Everyone onboard? Participation ratios as a metric for research activity assessments within young universities

Frode Eika Sandnes[1*,2]

[1]Department of Computer Science, Faculty of Technology, Art and Design, Oslo Metropolitan University, P.O. Box 4, St. Olavs plass, 0130 Oslo, Norway

[2]Faculty of Technology, Kristiania University College, Kirkegata 24-26, 0153 Oslo, Noway

Email: frodes@oslomet.no

Tel: +47 466 97 592

ORCID: 0000-0001-7781-748X

MSC: N/A

JEL: I21, I23, I28

Abstract

Research productivity metrics designed for international and national comparison of institutions may be uncritically adopted for research assessments, especially within young universities with short research traditions. It is argued that such measures can be misleading when assessing small groups of researchers such as within a department, and that they may provide limited usefulness for leaders at different levels of an organization. Instead, the ratio of participation is proposed as a research evaluation metric for smaller groups of researchers. Publication data from Norway are used as examples. The proposed metric is simple to implement and understand, and therefore holds potential for young universities that are strategically strengthening their research capabilities.

Keywords: Research assessment, aggregation, research productivity, higher education leadership, Herfindahl-Hirschman index, Simpson diversity index

Within-university research assessments

Data from research assessments is an important ingredient in the governance of a university. Publication statistics can be used to identify weak and strong areas, implement change, allocate resources, and track the effects of implemented actions. The higher education landscape is diverse (van Vught and Ziegele 2013); institutions range from research intensive universities to vocation-oriented university colleges of applied sciences with limited research traditions. Some of these institutions have strategically intensified their research capabilities such as some university colleges of applied sciences in the Netherland (Griffioen and De Jong 2013). Examples from the UK, Sweden, and Norway (Sandnes 2018) show that other institutions have strategically intensified their research activities to obtain full university status. There is thus a need for metrics that address issues relevant to the type of institution (van Vught and Ziegele 2013).

The publication is one of the physical units of bibliometrics (Broadus 1987). A publication count is a practical proxy for research intensity as publication statistics are often readily available via publication databases. Moreover, it can be relatively straightforward for university administrators to understand and process basic publication statistics without formal research training. Publication statistics are often replicable, and they facilitate straightforward comparisons. Although easy, comparing the research intensities of different groups using indicators derived from publication

counts can be misleading as publication cultures vary from infrequent publishing of long monographs (Verleysen and Ossenblok 2017) to "salami slicing" where least publishable units (LPU) are published frequently as short papers (Frandsen, Eriksen, Hammer, and Christensen 2019; Ding, Nguyen, Gebel, Bauman, and Bero 2020).

Many of the measures discussed in the academic literature are based on data aggregated at country, or institutional level with the purpose of benchmarking (see for instance King 1988; Maclean and Janagap 1993; Bayers 2005; Docampo 2011; Docampo and Cram 2017; Mingers, O'Hanley and Okunola 2017, Sandnes 2021a). Evaluations can also be more focused, for example, assessing gender balance in research (Hernández-Martín et al. 2019), how scholars' leadership duties affect research performance (Lou et al. 2018), degree of international collaboration (Narváez-Berthelemot 1995, Leite, Mugnaini and Leta 2011), timeliness of research (Klavans and Boyack 2008; Klavans and Boyack 2010; Sandnes 2021b) and author order (Abramo, D'Angelo and Rosati 2013).

Problems with means-based productivity measures

When comparing the publication output of different countries, institutions, organizational units, or disciplines it is common to normalize publication data. To normalize groups of different sizes it is common to divide the total publication count produced by the group by the number of researchers in the group resulting in a unit of publications per researcher over a given time-period (often one year, three years, etc). Such normalized measures, or similar derivations thereof, can be classified as *means-based measures*. Several studies have warned against the mean as an unsuitable measure for publication quantity due to Lotka's law and long-tailed skewed distributions (see for instance (Haitun 1986) as a small handful of scholars usually produces most of the publications. The mean has also been shown to be unsuitable for citations (Kiesslich et al. 2021).

