



Building consensus on societal wellbeing: a semantic synthesis of indicators to move beyond GDP

Ida Kubiszewski^{a,*}, Robert Costanza^a, Joseph Eastoe^a, Tianchu Lu^a, Kenneth Mulder^c, Grover Patteson Hernandez^c, Péter Benczúr^d, Sandrine Dixon-Dectève^b

^a Institute for Global Prosperity, University College London, UK

^b Club of Rome, Switzerland

^c Hampshire College, USA

^d Joint Research Centre, European Commission, Brussels, Belgium

ABSTRACT

To achieve sustainable wellbeing for both humanity and the rest of nature, we must shift from a narrow focus on Gross Domestic Product (GDP) to a broader understanding and measurement of sustainable wellbeing and prosperity within the planetary boundaries. Several hundred alternative indicators have been proposed to replace GDP, but their variety and lack of consensus have allowed GDP to retain its privileged status. What is needed now is broad agreement on shifting beyond GDP. We conducted a systematic literature review of existing alternative indicators and identified over 200 across multiple spatial scales. Using these indicators, we built a database to compare their similarities and differences. While the terminology for describing the components of wellbeing varied greatly, there was a surprising degree of agreement on the core concepts and elements. We applied semantic modelling to estimate the degree of similarity among the indicators' components and identified those that represented a broad synthesis. Results show that indicators with around 20 components capture a large share of the overall similarity across the indicators in the dataset. Beyond 20 components, adding additional components yielded diminishing returns in similarity. Based on this, we created a 20-component indicator to serve as a model for building consensus and mapped its relationship to several well-known alternative indicators. We aim for this database and synthesis to support broad stakeholder engagement toward the consensus we need to move beyond GDP.

1. Introduction

Humanity faces a convergence of social, environmental, and economic crises, driven by our addiction to an outdated development paradigm based on fossil-fuelled Gross Domestic Product (GDP) growth at all costs (Costanza, 2023; Dixon-Declève et al., 2020). We need a shift towards a development approach based on achieving sustainable and inclusive wellbeing for the entire integrated system of humans and the rest of nature (Benczur et al., 2024; United Nations Network of Economic Statisticians, 2024).

Governments have long used GDP growth as a proxy for national progress, even though it was never intended as a measure of societal welfare (Kuznets, 1934). However, post-WWII, GDP and its growth became the dominant metric of progress with the assumption that increased economic output leads to improvements in employment, living standards, and welfare (Costanza et al., 2014b). In the present-day context of the Anthropocene, overemphasis on GDP growth is having negative side effects on other contributions to societal wellbeing. This includes increasing inequality of income and wealth. This inequality

leads to a decrease in trust, erosion in social capital, and exacerbation of the public health crisis, amongst other societal problems (Wilkinson and Pickett, 2010).

Another side effect is degrading natural capital and ecosystem services, which manifests as climate disruption, biodiversity loss, air and water pollution, and other environmental impacts (Costanza et al., 2014a).

The problems with using GDP as a measure of societal wellbeing have been known for decades (Costanza et al., 2014c; Fioramonti, 2013; Fleurbaey and Blanchet, 2013; Hoekstra, 2019b; Kubiszewski et al., 2013; Stiglitz et al., 2009). For example, the UN Development Programme (Biggeri et al., 2023; UNDP, 1996) identified several types of GDP growth that undermine human and ecological wellbeing:

1. **Jobless growth** – economic expansion without corresponding employment gains;
2. **Voiceless growth** – growth achieved at the expense of civil liberties and democratic rights;
3. **Ruthless growth** – growth accompanied by rising inequality;

* Corresponding author.

E-mail address: i.kubiszewski@ucl.ac.uk (I. Kubiszewski).

<https://doi.org/10.1016/j.ecolind.2025.114076>

Received 5 August 2025; Received in revised form 13 August 2025; Accepted 17 August 2025

Available online 27 August 2025

1470-160X/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

4. **Rootless growth** – growth through economic globalisation with the destruction of local cultures;
5. **Futureless growth** – unsustainable consumption of finite natural resources, stealing our collective future;
6. **Healthless growth** – growth that undermines public health systems; and
7. **Peaceless growth** – growth that increases conflicts, instability, and wars.

These typologies underscore the need to refocus development goals toward sustainable and inclusive wellbeing, and to assess economic growth not as an end in itself, but only insofar as it contributes to broader social and ecological objectives.

The recognition of the need to move beyond GDP has grown significantly across local, national, and global scales. For example, at the recent UN Summit for the Future “Leaders decided on concrete next steps to develop measures of progress on sustainable development beyond GDP, capturing human and planetary wellbeing and sustainability” (United Nations, 2024). It turns out that hundreds of alternative indicators have already been proposed and implemented at various scales. This abundance of indicators reflects a broad recognition of the limitations of GDP and a broad interest in finding an alternative, but it also poses a barrier to consensus. It is easy to critique GDP; it is far more difficult to agree on what should replace it.

This paper aims to support the consensus-building process by identifying shared conceptual ground across existing alternatives. To explore this, we constructed a database of 213 global wellbeing indicators and used semantic modelling to determine the similarities amongst these indicators. This offers a foundation towards building a broad consensus on post-GDP metrics.

2. Methods

2.1. Definitions

The literature on wellbeing indicators often employs varied terminology. To reduce confusion, we use the following key terms and definitions throughout this paper:

- **Indicator:** A metric that reflects the condition or trend in the wellbeing of a socioecological system. (i.e. indices or metrics)
- **Component:** A statistic or measure contributing to a larger indicator (e.g. literacy rate, inequality, etc.).
- **Cluster:** A group of semantically similar components.
- **Summarising component:** A term that captures the semantic centre of a cluster.
- **Synthesising indicator:** An indicator constructed by combining summarising components with a pre-defined number of clusters (we used 10, 20, 40, and 80 clusters). Two types of synthesising indicators exist: derived and created.
- **Derived:** Constructed using the component with the highest semantic similarity to the centroid of each cluster. These components already exist in our database and are identified algorithmically.
- **Created:** Constructed using manually created summarising components that best represent each cluster’s semantic centre. These components may or may not already exist in the original database. These synthesising indicators employ subjective judgement.
- **Indicators in the database:** All wellbeing indicators included in our compiled database.

