Partnerships; however, this position should not

be interpreted as evidence of overall superiority or as implying the absence of important weaknesses in other dimensions.

Israel is located in the upper block because of its performance in the People, Planet, and Partnerships dimensions, although the Peace dimension introduces an internal tension. Mexico receives an important boost from People and Planet, but the Prosperity and Peace components moderate its score. Kuwait, for its part, clearly illustrates the compensatory bias of the index, since its results in Prosperity and Peace coexist with weak environmental performance. These patterns confirm that the aggregate score should be examined jointly with the dimensional decomposition rather than as a unidimensional attribute of countries [3,9,11,12].

The comparison between the Latin American reference cases and the Middle Eastern and neighbouring cases does not support a uniform regional hierarchy. Observed differences instead reflect country-specific combinations of human well-being, environmental performance, economic dynamism, institutional conditions, peace, and cooperation capacity. For Colombia and Mexico, the policy implications point to consolidating social and environmental achievements while strengthening peace, institutional quality, and partnerships. For Israel, Kuwait, the United Arab Emirates, and Oman, the index likewise shows that relatively high aggregate scores may coexist with environmental, institutional, or peace-related imbalances.

These findings are consistent with previous empirical applications of the 5P framework, which document substantial heterogeneity across the five sustainability dimensions and show that countries or territories do not necessarily perform uniformly across all pillars [40,41]. The present results extend this evidence by showing that such heterogeneity persists in a cross-regional comparison and that exact rank positions become less informative when dimensional trade-offs are substantial.

For countries affected by conflicts and cooperation constraints, shortfalls in Peace and Partnerships become particularly relevant. These implications must be understood as strictly descriptive and comparative; therefore, they do not demonstrate causal relationships between specific policies and the position attained [1,18,22,23,33–36].

Indicator composition remains an important limitation. Using established composite indices as inputs broadens thematic coverage but also creates a risk of conceptual overlap because the HDI, HPI, Green Growth Index, WEF Nexus Index, GPI, Democracy Matrix, and Concordia indicators contain partially related components.

On the other hand, the GDP growth rate captures short-term dynamism rather than structural prosperity and is particularly sensitive to the post-2020 recovery context. It is therefore not interpreted as a comprehensive well-being measure. The Partnerships dimension is also represented by a single indicator. These limitations do not invalidate the comparative exercise, but they narrow its construct coverage. The 5P Index should be read as a formative cross-country measure, and future versions should incorporate broader prosperity and cooperation indicators when comparable data become available [2,3,8,11,13].

The robustness assessment combines deterministic weighting scenarios, direct indicator-level PCA weighting, z-score standardization, a less-compensatory geometric mean, leave-one-dimension-out tests, constrained weight ranges, reference-base re-estimation, and Monte Carlo weight uncertainty. The recalculated coefficients show high global concordance under moderate changes, but the largest shifts occur under aggregation and omission scenarios rather than modest reweighting or reference-base rescaling. This distinction is important: overall rank correlation can remain high while specific countries move several positions. The evidence is therefore stronger for broad performance groups than for exact ordering, and the Monte Carlo analysis should be interpreted specifically as uncertainty in dimension weights rather than total uncertainty in the model [4–6,37].

The sensitivity exercise also has a clearly defined reproducibility boundary. Scenario E8 directly tests the influence of the normalization reference base by re-estimating min–max limits within the 16 ranked cases, but the p5–p95 winsorization cut points themselves are held fixed and are not varied. Appendix A makes the processed 16-case matrix transparent and permits independent reconstruction of the baseline ranking and the direct indicator-level PCA scenario. By contrast, the complete 32-case roster, exclusion ledger, and case-level winsorization audit remain in the authors' calculation records and are not publicly deposited at this revision stage. Independent replication of that auxiliary preprocessing and exploratory covariance stage therefore still requires access to the calculation workbook. This limitation is stated openly rather than being treated as evidence already tested by the robustness analysis.

