https://doi.org/10.51408/issi2025_098

Current Interdisciplinarity Measures Fail to Reflect Authors' Perspectives

Dag W. Aksnes[1], Henrik Karlstrøm[2], Fredrik N. Piro[3]

[1] dag.w.aksnes@nifu.no, [2] henrik.karlstrom@nifu.no, [3] fredrik.prio@nifu.no
NIFU – Nordic Institute for studies in Innovation, Research and Education, Økernveien 9, 0653 Oslo (Norway)

Abstract

The objective of this paper is to compare bibliometrically constructed indexes of interdisciplinarity with authors' self-assessments of the interdisciplinarity of their own papers, thereby providing knowledge into how well these indicators correspond with researchers' own perceptions. The bibliometric interdisciplinarity measures analyzed include the Shannon entropy, ${}^{2}D^{S}$, and DIV* indicators. The data analyzed in the study are derived from two separate questionnaire surveys, in which authors were asked about specific articles they had published. The results reveal that there is little agreement between the bibliometric measures and authors' assessments of interdisciplinarity, with correlations ranging from weak to very weak.

Introduction

Interdisciplinarity or interdisciplinary research (IDR) has become buzzwords in research policy (Cantone, 2024). This is not surprising as interdisciplinarity is increasingly characterizing contemporary research practices (Porter & Rafols, 2009). Moreover, funding agencies are frequently emphasizing IDR, for example through establishing interdisciplinary research centers and programs (Avila-Robinson, Mejia & Sengoku, 2021; Chen et al., 2021).

However, defining exactly what interdisciplinarity means is challenging (Miller, 2020). Several definitions and concepts of IDR exist in the literature (Laursen, Motzer & Andersson, 2022; von Wehrden et al., 2019). At the same time, numerous indicators attempting to measure interdisciplinarity bibliometrically have been developed over the years. While these indicators ostensibly aim to capture the same phenomenon, they often produce divergent results (Avila-Robinson, Mejia & Sengoku, 2021; Cantone, 2024). This means that the degree of interdisciplinarity for any given unit of analysis can vary significantly depending on the specific measures applied. Wang & Schneider fundamentally question the validity of the measures and argue that "the current measurements of interdisciplinarity should be interpreted with much caution" (2019, p. 239).

Against this background there is a need for more validity studies. While all proposed IDR indicators strive for accurate and valid measurement of interdisciplinarity (given the chosen definition), there is a risk that these indicators capture *bibliometric-derived interdisciplinarity* – patterns and connections discernible in bibliometric data – rather than *"real-life"* interdisciplinarity. We believe that it is important to understand how IDR measures align with perceptions of interdisciplinarity from the very producers of the research. Thus, the objective of this paper is to compare bibliometrically constructed indexes of IDR with self-assessments of

interdisciplinarity by the authors of the same papers, thereby providing knowledge into how well these indicators correspond with researchers' own perceptions.

Few prior studies have addressed the issue of construct validity. Zhang et al. (2018) examined over 150,000 PLoS One articles, comparing an IDR measure based on cited references with authors' departmental affiliations, and found a low correlation between the two. Roessner et al. (2013) conducted an ethnographic study of a single researcher, examining how bibliometric measures aligned with perceptions of knowledge integration but did not find a simple answer to the question of the validity of IDR indicators. Of particular relevance to the present study is the work of Avila-Robinson, Mejia, and Sengoku (2021), which, to our knowledge, is the only study that analyses a larger dataset to compare bibliometric IDR measures with authors' self-assessments of interdisciplinarity. Their findings based on analyses of a thousand publications reveal relative weak yet statistically significant associations (coefficients ranging from 0.20 to 0.27) between self-assessments of interdisciplinarity and four IDR measures.

In our study we analyse three indicators which are commonly used to analyse interdisciplinarity:

- Shannon's entropy. Originally introduced as a measure of "information uncertainty", Shannon entropy is often used to quantify the diversity of disciplines referenced in a given publication or set of publications (see e.g Leydesdorff, Wagner, & Bornmann 2019a).
- The ${}^{2}D^{S}$-measure (Zhang, Rousseau & Glänzel (2016) which is related to, but a further development of the Rao-Stirling measure. It quantifies the diversity (the range of disciplines) and disparity of disciplines (their unrelatedness) cited in a research publication.
- DIV*. Leydesdorff, Wagner & Bornmann (2019a) introduced the DIV measure, an alternative measure of interdisciplinarity that combines balance, variety, and disparity into a single metric. This measure was later refined into DIV* to correspond with Rousseau's (2019) principles for interdisciplinarity metrics (Leydesdorff, Wagner & Bornmann, 2019b).

