Toward greater consistency and validity in measuring interdisciplinarity: a systematic and conceptual evaluation

Sander Zwanenburg1 · Maryam Nakhoda1 · Peter Whigham1

Received: 31 October 2021 / Accepted: 14 February 2022 / Published online: 12 March 2022 © The Author(s) 2022

Abstract

While interdisciplinary research (IDR) has attracted much attention, this has not yet resulted in a coherent body of knowledge of interdisciplinarity. One of the impediments is a lack of consensus on its conceptualization and measurement. Some of the proposed measures have shown to misalign empirically, meaning that conclusions about IDR can differ across measures. To clarify this disagreement conceptually, and to stimulate better coherence in measurement, this paper starts with a review of the IDR definitions. From a synthesis of these definitions, we provide a conceptual definition and a logical structure of the construct, and derive evaluation criteria for its measures. We use these to evaluate 21 measures of IDR. The results show that measures vary widely in meeting the criteria, which can explain some of the observed inconsistencies in earlier studies. We discuss the most common limitations and present empirical analyses to gauge their severity. We present several suggestions for future measurement of the interdisciplinarity of research. We hope that with these suggestions, researchers can draw more consistent conclusions, aiding in the development of a coherent body of knowledge of this ever-important phenomenon.

Keywords Interdisciplinary score · Measures · Conceptualization · Validity

Introduction

Many complex problems arising in society require the integration of knowledge beyond the boundaries of a single discipline (Choi & Richards, 2017; MacLeod & Nagatsu, 2016; Okamura, 2019; Wang et al., 2017), such as climate change, food security, peace, social justice (Al-Suqri & AlKindi, 2017), and environmental issues (Morillo et al., 2003; Steele & Stier, 2000). Interdisciplinary research (IDR) is often seen as more innovative and has shown to be more impactful, at least in terms of citations (Leahey et al., 2017). Engaging in IDR, however, is challenging. A recent study found that it was associated with lower research productivity (Leahey et al., 2017). Specialised disciplines have defined research practices, languages, philosophies and communities that often seem inconsistent or incompatible.

* Sander Zwanenburg
sander.zwanenburg@otago.ac.nz

1 Department of Information Science, University of Otago, Dunedin, New Zealand
This problem of knowledge is not new; IDR has long been promoted to improve unity and synthesis of knowledge. The infrastructure of science has been shaped by a history of disciplines splitting into subspecialties, which has led to attempts to integrate or synthesize science from the middle of the last century (Klein, 1990). Now, with digital databases capturing the output of interdisciplinary research, studies on IDR has grown in the field of quantitative bibliometrics, often studying the diversity of cited disciplines by adapting diversity indicators from ecology and economics (Rao, 1982; Stirling, 1998). Such studies on IDR have an unprecedented potential to reveal insights into the practice, its success, and how its challenges are best overcome.

However, to date, studies on IDR have not constructed a coherent body of knowledge due to a lack of consistency in measuring interdisciplinarity (Leydesdorff & Rafols, 2011; Wang & Schneider, 2020). Choosing data sets and methodologies produce “inconsistent and sometimes contradictory” results (Digital Science, 2016; Wang & Schneider, 2020). Partly this may be due to different indicators capturing different aspects of interdisciplinarity (Leydesdorff & Rafols, 2011). For example, they may capture one or multiple of the three dimensions of diversity: variety, balance, and disparity (Leydesdorff et al., 2019). Some take into account the varying distances between fields; others do not. Some assume a network view, and others a more hierarchical structuralists perspective (Wagner et al., 2011). Yet even when measures purportedly capture similar aspects, they may still be empirically inconsistent (Wang & Schneider, 2020).

These inconsistencies stem from diversity in conceptualization and operationalization of interdisciplinarity. First, some authors (e.g., Karlqvist, 1999) have approached interdisciplinarity qualitatively, others quantitatively. The range of quantitative conceptualization of interdisciplinarity covers two distinct perspectives. One is considering the processes and dynamics that contribute to interdisciplinary research and the other one focuses on research publications and to what extent they integrate different research disciplines (Mugabushaka et al., 2016). Other operationalizations provide more details about the integration of different fields of knowledge, in terms of methods, tools, theories, and data (Porter et al., 2007; Rafols & Meyer, 2009). Each of these approaches can lead to different definitions, measures and indicators. Thus, the choice of a specific view and its operationalized indicators produce inconsistent results (Wang & Schneider, 2020).

If IDR would be quantified with more consistent measures, it is more likely that conclusions about this phenomenon, including how to maximise its success, will also be consistent, allowing for a more coherent body of knowledge.

In this paper we aim to take a first step towards this. We first define interdisciplinary research based on a synthesis of 25 definitions in the literature. From this definition, we derive its logical structure and evaluation criteria. We then use these to assess 21 measures of IDR. We then discuss the results and the common limitations of measures, and present empirical analyses to gauge their severity. Last we discuss how researchers can choose the best approaches, measures, and techniques toward more consistency and coherence in developing an understanding of IDR.
Definition of IDR

Our review of 25 definitions of interdisciplinarity provided in the literature reveals various similarities. Within the context of quantitative, each definition of interdisciplinarity either explicitly or implicitly refers to multiple disciplines, or similarly, fields, or bodies of knowledge or research practice. Regardless of the name given to them, they are treated as distinct collections (or systems) of ideas, knowledge, data, tools, theories, practices, specialists, or combinations thereof. Central to most definitions is a connection, communication, or exchange of such items across distinct collections, and/or a resulting intellectual synthesis, fusion, or integration. This result may serve different purposes, but typically relate to understanding issues or solving problems that are not contained to a single discipline.

Most definitions imply these collections must be distant, distinct, or diverse, either sufficiently so for interdisciplinarity to ‘occur’ or ‘apply’, or increasingly so for IDR to be more evident or pronounced. Similarly, some definitions imply that a condition for IDR is that the resulting synthesis or integration must be of a sufficient degree or depth. Most definitions do not carry assumptions or criteria about those conducting IDR, such as teams versus individuals, or specialists versus generalists. They also do not tend to constrain IDR as an attribute of particular classes of objects, such as authors, papers, departments, or fields of study. Based on this review, we believe that interdisciplinary research is best defined as the integration of knowledge from diverse disciplines.

This definition is consistent with past literature in four important ways. First, it allows for a variety of objects of interest. IDR has been studied in the context of individual papers (Chen et al., 2015; Porter et al., 2007), authors (Qin et al., 1997), groups of authors such as departments or universities (Bordons et al., 1999; Gowanlock & Gazan, 2013), journals (Leydesdorff, 2007) and disciplines (Rinia et al., 2002; Schummer, 2004). Any of these objects can all be described in terms of the extent they achieve or reflect the integration of knowledge from diverse disciplines.