2.2. Database Compilation

We constructed a database of wellbeing indicators using the following search terms in both Google and Google Scholar: wellbeing indicator, prosperity indicator, happiness indicator, socioeconomic indicator, life satisfaction indicator, socio-political indicator, development

indicator, quality of life indicator, flourishing indicator, welfare indicator, and progress indicator. We included only English-language indicators, both proposed and calculated. We also used existing compiled lists of indicators from the EU Horizon 2020 WISE and SPES projects, the OECD, and other major studies in this area (Benczur et al., 2025; Hoekstra, 2019a; Jansen et al., 2024; OECD, 2025).

For each indicator in the database, we recorded spatial and temporal scale, area coverage, units, whether it includes an aggregate index or consists of a dashboard, the indicator type (e.g. adjustment to GDP, direct survey, or index) (Costanza et al., 2014c), and components. The components listed were taken directly from the indicator descriptions at the original sources.

Initially, we identified over 400 indicators. After reviewing these, we excluded indicators that had a limited focus (e.g. looked at a small portion of the population, e.g. only immigrants), a very narrow aspect of wellbeing (e.g. only physical health), or wellbeing during a specific situation (e.g. during a pandemic). Our final database includes 213 indicators.

2.3. Clustering and synthesising indicators

We analysed the semantic content of components using the Sentence Bidirectional Encoder Representations from Transformers (S-BERT) model (Devlin et al., 2019; Reimers and Gurevych, 2019). S-BERT is a natural language processing model that generates semantically meaningful sentence embeddings into a Siamese network. It has been shown to outperform other natural language processing models, including in the context of sustainability assessment (Maibaum et al., 2024; Matsui et al., 2022). For our purposes, we used a 768-dimensional model¹ trained on over a billion sentence pairs developed by the Hugging Face Natural Language Process community (Wolf et al., 2020). The semantic similarities between two components are given by the cosine similarity of their embeddings (the cosine of the angle between the two representative vectors). Before embedding, we remove all quantifying terms that did not contribute to the semantic content of a component with regard to the theme or topic. These included words such as “rate”, “proportion”, “population”, “people”, and “percent”.

Once the components are embedded, we apply a hill-climbing clustering algorithm (Bird et al., 2009) using cosine similarity to group components into 10, 20, 40 and 80 clusters. For each cluster, we identify the centroid, determine the similarities between each component in the cluster and that centroid, and find the closest existing component to each centroid (Fig. 1). Table 1 lists the components from one of the clusters produced by the algorithm with 20 clusters. Only the 20 components with the greatest similarity to the centroid are shown in Table 1. Fig. 2 shows four word clouds representing the ‘education’ cluster produced by the 10-, 20-, 40-, and 80-cluster algorithms. This is to show the range of the components that may be included in clusters based on the total number of clusters. As the number of clusters increases, fewer terms appear in each cluster and the clusters become more focused.

The cluster centroids are the closest points to all of the components in a cluster and thus represent the semantic intent of the other components in the clusters. However, it is not possible to reverse the embedding and determine a phrase representing the centroid using S-BERT. Given this limitation, we generate a component for each cluster through two different processes: (1) we *derive* the existing component with the highest average similarity to all other components in its cluster, and (2) we *create* a word or phrase that captures the theme represented by the components. For example, for the cluster in Table 2, we *created* the term ‘Civic Engagement’ as a summarising component. While the *derived* component for that cluster, the existing component with the highest average similarity, is ‘Civic and Social Participation’.

¹ all-mpnet-base-v2 at <https://huggingface.co/sentence-transformers/all-mpnet-base-v2>.