It is indeed possible to have one researcher authoring all the publications of an organizational unit. This is an undesirable situation if the other researchers are inactive. A high publication mean may give stakeholders a false impression of broad and active research participation within an organizational unit. If such a parameter shows a mean that is within a normal range it may not trigger the needed attention from leaders. Moreover, a unit may be tempted to employ quick fixes by recruiting a very active researcher to patch the research performance of a unit, colloquially referred to as "blood doping".

In addition to being highly sensitive to outliers resulting from the production of the few very active researchers, publication means for small organizational units can fluctuate greatly from one year to the next. Yet, the mean is a widely understood quantity which may explain its prevalent use among administrators and leaders of academic institutions.

It has been suggested that other features, such as the speed to get research published may be more suitable measures of productivity than the number of publications (Wagner-Döbler 1995; Gupta and Karisiddappa 1997). A brief introduction to the debate on research productivity can be found in Brischoux and Angelier (2015).

Limitations of diversity measures

Some of the academic literature has described how diversity measures can be used to assess research (Yang et al. 2010; Moschini et al. 2020; Mugabushaka, Kyriakou, and Papazoglou 2016). One such measure is the *Herfindahl-Hirschman Index* (HHI) which was originally used to quantify the degree of competition among companies according to their market share (Rhoades 1993). The

Herfindahl-Hirschman Index is computed by summing the squares of all the "market share" portions, that is

$$HHI = \sum_{i=1}^{n} p_i^2$$

where p_i is the ith portion of n portions which sum should be 1. The squaring of the portions means that entities with a large portion get a stronger weight. A high HHI indicates imbalance, while a low HHI value signifies a more balanced distribution. Matsumoto, Merlone, and Szidarovszky (2012) discussed some limitations of the index, while Cracau and Lima (2016) discussed ways of normalizing the HHI as the lower bound of a portion $1/n$ depends on the number of entities n. A straightforward normalization approach includes the following:

$$HHI^N = \frac{HHI - 1/n}{1 - 1/n}$$

Yang et al. (2010) used the Herfindahl-Hirschman Index to observe changes in citation concentration among Chinese Scholars and found that citation patterns have become more diverse over time. Moschini et al. (2020) used the HHI to measure the concentration of subject areas in published research.

The inverse of the HHI is sometimes referred to as the *Simpson Diversity Index* (1 - *HHI*), where a low value indicates imbalance and a high value signals balance. Mugabushaka, Kyriakou, and Papazoglou (2016) used diversity measures to assess interdisciplinarity. They also discussed the Rao-Stirling index and Shannon entropy (Mugabushaka, Kyriakou, and Papazoglou 2016) as diversity measures. The widely used Shannon entropy can be described as a measure of spread for distributions of categorical data. It can be defined as:

$$H = -\sum_{i=1}^{n} p_i \ln p_i$$

It is probably harder for laypersons to interpret the entropy H compared to the HHI since it is not limited to a fixed range. More importantly, although diversity measures such as the HHI and entropy are capable of quantifying diversity, they do not incorporate zero portions as each portion p_i must be non-zero. Such measures are therefore unsuitable for analysing inactive researchers. To overcome this limitation the participation ratio is proposed where the focus is shifted from the characteristics of active researchers to the balance between inactive and active researchers.

Participation ratios

It is proposed that within-university research activity assessments should be made based on the ratio of active researchers, namely.

$$participation\ ratio = \frac{no.\ researchers\ with\ at\ least\ one\ publication}{total\ number\ of\ reserachers}$$