Second, the definition recognizes that integration of knowledge lies at the heart of the concept of IDR (e.g., Chang & Huang, 2012; Klein, 1990; Porter & Rafols, 2009; Rafols & Meyer, 2009; Rhoten & Pfirman, 2007). While the integration of knowledge is fundamentally a cognitive process (Wagner et al., 2011), it may emerge both from within individuals and from interactions of multiple individuals (Porter et al., 2007). Based on the review, integration must be broadly understood as any combination, fusion, synthesis, or even juxtaposition of two or more ideas, producing a whole idea, a more general idea, or a meta-idea. This differentiates IDR from multi-disciplinarity, where multiple disciplines are simply represented, without being integrated necessarily. While philosophically, it is possible to speak of the depth or degree with which two or more ideas are integrated (Rafols & Meyer, 2009), definitions of interdisciplinarity tend to be more consistent with a dichotomous view: ideas are assumed integrated, or they are not.

Third, the definition recognizes a link between knowledge and disciplines. Indeed, it is not the disciplines themselves that are being integrated in interdisciplinary research but some knowledge within them. A discipline is a distinct collection of facts, concepts, and methods (Barry et al., 2008; Braun & Schubert, 2003). Therefore, discipline refers to any

---

1 We searched for keywords such as “interdisciplinarity” and “interdisciplinary research” within academic databases (e.g., WOS, Scopus, and ProQuest). Around 100 sources including books, journal papers, and research reports were retrieved from which we retrieved 25 definitions from various sources. For space considerations, they are not reproduced here.
area of study that has had some level of coherence, by having defined communities, educational programmes, research culture, and a shared body of knowledge. Since the boundaries of disciplines can vary across communities and geographies, change over time, and be expressed at different levels of granularity, there is not universally accepted or persistently valid classification of disciplines.

Last, the definition recognizes that a collection of disciplines can be more or less diverse and that this degree of diversity is a common basis for attributing values on continuous scales of interdisciplinarity. In the context of disciplines, diversity refers to the variety, balance, and disparity of a set of disciplines (Leydesdorff & Rafols, 2011; Zhou et al., 2012). These dimensions or aspects of diversity are additive: the higher the number of different disciplines in a set, the more evenly their frequency is distributed, and the more disparate they are from each other, the more diverse is the set. By extension, the more diverse the disciplines from which research integrates knowledge, the more interdisciplinary the research.

These four aspects constitute a logical structure for the construct of IDR, as depicted in Fig. 1. In the following, we aim to evaluate how commonly used measures of IDR align with this definition and its implied logical structure.

Method

Our conceptual evaluation of commonly used measures vis-à-vis a synthesized definition of IDR will follow three stages. First, from the preceding review we will develop conceptual criteria for measures, such that measures that meet these criteria are in alignment with the definition of IDR. Second, we identify and summarize the measures of IDR. Third, we evaluate these measures according to the evaluation criteria.
As part of the last stage, we examined in particular the commonly used strategy to link knowledge to a set of disciplines by relying on the reference papers allocation to Web of Science Subject Classification. We also examined the strategy of quantifying diversity of a set of disciplines based on measures of discipline similarity.

**Developing evaluation criteria**

Good measures can be used where needed (usability) while validly quantifying the position of an object of interest on a scale (construct validity; MacKenzie et al., 2011). By evaluating the logical structure of the construct of IDR based on these qualities we will identify our evaluation criteria.²

1. IDR scores describing objects of interest.

   Using a common measure across studies on authors, journals, schools, disciplines etc. will help develop a common body of understanding of IDR. Therefore:

   1a; *Multi-object* The measure is applicable to a variety of objects of study, including individual papers, authors, institutions, journals, and disciplines. This criterion is satisfied if a measure is scalable from a small object of interest (e.g., a given paper) to a larger object of interest (e.g., a journal or discipline).

   Most research on IDR involves evaluations and comparisons within or across these objects of study (Leydesdorff, 2007; Morillo et al., 2003; Wang & Schneider, 2020). For example, one may wish to compare a new journal to an established journal, or a small department to a larger one. Given that they can vary greatly in the quantity of research they represent, such comparisons are only meaningful when controlling for these quantities.

   1b; *Size independent* The measure’s values for IDR of objects of study are independent of the amount of research (e.g., number of published papers) belonging to these objects.

2. Objects of interest integrating knowledge

   If the measure is to reflect knowledge integration, it must be based on evidence that knowledge is indeed integrated. This is a trade-off between strength of evidence and practicality:

   - When the integration of ideas is examined by hand, i.e. cognitively, this evidence can be strongest, but also most expensive. The degree to which ideas are integrated is an abstract conception; it will take much time to assess this cognitively, let alone arrive at an inter-rater consensus for all ideas presented in a given paper. To our knowledge, no such endeavor has been employed at a scale that allows for finding quantitative patterns in IDR.
   - When full text articles are used with automated citation analyses, strong evidence can be obtained (Craven et al., 2019; Karlovčec & Mladenić, 2015). However, retrieving full text articles that contain machine readable citations within the text

² In this paper we will not evaluate measures based on quantitative properties, such as scaling, distribution of scores, and discriminatory power across the scale. In our view, these considerations are secondary to ensuring a strong link between the meaning of IDR and how it is measured (Zwanenburg et al. 2019).
is fraught with practical difficulties and limitations, including various formats and levels of access across journals, other venues, and disciplines. None of the reviewed measures follow this approach.

- When the list of references is used for each paper, adequate evidence can be obtained. This approach would be less precise than the above approaches, since it assumes that each of the references contains ideas that are being integrated, and to the same extent. However, it is far more practical, since databases like Web of Science do include meta-data on the references, while they lack this data on individual citations.

- When all references belong to a set of papers (e.g., making up a journals or an entire discipline’s collection) the evidence is arguably less than sufficient. For example, a journal such as Nature contains many parallel conversations in separate disciplines, without integrating their ideas. In other words, when an object of interest is linked to diverse cited disciplines, it is not necessarily integrating knowledge from these disciplines, especially if these citations occurred across many integrated works of knowledge, such as a large set of papers.

- When no citations or reference lists are used, but only the (more readily available) assignments between papers (or other knowledge artefacts like journals) and disciplines are used, there is no evidence of integration at all. For example, if a discipline contains many journals that also belong to other disciplines, there may be an overlap of disciplines in terms of the ideas they cover. While this may well be consistent with definitions of multi-disciplinarity, it is inconsistent with the literature on interdisciplinarity since this does not point to integration of ideas from diverse disciplines.

Hence, the second criterion is:

2: Evidence of integration The measure is based on sufficient evidence that knowledge is integrated.