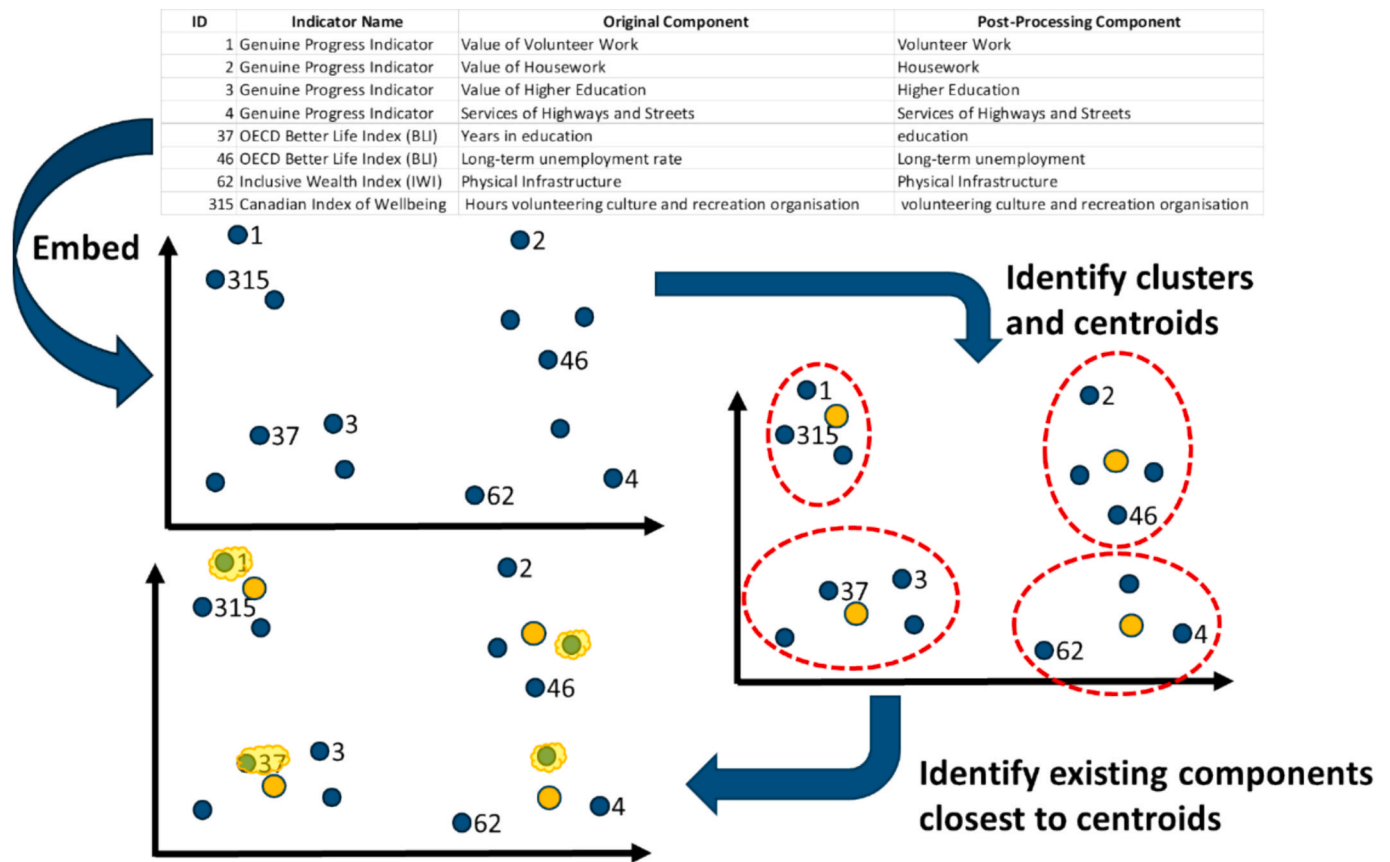


Fig. 1. Embedding and clustering of indicator components. Components are first embedded in a high-dimensional vector space based on semantic content. They are then clustered with each cluster having a high-dimensional centroid (orange circles). We then identify the existing component that is closest to the centroid (yellow clouds in the last figure). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Representative components from one cluster of the 20-cluster synthesising indicator. Repeated components come from different indicators in our database. Only the 20 components with the highest similarity to the centre point are shown.

Component	Similarity to Centroid
Civic and Social Participation	0.851
Social and civic participation	0.848
Social participation	0.846
Social participation	0.846
Social participation	0.846
Participation in community events and activities	0.846
community and civic participation	0.840
Participation in local activity	0.835
Participation in local activity	0.835
Community Participation	0.822
Community participation	0.822
Civic Participation	0.820
Civic participation	0.820
Civic and political participation	0.818
Cultural participation	0.789
Civic activity	0.781
Participation in formal voluntary activities	0.778
Social inclusion and participation	0.773
Civic engagement	0.770
Civic engagement	0.770

By utilising these two methods, we construct eight synthesising indicators (four derived and four created). Each derived synthesising indicator contains the same number of components as the number of clusters (10, 20, 40 and 80), as the algorithm always identifies the best-matching component for each cluster. For the *created* synthesising

indicators, some clusters appear to be ‘catch-all’ clusters that represent disparate components and do not fit any general themes. These we exclude from our created synthesising indicators as no clear representative theme existed. Additionally, when the total number of clusters is high, some clusters can be consolidated as they have largely identical themes, and we chose to represent these by a single component in our *created* indicators. Consequently, the *created* synthesising indicators have 10, 19, 36, and 73 components developed from 10, 20, 40 and 80 clusters.

2.4. Assessing synthesising potential

To determine how well a synthesising indicator (derived or created) captures the semantic content of all the indicators in our database, we used the following procedure.

Given two components and their corresponding embeddings as vectors, \vec{v} and \vec{w} , the similarity between them is given by the cosine of their angle and denoted by $\text{sim}(\vec{v}, \vec{w})$. Consider two indicators and denote them according to the embeddings of their components: $I_1 = \{\vec{v}_i\}$ and $I_2 = \{\vec{w}_j\}$. To determine the degree to which indicator I_1 captures the content of indicator I_2 , we first determine how well each component in I_2 is represented by a component in I_1 by calculating $\text{sim}(I_1, \vec{w}_j) = \max_{\vec{v} \in I_1} \text{sim}(\vec{v}, \vec{w}_j)$, the maximum similarity to \vec{w}_j over all components in I_1 . This gives the similarity with the best-matching component in the proposed indicator. We then define $\text{sim}(I_1, I_2) = \text{avg}\{\text{sim}(I_1, \vec{w}_j) | \vec{w}_j \in I_2\}$. This metric tells us how well, on average, components in I_1 are captured by components in I_2 . The closer this value is to 1, the better I_1 captures the variables measured by I_2 .