This ratio can be applied for any time-window. A time window of one year may be practical if reporting follows the financial year. The ratio reveals to what degree the scholars in an organizational unit participate and contribute with active research. Wide participation is viewed as desirable, while sparse participation is viewed as undesirable. The high activity of a few researchers is not rewarded by the radio.
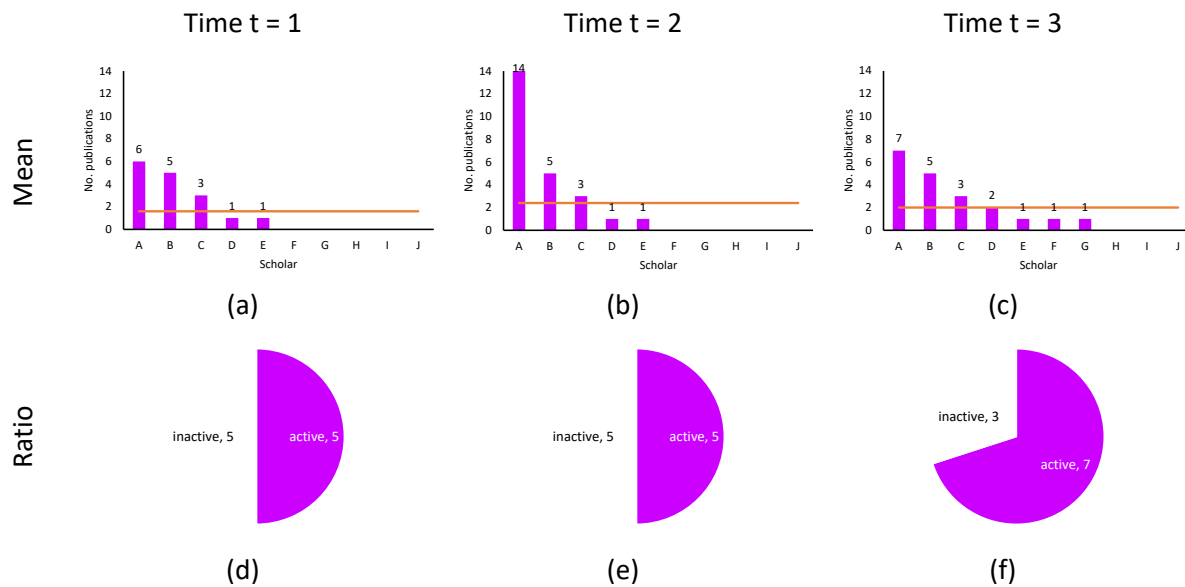
Figure 1. Tracking change with mean publications counts and participation ratios over time. The orange lines indicate the mean number of publications per scholar. Means are sensitive to individual variations. The only difference between time t =1 and t = 2 is that researcher A has 14 and 6 publications, resulting in a mean of 2.4 and 1.6, respectively. In other words, one scholar affects the indicator by 33%. In both cases only 50% of the scholars are active. Time t = 3 yields a lower mean than t = 1, yet 70% of the scholars is active.

The productivity of a single author may fluctuate considerably over time, while the ratio of active scholars may be comparable more stable in an organizational unit with an established culture for research. Figure 1 illustrates how the ratio-based measure is unaffected by fluctuations in research output by individual scholars, thereby making it easier to observe collective trends and identify changes over time.

The participation ratio is easy to compute, easy to understand and easy to aggregate to different levels in the organization such as research group-level, department-level, or faculty-level. Aggregation is achieved by computing the ratio for all the scholars in the given group. It does not focus on individual scholars, and its simplicity means that it may be well-suited as a dialogue tool.

One may speculate that such a ratio-based measure would be generally more motivating for most scholars than a means-based measure as all members of a unit know that their contribution matter, even if the contribution is small. With means-based measures the productivity of the average scholar may be overshadowed by the extreme productivity of a few highly productive colleagues. Another advantage of the ratio measure is that it does not reward "salami slicing" behaviours to the same degree as publication counts do.

Traditional activity measures need to be normalized before comparing different disciplines due to their different publication traditions. However, activity ratios can be used to make direct comparisons of different groups and across different disciplines.

One may argue that a ratio-based activity measure is easily manipulated. If for instance all the members of the unit are added as a co-author to a single publication the unit will achieve a maximum score. However, for a ratio-based scheme to work we need to trust that researchers adhere to ethical norms for co-authorship such as those outlined in the Vancouver Recommendations.

