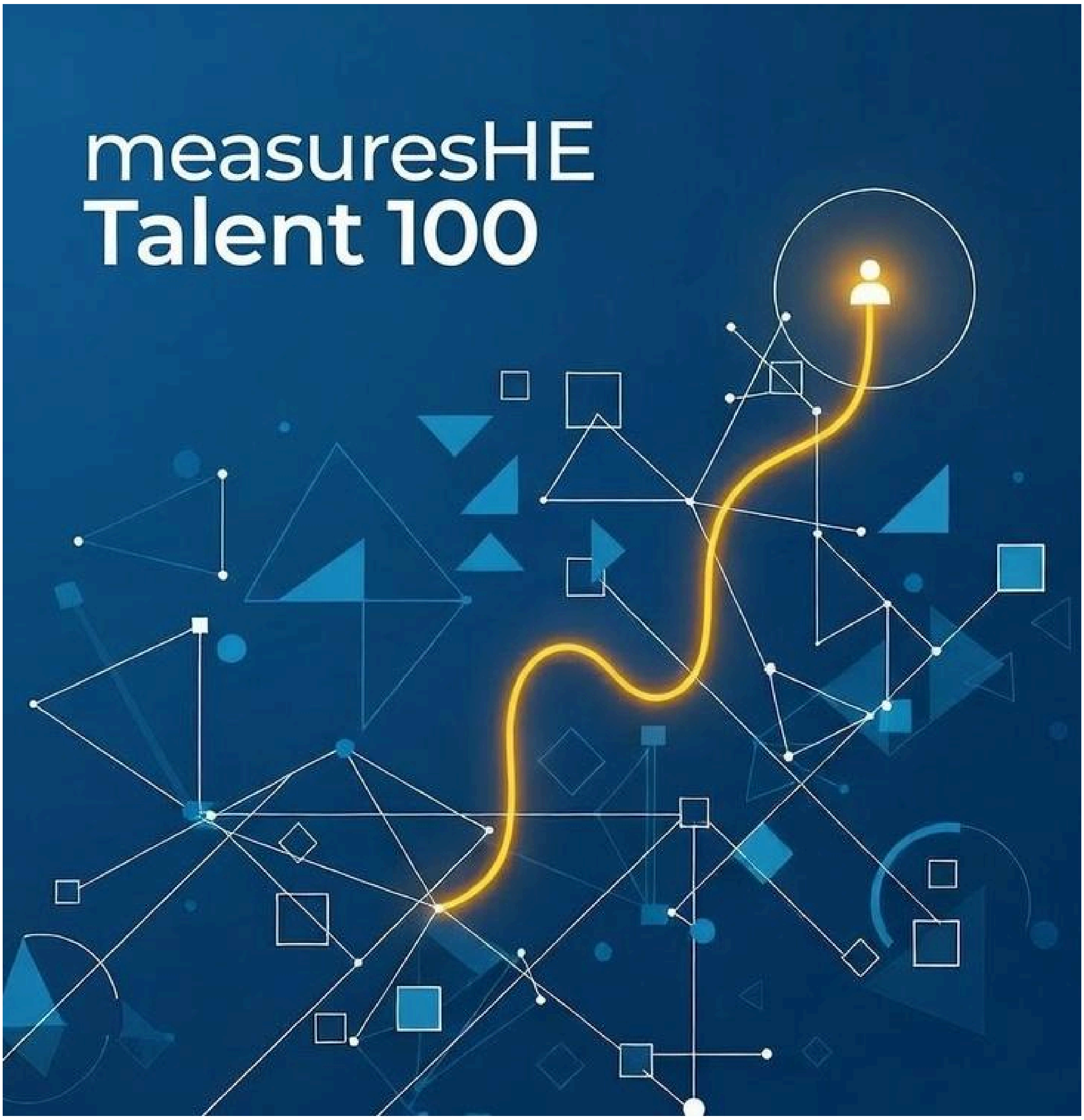
measuresHE
Higher Education Analytics



Methodology Document

Talent 100 Methodology 2026

Welcome to the inaugural edition of **measuresHE** Talent 100. This document outlines the comprehensive and robust methodology used by **measuresHE** to evaluate and rank academic authors based on their research quality, productivity, and collaborative impact. This framework is designed to provide a transparent, fair, and field-normalized assessment of scholarly talent across diverse academic domains.