3. Knowledge being allocated to a set of disciplines.

Based on heterogenous nature of interdisciplinary papers (Xu et al., 2021), measuring the integration of knowledge from source disciplines also requires an appropriate allocation of knowledge to their disciplines. Like the integration criterion above, evaluating this allocation manually can be valid but is time consuming. Any quantitative, systematic, and scalable approach will need to rely on identifying knowledge with explicit knowledge artefacts, e.g., phrases, paragraphs, papers, journals, and books, and in turn linking these artefacts with disciplines. Regardless of which type of knowledge artefact is selected, there will be assumptions and limitations in the method of linking these to disciplines. For example, a source paper may be linked to a discipline if that source paper is part of a journal which in turn is categorized into one or multiple disciplines. Not all papers are journal papers, and some journal papers may be at the fringes of a journal’s domain and not necessarily be appropriately classed via its journal. Hence, the method of linking knowledge artefacts to disciplines has implications for validity of each allocation and the completeness of these allocations:

3a; Valid discipline identification each allocation between source knowledge and its discipline is valid.

3b; Identification of all source disciplines all source knowledge is identified and allocated to disciplines.
Measures of IDR have leveraged various classifications of disciplines, such as the Web of Science Subject Classification (Wagner et al., 2011), the All Science Journal Classification (Leydesdorff et al., 2015), the Leuven-Budapest classification (Thijs et al., 2015), or proxies of disciplines, such as journal title, or citation-based fields (Hernández & Dorta-González, 2020). All of these have different granularities. While some have presented evidence of satisfactory granularity of some of these classifications (Huang et al., 2021), it is unlikely that consistency in classification use will be achieved. Communities and institutions have been using different discipline classifications (or categorizations) (e.g., Zhang et al., 2015), and these classifications may change over time (Huutoniemi et al., 2010). To accommodate multiple classifications within consistent measurement, measures should be able to be applied to any of this minimizing bias. In particular, measures should not result in widely different scores of IDR if classifications of different granularities are used. For example, relying on simple counts of the number of disciplines without controlling for their disparity will introduce a bias and complicate the comparison of applications across classifications.

3c; No or low classification bias: The measure is applicable to any discipline classification and is independent of its granularity.

4. A set of disciplines being diverse

The content of the diversity construct, as used in the IDR context, consists of disparity, variety, and balance (Stirling, 1998, 2007; cf. Rao, 1982). Valid measures of IDR thus co-vary with variations in these three components.

4a; All diversity aspects captured The measure is sensitive to differences in the disparity, variety, and balance of source disciplines, such that the measure produces higher scores with more disparate source disciplines, more source disciplines, and when source knowledge is more evenly distributed across disciplines.

When measures capture each of the diversity aspects separately, IDR scores can be decomposed. While these measures are not more valid per se, they are more helpful as they allow for attributing IDR scores to these components, potentially enhancing transparency and understanding of these scores.

4b; Each diversity aspect captured The measure can be decomposed into scores of disparity, variety, and balance of source disciplines.

How a measure of IDR is best formed by these scores, for example as their sum or their product, cannot be directly inferred from the definition of IDR. This topic has received some recent attention (e.g. Mutz, 2021), but this has not yet produced a clear, well-established consensus to merit inclusion of this aspect into our conceptual criteria.

Identifying the measures

For our evaluation, we included all measures included in the empirical evaluation by Wang and Schneider (2020). We updated this set by including measures recently proposed in the literature, regardless of them being empirically deployed or not, the intent or motivation behind their proposal, or any particular object of study (Wang et al., 2015). To prevent redundancy, we did not include general or base measures when more specific applications or refinements of these measures were included (for example, Shannon entropy and Simpson index have been applied in the Overall Diversity Indicator and the Rao-Stirling measure respectively).
1. **P_multi** This measure indicates the percentage of journals with multiple categories in the category of interest, thus reflecting how closely related a discipline is to other disciplines (Morillo et al., 2001).

2. **p_outside** The multi-assignation indicator for a subject category is specified by the percentage of journals assigned to more than one category outside the research area (Morillo et al., 2001), which is in turn composed of several WOS SCs. (Wang & Schneider, 2020).

3. **Pro** This measure is calculated as the percentage of references/citations from categories different to the subject categories of references/citations of the focal journal (Morillo et al., 2001; Porter & Chubin, 1985).

4. **D_links** Diversity of links, for journals of a given subject category, is the number of different categories that belong to the same journals. The link is the number of related categories and established between the research areas or subject categories, with properties like diversity and strength (Morillo et al., 2003).

5. **Pratt index** This measure is an index to assess frequency of distribution to compare journal and subject concentration in different fields, as the interdisciplinarity of a subject field is acquired by having the greater proportion of cited references over different categories (Pratt, 1977).

6. **Specialization index (Spec)** This measure reflects on the distribution of cited references of a paper in each category over all other subject categories (Wang & Schneider, 2020). This assumes that integration and specialization (a researcher’s publications) give a more complete picture of interdisciplinarity of research publications (Porter et al., 2007).

7. **Brillouin index** It is also regarded as a measure of uncertainty. High uncertainty equals to high diversity. This measure can be applied to authorship, subject fields, and cited references, and considers the “richness”, the number of observations, and “relative abundance” which is their spread over categories (Steele & Stier, 2000).

8. **Gini coefficient** The concepts of uncertainty and inequity are represented in this measure as it can be related to Shannon entropy in interpretation. It considers the distribution of references over SCs for a group of publications (Wang et al., 2015). At journal level, the distribution of cited references has maximum inequity in case it cites the papers in the journal itself. Thus, the entropy is zero and the journal is mono-disciplinary (Leydesdorff & Rafols, 2011).

9. **Rao-Stirling (RS) diversity index** Also known as integration score, this is a measure of diversity of knowledge sources of papers, including these dimensions: the number of distinct categories (variety), the evenness of the distribution of citations among categories (balance), and the degree to which categories are similar or dissimilar (disparity) (Porter & Rafols, 2009). Increase in any of these attributes leads to increase in the diversity of the system (Rafols & Meyer, 2009).

10. **Hill-type true diversity index** This measure is related to the RS diversity index, and gives an indicator for similarity, variety, and balance (Zhang et al., 2015). This measure satisfies properties such as decrease in diversity with growth of similarity of cited publications, and having a range for diversity between one and number of species not limited between 0 and 1 (Leydesdorff & Ivanova, 2021).

11. **Coherence** This indicator calculates knowledge integration at the level of subject categories (Soós & Kampis, 2012). To operationalize, normalized Salton’s cosine is measured for bibliographic couplings, using “two indicators of network coherence: mean linkage strength and mean path length” (Rafols & Meyer, 2009).
12. **Betweenness-centrality (BC)** A measure proposed to assess interdisciplinarity of journals. Betweenness measures the degree of centrality that a node (representing e.g., an article or journal) is located on the shortest path between two other nodes in a network (Leydesdorff, 2007).

13. **Cluster Coefficient (CC):** As a network analysis measure and an indicator of intermediation, it refers to “the proportion of observed links between journals over the possible maximum number of links”, which is weighted by proportion of references/citations per journal (pi) (Rafols et al., 2012).

14. **Average Similarity (AS)** In this indicator of intermediation, average similarity of a focal journal to N other journals is assessed and weighted by distribution of publications across categories (Rafols et al., 2012).