Given this context, external benchmarking is useful as an additional diagnostic. The moderate correlations with the SDG Index indicate partial convergence, but the bootstrap intervals are wide and the country-level gaps are systematic rather than random. The proposed 5P Index places comparatively more emphasis on Peace, Partnerships, and the selected Prosperity proxy than the broader SDG architecture, which helps explain why conflict-affected cases tend to score lower in the 5P ranking. External comparison therefore informs interpretation but does not establish functional equivalence [5,15,16].

At the operational level, the instrument helps identify gaps when the aggregate score is interpreted jointly with dimensional subindices and the documented uncertainty of the ordering. For public management, its main value lies in showing where shortfalls are concentrated, how compensation across dimensions affects the aggregate result, and which conclusions remain stable under reasonable methodological changes.

6. Conclusions

In response to the general objective, the research constructed a comparable 5P Composite Index for Colombia, Mexico, and 14 countries and territories in the Middle East and its regional environment, and evaluated its stability through conceptual, statistical, ordinal, sensitivity, uncertainty, and external-convergence contrasts. This approach explicitly distinguishes instrument construction from the assessment of its robustness and demonstrates that evaluation is not limited to reporting an aggregate score.

The contribution can be summarized at three levels. Conceptually, the study jointly operationalizes People, Planet, Prosperity, Peace, and Partnerships under a formative logic consistent with the 2030 Agenda [1]. Methodologically, it distinguishes the 16-case applied sample from the expanded database used only for exploratory covariance diagnostics and alternative PCA weighting, and it combines explicit normalization rules with ordinal concordance, leave-one-dimension-out analysis, alternative aggregation, reference-base sensitivity, constrained weights, and Monte Carlo simulation with 10,000 iterations per scenario [2–6,9]. At the applied level, the index produces country profiles that reveal strengths, structural imbalances, and compensation mechanisms that are not visible in an isolated aggregate measure.

Regarding the limitations, the external validity of the study is constrained by the lack of exact temporal synchronization among the indicators, the 2020–2023 availability window, conceptual overlap among some source indices, the representation of Prosperity and Partnerships by a single indicator, and the small expanded sample used for exploratory PCA. The covariance diagnostics do not validate the formative specification. Rather, the low global and People alpha values and the negative Peace alpha show that a reflective interpretation would be inappropriate for several components, while the very high Planet alpha signals possible redundancy. The formative specification itself rests on the conceptual 5P architecture and on the non-interchangeable role assigned to its dimensions, not on

Cronbach's alpha. Reproducibility is also incomplete for the expanded 32-case preprocessing stage because the full reference-base roster, exclusion ledger, and exact winsorization cut points are not publicly deposited.

These conditions do not invalidate the formative specification, but they require interpretation beyond the aggregate score. Dimensional profiles and sensitivity analyses are therefore essential complements to the final ranking. Because the baseline uses linear aggregation, strong performance in one dimension can compensate for weakness in another, so a high aggregate score may coexist with specific shortfalls [9,12].

Future research should extend temporal and geographic coverage, incorporate direct measures of structural prosperity and international cooperation, revisit redundancy within Planet and the internal tension within Peace, and test non-compensatory aggregation rules in independent samples. Future work should also publicly deposit the complete reference-base roster, exclusion ledger, processed data, and formula-driven calculation files, and should vary winsorization limits jointly with normalization and indicator selection. Within its current scope, the index is best interpreted as a transparent comparative diagnostic of 5P profiles rather than as a definitive country hierarchy.

Author Contributions: Conceptualization, A.A.A. and J.G.D.l.V.M.; methodology, A.A.A., J.G.D.l.V.M. and H.F.P.; formal analysis, A.A.A. and H.F.P.; investigation, A.A.A., J.G.D.l.V.M., H.F.P., J.D.F.R., J.J.R.A. and Z.R.P.; data curation, A.A.A., J.D.F.R., J.J.R.A. and Z.R.P.; writing—original draft preparation, A.A.A.; writing—review and editing, A.A.A., J.G.D.l.V.M., H.F.P., J.D.F.R., J.J.R.A. and Z.R.P.; visualization, A.A.A. and J.D.F.R.; supervision, A.A.A.; project administration, A.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. The study is based exclusively on aggregated, publicly available country-level data and did not involve human participants or animals.

Informed Consent Statement: Not applicable. The study did not involve individual participants or personally identifiable information.