Purpose and Scope

The primary purpose of the **measuresHE** Talent 100 is to serve as a robust and ethical tool for recognizing and evaluating high-quality academic authors based on their research thought leadership, productivity, and collaborative impact. In an era of ever-increasing academic output, discerning true scholarly talent is crucial, and this ranking provides a reliable, transparent, and data-driven mechanism to achieve this goal. Crucially, the ranking is structured to accommodate the diversity of the academic landscape. A separate, specialized ranking is conducted for each of the top-level subject domains defined by the **measuresHE** classification system except Arts and Humanities, ensuring that authors are compared fairly within their respective fields, acknowledging the distinct publication and citation cultures across different disciplines.

Six subject areas, ranging from the Physical Sciences to Business and Economics, have yielded robust, statistically significant candidate pools. However, the Arts and Humanities dataset presented unique challenges that we believe would compromise the fairness and utility of a "Top 100" list at this time.

The current eligibility criteria require a minimum of 20 works within the analysis period. While this ensures a robust baseline for journal-heavy disciplines, it does not sufficiently account for the scholarly traditions of the Arts and Humanities, where high-impact contributions are frequently centered around comprehensive monographs and long-form chapters rather than high-frequency article outputs.

measuresHE is committed to celebrating excellence across all disciplines. Rather than releasing a potentially unrepresentative ranking for Arts and Humanities, we continue to refine our methodology to develop a more nuanced framework, possibly incorporating other data sources and longer-range citation windows, that reflects the true intellectual leadership within this vital field.

The full evaluation for Talent 100 assesses hundreds of thousands of active academics. From this extensive analysis, the publicly available dataset highlights the Top 100 academics across each of our seven top-level subject domains.

Core Principles of the Methodology

The methodology underpinning the **measuresHE** Talent 100 is fundamentally designed to reward genuine scholarly contribution, recognising an author's **research quality**, **research productivity**, and **collaborative impact**. Our guiding principle is to create a system that is not only transparent and data-driven but also highly resistant to common manipulative practices that undermine the integrity of academic metrics, thereby ensuring an ethical and reliable evaluation of individual scholarly talent. This focus on robustness ensures that the ranking provides the most accurate and fair measure of an author's true impact on the academic community.

Key Design Robustness Features:

A significant focus of this methodology is its resilience against common forms of metric manipulation in author-level evaluation. The ranking algorithms are engineered to neutralise or minimise the impact of the following techniques:

- **Self-Citations and Citation Cartels**: The **Research gravitas** metric, which is based on the PageRank algorithm, is inherently resistant to these practices. It discounts the value of citations from low-impact authors (which would include cartel participants) and explicitly **ignores self-citations** in its graph formulation.
- **Outlier Distortion in Quality**: The **Typical research quality** metric employs an **Olympic mean** (a trimmed mean) of Field-Weighted Citation Impact (FWCI) scores. This mathematically removes the 5% of an author's works with the lowest FWCI and the 5% with the highest FWCI, preventing a small number of extraordinarily high- or low-impact papers (including potentially fraudulent or bought citations) from disproportionately skewing the author's overall typical quality assessment.
- **Hyper-Authorship Inflation**: The **Collaboration intensity** metric is designed to strictly penalize the practice of inflating collaboration scores through being a nominal author on papers with hundreds of co-authors. Its interaction credit formula ($I_w = \frac{2}{|A_w|}$) ensures that a significant, meaningful contribution on a paper with a small number of authors is weighted substantially higher than being one of many authors on a large-scale collaboration.
- **Data Manipulation**: The use of **OpenAlex** as a single, open-source, and community-driven data backbone ensures a transparent and auditable

foundation, overcoming limitations and data biases often associated with closed-source systems.