Data & methods

Survey data

The data analyzed in this study are derived from two separate questionnaire surveys, in which authors were asked about specific articles they had published. The surveys addressed various dimensions of the articles. Survey 1 focused on quality aspects and the research process, with interdisciplinarity being one of the dimensions examined (see Aksnes et al. (2023) for survey findings related to research quality and citation rates). Survey 2 focused on publication practices in environmental sciences, where the authors were asked to evaluate the articles and their experiences with the publication process. The survey particularly targeted issues related to interdisciplinarity. Both surveys included numerous questions, but for the purpose of this paper only a few were used.

In Survey 1, authors were simply asked to rate the extent to which each of three specific articles was interdisciplinary, using a scale from 1 (low) to 5 (high). No definition of interdisciplinary was provided, allowing respondents to apply their own interpretations. In Survey 2, the authors were asked to indicate whether the following characterizes the paper: a) Based on research from multiple academic disciplines (multidisciplinary), b) Combines and integrates research across academic disciplines (interdisciplinary). The following alternatives were given: Yes, Partly, No, Not Applicable/Don't Know. In another question they were asked to describe who was involved in the research that the paper builds on. One of the options provided was: "Researchers from other fields were involved in the research". Here a binary response option was used (Selected/Not Selected). Thus, this survey applies short definitions and distinguishes between multidisciplinary and interdisciplinary research.

Survey 1 one was sent to a stratified sample of researchers in Norway. The survey was conducted in January 2022 and questionnaires were distributed to a sample of 1,250 researchers. The survey asked researchers to assess three of their own papers, selected through stratified random sampling based on citation metrics (one paper from each of the following citation rank percentiles: top 10%, 10-50% and 50-100%). With a response rate of 47%, the final sample included 592 researchers, each contributing three publications. As a result, the study encompasses assessments of 1,780 publications, of which 1,695 included responses to the question on interdisciplinarity.

Survey 2 was a global survey conducted in the spring of 2023. It was distributed to corresponding authors of articles in 21 environmental science journals, including the mega-journal *Sustainability*. Out of an initial sample of approximately 14,500 selected authors, the survey achieved a response rate of 12.5%, resulting in 1,800 responses. Consequently, the survey encompasses the evaluation of 1,800 articles, of which approximately 1,510 included responses to the questions on interdisciplinarity.

Bibliometric data

The study relies on bibliometric data from the Web of Science (WoS) database, which has been used to calculate bibliometric interdisciplinarity measures. We applied a local version of WoS maintained by the Norwegian Agency for Shared Services in Education and Research. We applied data on the references of the publications to calculate the three interdisciplinarity scores used in the study. Only publications with at least 10 WoS-indexed references were included in the comparative analysis, as a minimum number of references is required to reliably calculate interdisciplinarity scores. This threshold reduced the number of articles by approximately 7%, from 1,559 articles in Survey 1 to 1,756-1,455 in Survey 2 (depending on question).

Results

As previously described, we have two different surveys addressing the interdisciplinarity of publications. Below we report the results of both of them. The analysis reveals a limited correspondence between the authors' assessments of

interdisciplinarity and the bibliometric interdisciplinarity scores. Figure 1 shows a Box-Whisker plot comparing author ratings of interdisciplinarity (survey 1) with DIV*-values. As can be seen there is a tendency that publications rated as having high interdisciplinarity generally have higher DIV* values. For example, publications rated with the lowest level of interdisciplinarity (1) have a median DIV* value of 0.021, whereas those rated at the highest level (5) obtained a median DIV* value of 0.027. However, the most striking result is the large disparity in results. Many publications rated by authors as highly interdisciplinary do not exhibit high scores on the bibliometric measures, and vice versa. Very similar plots and patterns were observed in Survey 2, and to avoid redundancy, separate figures are not presented.