Data Availability Statement: The international source indicators used in this study are publicly available from the databases identified in the methodology and references. Appendix A reproduces the processed indicator-level matrix used to calculate the 16-case baseline ranking and the direct PCA sensitivity scenario, thereby allowing those results to be independently reconstructed from the manuscript. The formula-driven calculation workbook and the expanded-reference-base audit records used for p5–p95 preprocessing, min–max reference parameters, KMO/Bartlett diagnostics, and exploratory PCA are available from the corresponding author upon reasonable request and can also be provided to the editorial office for verification. The complete 32-case roster, exclusion ledger, and case-level winsorization audit are not currently deposited in a persistent public repository; consequently, full independent replication of that auxiliary stage is not claimed and is identified as a limitation in the Discussion.

Acknowledgments: The authors acknowledge the institutional support of their respective universities and research groups during the development of this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Processed Indicator-Level Data for the Main Ranking

Table A1 reproduces the processed 0–100 indicator values used for the 16 ranked cases. These values enable independent reconstruction of the baseline dimensional subindices and, together with the indicator weights in Table 6, the direct indicator-level PCA sensitivity scenario E3. GPI (inv.) denotes the directionally inverted Global Peace Index, so higher values consistently indicate better performance.

Table A1. Processed indicator-level matrix used for the 16-case ranking (0–100).

Country	HDI	HPI	GGI	WEF Nexus
Colombia	87.00	77.00	86.00	94.00
Israel	96.00	84.00	86.00	78.00
Mexico	78.00	70.00	86.00	74.00
Kuwait	85.00	67.00	31.00	45.00
United Arab Emirates	92.00	72.00	39.00	51.00
Oman	79.00	65.00	45.00	55.00
Türkiye	75.00	65.00	52.00	58.00
Saudi Arabia	83.00	67.00	26.00	44.00
Jordan	69.00	61.00	46.00	50.00
Qatar	91.00	69.00	20.00	40.00
Pakistan	43.00	47.00	43.00	45.00
Egypt	51.00	49.00	32.00	38.00
Bahrain	81.00	63.00	24.00	40.00
Iran	66.00	58.00	38.00	42.00
Palestine	54.00	56.00	43.00	41.00
Iraq	46.00	50.00	34.00	36.00
Country	**GDP growth**	**GPI (inv.)**	**Democracy**	**Concordia**
Colombia	84.00	76.00	64.00	91.50
Israel	75.00	35.00	55.00	88.00
Mexico	52.00	43.00	53.00	67.90
Kuwait	78.00	83.00	67.00	34.90
United Arab Emirates	68.00	61.00	55.00	20.10
Oman	62.00	65.00	57.00	27.50
Türkiye	45.00	38.00	46.00	43.00
Saudi Arabia	65.00	61.00	49.00	8.10
Jordan	35.00	55.00	49.00	31.90
Qatar	58.00	29.00	41.00	5.75
Pakistan	42.00	26.00	34.00	27.75
Egypt	40.00	36.00	40.00	24.50
Bahrain	45.00	18.00	32.00	5.40
Iran	38.00	10.00	26.00	20.10
Palestine	20.00	10.00	20.00	36.10
Iraq	30.00	12.00	24.00	30.90

Note. Values are the processed indicator-level scores used in the analytical ranking. People = mean (HDI, HPI); Planet = mean (GGI, WEF Nexus); Prosperity = GDP growth; Peace = mean (GPI inverted, Democracy); Partnerships = Concordia. The baseline index applies equal 20% weights to the five dimensions. E3 instead applies the eight PCA weights reported in Table 6 directly to these indicator-level values. Source: Authors' processed analytical matrix.