Example case: A young Norwegian university

The Norwegian publication point metric was introduced in 2004 over concerns that researchers in Norway were comparatively too inactive compared to researchers in neighbouring countries, and incentives were introduced to stimulate a higher activity level (Sandnes 2018). At the national level, the metric has indeed served its intentions as there has been a well-documented increase in publication activity over the last 15 years. The publication point is specific to Norway. It is computed based on the type of publication (journal paper, book chapter/proceedings or monograph), journal rank (regular or high), the number of authors, and presence of international co-authors (Haugen and Sandnes 2016). Although the national publication metric takes several factors into consideration it is in its essence a weighted mean publication count when aggregated.

According to Haugen and Sandnes (2016) the regime for research assessment in Norway is influenced by the measures employed by the Ministry of Education. The Norwegian Centre for Research publishes annual statistics and visualisations (Sandnes 2012) for each university and the organizational units within the university (https://dbh.nsd.uib.no/). Research metrics include publication points per scholar, where a scholar is defined as someone with a PhD-qualification, or similar.

The Ministry of Education regularly hosts formal appraisal/dialogue meetings with each of the universities where the performance of a given university is the topic of discussion. The discussions are often structured around parameters with scores below the norm such as too low research output, too few successful graduates, low scores on national evaluations, etc. Instructions to make action plans for improvements are usually given with expectations of growth. The Ministry level process and its indicators are often copied and used within the universities where the top leadership group (the rectorates) host formal appraisal meetings with the faculty deans and the deans host similar formal meetings with the department heads. Often, the same types of statistics are used at lower levels in organizations since such data are readily available in the national database. These statistics, and corresponding metrics, thus serve as a de-facto national standard.
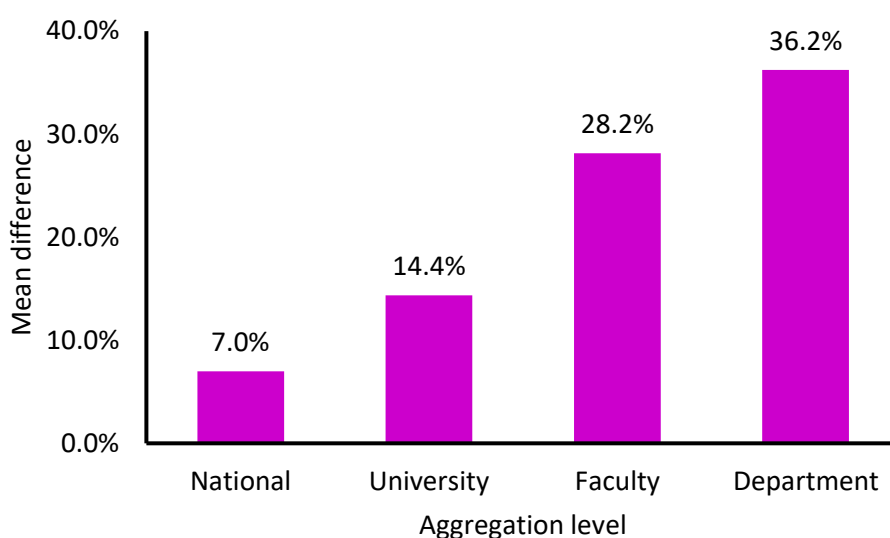

Figure 2. The mean percentage yearly difference in publication points per scholar at different aggregation levels during 2012-2019 extracted from the Norwegian database of statistics for higher education (DBH). This example shows national level (Norway), university level (OsloMet), Faculty Level (Technology, Art and Design) and department level (Department of Computer Science).

Figure 3. Variability of the publication points per scholar at department level (Department of Aesthetics, OsloMet) extracted from the Norwegian database of statistics for higher education (DBH).

Figure 2 illustrates the instability of the publication metric used in Norway over a period of 7 years. Even at the national level the mean percentage difference was 7% and the maximum was more than 40% for more than 18,000 scholars. At university level the variation is much larger with a mean difference of 14.4%, which is as much as 36.2% at department level with a maximum yearly difference of 96.7%. Here the largest department with the longest and strongest research traditions was chosen which also had the smallest yearly differences of the departments in the faculty.