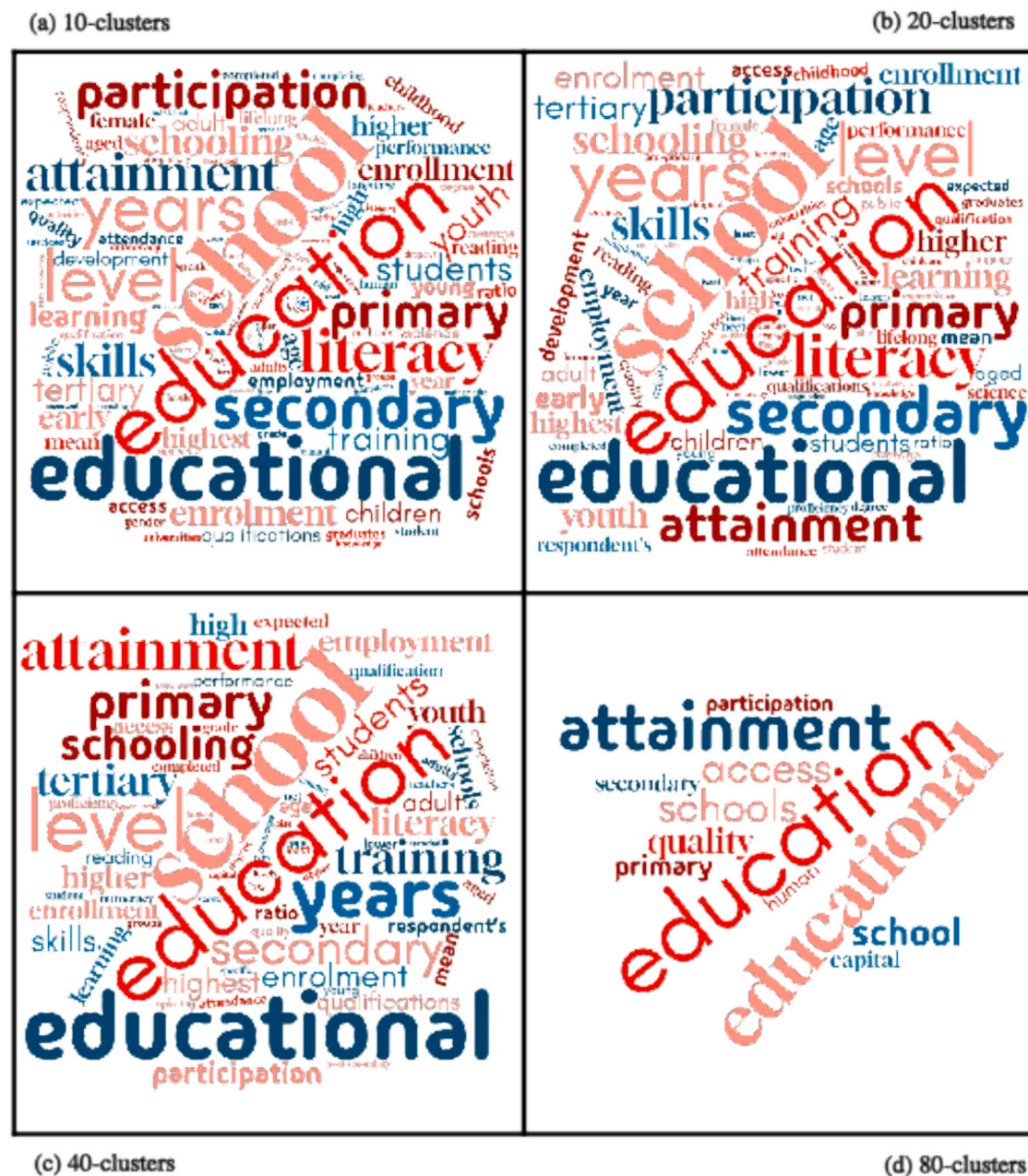


Fig. 2. Word clouds derived from the clusters of components around the topic of education from the different clustering algorithms. Here we show components from the (a) 10-, (b) 20-, (c) 40- and (d) 80-cluster algorithms. This gives a visual representation of the change in the extent and focus of terms based on the number of clusters. We dropped all words appearing less than 4 times for clarity.

To determine how well an indicator I_j serves to capture all other indicators, we calculate its synthesising potential (SP): $SP(I_j) = \text{avg}\{sim(I_j, I_k) | k \neq j\}$. SP is calculated for all the indicators in the database as well as the synthesising indicators. A higher value implies a greater ability to represent the other indicators in the database.

SP is in part determined by the number of components an indicator has, since more assessed variables means both more breadth and depth and thus a greater ability to match the content of other indicators. To take this into consideration when assessing the SP of an indicator, we fit a logarithmic regression model to predict SP based on the number of components. We then identified the best-performing indicators based on the residuals from this model. Indicators with high residuals excel at capturing the content of all other indicators relative to their complexity (number of components).

2.5. Optimal number of components

To determine an 'optimal' number of components, we found the natural breakpoint in the relationship between the number of components and SP. To determine this number of components, we fit a piecewise linear function with a single breakpoint using maximum likelihood estimation. This breakpoint indicates the point of maximum return of SP on component inclusion.

2.6. Popularity scores

To approximate indicator popularity, we collected the number of search results from Google and Google Scholar for each indicator. We used Google Custom Search API and Python scripting to systematically extract the number of Google search results for each indicator using

Table 2

Components for the *derived* and *created* synthesising indicators with 20 components. The components are related to: human capital (blue), social capital (purple), built/financial capital (orange), and natural capital (green). The ‘number of terms’ column indicates how many terms make up each cluster, including repetitions.

DERIVED SYNTHESISING INDICATOR	CREATED SYNTHESISING INDICATOR	NUMBER OF TERMS
Consumption	(Catch-all cluster. <i>Dropped.</i>)	391
Satisfaction with life	Life satisfaction	450
Health	Health	484
Life expectancy at birth	Life expectancy	261
Secondary Education	Education	487
Crime victimization	Crime	351
Civic and social participation	Civic engagement	437
Government Integrity	Governance	486
Income poverty	Inequality	324
Gender equality	Gender equality	223
Housing conditions	Housing	273
Public transport	Infrastructure	310
Financial	Financial security	226
Employment	Employment	399
GDP (per capita)	Per Capita GDP	389
International trade	Business health	343
Protected natural areas	Natural capital	449
Water sanitation	Water quality	345
Air pollution	Air quality	224
Greenhouse gas emissions	Greenhouse gas emissions	467

predefined search strings. Due to restrictions on automated extraction from Google Scholar, we manually retrieved the number of search results for each indicator using Google Chrome’s Incognito mode to minimise personalisation effects. All searches were conducted between 24 and 26 February 2025, to maintain consistency across queries.

2.7. Using popularity scores as weights

We sought to take into account the public and scientific popularity of the indicators both in terms of the clustering algorithm and by determining SP. To account for varying popularity between indicators, we used our two measures of popularity to repeat our assessments, giving greater weight to more popular indicators. Given the significant span of values in our popularity scores (up to eight orders of magnitude), we scaled the popularity scores to make two modified sets of weights, one by log transforming and the other by taking the square root. We used the log-transformed weights in the clustering algorithm for 10 and 20 clusters by multiplying the representation of an indicator in the database by the weights (rounded to whole numbers). We also used both weights (log-transformed and square-rooted) to produce weighted versions of the SP by calculating a weighted average. We also assessed the correlation between popularity and the performance of an indicator as a synthesising indicator.