15. **Journal interdisciplinarity** This measure quantifies the degree of specialization versus interdisciplinarity exhibited by a journal based on the bipartite network between scholars and journals (Carusi & Bianchi, 2020).

16. **Author interdisciplinarity** This measure extends Rao-Stirling diversity measure, with co-author networks and from the perspective of disparity (Zhang et al., 2020).

17. **Diversity measure (DIV)** This measure is the product of (a) variety, (b) balance, and (c) disparity, where variety is controlled for the number of classes in the set, balance is given by 1-Gini, and disparity based on a normalized distance between categories (Leydesdorff et al., 2019). The measure is implemented by linking a paper’s references to Web of Science categories.

18. **Interdisciplinary Research Index** This measure is based on a citation network, where a given paper’s references are evaluated in terms of their co-occurrence in reference lists in other papers, using motives (Hernández & Dorta-González, 2020). The index indicates the inverse of the degree a paper’s reference set is commonly cited together, and thus a difficulty of attributing it to a specific research area. It is implemented using the Scopus database and its recognized citations.

19. **Overall diversity indicator div** This is a modification of the diversity measure DIV (numbered 17). Based on Shannon’s probabilistically based, it reconceptualizes the three components of diversity as entropy masses and adds these together rather than multiplies them (Mutz, 2021).

20. **Reverse Simpson index (RSI)** This measure equals 1 minus the sum of the squared proportions of each discipline in a given reference set, thus being reflective of how concentrated vs distributed a reference set is (Xu et al., 2021).

21. **Refined Diversity measure (DIV*)** This measure equals N * DIV (Rousseau, 2019) – as such, it allows for values up to N for a set of N equally abundant, totally dissimilar items, which is seen as a true diversity measure (Jost, 2009).

**Results**

The outcome of the evaluation is summarized in Table 1. Our findings indicate that the measures of IDR vary widely in meeting the conceptual criteria. We will discuss each of the criteria and pay particular attention to those criteria that are met least. In our discussion section we highlight several approaches to meet them with refinements or future work.
| Measure                                                                 | 1a: Multi-object | 1b: Size independent | 2: Evidence of integration | 3a: Valid discipline allocation | 3b: All sources identified | 3c: No/low classification bias | 4a: All diversity aspects captured* | 4b: Each diversity aspect captured separately |
|------------------------------------------------------------------------|------------------|-----------------------|----------------------------|--------------------------------|-----------------------------|----------------------------------|-----------------------------------------|------------------------------------------|
| 1. Percentage of multi-assigned journals (p_multi)                     | x                | x                     | ?                          | x                              | x                           | x                                | x                                       | x                                        |
| 2. Percentage of journals outside the area (p_outside)                 | x                | ✓                     | x                          | ?                              | x                           | x                                | x                                       | x                                        |
| 3. Percentage of references outside the category (pro)                 | x                | ✓                     | x                          | ?                              | x                           | x                                | x                                       | x                                        |
| 4. Diversity of links (d_links)                                       | x                | ✓                     | x                          | ?                              | x                           | x                                | 1                                       | x                                        |
| 5. Pratt index                                                         | x                | ✓                     | x                          | ?                              | 1                           | x                                | 1                                       | x                                        |
| 6. Specialization index (Spec)                                         | ✓                | ✓                     | x                          | ?                              | x                           | 1                                | 1                                       | x                                        |
| 7. Brillouin diversity index                                           | ✓                | x                     | x                          | ?                              | x                           | 2                                | x                                       | x                                        |
| 8. Gini coefficient                                                    | ✓                | ✓                     | x                          | ?                              | x                           | 1                                | x                                       | x                                        |
| 9. Rao-Stirling (RS) diversity index                                   | ✓                | ✓                     | ✓                          | ?                              | ✓                           | ✓                                | ✓                                       | ✓                                        |
| 10. Hill-type measure                                                  | ✓                | ✓                     | ✓                          | ?                              | ✓                           | ✓                                | ✓                                       | ✓                                        |
| 11. Coherence measure                                                 | ✓                | ✓                     | x                          | x                              | ✓                           | x                                | x                                       | x                                        |
| 12. Betweenness-centrality                                             | ✓                | ✓                     | x                          | x                              | x                           | x                                | x                                       | x                                        |
| 13. Cluster Coefficient (CC)                                           | x                | x                     | x                          | x                              | x                           | x                                | x                                       | x                                        |
| 14. Average Similarity (AS)                                            | x                | x                     | x                          | x                              | x                           | x                                | x                                       | x                                        |
| 15. Journal interdisciplinarity                                         | x                | ✓                     | x                          | ?                              | x                           | ✓                                | x                                       | x                                        |
| 16. Author interdisciplinarity                                         | ✓                | ✓                     | x                          | ?                              | x                           | N/A                              | ✓**                                     | x                                        |
| 17. Diversity measure (DIV)                                            | ✓                | ✓                     | ✓                          | ?                              | ✓                           | ✓                                | ✓                                       | ✓                                        |
| 18. Interdisciplinary Research Index                                   | ✓                | ✓                     | ✓                          | ?                              | ✓                           | x                                | x                                       | x                                        |
| 19. Overall diversity indicator (div_v)                                | ✓                | ✓                     | ✓                          | ?                              | ✓                           | ✓                                | ✓                                       | ✓                                        |
| 20. Reverse Simpson Index (RSI)                                        | ✓                | ✓                     | ✓                          | ?                              | x                           | 1                                | x                                       | x                                        |
| Refined diversity indicator (DIV*)                                     | ✓                | ✓                     | ✓                          | ?                              | ✓                           | ✓                                | ✓                                       | ✓                                        |

* ‘1’ and ‘2’ refer to one or two of three criteria being met

**The three dimensions are tapped into, but not as applied to the diversity of disciplines but that of authors
Evaluating Criterion 1a: Multi-object

About two-thirds of the measures were found to allow for a variety of objects of study typically because they allow for scaling from one paper to any set of papers, while some are only applicable to limited objects of interest (e.g., a set of journals such as an entire discipline).

Evaluating Criterion 1b: Size Independent

Nearly all measures are size independent, such that objects of interest do not tend to get higher scores when they belong to more research.

Evaluating Criterion 2: Evidence of Integration

Most measures do not leverage evidence of knowledge integration, or insufficiently so. The typical reason for this was that the consideration of multiple disciplines was not in the context of disciplinary sources. For example, the journal interdisciplinarity measure (numbered 15) taps into clusters of communities, without directly tapping into integration of knowledge. Similarly, other indicators are also more indicative of how clearly a given work is attributable to a discipline (notably 18). Another example is leveraging entire collections of citations at an aggregate level (such as journal level), ignoring which of these citations come from which paper. Arguably, many measures that do not meet this criterion tap into ‘multi-disciplinarity’ where multiple disciplines are represented in a set. While network-based measures may rely on citation counts, if the network is specified at a level higher than paper level (e.g., journal, field, or discipline) they do not rely on evidence that knowledge from multiple disciplines are integrated. The six measures that met this criterion all relied on paper-based references.