To visualise the problem for small groups, Figure 3 shows how the publication points per scholar can oscillate strongly for a less research-intensive department (mean yearly percentage difference of 106%, maximum yearly difference 280%. As the publication points per scholar metric incorporate several elements, including journal rank, number of authors, international collaboration, and publication type, it is challenging to derive the causes of these oscillations.
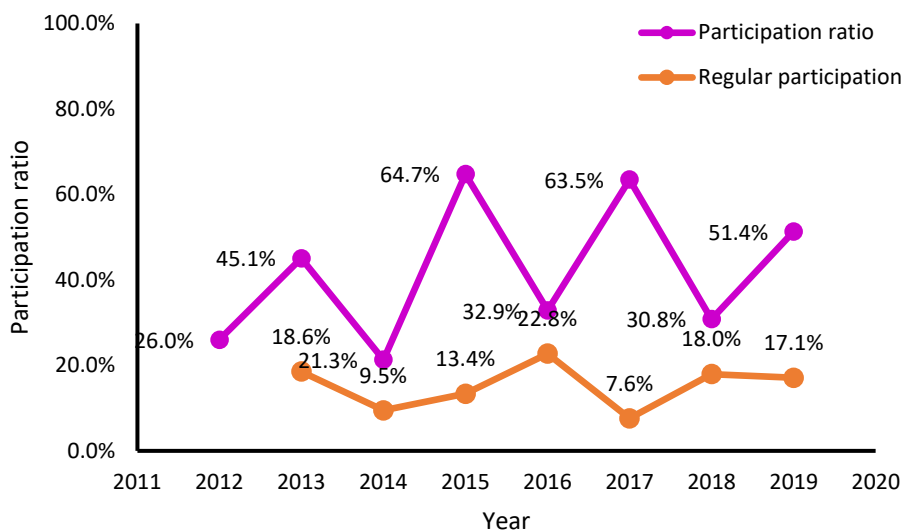
Figure 4. Example of participation ratios over time, including regular participants (Department of Aesthetics, OsloMet). Regular participations are defined as those who also have authored a paper in the preceding year.

Figure 4 illustrates an example of participation ratios over time for the Department of Aesthetics. The chart helps explain that the fluctuations in Figure 3 are caused by variations in participation from 21% to 64%. Most scholars in the unit publish less frequently, seemingly every two years. Regular participation is less variable within a smaller range of 8% to 23% participation and the curve does not follow the overall participation curve. Moreover, the chart reveals that about half of the scholars in the department are sporadically involved in publication activities. Overall, the research output in this unit appears stable and low as the chart does not reveal much drift over time.

Conclusions

Ratios of active researchers in terms of publications has been proposed as a simple research activity metric for young universities that are building up their research activity. Although other measures may be more appropriate for highly active and mature research units it is still also common for established research-intensive universities to have a sizable proportion of scholars that do not publish. Ratios may also be applied to quality measures such as citations or international collaboration. For instance, one may compute ratios of how many of the papers published in the unit the last five years have been cited, or variations on this theme. The focus is then shifted from absolute citation counts to a collective contribution of citable research. Obviously, such metrics should exclude self-citations (Sandnes 2020). Although participation ratios were exemplified through the situation in Norway, one may expect that the participation ratio proposed herein is also relevant outside of Norway. Several countries are experiencing intensification of research efforts in higher education institutions that previously did not conduct much research (such as universities of applied sciences in the Netherlands, see Griffioen and De Jong 2013). Processes can be observed in several countries where the status of higher education institutions are elevated to university level (for instance polytechnics in the UK in the 90s and more recently the elevation of colleges). In such contexts the participation ratio may provide valuable insights that complement mean publication counts and similar traditional measures to help institutional leaders with making informed decisions.

