

<https://doi.org/10.62405/osiconf.2026.20>

Маслов Дмитро,

к.е.н.,

Державна установа «Інститут ринку і економіко-екологічних досліджень

Національної академії наук України», Одеса, Україна,

ORCID: 0000-0002-1721-8319, e-mail: dmitrymaslov@ukr.net

ПІДРАХУНОК ЧИ ВИМІРЮВАННЯ: МЕЖІ ТРАДИЦІЙНОЇ БІБЛІОМЕТРІЇ ТА ІНТЕРСУБ'ЄКТИВНА АЛЬТЕРНАТИВА

АНОТАЦІЯ. Мета роботи — обґрунтувати інструмент оцінювання стану наукової системи, який не спирається на кількісні показники публікаційної активності. Використано концептуальний аналіз літератури, метрологічний аналіз поняття вимірювання, аналіз розподілу робочого часу дослідника та описовий кількісний аналіз динаміки відкликаних публікацій за даними Retraction Watch і SCImago за 1996–2025 роки. Показано, що різномірні публікації та цитування не утворюють величини, придатної до вимірювання, а частка відкликаних статей за два десятиліття зросла приблизно на порядок. Запропоновано категоризацію робочого часу на дійсну наукову, корисну ненаукову та безглузду діяльність і систему інтерсуб'єктивних індексів, побудованих за методикою дифузійних індексів економічних настроїв. Зроблено висновок, що ефективність наукової системи доцільно визначати через частку часу, збережену для дійсної наукової діяльності, а не через обсяг результату на одиницю часу.

КЛЮЧОВІ СЛОВА: оцінювання наукової діяльності; відповідальна наукометрія; інтерсуб'єктивні індикатори; структура робочого часу; закон Гудхарта; відкрита наука.

ВСТУП. Реформування систем оцінювання досліджень є однією з опор порядку денного відкритої науки. Проте предметом реформи залишається переважно оцінювання окремих дослідників, тоді як стан наукової системи в цілому лишається без власного інструменту вимірювання.

Актуальність зумовлена парадоксом «наукової продуктивності». За традиційними бібліометричними показниками наука ніколи не була продуктивнішою, однак той самий період приніс кризу відтворюваності (Open Science Collaboration, 2015), паперові фабрики (Else & Van Noorden, 2021) та перевантажене рецензування. Ця проблема стосується України безпосередньо: з українською мережею пов'язують близько 1500 статей, продукованих паперовими фабриками (Naddaf, 2025).

Фундаментальна причина цього парадоксу не в неточності показників, які застосовують для оцінки науковців, а в їхній невалідності: D. Aksnes, L. Langfeldt і P. Wouters (Aksnes et al., 2019) демонструють, що цитування накопичуються з багатьох причин, здебільшого не пов'язаних із науковим внеском. Другий шар проблеми виникає, коли невалідний показник набуває високих ставок: D. Campbell (Campbell, 1979) попереджав, що кількісні індикатори підпадають під тиск спотворення, M. Fire і C. Guestrin (2019) документують дію закону Гудхарта в академічному видавництві, а P. Smaldino і R. McElreath (Fire & Guestrin, 2016) моделюють, як така структура стимулів відбирає тих, хто ефективно виробляє мінімально достатні для публікації одиниці незалежно від їхньої наукової цінності.

У відповідь на ці проблеми виникли реформаторські ініціативи. Лейденський маніфест (Hicks et al., 2015), Коаліція CoARA (CoARA, 2022) говорять про «відповідальніше» використання метрик. Проте невалідний за побудовою показник не стає валідним від обережнішого поводження з ним. Так само ці показники адресовані оцінюванню осіб та установ, жоден з них не дає сталого індикатора стану системи і не описує, як витрачається її головний ресурс — робочий час дослідників.

Мета роботи — обґрунтувати систему оцінювання стану наукової системи, яка спирається не на кількісні показники публікаційної активності, а на категоризацію робочого часу дослідників та агрегування їхніх експертних суджень.