References

1. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. 2015. Available online: https://sdgs.un.org/2030agenda (accessed on 9 October 2026).
2. Saisana, M.; Saltelli, A.; Tarantola, S. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2005**, *168*, 307–323. [CrossRef]
3. Becker, W.; Saisana, M.; Paruolo, P.; Vandecasteele, I. Weights and importance in composite indicators: Closing the gap. *Ecol. Indic.* **2017**, *80*, 12–22. [CrossRef] [PubMed]
4. Permanyer, I. Assessing the robustness of composite indices rankings. *Rev. Income Wealth* **2011**, *57*, 306–326. [CrossRef]
5. Seth, S.; McGillivray, M. Composite indices, alternative weights, and comparison robustness. *Soc. Choice Welf.* **2018**, *51*, 657–679. [CrossRef]
6. Foster, J.E.; McGillivray, M.; Seth, S. Composite indices: Rank robustness, statistical association, and redundancy. *Econom. Rev.* **2013**, *32*, 35–56. [CrossRef]
7. Cherchye, L.; Moesen, W.; Rogge, N.; Van Puyenbroeck, T.; Saisana, M.; Saltelli, A.; Liska, R.; Tarantola, S. Creating composite indicators with DEA and robustness analysis: The case of the Technology Achievement Index. *J. Oper. Res. Soc.* **2008**, *59*, 239–251. [CrossRef]
8. Kelemen, A.; Szabó, Z.K.; Bozóki, S.; Szádoczki, Z.; Hartvig, Á.D. A sensitivity analysis of composite indicators: Min/max thresholds. *Environ. Sustain. Indic.* **2024**, *23*, 100453. [CrossRef]
9. Greco, S.; Ishizaka, A.; Tasiou, M.; Torrisi, G. On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Soc. Indic. Res.* **2019**, *141*, 61–94. [CrossRef]
10. Dobbie, M.J.; Dail, D. Robustness and sensitivity of weighting and aggregation in constructing composite indices. *Ecol. Indic.* **2013**, *29*, 270–277. [CrossRef]
11. Schlossárek, M.; Syrovátka, M.; Vencálek, O. The importance of variables in composite indices: A contribution to the methodology and application to development indices. *Soc. Indic. Res.* **2019**, *145*, 1125–1160. [CrossRef]
12. Mazziotta, M.; Pareto, A. Methods for constructing composite indices: One for all or all for one? *Riv. Ital. Econ. Demogr. Stat.* **2013**, *67*, 67–80.
13. Stevens, S.M. Composite environmental indices—A case of rickety rankings. *PeerJ* **2023**, *11*, e16325. [CrossRef] [PubMed]
14. Sachs, J.; Schmidt-Traub, G.; Kroll, C.; Lafortune, G.; Fuller, G. *SDG Index and Dashboards Report 2018*; Bertelsmann Stiftung and Sustainable Development Solutions Network: New York, NY, USA, 2018.
15. Strezov, V.; Evans, A.; Evans, T.J. Assessment of the economic, social and environmental dimensions of the indicators for sustainable development. *Sustain. Dev.* **2017**, *25*, 242–253. [CrossRef]
16. Patrick, R.; Henderson-Wilson, C.; Lawson, J.; Capetola, T.; Shaw, A.; Davison, M.; Freeman, A. Planetary health indicators for the local level: Opportunities and challenges in applying the happy planet index in Victoria, Australia. *Glob. Health Promot.* **2022**, *29*, 14–23. [CrossRef] [PubMed]
17. Wilson, J.; Tyedmers, P.; Pelot, R. Contrasting and comparing sustainable development indicator metrics. *Ecol. Indic.* **2007**, *7*, 299–314. [CrossRef]
18. WEF Nexus Index. WEF Nexus Index Website. 2023. Available online: https://wefnexusindex.org/ (accessed on 9 October 2026).
19. World Bank. World Bank Open Data: GDP Growth (Annual %). 2022. Available online: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG (accessed on 9 October 2026).
20. Yoon, J. Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Comput. Econ.* **2021**, *57*, 247–265. [CrossRef]
21. Bakker, V.; Verburg, P.H.; Van Vliet, J. Trade-offs between prosperity and urban land per capita in major world cities. *Geogr. Sustain.* **2021**, *2*, 134–138. [CrossRef]
22. Ahmad, M.; Ahmed, N.; Jabeen, M.; Jabeen, G.; Qamar, S.; Chandio, A.A.; Rauf, A. Empirics on heterogeneous links among urbanization, the intensity of electric power consumption, water-based emissions, and economic progress in regional China. *Environ. Sci. Pollut. Res.* **2020**, *27*, 38937–38950. [CrossRef] [PubMed]
23. Ming, J.; Liao, X.; Zhao, X. Grey water footprint for global energy demands. *Front. Earth Sci.* **2020**, *14*, 201–208. [CrossRef]
24. Lestari, E.P.; Rahayu, H.C.; Retnaningsih, T.K.; Suhartono, S. Significant role of the human development index in alleviating poverty. *J. Soc. Econ. Res.* **2022**, *9*, 147–160. [CrossRef]
25. United Nations Development Programme. Human Development Insights. 2021. Available online: https://hdr.undp.org/data-center/country-insights#/ranks (accessed on 9 October 2026).
26. Hot or Cool Institute. Happy Planet Index. 2020. Available online: https://happyplanetindex.org/accessible-hpi/ (accessed on 20 October 2026).
27. Global Green Growth Institute. Green Growth Index. 2022. Available online: https://ggindex-simtool.gggi.org/SimulationDashBoard/global_overview (accessed on 9 October 2026).