References

Abramo, G., D'Angelo, C. A., & Rosati, F. (2013). Measuring institutional research productivity for the life sciences: the importance of accounting for the order of authors in the byline. Scientometrics, 97(3), 779-795. https://doi.org/10.1007/s11192-013-1013-9

Bayers, N. K. (2005). Using ISI data in the analysis of German national and institutional research output. Scientometrics, 62(1), 155-163. https://doi.org/10.1007/s11192-005-0010-z

Brischoux, F., & Angelier, F. (2015). Academia's never-ending selection for productivity. Scientometrics, 103(1), 333-336. https://doi.org/10.1007/s11192-015-1534-5

Broadus, R. N. (1987). Toward a definition of "bibliometrics". Scientometrics, 12(5-6), 373-379. https://doi.org/10.1007/BF02016680

Cracau, D., & Lima, J. E. D. (2016). On the normalized herfindahl-hirschman index: a technical note. International Journal on Food System Dynamics, 7(4), 382-386. https://doi.org/10.18461/ijfsd.v7i4.748

Ding, D., Nguyen, B., Gebel, K., Bauman, A., & Bero, L. (2020). Duplicate and salami publication: a prevalence study of journal policies. International journal of epidemiology, 49(1), 281-288. https://doi.org/10.1093/ije/dyz187

Docampo, D. (2011). On using the Shanghai ranking to assess the research performance of university systems. Scientometrics, 86(1), 77-92. https://doi.org/10.1007/s11192-010-0280-y

Docampo, D., & Cram, L. (2017). Academic performance and institutional resources: a cross-country analysis of research universities. Scientometrics, 110(2), 739-764. https://doi.org/10.1007/s11192-016-2189-6

Frandsen, T. F., Eriksen, M. B., Hammer, D. M. G., & Christensen, J. B. (2019). Fragmented publishing: a large-scale study of health science. Scientometrics, 119(3), 1729-1743. https://doi.org/10.1007/s11192-019-03109-9

Griffioen, D. M., & De Jong, U. (2013). Academic drift in Dutch non-university higher education evaluated: A staff perspective. Higher Education Policy, 26(2), 173-191. https://doi.org/10.1057/hep.2012.24

Gupta, B., & Karisiddappa, C. (1997). Productivity of authors as reflected byduration of their scientific participation and speed of publication. Scientometrics, 39(3), 281-291. https://doi.org/10.1007/BF02458531

Haitun, S. (1986). Problems of quantitative analysis of scientific activities: The non-additivity of data. Part II. Corollaries. Scientometrics, 10(3-4), 133-155. https://doi.org/10.1007/BF02026038

Haugen, K. K., & Sandnes, F. E. (2016). The new Norwegian incentive system for publication: from bad to worse. Scientometrics, 109(2), 1299-1306. https://doi.org/10.1007/s11192-016-2075-2

Hernández-Martín, E., Calle, F., Dueñas, J. C., Holgado, M., & Gómez-Pérez, A. (2019). Participation of women in doctorate, research, innovation, and management activities at Universidad Politécnica de Madrid: analysis of the decade 2006–2016. Scientometrics, 120(3), 1059-1089. https://doi.org/10.1007/s11192-019-03179-9

King, J. E. A. N. (1988). The use of bibliometric techniques for institutional research evaluation: A study of avian virology research. Scientometrics, 14(3-4), 295-313. https://doi.org/10.1007/BF02020081

Kiesslich, T., Beyreis, M., Zimmermann, G., & Traweger, A. (2021). Citation inequality and the Journal Impact Factor: median, mean,(does it) matter?. Scientometrics, 126, 1249–1269. https://doi.org/10.1007/s11192-020-03812-y

Klavans, R., & Boyack, K. (2008). Thought leadership: A new indicator for national and institutional comparison. Scientometrics, 75(2), 239-250. https://doi.org/10.1007/s11192-007-1854-1

Klavans, R., & Boyack, K. (2010). Toward an objective, reliable and accurate method for measuring research leadership. Scientometrics, 82(3), 539-553. https://doi.org/10.1007/s11192-010-0188-6

Leite, P., Mugnaini, R., & Leta, J. (2011). A new indicator for international visibility: exploring Brazilian scientific community. Scientometrics, 88(1), 311-319. https://doi.org/10.1007/s11192-011-0379-9