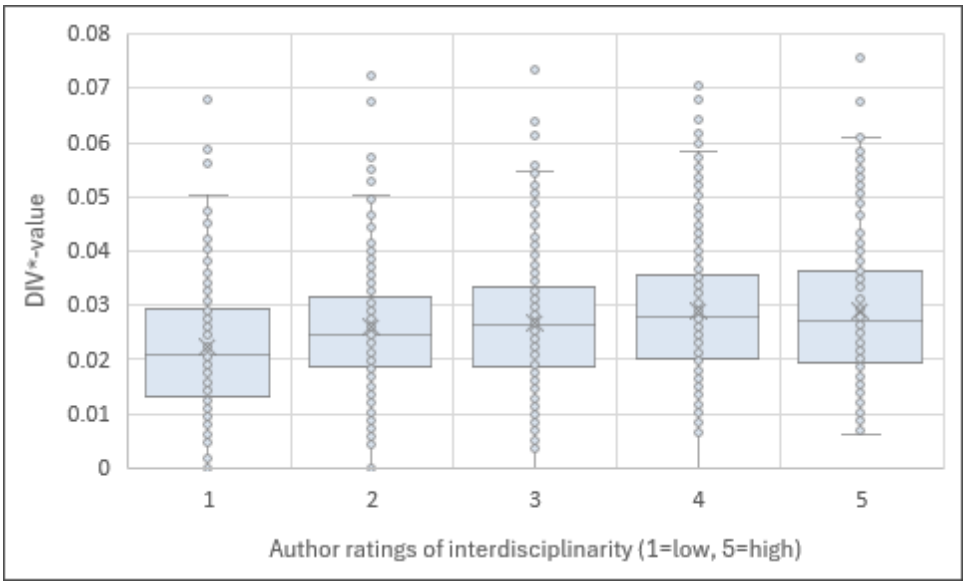

Figure 1. Box-Whisker plot of author ratings of interdisciplinarity (survey 1) and DIV*-values (interquartile range (1st-3rd), mean (cross), median (line within the box).

Table 1 shows the results of a correlation analysis using Spearman's rank-order method, meaning it is based on ranks rather than raw values. This is due to survey results measured on an ordinal scale, specifically for Survey 2, the variables are either trinary (e.g., Yes/Partly/No) or binary (e.g., Selected/Not Selected), necessitating the use of a non-parametric method.

The results on Survey 1 and the results on two of the questions in Survey 2 (Research from multiple academic disciplines (multidisciplinary) - Combines and integrates research across academic disciplines (interdisciplinary) are very similar. For these variable Spearmans's rho is in the range of 0.130-0.175 for all three bibliometric measures. This is a weak/very weak correlation. However, the p-values for these correlations are extremely low ($p < 0.00001$), indicating strong statistical significance, partly explained by the large number of observations. Among the associations, the correlations with Integrated disciplines are slightly higher, suggesting that this dimension aligns more strongly with the interdisciplinarity

measures. Conversely, the coefficients for the DIV* metric are marginally higher than those for the other measures, suggesting that DIV* is slightly more reflective of the authors' views.

Table 1. Correlation analysis: Relationship between interdisciplinarity measures and the authors' assessments (Spearman's rank correlation coefficients and p-values).

	Shannon entropy		$^2D^S$		*DIV**		
	Spear-man's rho	*P-value*	*Spear-man's rho*	*P-value*	*Spear-man's rho*	*P-value*	*N*
Interdisc. rank (1)	0.152	1.5×10^{-9}	0.130	2.6×10^{-7}	0.160	2.1×10^{-14}	1559
Multiple disc. (2)	0.160	7.3×10^{-10}	0.155	2.2×10^{-9}	0.158	1.1×10^{-9}	1466
Integrated disc. (2)	0.173	3.0×10^{-11}	0.155	2.8×10^{-9}	0.175	1.7×10^{-11}	1455
Other field researchers (2)	0.056	0.019	0.027	0.25	0.053	0.027	1756

For the indicator "Other field researchers," correlations are generally negligible (rho < 0.1) and have higher p-values. While the correlation with Shannon entropy and DIV* is marginally significant ($p < 0.05$), the correlation with $^2D^S$ is not statistically significant ($p = 0.25$). This suggests that the interdisciplinarity measures are less relevant for capturing this dimension.

Discussion & conclusions

The findings of this study contribute to the ongoing debate regarding the validity of bibliometric measures of interdisciplinarity by juxtaposing these with researchers' self-assessments. The results reveal that there is little agreement between bibliometric measures (Shannon entropy, $^2D^S$, and DIV*) and authors' assessments of interdisciplinarity: the correlations are weak to very weak. This suggests that bibliometric measures only to a very little extent capture researchers' perceptions of the interdisciplinarity of their own work. In particular, we got poor correspondence for the dimension "Other field researchers," indicating that bibliometric measures are not capable of reflecting the involvement of researchers from different fields. These findings are consistent with earlier studies, such as those by Avila-Robinson, Mejia, and Sengoku (2021), which also reported relatively weak correlations between bibliometric interdisciplinarity metrics and authors' views, albeit slightly stronger than observed in our study.

Another notable finding is that, despite variations in their construction and methodological foundation, the three bibliometric measures examined show a similar level of correspondence. The DIV* metric appears to align marginally better with

authors' self-assessments compared to Shannon entropy and ${}^{2}D^{S}$. This is reflected in the slightly higher correlation coefficients for DIV* across most variables.

The study is framed as a kind of validation exercise of bibliometric interdisciplinarity measures.

Our findings suggest that these metrics have inherent limitations and lack validity, at least insofar as researchers' perceptions can be used as benchmarks to assess the issue. This divergence naturally leads to the conclusion that one should not rely solely on bibliometric indicators to assess interdisciplinarity.