By being robust against these practices, the **measuresHE** Talent 100 aims to provide a more accurate and ethical measure of scholarly contribution.

Applicability and Adaptability

While the default ranking system utilises **measuresHE's** established top-level subject domains, the methodology is inherently flexible and highly adaptable. This same core set of principles and algorithms can be reapplied to generate rankings for a variety of other subject groupings and research classifications, such the UN Sustainable Development Goals, and specific granular research topics such as Artificial Intelligence and Vaccine Development.

This inherent flexibility allows the **measuresHE** methodology to be a powerful tool for research administrators, funding bodies, librarians, and academics seeking to evaluate impact across both broad and specialised scientific landscapes.

Data Source

The ranking data is sourced from OpenAlex, an open-source, community-driven bibliometric database. OpenAlex's global perspective and open nature overcome many of the limitations associated with closed-source systems. Specifically, the calculations for this ranking utilised a data snapshot extracted on November 1st, 2025. This ranking evaluates academic works published from 2020 to 2024.

Metrics

- Publication volume
- Typical research quality
- Best works
- Research gravitas
- Collaboration intensity

Publication and citation patterns vary within different subfields of a subject domain, and these dynamics can evolve over time. To ensure fair comparison, where necessary, metrics are normalised based on both the subfield and the year of publication. The subsequent sections detail the specific normalisation methods employed.

Publication Volume

This metric measures the publication volume for authors, which essentially measures an author's share of research output within broader subject domains.

Mathematical Explanation

The metric calculates an aggregated work count (N) for each author (a) within a given academic subject domain (s). This metric accounts for the multidisciplinary nature of works by distributing their value across relevant subject domains.

1. Fractional Work Value

First, for any given academic work (w), its value is distributed equally across the set of subfields (F_w) to which it belongs. The fractional value (f) of a work w for a specific subfield i is

$$f_{w,i} = \frac{1}{|F_w|} \text{ if subfield } i \in F_w\text{, 0 otherwise}$$

where $|F_w|$ is the total number of subfields associated with work w.

2. Subject Aggregation

Next, we calculate the total publication volume (N) for an author (a) in a specific subject (s). This is the sum of the fractional values of all works authored by a for subfields that map to subject domain s.

$$N_{a,s} = \sum_{w \in W_a} \sum_{i \in F_w \cap S_s} f_{w,i}$$

Where:

- W_s is the set of relevant works authored by a
- S_s is the set of subfields belonging to subject s
- F_w is the set of subfields associated with work w

Typical Research Quality

This metric is based on Field Weighted Citation Impact (FWCI), which compares the number of citations a publication receives with the average number of citations of the publications of the same type, in the same subject and published in the same year. This metric assesses the “typical” quality of an author's research output.

Mathematical Explanation

The metric measures an outlier-trimmed arithmetic mean of the FWCI for each author-subject pair. This specific type of trimmed mean is often called an Olympic mean.

FWCI as a measure is bounded at the lower end by zero, but is unbounded at the top end. Its global average is one. Thus, a simple arithmetic mean can sometimes be distorted by a small number of outliers at the top end. We use the Olympic mean to remove this distorting effect.

1. Define the Initial Dataset

First, for a given author (a) and subject domain (s), we define the set of all their published articles, which we will call $W_{a,s}$. Each work w in this set has an associated Field-Weighted Citation Impact score, $FWCI(w)$.

2. Order and Trim the Data

For the set $W_{a,s}$ with $n = |W_{a,s}|$ works, the works are ordered based on their FWCI scores from the lowest to the highest.

A new, trimmed set of works $W'_{a,s}$ is created by including only those works whose position $R(w)$ falls within the central 90% of the distribution. The condition for inclusion is

$$n \times 0.05 < R(w) \leq n \times 0.95$$

This effectively removes the 5% of works with the lowest FWCI and the 5% of works with the highest FWCI scores.

3. Calculate the Olympic Mean

The olympic mean of the author a for subject domain s is the simple arithmetic mean of the FWCI scores of the works in the trimmed set $W'_{a,s}$.

$$FWCI_{omean}(a, s) = \frac{\sum_{w \in W'_{a,s}} FWCI(w)}{|W'_{a,s}|}$$

This calculation gives a measure of the central tendency of the author's citation impact that is not skewed by the most or least successful publications.