3. Results

3.1. The database

The database includes 213 indicators. Of these, 85 were proposed by academic researchers, 14 by businesses, 74 by governmental bodies, and 40 by NGOs. In terms of spatial scale, 70.9 % of the indicators we found focused on the national level, 6.1 % on the regional level, 14.6 % on the local level, and 6.1 % were at mixed scales. The majority are composite indicators (68.5 %), while a smaller number are a dashboard (15.5 %),

adjusted GDP (2.3 %), direct surveys (10.8 %), or a mix (2.3 %).

The indicators in the database have, on average, 33 components (max: 295, min: 3).

3.2. The synthesising indicators

We develop two sets of four synthesising indicators, *derived* and *created*. Table 2 shows the summarising components for the 20-component *derived* and *created* synthesising indicators. The summarising components for the other three synthesising indicators (10, 40, 80) are in Table S2. Table 3 illustrates the relationships between the summarising components for the *created* synthesising indicators with 10, 20, and 40 components. Specifically, it shows how the 10-component indicator groups the summarising components from the larger indicators (20 and 40 components) into thematic categories.

3.3. Performance based on synthesising potential

We evaluated the performance of each synthesising indicator using residuals from the predicted SP. For the *created* synthesising indicators, residuals for the 10-, 19-, 36-, and 73-component indicators (Table S1) are 0.105, 0.120, 0.124, and 0.118, respectively. These are generally lower than those of the *derived* synthesising indicators, which have an average residual of 0.130. Table 4 lists the 15 indicators from our database with the highest SP based on the residuals, also showing their number of components.

In general, indicators with more components have a higher SP. Fig. 3 plots the SP against the number of components for *all the indicators in our database with 180 or fewer components* (blue dots and orange diamonds). SP increases rapidly up to 20 components, after which the growth rate decreases significantly. Fitting a piece-wise linear function using maximum likelihood analysis identifies a breakpoint at 22 components. SP continues to increase more gradually up to between 60 and 80 components.

Table 3
Relationships between components of the created indicators.

Number of Clusters		
10	20	40
Life Satisfaction	Life Satisfaction	Life Satisfaction Leisure
Life Expectancy	Crime	Violent Crime
	Health	Mental Health Physical Health Healthcare Child Welfare
Housing	Life Expectancy	Life Expectancy Mortality
		Housing
Protected Areas	Natural Capital	Natural Areas Biodiversity Environmental Sustainability
Infrastructure	Water Quality	Water Quality
	Infrastructure	Transportation Internet Access
Civic Engagement	Governance	Trust in Institutions
	Civic Engagement	Civic Engagement Sense of Community Community Participation
Employment	Gender Equality	Gender Equality Civil Liberties
		Employment
Education	Education	Education
Greenhouse Gas Emissions	Greenhouse Gas Emissions	Greenhouse Gas Emissions Renewable Energy Waste Management
Per Capita GDP	Air Quality	Air Pollution
	Per Capita GDP	Per Capita GDP Household Consumption Household Income Income Inequality
	Business Health	Research and Development Agriculture
		Poverty
	Inequality	Trade
	Financial Security	

Table 4
Top 15 existing indicators ranked by their ability to synthesise the semantic content of other indicators as a function of their number of components.

Name	Number of Components	Synthesizing Potential	Residual
Canada's Quality of Life Framework	19	0.442	0.083
Measuring What Matters Framework	50	0.499	0.076
The European Quality of Life Survey	12	0.400	0.070
Iceland's Indicators for measuring Wellbeing	39	0.474	0.067
Happy City Index	23	0.437	0.065
Iceland Wellbeing Framework	39	0.471	0.065
The Scottish Trends Index of Social and Economic Wellbeing	4	0.322	0.064
Sustainable Society Index (SSI)	22	0.432	0.062
Multidimensional index of sustainability (EU)	10	0.380	0.062
Community Indicators Victoria (CIV)	74	0.509	0.061
THE HAPPINESS IN PENANG (HIP) INDEX	25	0.437	0.059
Indicators Aotearoa New Zealand	128	0.543	0.059
City Prosperity Index (CPI)	25	0.436	0.058
Statistics Portugal Wellbeing Index	9	0.368	0.057
OECD Better Life Index (BLI)	24	0.428	0.053

Fig. 3 also shows the location of our *derived* and *created* indicators, and the five 'high performing' existing indicators. For example, Canada's Quality of Life Framework performs well with 19 components, close to what we identified as an 'optimal' number of components. We identified five indicators from our database as 'high-performing' (Table 5; red diamonds in Fig. 3). These five indicators have some of the highest SP based on the residuals relative to their number of components (Table S2). Fig. 3 also shows that both the created (grey circles) and the derived (black squares) synthesising indicators outperform all other indicators in the database with similar numbers of components.

To give a sense of the overlap with other indicators, we identified five indicators from the database that have the highest average semantic similarity to each of the five high-performing indicators (Table 6).

We found no significant differences in thematic content or synthesising potential based on the type of organisation proposing the indicators (e.g., academic, governmental, or NGO), suggesting a broad conceptual convergence across sectors.

3.4. Popularity scores

Indicator popularity varies widely. The Sustainable Development Goals (SDGs) receive the highest number of Google hits (275 million), while 13 indicators in our database receive no hits at all. On Google Scholar, the Human Development Index has the greatest number of hits (1.29 million), while 14 indicators in our database receive none.

When we reran our analysis, weighting our analysis by these popularity scores, no significant change was seen in the results. Thematic clusters from the weighted clustering algorithm (for both 10 and 20 clusters) closely resemble those from the unweighted version. While SP values differ slightly between the weighted and unweighted analyses (correlation between the two sets of values given by $r = 0.94$), the rank order remains nearly identical ($s = 0.99$, where s is Spearman's rank correlation coefficient). Similarly, there was no correlation between popularity and SP ($r^2 = 0.06$). This result suggests that popular indicators do not necessarily represent other indicators more comprehensively. It also suggests that weighting by popularity does not alter thematic structures.