Evaluating Criterion 3a: Valid discipline allocation

Most measures themselves do not explicitly state how disciplines are to be identified. In these cases, we examined implementations of the measures, or typical implementations. In many cases the validity of discipline allocation is ambiguous and/or dependent on the data—in the table, we marked this with question marks.

Some implementations relied on the dataset itself (notably 15 and 18), with authorship or citation data being aggregated to define communities or clusters and to allocate knowledge to these pseudo-disciplines. These approach is potentially elegant and versatile, yet also introduces various assumptions. First, it assumed that papers have amassed sufficient data (being either citations or authorships in its network) to carry meaningful information about their discipline. Many papers, especially new ones, are not cited at all. Further, this approach is heavily dependent on the choice of dataset, which may be problematically arbitrary at small scales. At large scales, this approach can be computationally intensive. Papers may be grouped differently over time and with different datasets, resulting in various scores and potentially limiting the generalizability of the
findings. Future work is needed to validate the findings and examine their robustness with diverse selections of papers to be analysed.

The degree to which these clusters or communities indeed represent disciplines will be best when large datasets are used. A disadvantage of this approach is that the identity of disciplines is not discerned and will vary from study to study. For a given object of interest, such as a paper, multiple values of interdisciplinarity may be obtained depending on which other objects are included in the dataset.

An entirely different approach to overcome this limitation is to not rely on pre-specified disciplinary classifications, but to define the disciplines based on the dataset itself. For example, when a paper is citing two works the relative frequency of these two works being co-cited in the literature can constitute an indicator of their similarity or their likelihood of belonging to the same discipline (Hernández et al., 2020). This approach is potentially elegant and versatile, yet also introduces various assumptions. First, it assumed that cited papers have amassed sufficient citations to carry meaningful information about their discipline. Many papers, especially new ones, are not cited at all. Further, this approach is heavily dependent on the choice of dataset, which may be problematically arbitrary at small scales. At large scales, this approach can be computationally intensive. Papers may be grouped differently over time and with different datasets, resulting in various scores and potentially limiting the generalizability of the findings. Future work is needed to validate the findings and examine their robustness with diverse selections of papers to be analysed.

Most implementations relied on external allocations connecting papers or publication venues to disciplines from a classification such as the Web of Science Subject Categories or the All Science Journal Classification. This popular approach relies on the assumption that the venue’s allocated discipline accurately describes the discipline of the papers within this venue. However, classifications change over time with allocations not always retrospectively applied (Thijs et al., 2021). Further, some venues may span multiple disciplines, such as Science or Nature, or exist at the fringe of a recognized discipline with some of its papers falling outside of the discipline. The Web of Science Subject Classification recognizes this by allowing journals to be allocated 1) multiple categories, and 2) categories that are termed ‘multidisciplinary’, whereas the All Science Journal Classification recognizes this by having some disciplines be termed ‘miscellaneous. How these properties are handled will have an impact on validity and coverage.

Some implementations have simply removed categories with unspecific names (Huang et al, 2021), sacrificing allocations in the dataset may be detrimental to the validity of IDR scores, especially for those papers with few remaining data points of source disciplines. Keeping these categories and treating them like any other is not necessarily problematic since any set of disciplines will naturally vary in terms of scope and specificity. Further, if citation-based similarity scores are used to quantify diversity, the substantive identity of these broader disciplines is taken into account quantitatively.

Most implementations that involve or could involve multiple disciplines per cited paper did not contain details on how these were treated. Some reference-based approaches have simply included all these disciplines into the set of source disciplines, including applications of the Rao-Stirling and Coherence measures (Rafols & Meyer, 2009). While procedurally convenient, this approach can inflate IDR scores. To illustrate, a paper may cite a multi-disciplinary journal for just a single idea or quote, thus not integrating knowledge from the disciplines associated with that journal.

To clarify the importance of employing a sound strategy in dealing with multiple disciplines per paper, we examined the occurrence of multiple Web of Science categories, based on the research output database at our institution. Of those where at least one broad subject
category could be identified \( (n=42,314) \), 30% were assigned multiple broad categories. This portion is stable over time and illustrated in more detail in Fig. 1. Naturally, when considering the specific categories rather than the five broad disciplines, this portion is much higher. This underlines that measures that rely on Web of Science Subject Categories need special care to avoid biasing the results. In the discussion section we highlight an approach to do so, and alternatives proposed in the literature.

**Evaluation Criterion 3b: Identification of all source disciplines**

Some measures did not focus on disciplines being sources per se, including concentration-like measures (number 1–3) and most network-based measures. Since other measures do not explicitly state how source disciplines are identified, we again examined their implementations. We concluded that many implementations did not provide evidence of coverage, or mechanisms to ensure completeness. Most measures relied on standard classifications of disciplines, typically based on the venue of the cited work.

To examine coverage, we evaluated the degree with which disciplines of references can be identified, using a large set of recent publications \( (n=42,315) \) authored by staff at our institution that were included in any of the databases of Web of Science. In particular, we wanted to examine if publication year and broad category were associated with coverage and may present a bias in measures of IDR that rely on this approach. Over time and per broad category, we plotted the portion of the references of these 42,315 citing publications that we could allocate a discipline based on the journal-discipline mapping by Web of Science. The result is given in Fig. 2.

We found that coverage improved over time, approaching or slightly exceeding 75% across all non-Arts disciplines. However, in the Arts discipline, coverage was generally poor, albeit improving. In the Arts, a higher portion of references were types different from
journals thus not being included in the discipline mapping of Web of Science. Our findings indicate that IDR scores on measures that rely on Web of Science Subject Classification are (1) less reliable for Arts-heavy papers, (2) negatively biased for those papers that integrate knowledge from Arts disciplines with other disciplines, and (3) generally less reliable for papers citing older references, regardless of the discipline. We expect similar tendencies in the All Science Journal Classification.

A few measures did, however, meet the criterion. For example, when relying on author communities or citation networks it is possible to identify disciplines for more, if not all papers. This requires for each paper a sufficient number of either authorships or citations in the network around this paper.

**Evaluating Criterion 3c: No or low classification bias**

Measures that did not rely on standard classifications met this criterion, while for those that did rely on these, we found mixed results. Some were more sensitive to the choice of classification than others, often because they did not account for the disparity of disciplines. This would result in higher scores when classifications at finer levels of granularity are employed. While this is a general effect among standard classification measures, it is dampened when the measure controls for disparity, as ‘splitting up’ a field will result in subfields that are similar and therefore do not contribute as much to IDR scores on these disparity-controlling measures.