МЕТОДИ. Дослідження є концептуальним і поєднує чотири методи. Критичний аналіз і синтез літератури встановлюють стан проблеми та межі наявних підходів. Метрологічний аналіз застосовує формальне означення вимірювання до підрахунку публікацій і цитувань, розмежовуючи лічбу та вимірювання. Аналіз розподілу робочого часу, базований на економічній теорії алокації часу (Becker, 1965), дає підставу для категоризації видів діяльності науковця. Описовий кількісний аналіз використано для динаміки відкликів: записи бази Retraction Watch (Retraction Watch, n.d.) зіставлено з річними обсягами публікацій за порталом SCImago (SCImago, n.d.) за 1996–2025 роки та виражено як кількість відкликів на 10000 опублікованих статей за роком публікації. Систему індексів побудовано перенесенням методики дифузійного індексу економічних настроїв (Koenig, 2002; Curtin, 2007) на опитування наукового співтовариства.

РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ.

Підрахунок публікацій не є вимірюванням

Фундаментальне заперечення проти традиційного бібліометричного оцінювання є метрологічним, а не емпіричним. У метрології вимірювання приписує значення величині, екземпляри якої однорідні і порівнювані, що дозволяє їх додавати як одиниці вимірювання. Наукові внески такими не є: дослідження, що визначає напрям галузі, і фрагмент, відрізаний від більшої роботи заради ще однієї публікації, однаково рахуються як «одна стаття», проте не є однією одиницею жодної спільної величини під назвою «науковий результат». Кількість публікацій не є вимірюванням наукового результату, бо немає сумірної величини, яку вона вимірювала б. З цитуваннями та похідними від них показниками ситуація така сама: посилання на хибну статтю рахується так само, як посилання на основоположну роботу в галузі. Індекс Гірша чи імпакт фактор не виправляють недоліки підрахунку кількості статей, вони поєднують дві невалідних метрики.

Свідчення бібліометричного тиску

Ерозія якості досліджень здебільшого не лишає сліду, бо не існує реєстру робіт, які виявилися слабшими, ніж мали бути. Але одним з непрямих свідчень можна вважати відкликання статей. За розрахунком автора на основі бази Retraction Watch (Retraction Watch, n.d.) та порталу SCImago (SCImago, n.d.) частка відкликаних згодом статей зросла з 0,8 на 10000 опублікованих у 1996 році до 9,6 у 2020-му, тобто приблизно на порядок. Два проміжні піки — 20,5 для 2010 року та 18,2 для 2011-го — припадають на концентровані розслідування великих випадків шахрайства. Дані від 2021 року ще накопичують відкликання, проте вже зафіксовані 14,9 для 2021 року та 20,7 для 2022-го відображають масові відкликання контенту паперових фабрик, опублікованого саме в ці роки.

Частину зростання пояснює краще виявлення, але показник є нижньою межею: він охоплює лише виявлені випадки і лише достатньо серйозні для видалення статті. Механізм ерозії якості можна представити наступним чином. Оцінювання винагороджує кількість, винагорода робить маніпулювання показниками раціональною, а ця гра живить ринки хижацьких видань, паперових фаюрик, маніпуляцій із цитуваннями та перевантажує рецензування, яке є єдиним якісним інструментом оцінки наукових результатів.

Категоризація робочого часу

Перевантаження науковців як авторів, так і рецензентів гостро ставить питання про робочий час дослідника як критичний ресурс для наукової роботи. Робочий день, поділений між дослідженням, адмініструванням, викладанням і звітуванням, близький до гри з нульовою сумою: коли висококваліфікований дослідник витрачає день на переформатування звітів, салями-публікації і т.п., справжньою ціною є не просто час, а дослідження, яке в цей час не відбувається.

Щоб зробити цю втрату видимою, адаптовано поняття «безглуздої роботи» D. Graeber (2018) — праці, яку ті, хто її виконує, вважають позбавленою сенсу. Йдеться не про те, що науковці вважають свою

роботу безглуздою, а про те, що одні заняття безпосередньо створюють знання, інші підтримують це створення, а треті нав'язані ззовні й діють проти створення нових знань. Таким чином, професійний час поділяється на три категорії:

- *дійсна наукова діяльність* (ДНД) — просуває дослідника до наукового результату, наприклад, обмірковування проблеми, опрацювання літератури, формулювання гіпотез, проведення досліджень;
- *корисна ненаукова діяльність* (КННД) — не є науковою, але підтримує науку або приносить користь суспільству, наприклад, виступи на конференціях, викладання. Категорія умовна: обслуговуючи безглузду мету, така робота теж може стати безглуздою;
- *безглузда діяльність* (БД) — не дає жодної користі для наукового результату; її виконують під примусом бюрократичних правил і вимог керівних органів, наприклад, переоформлення документів під змінені вимоги, реагування на пропозиції хижацьких видань, маніпулювання показниками через дроблення публікацій.