Nevertheless, the issue is complex. The relationship between bibliometric metrics and author assessments can also be interpreted in the opposite direction. Systematic and objective measures of interdisciplinarity are compared with subjective assessments by the authors. After all, measures like Shannon entropy, 2DS, and DIV* might provide useful insights into the diversity and integration of disciplines referenced in a publication. However, this dimension does not seem to reflect the "real-life" interdisciplinarity experienced by researchers. Perhaps researchers emphasize a more composite set of factors in their assessments, such as disciplinary norms, collaboration dynamics, methodological and theoretical approaches, etc. This suggests that researchers' perceptions are likely influenced by factors not readily captured by the bibliometric indicators, leading to limited comparability. Generally, the two approaches can be expected to correlate positively only if the aspects assessed by the authors correspond to those reflected in the bibliometric metrics.

Moreover, authors' assessments of their own publications may not be entirely inter-subjective. Different authors could rank the interdisciplinarity of the same publication differently due to varying perspectives, biases, or even memory limitations.

Thus, more studies are needed to determine the extent of inter-subjectivity in authors' assessments of interdisciplinarity, as well as to explore factors that shape researchers' perceptions of interdisciplinarity. Future studies could also expand the range of bibliometric measures to assess whether alternative metrics align better with researchers' perceptions (see, e.g. Abramo, D'Angelo & Di Costa (2012).

Acknowledgments

This work was supported by the Research Council Norway (RCN) [grant number 256223] (R-QUEST).

References

Abramo, G., D'Angelo, C.A., & Di Costa, F. (2012). Identifying interdisciplinarity through the disciplinary classification of coauthors of scientific publications. *Journal of the American Society for Information Science and Technology*. 63: 2206-2222.

Aksnes, D. W., Piro, F. N., & Fossum, L. W. (2023). Citation metrics covary with researchers' assessments of the quality of their works. *Quantitative Science Studies*, 4(1), 105-126.

Avila-Robinson, A., Mejia, C. & Sengoku, S. (2021). Are bibliometric measures consistent with scientists' perceptions? The case of interdisciplinary research. Scientometrics, 126, 7477-7502.

Cantone, G.G. (2024). How to measure interdisciplinary research? A systemic design for the model of measurement. *Scientometrics*, 129, 4937-4982.
Chen, S., Qiu, J., Arsenault, C. & Lariviere, V. (2021). Exploring the interdisciplinarity patterns of highly cited papers. *Journal of Informetrics*, 15, 101124.
Chen, S., Zhang, K., Qiu, J. & Chai, J. (2024). Interdisciplinary and expert rating: an analysis based on faculty opinions. *Scientometrics*, 129, 6597-6628.
Laursen, B.K., Motzer, N. & Anderson, K.J. (2022). Pathways for assessing interdisciplinarity: A systematic review. *Research Evaluation*, 31(3), 326-343.
Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019a). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal Informetrics*, 13 (1), 255–269.
Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019b). Diversity measurement: Steps towards the measurement of interdisciplinarity? *Journal Informetrics*, 13(3), 904–905.
Miller, R. (2020). Interdisciplinarity: Its Meaning and Consequences. *Oxford Research Encyclopedia of International Studies*. Retrieved 10 Oct. 2024, from https://oxfordre.com/internationalstudies/view/10.1093/acrefore/9780190846626.001.0001/acrefore-9780190846626-e-92.
Porter A. L. & Rafols I. (2009). Is Science Becoming More Interdisciplinary? Measuring and Mapping Six Research Fields over Time'. *Scientometrics*, 81, 719–45.
Roessner, D., Porter, A.I., Neresessian, N.J. & Carley, S. (2013). Validating Indicators of Interdisciplinarity: Linking Bibliometrics Measures to Studies of Engineering Research Labs. *Scientometrics*, 94, 439-468.
Rousseau, R. (2019). On the Leydesdorff-Wagner-Bornmann proposal for diversity measurement. *Journal Informetrics*, 13(3), 906.
Wang, Q. & Schneider, J.W. (2019). Consistency and validity of interdisciplinary measures. *Quantitative Science Studies*, 1, 239-263.
von Wehrden, H., Guimaraes, M.H., Bina, O., Varanda, M., Lang, D.J., John, B., Gralla, F., Alexander, D., Raines, D., White, A. & Lawrence, R.J. (2019). Interdisciplinary and transdisciplinary research: finding the common ground of multi-faceted concepts. *Sustainability Science*, 14(6).
Zhang, L., Rousseau, R., & Glänzel, W. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal Association for Information Science Technology*, 67, 1257–1265
Zhang, L., Sun, B., Chinchilla-Rodriguez, Z. & Huang, Y. (2018). Interdisciplinarity and collaboration: On the relationship between disciplinary diversity in departmental affiliations and reference lists. *Scientometrics*, 117, 271-291.