**Evaluating Criterion 4a: All diversity aspects captured**

There was considerable variability in tapping into, or being sensitive to, the different dimensions of diversity. Some measures failed to meet the criterion entirely because they focused on how often disciplines were in or outside a set (e.g. measures 1, 2, and 3). As they ignored the specific identity of these disciplines, they did not capture any of the diversity aspects.

Of all aspects of diversity, variety was most often informing measures, either explicitly or implicitly, whereas disparity did so least often. This is unfortunate, not just because disparity is an established aspect of diversity per se, also because accounting for disparity helps protect measurement against bias from the choice of discipline classification. In constructing a common and consistent body of knowledge on IDR, we believe this is critical given that (1) defining any classification of disciplines involves a degree of arbitrariness, and (2) that multiple classifications are used in the literature and around the world, and (3) that classifications change over time.

Various measures did capture all aspects of diversity include Rao-Stirling diversity index, Hill-type (true diversity) measure, Journal interdisciplinarity, DIV, and div.ε. While there are a variety of ways to quantify disparity (Bromham et al., 2016; Leydesdorff & Ivanova, 2021; Zhou et al., 2021), one approach stood out. The typical way these measures or implementations of them quantified disparity was by equating it to 1 minus the cosine similarity of the discipline-to-discipline citation counts using large datasets, such that disciplines that tend to cite the same disciplines are deemed more similar and less disparate than disciplines that have more diverging citation patterns (Leydesdorff & Ivanova, 2021). When these datasets are large enough, for example all published papers in a discipline in a given year, they are sufficiently stable and reliable for most analyses (Huang et al., 2021).
Evaluating Criterion 4b: Each diversity aspect captured separately

Out of 21 investigated measures, only two indicators met this criterion. Some of the measures that were sensitive to variations in the three aspects did so implicitly, i.e. without explicitly capturing each of the three aspects.

The two measures that did so expressed IDR scores as either the product of the three components (Leydesdorff et al., 2019) or their sum (Mutz, 2021). While Mutz (2021) presents an argument to sum these components rather than multiply them as was done in earlier proposals, our evaluation is that the definition of IDR, based on a synthesis of the literature, is consistent with either approach. A conceptual for either approach may emerge from a thorough conceptual evaluation of the diversity concept, which is beyond the scope of this research.

What both measures have in common, is that IDR can be decomposed into its constituent factors or components, enabling deeper study of the phenomenon of interdisciplinarity with the measure itself. For example, this decomposition can shed light on forms or types of IDR, or the relationships between the aspects of diversity.

Discussion

In this paper, we reviewed definitions of IDR, synthesized these into one definition, and derived from this its logical structure and evaluation criteria. We then evaluated 21 measures of IDR according to this definition, and discussed the variability and limitations.

Our findings confirm that measures of IDR vary widely in their consistency with the definition of IDR and its logical structure. We believe that these differences can go a long way in explaining the empirical disagreement between measures of IDR as observed earlier (Wang & Schneider, 2020).

While none of the measures clearly satisfied all of the criteria, some of the current measures, in particular those that capture all diversity aspects, provide a good platform for future measurement of interdisciplinarity that is more valid and coherent. The use of study-specific classifications or groupings of disciplines based on citation or authorship networks is in its infancy, has been executed only few times, and carries a number of assumptions and limitations discussed earlier. While adoption of these novel approaches may enrich the methodological toolbox, we expect that these approaches will not contribute to more empirical agreement of measures until they have matured and are better understood.

We expect the reliance on standard classifications will continue. Care must be taken in leveraging them. In our study we identified that key limitations around their use include unclear or inconsistent treatment of multiple categories per paper and an incomplete coverage.

First, to deal with multiple categories per paper, as is relatively common when using the Web of Science Subject Classification, some have simply included all disciplines as source disciplines. However, since citing papers do not necessarily integrate ideas from multiple disciplines belonging to a cited paper, this approach is introducing a bias. We suggest an alternative approach of evaluating disciplinary distance or similarity for each pair of papers. For example, in an adapted form of the Rao Stirling measure, the similarity component $S_{ij}$ may need to be respecified at the level of papers rather than disciplines, for example using a simple average of each pair of disciplines belonging to two papers:
where $PS_{ij}$ is paper similarity, $i$ and $j$ are two distinct cited papers, $k$ and $l$ are subject categories of papers $i$ and $j$ respectively; $n$ and $m$ are the number of subject categories of papers $i$ and $j$, $SS_{k,l}$ is the Subject Similarity of subject categories $k$ and $l$ (often the cosine similarity of the arrays of category-to-category citation counts). Using paper similarity, no separate proportions of occurrence need to be calculated, reducing the IDR equation to:

$$IDR = 1 - \sum_{i,j(i\neq j)} PS_{ij}$$

Second, considering the coverage of discipline identification, our analysis indicated that relying on venue-based classification approaches will be limited and introduces biases related to certain disciplines and year of publication. Three responses are possible:

- To simply remove uncategorized items from the analysis, accepting any bias stemming from this.
- To augment the venue-based classification with another classification approach. For example, supervised machine learning techniques can be employed to learn patterns from those venue-based classification and apply them to unclassified items.
- To switch to alternative classification approaches. One approach is to rely on networks of citations and/or authorships, to infer category membership, similar to some of the proposed measures. Another is to link knowledge to disciplines not via papers but via knowledge memes, as has been proposed in the area of knowledge diffusion (Mao et al., 2020) or other techniques that leverage text mining Karlovčec & Mladenić, 2015; Craven et al., 2019). A third approach is to infer disciplines by examining the references of the cited and identify the most dominant discipline in this set (Glänzel et al., 2021).

**Conclusion**

The literature has shown measures of ostensibly the same construct, interdisciplinarity, do not align empirically (Wang & Schneider, 2020). Our results take a step in identifying the underlying conceptual reasons for this disagreement. There are several take-aways for researchers interested in operationalizing interdisciplinarity:

First, to develop a coherent body of understanding on IDR caution is to be exercised in terms of the selection of measures. Our results suggest that the observed empirical disagreement is not just noise but there are significant conceptual differences underlying these measures. The danger of picking a convenient measure is that its conceptual foundations do not align with much of the extant literature and that the results will not be able to contribute to it.

Second, to exercise such caution means clarifying the conceptual definition of interdisciplinarity. Our review has synthesized 25 definitions found in the literature; this synthesis seems consistent with most of the meaning attributed to interdisciplinarity in the quantitative literature on this phenomenon.

Third, whether this definition is relied on, or another one, it is imperative to be aware of and communicate the assumptions that this choice implies. For example, by emphasizing
variety over other dimensions of diversity will imply tie scores on the measure more strongly to a particular classification of disciplines.

Arguably, by relying on other definitions, or by proposing a different orientation in focus in research on IDR—(e.g., Marres & Rijcke, 2020), we will continue to see disagreement across measures, which may prevent us from drawing consistent conclusions about IDR.