Best Works

This metric measures the quantity of an author's research output that achieves a high level of citation impact within a specific subject. It specifically counts the number of works authored by an individual that fall into the top 10% of global research for that subject domain based on Field-Weighted Citation Impact (FWCI).

Mathematical Explanation

The metric calculates the absolute count of "top works" (N_{top}) for each author (a) within a given academic subject (s).

1. Define the High-Impact Threshold

First, for a given subject domain (s), we examine the distribution of FWCI scores for all relevant works published in the analysis period. We calculate the 90th percentile threshold ($T_{90,s}$) for that subject domain.

$$T_{90,s} = Percentile_{90}(\{FWCI(w) | w \in W_s\})$$

where W_s is the set of works in subject domain s.

2. Identify and Count Top Works

For each author (a) in subject domain (s), we identify the set of their works ($W_{a,s}$) and count how many equal or exceed the subject's high-impact threshold.

$$N_{top}(a, s) = \sum_{w \in W_{a,s}} I(FWCI(w) \geq T_{90,s})$$

where $I(.)$ is an indicator that returns 1 if the condition is true and 0 otherwise.

Research Gravitas

This metric measures an author's ability to lead the conversation within an academic community. One can think of citations as a conversation. The cited author speaks and the citing authors listen and integrate the information. If that citing author themselves is listened to by others, then the contents from the first author can spread to a wider audience.

This metric differs from traditional citation-based metrics in that it treats each citation differently based on the importance of the citing author, utilizing the **PageRank** algorithm.

1. Graph Formulation

For each academic domain s, a weighted, directed graph $G_s = (V_s, E_s)$ is constructed from the relevant works.

- **Vertices** (V_s): The set of vertices represents all eligible authors active within domain s
- **Edges** (E_s): A directed edge (i, j) exists from a citing author i to a cited author j if author i published a work w_{citing} that cites a work w_{cited} authored by j, and both works are associated with domain s. Self-citations (where $i = j$) are ignored.
- **Weights** (w_{ij}): Each edge is assigned a weight based on the fractional attribution of the citation.

2. Fractional Citation Weight

The weight of a single citation link is distributed among the domains and authors involved to ensure fair credit allocation.

Let $N_{auth}(w)$ be the number of eligible authors on work w. Let $N_{dom}(w)$ be the number of domains associated with work w. Also, Let D_w be the set of domains that are covered by work w.

For a specific citation from work w_{citing} to work w_{cited}, and for a specific domain s common to both works ($s \in D_{w_{citing}} \cap D_{w_{cited}}$), the contribution to the edge weight between a citing author i (on w_{citing}) and a cited author j (on w_{cited}) is:

$$\delta_{w_{citing}, w_{cited}} = 1/\left(N_{dom}(w_{citing}) \times N_{auth}(w_{citing}) \times N_{dom}(w_{cited}) \times N_{auth}(w_{cited})\right)$$

The total weight of the edge from author i to author j in the graph for domain s is the sum of these contributions over all citation instances:

$$W_s(i \rightarrow j) = \sum_{(w_{citing}, w_{cited}) \in C_{i,j,s}} \delta_{w_{citing}, w_{cited}}$$

where $C_{i,j,s}$ is the set of all citations where author i on the citing work, author j is on the cited work, and both works belong to domain s.

3. PageRank Algorithm

The PageRank for an author (node) j in domain s is calculated using the iterative algorithm.

$$PR_s(j) = \frac{1-\alpha}{N} + \alpha \sum_{i \in M(j)} \frac{PR_s(i) \times W_s(i \rightarrow j)}{L(i)}$$

Where:

- α is the damping factor, set to **0.8** (reset probability = 0.2)
- $M(j)$ is the set of authors that cite author j
- $L(i)$ is the toal weighted out-degree of author i
- N is the total number of authors in the graph for the domain

The algorithm iterates until convergence (tolerance 10^{-10})

Collaboration Intensity

This metric measures the intensity of an author's collaborative interactions, normalized to reward significant contributions to collaborative teams. Unlike a simple count of co-authors (which can be inflated by "hyper-authorship" on papers with hundreds or thousands of authors), this metric attributes a fixed "collaboration credit" to each paper, which is then shared among the authors.