28. Jadoon, I.A.; Mumtaz, R.; Sheikh, J.; Ayub, U.; Tahir, M. The impact of green growth on financial stability. *J. Financ. Regul. Compliance* **2021**, *29*, 533–560. [CrossRef]
29. Kararach, G.; Nhamo, G.; Mubila, M.; Nhamo, S.; Nhemachena, C.; Babu, S. Reflections on the Green Growth Index for developing countries: A focus of selected African countries. *Dev. Policy Rev.* **2018**, *36*, 432–454. [CrossRef]
30. Matrix, D. Ranking of Countries by Quality of Democracy. 2020. Available online: https://www.democracymatrix.com/ranking (accessed on 9 October 2026).
31. Lauth, H.J.; Schlenkrich, O. Making trade-offs visible: Theoretical and methodological considerations about the relationship between dimensions and institutions of democracy and empirical findings. *Politics Gov.* **2018**, *6*, 78–91. [CrossRef]
32. Schlenkrich, O. Identifying Profiles of Democracy. In *Origin and Performance of Democracy Profiles*; Vergleichende Politikwissenschaft; Springer VS: Wiesbaden, Germany, 2021; pp. 17–41. [CrossRef]
33. Concordia. Concordia Partnership Index. 2022. Available online: https://concordia.net/annualsummit/2022annualsummit/2022-annual-summit-report/ (accessed on 9 October 2026).
34. Faling, M.; Garard, J.; Schouten, G.; Schulz, K.; Veeger, M.; Vervoort, J.; Rutting, L. Innovative participatory research methods in earth system governance. *Annu. Rev. Soc. Partnersh.* **2017**, *12*, 76–80. Available online: https://research.rug.nl/en/publications/innovative-participatory-research-methods-in-earth-system-governa (accessed on 20 October 2026).
35. Filatova, I.; Nikolaichuk, L.; Zakaev, D.; Ilin, I. Public-private partnership as a tool of sustainable development in the oil-refining sector: Russian case. *Sustainability* **2021**, *13*, 5153. [CrossRef]
36. Vision of Humanity. Global Peace Index 2023. 2023. Available online: https://www.visionofhumanity.org/wp-content/uploads/2023/06/GPI-2023-Web.pdf (accessed on 9 October 2026).
37. OECD; European Union; European Commission; Joint Research Centre. *Handbook on Constructing Composite Indicators: Methodology and User Guide*; OECD Publishing: Paris, France, 2008. [CrossRef]
38. Hajkowicz, S. Multi-attributed environmental index construction. *Ecol. Econ.* **2006**, *57*, 122–139. [CrossRef]
39. Singh, R.K.; Murty, H.R.; Gupta, S.K.; Dikshit, A.K. An overview of sustainability assessment methodologies. *Ecol. Indic.* **2009**, *9*, 189–212. [CrossRef]
40. Marti, L.; Cervelló-Royo, R. Disparities in sustainable development goals compliance and their association with country risk. *Sustain. Dev.* **2023**, *31*, 3038–3051. [CrossRef]
41. de la Vega Meneses, J.G. The Situation of the Mexican Federal Entities under the 5P Approach of Sustainable Development. *Rev. Act.* **2024**, *21*, 61–73. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.