This study provides a handle for researchers wishing more conceptual guidance in the concept and measurement of IDR through the synthesized definition, the identification of evaluation criteria, and the application of these to measures of IDR. While these criteria are not exhaustive in the selection of measures to employ—one can think of practical considerations like the ease with which meta-data is obtained—they do cover the conceptual domain of the construct and are able to point to important differences in existing measures.

This study does not advocate for any one measure in particular or show that any of the measures are invalid in themselves. It shows that many do not meet most criteria that are based on a synthesis of the literature. This synthesis is not a consensus on the definition of IDR per se: one can argue for different ‘flavours’ of IDR, and place more or less importance on its dimensions of integration of knowledge, and the dimensions of diversity of disciplines. However, given the empirical disagreement, a continued emphasis on idiosyncratic definitions, conceptually or operationally, will deter the development of a common body of understanding on IDR in which results, and conclusions are intercompatible.

We hope that our synthesis and evaluation help in fostering more consistency in measurement and thus coherence in our body of understanding, such that practitioners and academics alike can unlock more of the great potential of interdisciplinary research.

Acknowledgements This paper is a substantially extended version of Zwanenburg et al. (2021). We thank Nancy Wang for an initial literature review in this area, and Adeel Akmal for his help in evaluating definitions of interdisciplinarity.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. No funding was received to assist with the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Al-Suqri, M. N., & AlKindi, S. S. (2017). The theoretical foundations and historical development of interdisciplinarity. M. N. Al-Suqri, A. K. Al-Kindi, S. S. AlKindi, & N. E. Saleem (Eds.), Promoting interdisciplinarity in knowledge generation and problem solving. IGI Global, pp. 1–13.
Barry, A., Born, G., & Weszkalnys, G. (2008). Logics of interdisciplinarity. Economy and Society, 37(1), 20–49.
Bordons, M., Zulueta, M., Romero, F., & Barrigón, S. (1999). Measuring interdisciplinary collaboration within a university: The effects of the multidisciplinary research programme. *Scientometrics, 46*(3), 383–398.

Braun, T., & Schubert, A. (2003). A quantitative view on the coming of age of interdisciplinarity in the sciences. *Scientometrics, 58*(1), 183–189.

Bromham, L., Dunnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature, 534*(7609), 684–687.

Carusi, C., & Bianchi, G. (2020). A look at interdisciplinarity using bipartite scholar/journal networks. *Scientometrics, 122*(2), 867–894. https://doi.org/10.1007/s11192-019-03309-3

Chang, Y. W., & Huang, M. H. (2012). A study of the evolution of interdisciplinarity in library and information science: Using three bibliometric methods. *Journal of the American Society for Information Science and Technology, 63*(1), 22–33.

Chen, S., Arsenault, C., Gingras, Y., & Larivière, V. (2015). Exploring the interdisciplinary evolution of a discipline: The case of Biochemistry and Molecular Biology. *Scientometrics, 102*(2), 1307–1323. https://doi.org/10.1007/s11192-014-1457-6

Choi, S., & Richards, K. (2017). Interdisciplinary discourse: Communicating across disciplines. In Interdisciplinary discourse: Communicating across disciplines (Issue 2008). https://doi.org/10.1057/978-1-137-47040-9

Craven, D., Winter, M., Hotzel, K., Gaikwad, J., Eisenhauer, N., Hohnmuth, M., König-Ries, B., & Wirth, C. (2019). Evolution of interdisciplinarity in biodiversity science. *Ecology and Evolution, 9*(12), 6744–6755.

Digital Science. (2016). Interdisciplinary research: Methodologies for identification and assessment. Retrieved October 15, 2021, from https://www.mrc.ac.uk/documents/pdf/assessment-of-interdisciplinary-research/

Glänzel, W., Thijs, B., & Huang, Y. (2021). Improving the precision of subject assignment for disparity measurement in studies of interdisciplinary research. W. Glänzel, S. Heeffer, P. Chi, & R. Rousseau (Eds.), Proceedings of the 18th conference of the International Society for Scientometrics and Informetrics (ISSI 2021), virtual event, 12–15 July 2021, KU Leuven, Belgium (pp. 453–464).

Gowanlock, M., & Gazan, R. (2013). Assessing researcher interdisciplinarity: A case study of the University of Hawaii NASA Astrobiology Institute. *Scientometrics, 94*(1), 133–161.

Hernández, J. M., & Dorta-González, P. (2020). Interdisciplinarity metric based on the co-citation network. *Mathematics, 8*(4), 544.

Huang, Y., Glänzel, W., Thijs, B., Porter, A. L., & Zhang, L. (2021). The comparison of various similarity measurement approaches on interdisciplinary indicators. *FEB Research Report MSI_2102, 1–24.

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy, 39*(1), 79–88.

Jost, L. (2009). Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008). *Ecological Economics, 68*(4), 925–928.

Karlovéè, M., & Mladenèè, D. (2015). Interdisciplinarity of scientific fields and its evolution based on graph of project collaboration and co-authoring. *Scientometrics, 102*(1), 433–454.

Karlvqvist, A. (1999). Going beyond disciplines: The meanings of interdisciplinarity. *Policy Sciences, 32*(4), 379–383.

Klein, J. T. (1990). *Interdisciplinarity: History, theory, and practice*. Wayne State University Press.

Leahy, E., Beckman, C. M., & Stanko, T. L. (2017). Prominent but less productive: The impact of interdisciplinarity on scientists’ research. *Administrative Science Quarterly, 62*(1), 105–139.

Leydesdorff, L., Wagner, C., & Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics, 13*(1).

Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology, 58*(9), 1303–1319.

Leydesdorff, L., de Moya-Anegón, F., & Guerrero-Bote, V. P. (2015). Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of Scopus data (1996–2012). *Journal of the Association for Information Science and Technology, 66*(5), 1001–1016.

Leydesdorff, L., & Ivanova, I. (2021). The measurement of “interdisciplinarity” and “synergy” in scientific and extra-scientific collaborations. *Journal of the Association for Information Science and Technology, 72*(4), 387–402.

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics, 5*(1), 87–100.
MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly, 35*(2), 293–334.

MacLeod, M., & Nagatsu, M. (2016). Model coupling in resource economics: Conditions for effective interdisciplinary collaboration. *Philosophy of Science, 83*(3), 412–433.

Mao, J., Liang, Z., Cao, Y., & Li, G. (2020). Quantifying cross-disciplinary knowledge flow from the perspective of content: Introducing an approach based on knowledge memes. *Journal of Informetrics, 14*(4), 101092.

Marres, N., & de Rijcke, S. (2020). From indicators to indicating interdisciplinarity: A participatory mapping methodology for research communities in-the-making. *Quantitative Science Studies, 1*(3), 1041–1055.

Morillo, F., Bordons, M., & Gómez, I. (2001). An approach to interdisciplinarity through bibliometric indicators. *Scientometrics, 51*(1), 203–222. https://doi.org/10.1023/A:1010529114941

Morillo, F., Bordons, M., & Gómez, I. (2003). Interdisciplinary in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology, 54*(13), 1237–1249. https://doi.org/10.1002/asi.10326

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.