4. Discussion

The proliferation of wellbeing indicators, characterised by wide variation in topics, number of components, spatial and temporal scales, and measurement units, complicates efforts to reach consensus on a unified indicator. Despite the diversity of indicators, our analysis reveals a high degree of conceptual agreement among existing indicators. Making this underlying similarity more explicit could facilitate the broad consensus needed to transition beyond GDP.

Fig. 3 shows that indicators with more components tend to capture a greater portion of the overall semantic content that exists across all indicators. However, this comes at the cost of increased complexity and a greater demand on data collection resources. Further, this relationship shows diminishing returns: SP increases rapidly up to approximately 20 components, after which the rate of increase slows considerably. This pattern suggests that while increasing the number of an indicator's components can enhance its integrative capacity and comprehensiveness, it also introduces practical trade-offs in terms of complexity, cost, interpretability, and usability. Identifying a "sweet spot," a balance between inclusivity and operational feasibility, may help in designing more effective indicators.

Our synthesising indicators, shown in Fig. 3, Tables 2 and 3, offer a promising approach. By distilling key themes into summarising components, these indicators aggregate the information present in the broader database and produce integrated, balanced representations of wellbeing. In particular, the 20-component synthesising indicator captures the diverse subject areas present across existing indicators while remaining concise, measurable, and usable. This indicator effectively

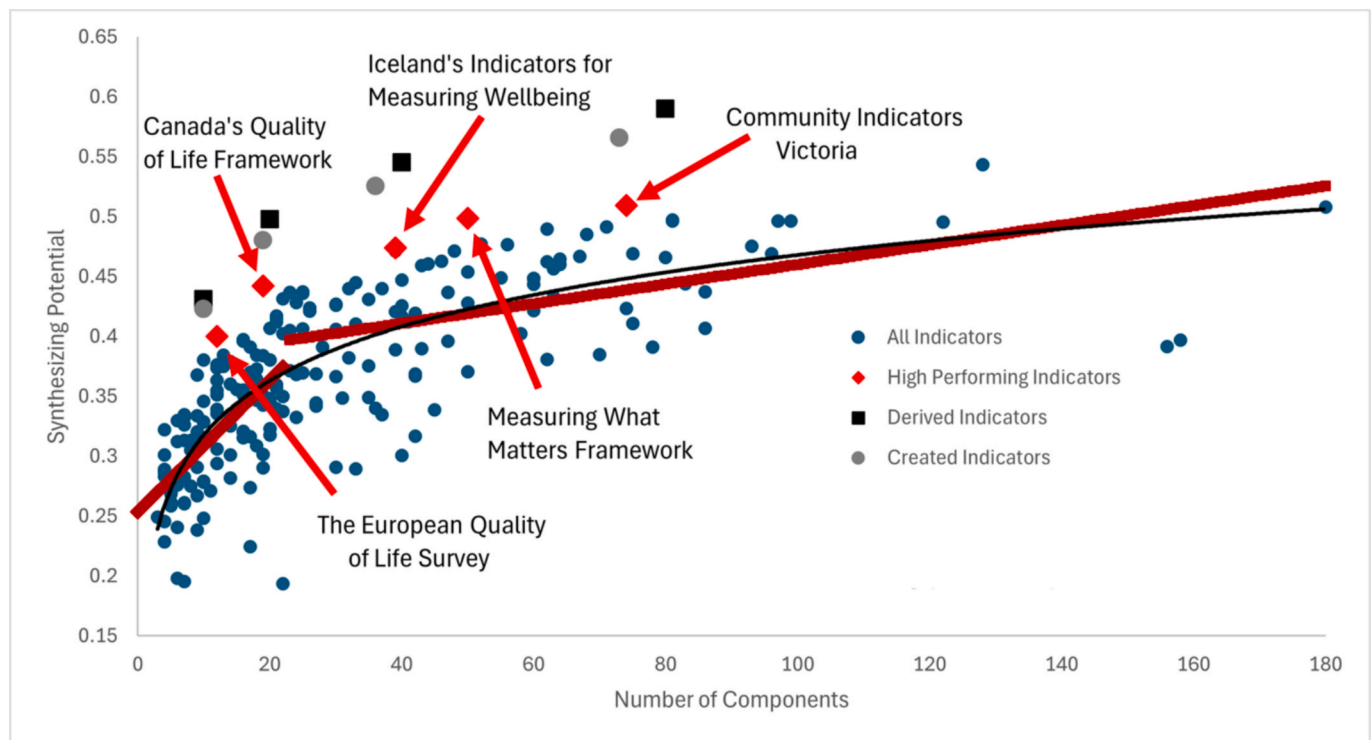


Fig. 3. Synthesizing potential versus number of components for the indicators in the database (blue dots), as well as the identified synthesising indicators, derived (black squares) and created (grey circles). The red diamonds represent the five indicators that we have labelled high-performing, which are reported in Table 5. The black curve is a log function used to predict synthesising potential based on the number of indicators ($R^2 = 0.64$). The two red lines represent the best fit of a piecewise linear function showing the point of greatest difference in synthesising potential as a function of the number of components occurs at $n = 22$ components. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5

Number of components and synthesising potential for five high-performing indicators based on the residuals.

Indicator	Number of Components	Average Similarity	Residual
Canada's Quality of Life Framework	19	0.442	0.083
Measuring What Matters Framework	50	0.499	0.076
The European Quality of Life Survey	12	0.400	0.070
Iceland's Indicators for Measuring Wellbeing	39	0.474	0.067
Community Indicators Victoria (CIV)	74	0.509	0.061

synthesises the core ideas that other indicators attempt to include, offering a comprehensive and manageable solution.