Робочий час ЧР є сумою трьох складників: $ЧР = ДНД + КННД + БД$. Оскільки він обмежений, категорії конкурують, а БД не виконує жодної законної функції, тож її частка фактично є прямою мірою організаційної неефективності.

Рамка може застосовуватися для оцінки ефективності наукової системи, для оцінки успішності наукових реформ. Питання в тому, яка частка сукупної спроможності системи припадає на ДНД. Тиск традиційних бібліометричних показників створює БД безпосередньо, а також може перетворювати корисну ненаукову роботу на безглузду, якщо підготовка публікації не ґрунтується на отриманих наукових результатах, а зумовлена необхідністю виконання норм. Це переозначає ефективність: система ефективна тоді, коли зберігає найбільшу частку часу для ДНД — необхідну умову доброї науки.

Система інтерсуб'єктивних індексів

Критика бібліометрії зазвичай відбувається через використання інструментів бібліометрії, але жодне з цих чисел не показує, чому це погано для науки; це судження виносять люди, і висновок про негаразди системи сам є інтерсуб'єктивним. Колективне судження науковців історично було головним механізмом контролю якості (через рецензування), тож пропозиція полягає в тому, щоб розгорнути це судження на систему: опитувати науковців як експертних свідків стану власної галузі чи національної системи. Економіка робить це десятиліттями через Індекс менеджерів із закупівель (Koenig, 2002) або Індекс споживчих настроїв (Curtin, 2007).

Система охоплює чотири групи індексів: Індекс наукових можливостей (ІНМ) — забезпеченість фінансуванням, обладнанням, доступом до інфраструктури й даних та часом; індекси структури часу — Індекс наукового часу (ІНЧ), Індекс корисного ненаукового часу (ІКННЧ) та Індекс безглузвих витрат часу (ІБВЧ), де зниження ІНЧ або зростання ІБВЧ означає втрату дослідницької спроможності; Індекс якості наукових результатів (ІЯНР) — оцінка якості власних і прочитаних робіт та зміни якості в часі; Індекс очікувань розвитку науки (ІОРН).

Кожен індекс будується однаково: респонденти відповідають за п'ятибальною шкалою, а частки відповідей поєднуються за формулою

$$\text{Індекс} = (Ч_1 \times 1) + (Ч_2 \times 0,5) + (Ч_3 \times 0) + (Ч_4 \times -0,5) + (Ч_5 \times -1) + 100,$$

де $Ч_1 \dots Ч_5$ — відсотки респондентів у кожній категорії від найпозитивнішої до найнегативнішої. Індекс змінюється від 0 до 200 з нейтральною точкою 100: значення ІНЧ 92 читається безпосередньо як повідомлення про те, що часу на дійсну науку меншає. Зіставність у часі вимагає стандартизованих формулювань, сталої вибірки та регулярного проведення, а вибірка має бути репрезентативною за етапами кар'єри, галузями, типами установ і регіонами.

ВИСНОВКИ. Традиційні бібліометричні показники як кількість публікацій, цитувань та похідні індекси рахують різні речі й видають підрахунок за міру дослідження, якою він бути не може. Щойно підрахунок несе високі ставки, на ньому починають грати, і значення традиційних індексів можуть зростати, коли умови для справжнього дослідження руйнуються.

Запропоновано спосіб бачити проблему через категоризацію робочого часу на дійсну наукову, корисну ненаукову та безглузду діяльність, та відстежувати динаміку через систему інтерсуб'єктивних індексів. Головним ризиком такого інструменту є колективне стратегічне відповідання, спільне для нього й для економічних індексів настроїв. Проте цей ризик обмежується через багатовимірність вимірювання, орієнтація на динаміку за незмінної методики та інституційна незалежність з оприлюдненою методикою.

Традиційні бібліометричні показники слід вилучити з оцінювання окремих дослідників і статей, де єдиною надійною підставою лишається якісне судження фахових колег. Кожну публікаційну та звітну вимогу варто перевірити простим випробуванням: чи служить вона дослідженню, чи адміністративній зручності, і яку безглузду діяльність породжує.