Mutz, R. (2021). Diversity and interdisciplinarity—should variety, balance and disparity be combined as a product or better as a sum? A probability-theoretical approach. W. Glänzel, S. Heeffer, P. Chi, & R. Rousseau (Eds.), *Proceedings of the 18th conference of the International Society for Scientometrics and Informetrics (ISSI 2021), virtual event*, 12–15 July 2021, KU Leuven, Belgium (pp. 825–836).

Okamura, K. (2019). Interdisciplinarity revisited: Evidence for research impact and dynamism. *Palgrave Communications, 5*(1), 1–9. https://doi.org/10.1057/s41599-019-0352-4

Porter, A. L., Cohen, A. S., D. Roessner, J., & Perreault, M. (2007). Measuring researcher interdisciplinarity. In *Scientometrics* (Vol. 72, Issue 1). https://doi.org/10.1007/s11192-007-1700-5

Porter, A. L., & Chubin, D. E. (1985). An indicator of cross-disciplinary research. *Scientometrics, 8*(3–4), 161–176. https://doi.org/10.1007/BF02016934

Porter, A., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics, 81*(3), 719–745.

Pratt, A. D. (1977). A measure of class concentration in bibliometrics. *Journal of the American Society for Information Science, 28*(5), 285–292.

Qin, J., Lancaster, F. W., & Allen, B. (1997). Types and levels of collaboration in interdisciplinary research in the sciences. *Journal of the American Society for Information Science, 48*(10), 893–916.

Rafols, I., Leydesdorff, L., O’Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy, 41*(7), 1262–1282.

Rafols, I., & Meyer, M. (2009). Diversity and network coherence as indicators of interdisciplinarity: Case studies in biomanoscience. *Scientometrics, 82*(2), 263–287.

Rao, C. R. (1982). Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhy: The Indian Journal of Statistics, Series A, 44*(1), 1–22.

Rhoten, D., & Pfirman, S. (2007). Women in interdisciplinary science: Exploring preferences and consequences. *Research Policy, 36*(1), 56–75.

Rinia, E., van Leeuwen, T., & van Raan, A. (2002). Impact measures of interdisciplinary research in the sciences: A forestry case study. *Journal of Informetrics, 28*(5), 285–292.

Rousseau, R. (2019). On the Leydesdorff-Wagner-Bornmann proposal for diversity measurement. *Journal of Informetrics, 13*(3), 906–907. https://doi.org/10.1016/j.joi.2019.03.015

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics, 59*(3), 425–465.

Soós, S., & Kampis, G. (2012). Beyond the basemap of science: Mapping multiple structures in research portfolios: Evidence from Hungary. *Scientometrics, 93*(3), 869–891.

Steele, T. W., & Stier, J. C. (2000). The impact of interdisciplinary research in the environmental sciences: A forestry case study. *Journal of the American Society for Information Science, 51*(5), 476–484.

Stirling, A. (1998). On the economics and analysis of diversity. Science Policy Research Unit (SPRU), Electronic Working Papers Series, Paper, 28, 1–156.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface, 4*(15), 707–719.
Thijs, B., Huang, Y., & Glänzel, W. (2021). Comparing different implementations of similarity for disparity measures in studies on interdisciplinarity. W. Glänzel, S. Heeffer, P. Chi, & R. Rousseau (Eds.), Proceedings of the 18th conference of the International Society for Scientometrics and Informetrics (ISSI 2021), virtual event, 12–15 July 2021, KU Leuven, Belgium (pp. 1097–1107).

Thijs, B., Zhang, L., & Glänzel, W. (2015). Bibliographic coupling and hierarchical clustering for the validation and improvement of subject-classification schemes. Scientometrics, 105(3), 1453–1467.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I., & Börner, K. (2011). Approaches to understanding and measuring interdisciplinlary scientific research (IDR): A review of the literature. Journal of Informetrics, 5(1), 14–26. https://doi.org/10.1016/j.joi.2010.06.004

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. PloS One, 10(5).

Wang, Q., & Schneider, J. W. (2020). Consistency and validity of interdisciplinarity measures. Quantitative Science Studies, 1(1), 239–263. https://doi.org/10.1162/qss_a_00011

Wang, X., Wang, Z., Huang, Y., Chen, Y., Zhang, Y., Ren, H., Li, R., & Pang, J. (2017). Measuring interdisciplinarity of a research system: Detecting distinction between publication categories and citation categories. Scientometrics, 111(3), 2023–2039. https://doi.org/10.1007/s11192-017-2348-4

Xu, J., Min, C., Huang, W., and Bu, Y. (2021). Interdisciplinarity versus Unidisciplinarity: A Structural Comparison of Multi-generation Citations and References. W. Glänzel, S. Heeffer, P. Chi, & R. Rousseau (Eds.), Proceedings of the 18th conference of the International Society for Scientometrics and Informetrics (ISSI 2021), virtual event, 12–15 July 2021, KU Leuven, Belgium (pp. 1259–1265).

Zhang, L., Rousseau, R., & Glänzel, W. (2015). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. Journal of the Association for Information Science and Technology, 67(5), 1257–1265.

Zhang, W., Shi, S., Huang, X., Zhang, S., Yao, P., & Qiu, Y. (2020). The distinctiveness of author interdisciplinarity: A long-neglected issue in research on interdisciplinarity. Journal of Information Science. https://doi.org/10.1177/0165551520939499

Zhou, H., Guns, R., & Engels, T. C. (2021). The evolution of interdisciplinarity in five social sciences and humanities disciplines: relations to impact and disruptiveness. W. Glänzel, S. Heeffer, P. Chi, & R. Rousseau (Eds.), Proceedings of the 18th conference of the International Society for Scientometrics and Informetrics (ISSI 2021), virtual event, 12–15 July 2021, KU Leuven, Belgium (pp. 1381–1392).

Zhou, Q., Rousseau, R., Yang, L., Yue, T., & Yang, G. (2012). A general framework for describing diversity within systems and similarity between systems with applications in informetrics. Scientometrics, 93(3), 787–812. https://doi.org/10.1007/s11192-012-0767-9

Zwanenburg, S. P., & Qureshi, I. (2019). Anticipating, avoiding, and alleviating measurement error: A synthesis of the literature with practical recommendations. Australasian Journal of Practical Information Systems, 23.

Zwanenburg, S., Nakhoda, M., & Whigham, P. (2021). A conceptual evaluation of measures of interdisciplinarity. In W. Glänzel, S. Heeffer, P.-S. Chi & R. Rousseau (Eds.), Proceedings of the 18th International Conference on Scientometrics & Informetrics (ISSI). (pp. 1423–1431). International Society for Scientometrics and Informetrics (I.S.S.I.).