The summarising components from the 20-component synthesis indicator (Table 2) reflect a relatively balanced distribution across capital types or their associated flows: built/financial (6), natural (4), human (4), and social (5). The size of each cluster, shown by the 'Number of Terms' in Table 2, further illustrates how frequently each subject area appears among the components of all indicators in the database.

We also identify several high-performing indicators (see Fig. 3, Tables 5 and 6) that demonstrate strong SP. These cases highlight that conceptual representativeness does not necessarily require extensive detail. Conversely, some indicators with many components exhibit relatively low similarity, suggesting that quantity alone does not guarantee comprehensiveness.

Interestingly, popularity, measured by Google and Google Scholar hits, does not correlate with SP. Popular indicators such as the Sustainable Development Goals (SDGs) and Human Development Index

(HDI) do not necessarily score higher on thematic synthesis. This disconnect indicates that visibility does not guarantee conceptual comprehensiveness, nor that the indicators that are the most synthesising or comprehensive become the most popular.

4.1. Measuring components

Our semantic model and clustering approach effectively synthesise the conceptual structure of existing wellbeing indicators. However, this method does not stipulate how to measure the summarising components. Selecting appropriate measurement strategies will require public discourse and consultation with subject-matter experts for each domain. These strategies may vary based on spatial and temporal scale, population demographics, statistical appropriateness, and data availability. It is also essential to ensure that the indicators are measurable without requiring excessive resources. Emerging technologies may enable more dynamic, real-time, and context-specific measurement approaches that improve feasibility and responsiveness.

4.2. Single index or dashboard

There is an ongoing debate between the use of dashboards of components and single indices that summarise all the components into a single number. A dashboard provides an array of components that can be tracked and prioritised individually. A single index, on the other hand, provides an aggregated number to summarise progress towards an overall goal. In our database, roughly 30 % of indicators use dashboards, while 70 % aggregate those components into a composite index. These approaches are not mutually exclusive. Single indices rely on the underlying components included in a dashboard.

Dashboards and indices both play essential roles in assessing sustainable and inclusive wellbeing across multiple scales. This study

Table 6

Five most thematically similar indicators for each of the five high-performing indicators.

The European Quality of Life Survey	
Statistics Portugal Wellbeing Index	0.660
Australian National Development Index	0.642
Finland Economy of Wellbeing	0.584
Community Wellbeing Index (CWB)	0.568
The Italian Composite Subjective Wellbeing Index	0.556
Canada's Quality of Life Framework	
The Scottish Trends Index of Social and Economic Wellbeing	0.621
Community Wellbeing Index (CWB)	0.593
OECD Better Life Index (BLI)	0.572
Regional Quality of Life Index	0.570
Sustainable Society Index (SSI)	0.569
Iceland's Indicators for Measuring Wellbeing	
Iceland Wellbeing Framework	0.983
The Scottish Trends Index of Social and Economic Wellbeing	0.721
Canada's Quality of Life Framework	0.629
Index of Child Care Environment	0.626
Community Wellbeing Index (CWB)	0.623
Measuring What Matters Framework	
The Scottish Trends Index of Social and Economic Wellbeing	0.703
Canada's Quality of Life Framework	0.699
The Progress Index	0.686
Wellbeing Index (WBI)	0.672
Inequality-adjusted Human Development Index (IHDI)	0.662
Community Indicators Victoria (CIV)	
The Scottish Trends Index of Social and Economic Wellbeing	0.730
Statistics Portugal Wellbeing Index	0.707
The Progress Index	0.698
The European Quality of Life Survey	0.669
Happy City Index	0.657

demonstrates how synthesising indicators can support the development of effective dashboards. However, transforming these dashboards into meaningful indices remains a critical area for further research. Common approaches, such as using weighted or unweighted averages, likely fall short, as they overlook the reality that components within an indicator often function as limiting factors for overall sustainable and inclusive wellbeing (Costanza et al., 2016).

5. Conclusions

This paper analyses 213 existing wellbeing indicators using semantic similarity and clustering to synthesise their thematic content. These indicators, developed by a wide range of organisations using diverse methodological approaches and covering various spatial and temporal scales, nonetheless reveal substantial conceptual overlap. Despite their apparent diversity, this overlap suggests that achieving international consensus on a new indicator to move beyond GDP is both possible and desirable.

While critical challenges remain around measurement, spatial scale, and aggregation methods, targeted dialogue and cross-sector collaboration can address them. Future research needs to prioritise developing appropriate measurement approaches for each of the components and testing the usability of synthesised indicators in diverse contexts.

Many countries already utilise multiple indicators, though often in a limited, uncoordinated manner or with uneven emphasis. Achieving consensus around the summarising components of the 20-component synthesis indicator would improve coherence, balance, and the likelihood of policy uptake.

But moving beyond GDP requires more than a new indicator. It requires a shared understanding of what constitutes a good life. Changing the measures we use alone cannot shift the dominant paradigm, but it is

a necessary step. The current system persists because GDP-based metrics and models (i.e. the System of National Accounts and macroeconomic models) interact with the policies, institutions, rules, and norms to reinforce the narrow goal of GDP growth. Transforming this system requires more than isolated new metrics, models, or policies. It requires an integrated system of all three, aligned toward the overarching objective of Sustainable and Inclusive Wellbeing (SIW) rather than the singular pursuit of GDP growth (Costanza et al., 2024) (Van Eynde et al., 2024).

Our work contributes to this conversation by proposing a synthesised indicator that builds on extensive prior work on wellbeing indicators. Through continued consensus building, we can help drive a global transition toward a more holistic and equitable indicator for measuring and achieving sustainable and inclusive wellbeing.