Нові реформи можна пілотувати на частині установ, відмовляючись від них у разі зростання ІБВЧ або зниження ІНЧ.

Перспективи використання результатів полягають у пілотній апробації інструментарію та проведенні його на репрезентативних вибірках і повторних хвилях, що дозволить встановити надійність, валідність і прогностичну здатність індексів. Підхід не є специфічним для оцінювання наукової системи. Освіта, охорона здоров'я, державна служба та інші сфери так само можуть потерпати від безглуздої діяльності, яка ніде не обліковується.

ПОДЯКИ. Роботу виконано за підтримки Національної академії наук України в межах державної бюджетної науково-дослідної роботи «Інноватизація регіональної освіти, науки та виробництва (на прикладі Південного регіону України)» (номер державної реєстрації 0123U104920).

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1). <https://doi.org/10.1177/2158244019829575>
- Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, 75(299), 493–517. <https://doi.org/10.2307/2228949>
- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67–90. [https://doi.org/10.1016/0149-7189\(79\)90048-X](https://doi.org/10.1016/0149-7189(79)90048-X)
- CoARA. (2022). Agreement on reforming research assessment. Coalition for Advancing Research Assessment. <https://coara.eu/agreement/the-agreement-full-text/>
- Curtin, R. (2007). Consumer sentiment surveys: Worldwide review and assessment. *Journal of Business Cycle Measurement and Analysis*, 2007(1), 7–42. <https://doi.org/10.1787/jbcm-a-v2007-art2-en>
- Else, H., & Van Noorden, R. (2021). The fight against fake-paper factories that churn out sham science. *Nature*, 591(7851), 516–519. <https://doi.org/10.1038/d41586-021-00733-5>
- Fire, M., & Guestrin, C. (2019). Over-optimization of academic publishing metrics: Observing Goodhart's law in action. *GigaScience*, 8(6). <https://doi.org/10.1093/gigascience/giz053>
- Graeber, D. (2018). *Bullshit jobs: A theory*. Simon & Schuster. <https://openlibrary.org/works/OL20153626W>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>
- Koenig, E. F. (2002). Using the Purchasing Managers' Index to assess the economy's strength and the likely direction of monetary policy. *Economic and Financial Policy Review*, 1(6). <https://ideas.repec.org/a/fip/fedder/y2002nv.1no.6.html>

- Naddaf, M. (2025). Europe's largest paper mill? 1,500 research articles linked to Ukrainian network. *Nature*. <https://doi.org/10.1038/d41586-025-02809-y>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- The Retraction Watch, (n.d.). The Retraction Watch Database [Data set]. Retrieved August 9, 2026. <http://retractiondatabase.org/>
- SCImago. (n.d.). SJR — SCImago Journal & Country Rank [Portal]. Retrieved August 9, 2026. <https://www.scimagojr.com/>
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), 160384. <https://doi.org/10.1098/rsos.160384>

Maslov Dmytro,

PhD in Economics,

State Institution «Institute of Market and Economic & Ecological Researches
of the National Academy of Sciences of Ukraine», Odesa, Ukraine,
ORCID: 0000-0002-1721-8319, e-mail: dmitrymaslov@ukr.net

COUNTING OR MEASUREMENT: THE LIMITS OF TRADITIONAL BIBLIOMETRICS AND AN INTERSUBJECTIVE ALTERNATIVE

ABSTRACT. *The purpose of the paper is to substantiate an instrument for assessing the state of a research system that does not rely on quantitative indicators of publication activity. The methods are conceptual analysis of the literature, metrological analysis of the notion of measurement, analysis of the allocation of researchers' working time, and descriptive quantitative analysis of retraction dynamics based on Retraction Watch and SCImago data for 1996–2025. Heterogeneous papers and citations are shown not to form a quantity that can be measured, while the share of articles later retracted has grown roughly tenfold over two decades. A categorisation of working time into actual scientific, useful non-scientific and meaningless activity is proposed, together with a system of intersubjective indices built on the diffusion-index methodology of economic sentiment indicators. Effectiveness is redefined as the share of time preserved for actual scientific activity.*

KEYWORDS: *research assessment; responsible metrics; intersubjective indicators; working time structure; Goodhart's law; open science.*