Lou, W., Zhao, Y., Chen, Y., & Zhang, J. (2018). Research or management? An investigation of the impact of leadership roles on the research performance of academic administrators. Scientometrics, 117(1), 191-209. https://doi.org/10.1007/s11192-018-2842-3

Maclean, J., & Janagap, C. (1993). The publication productivity of international agricultural research centers. Scientometrics, 28(3), 329-348. https://doi.org/10.1007/BF02026514

Matsumoto, A., Merlone, U., & Szidarovszky, F. (2012). Some notes on applying the Herfindahl–Hirschman Index. Applied Economics Letters, 19(2), 181-184. https://doi.org/10.1080/13504851.2011.570705

Mingers, J., O'Hanley, J. R., & Okunola, M. (2017). Using Google Scholar institutional level data to evaluate the quality of university research. Scientometrics, 113(3), 1627-1643. https://doi.org/10.1007/s11192-017-2532-6

Moschini, U., Fenialdi, E., Daraio, C., Ruocco, G., & Molinari, E. (2020). A comparison of three multidisciplinarity indices based on the diversity of Scopus subject areas of authors' documents, their bibliography and their citing papers. Scientometrics, 1-14. https://doi.org/10.1007/s11192-020-03481-x

Mugabushaka, A. M., Kyriakou, A., & Papazoglou, T. (2016). Bibliometric indicators of interdisciplinarity: the potential of the Leinster–Cobbold diversity indices to study disciplinary diversity. Scientometrics, 107(2), 593-607. https://doi.org/10.1007/s11192-016-1865-x

Narváez-Berthelemot, N. (1995). An index to measure the international collaboration of developing countries based on the participation of national institutions: the case of Latin America. Scientometrics, 34(1), 37-44. https://doi.org/10.1007/BF02019171

Rhoades, S. A. (1993). The herfindahl-hirschman index. Fed. Res. Bull., 79, 188.

Sandnes, F. E. (2012). On the truthfulness of petal graphs for visualisation of data. In: Proceedings of The Norwegian Informatics Conference (NIK 2012). Tapir Academic Publishers, Trondheim, Norway, 225-235.

Sandnes, F. E. (2018). Do Norwegian academics who publish more earn higher salaries?. Scientometrics, 115(1), 263-281. https://doi.org/10.1007/s11192-018-2639-4

Sandnes, F. E. (2020). A simple back-of-the-envelope test for self-citations using Google Scholar author profiles. Scientometrics, 124(2), 1685-1689. https://doi.org/10.1007/s11192-020-03521-6

Sandnes, F.E. (2021a). A Bibliometric Study of Human-Computer Interaction Research Activity in the Nordic-Baltic Eight Countries. Scientometrics. (Accepted, Feb., 2021).

Sandnes, F. E. (2021b). Is there an imbalance in the supply and demand for universal accessibility knowledge? Twenty years of UAIS papers viewed through the lens of WCAG. Universal Access in the Information Society (in press). https://doi.org/10.1007/s10209-020-00712-z

van Vught, F., & Ziegele, F. (2013). U-Multirank: A user-driven and multi-dimensional ranking tool in global higher education and research. Rankings and Accountability in Higher Education – Uses and Misuses (Eds: Marope, P.T.M., Wells, P.J., and Hazelkorn, E.), UNESCO publishing, 256-277.

Verleysen, F. T., & Ossenblok, T. L. (2017). Profiles of monograph authors in the social sciences and humanities: an analysis of productivity, career stage, co-authorship, disciplinary affiliation and gender, based on a regional bibliographic database. Scientometrics, 111(3), 1673-1686. https://doi.org/10.1007/s11192-017-2312-3

Wagner-Döbler, R. (1995). Were has the cumulative advantage gone? Some observations about the frequency distribution of scientific productivity, of duration of scientific participation, and of speed of publication. Scientometrics, 32(2), 123-132. https://doi.org/10.1007/BF02016890

Yang, S., Ma, F., Song, Y., & Qiu, J. (2010). A longitudinal analysis of citation distribution breadth for Chinese scholars. Scientometrics, 85(3), 755-765. https://doi.org/10.1007/s11192-010-0245-1