CRedit authorship contribution statement

Ida Kubiszewski: Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Robert Costanza:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization. **Joseph Eastoe:** Writing – review & editing, Project administration, Methodology, Data curation. **Tianchu Lu:** Writing – review & editing, Project administration, Data curation. **Kenneth Mulder:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Formal analysis, Data curation. **Grover Patteson Hernandez:** Writing – review & editing, Visualization, Methodology, Formal analysis, Data curation. **P  ter Bencz  r:** Writing – review & editing, Validation. **Sandrine Dixon-Dect  ve:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to thank the class of UCL students who helped put together the database. This includes Urban Luka Jurc Blagovic, Berlin Chen, Qiufan Chen, Xuzheng Chen, Tristen Taylor Chow, Justin Dempsey, Jacob Harvey-McMahon, Ragini Kayal, Nadira Luddin, Sara Mahdi, Peter-Paul Mbele, Ornella Rincones Medina, Thomas Metzger, Bri Powell, Imogen Stewart-Green, Chunhong Wan, Jing Wang, and Wenxin Zhang. We also thank the anonymous reviewer for helpful comments on an earlier draft. We acknowledge support from the MERGE project, funded by the European Union (number 101132524).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2025.114076>.

Data availability

Data will be made available on request.

References

- Benczur, P., Boskovic, A., Cariboni, J., Chevallier, R., Le Blanc, J., Sandor, A., Zec, S., 2024. Sustainable and Inclusive Wellbeing, the Road Forward. Publications Office of the European Union, Luxembourg.
- Benczur, P., Boskovic, A., Giovannini, E., Pagano, A., Sandor, A.-M., 2025. Measuring Sustainable and Inclusive Wellbeing: A Multidimensional Dashboard Approach. Publications Office of the European Union, Luxembourg.

- Biggeri, M., Ferrannini, A., Lodi, L., Cammeo, J., Francescutto, A., 2023. The “winds of change”: the SPES framework on Sustainable Human Development. SPES Working paper.
- Bird, S., Klein, E., Loper, E., 2009. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media Inc.
- Costanza, R., 2023. Addicted to Growth: Societal Therapy for a Sustainable Wellbeing Future. Routledge/Earthscan.
- Costanza, R., de Groot, R., Sutton, P.C., van der Ploeg, S., Anderson, S., Kubiszewski, I., Farber, S., Turner, R.K., 2014a. Changes in the global value of ecosystem services. *Glob. Environ. Chang.* 26, 152–158.
- Costanza, R., Fioramonti, L., Kubiszewski, I., 2016. The UN Sustainable Development Goals and the dynamics of well-being. *Front. Ecol. Environ.* 14, 59.
- Costanza, R., Hart, M., Kubiszewski, I., Talberth, J., 2014b. A short history of GDP: moving towards better measures of human well-being. *Solutions* 5, 91–97.
- Costanza, R., Hoekstra, R., Eastoe, J., Rum, I., Kubiszewski, I., O'Neill, D., Biggeri, M., Pollio, C., Zinutti, M., 2024. A synthesis of models, metrics, and policies for a sustainable and inclusive wellbeing future-deliverable D1. 3-Merge Project.
- Costanza, R., Kubiszewski, I., Giovannini, E., Lovins, H., McGlade, J., Pickett, K.E., Ragnarsdottir, K.V., Roberts, D., De Vogli, R., Wilkinson, R., 2014c. Time to leave GDP behind. *Nature* 505, 283–285.
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186.
- Dixon-Declevé, S., Gaffney, O., Ghosh, J., Randers, J., Rockström, J., Stoknes, P.E., 2020. *Earth for All: A Survival Guide for Humanity*. New Society Publishers, British Columbia, Canada.
- Fioramonti, L., 2013. *Gross Domestic Problem: The Politics Behind the World's Most Powerful Number*. Zed Books Ltd.
- Fleurbay, M., Blanchet, D., 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford University Press.
- Hoekstra, R., 2019a. Replacing GDP by 2030: Towards a Common Language for the Well-being and Sustainability Community. Cambridge University Press, Cambridge.
- Hoekstra, R., 2019b. Replacing GDP by 2030: Towards A Common Language for the Well-Being and Sustainability Community. Cambridge University Press.
- Jansen, A., Wang, R., Behrens, P., Hoekstra, R., 2024. Beyond GDP: a review and conceptual framework for measuring sustainable and inclusive wellbeing. *Lancet Planet. Health* 8, e695–e705.
- Kubiszewski, I., Costanza, R., Franco, C., Lawn, P., Talberth, J., Jackson, T., Aylmer, C., 2013. Beyond GDP: measuring and achieving global genuine progress. *Ecol. Econ.* 93, 57–68.
- Kuznets, S., 1934. National Income, 1929-1932, A report to the U.S. Senate, 73rd Congress, 2nd Session. US Government Printing Office, Washington, DC.
- Maibaum, F., Kriebel, J., Foege, J.N., 2024. Selecting textual analysis tools to classify sustainability information in corporate reporting. *Decis. Support Syst.* 183, 114269.
- Matsui, T., Suzuki, K., Ando, K., Kitai, Y., Haga, C., Masuhara, N., Kawakubo, S., 2022. A natural language processing model for supporting sustainable development goals: translating semantics, visualizing nexus, and connecting stakeholders. *Sustain. Sci.* 17, 969–985.
- Oecd, 2025. *Well-Being and Beyond GDP*. OECD, Paris, France.
- Reimers, N., Gurevych, I., 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.
- Stiglitz, J.E., Sen, A., Fitoussi, J.P., 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress. OECD.
- Undp, 1996. *Human Development Report*. Nairobi, Kenya.
- United Nations, 2024. *Pact for the Future: What it Delivers*. United Nations, New York.
- United Nations Network of Economic Statisticians, 2024. Completing the Statistical Architecture: Developing the System of Population and Social Statistics and the Framework for Inclusive and Sustainable Wellbeing - A Roadmap. UN Statistical Commission, New York.
- Van Eynde, R., Greenford, D.H., O'Neill, D.W., Demaria, F., 2024. Modelling what matters: how do current models handle environmental limits and social outcomes? *J. Clean. Prod.* 476, 143777.
- Wilkinson, R., Pickett, K., 2010. *The Spirit Level: Why Equality is Better for Everyone*. Penguin UK.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A., 2020. Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, pp. 38–45.