Mathematical Explanation

The metric calculates a subfield-normalised collaboration score for each author (a) within a given academic subject domain (s).

1. Interaction Credit

For any given academic work (w), a collaboration credit (I_w) is assigned. The logic aims to normalize collaboration such that each collaboration relationship between each author pair is assumed to contribute equally to the work. For example, if there are exactly two authors, there is exactly one author-pair relationship, than that relationship takes the complete credit for that work; if there are 3 authors, there are 3 author-pair relationships with each author involved in two relationships, thus each author's relationships account for ⅔ of the work etc.
This ensures that being one of two authors on a paper is weighted significantly higher than being one of 100 authors.

$$I_w = \frac{2}{|A_w|} \text{ if } |A_w| > 1, \text{ 0 otherwise}$$

where $|A_w|$ is the total number of authors on work w.

2. Fractional Work Value

The interaction credit is distributed equally across the set of subfields (F_w) to which the work belongs. The fractional value (f) of a work w for a specific subfield i is:

$$f_{w,i} = \frac{I_w}{|F_w|} = \frac{2}{|A_w| \times |F_w|} \text{ if subfield } i \in F_w, \text{ 0 otherwise}$$

3. Author Subfield Network Score

We calculate the total Collaboration Network Score (N) for an author (a) in subfield i and year y.

$$N_{a,i,y} = \sum_{w \in W_{a,y}} f_{w,i}$$

4. Normalisation and Aggregation

To allow for fair comparison across disciplines, we normalise the author's score by the average network score in that subfield.
First, calculate the average score ($\mu_{i,y}$) for all active authors in subfield i and year y.
The the Relative Network Score (R) for the author a is:

$$R_{a,i,y} = \frac{N_{a,i,y}}{\mu_{i,y})}$$

Finally, the score for an author a in a broader subject domain s is the arithmetic mean of their relative scores across all relevant subfields and years:

$$CollabIntensity_{a,s} = \frac{\sum_{i \in s} \sum_{y} R_{a,i,y}}{|s| \times 5}$$

Metric Scoring

The previous section describes how the values of each metric is calculated. This section describes how the metric values are transformed into metric scores that range from zero to 100.

Metric	Higher is better	Scoring Algorithm	Weight
Publication volume	True	Log Normal CDF	20%
Typical research quality	True	Exponential CDF	35%
Best works	True	Log Normal CDF	15%
Research gravitas	True	Exponential CDF	20%

Metric	Higher is better	Scoring Algorithm	Weight
Collaboration intensity	True	Log Normal CDF	10%

The methodology utilizes the well-established statistical functions of the Normal Cumulative Distribution Function (CDF) and the Exponential CDF. The Log Normal CDF is calculated by first applying a logarithmic function before the Normal CDF. Specifically, for the "Best works" and "Collaboration intensity" metrics, the metric value is incremented by one before the logarithmic function is applied. This adjustment is necessary to appropriately handle instances where the metric value is zero.

The overall score of each journal j in each subject s is defined as the weighted sum of the metric scores.

$$overall_{j,s} = \sum_{metric} score_{metric,j,s} \times weight_{metric,j,s}$$

Eligibility Criteria

- The author must have at least 20 works associated with a subject domain to be eligible to be ranked in that subject domain.
- The author's earliest publication year must be no earlier than 60 years prior to the ranking year i.e. on or after 1966 for the 2026 ranking.
- The author is currently active. Their most recent publication must be within the last 3 years i.e. 2023, 2024 or 2025 for the 2026 ranking.

Appendix

Subfield to domain mapping

The details of the mapping of OpenAlex subfields to **measuresHE** subjects and domains is available here.

Why OpenAlex

Powering the Next Generation of Research Intelligence