




How qualitative criteria can improve the assessment process of interdisciplinary research proposals

Anne-Floor Schölvinck ¹, Duygu Uygun-Tunç^{2,3}, Daniël Lakens ⁴, Krist Vaesen²,
Laurens K. Hessels ^{1,5,*}

¹Rathenau Instituut, KNAW, The Hague 2593 HW, The Netherlands

²Philosophy & Ethics Group, Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands

³Department of Philosophy, University of Chicago, Chicago 60615, Illinois, USA

⁴Human-Technology interaction group, Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands

⁵Centre for Science and Technology Studies (CWTS), Leiden University, Leiden 2333 BN, The Netherlands

*Corresponding author. Rathenau Instituut, KNAW, The Hague 2593 HW, The Netherlands and Centre for Science and Technology Studies (CWTS), Leiden University, Leiden 2333 BN, The Netherlands. E-mail: l.hessels@rathenau.nl

Abstract

Despite the increasing recognition for the scientific and societal potential of interdisciplinary research, selection committees struggle with the evaluation of interdisciplinary proposals. Interdisciplinary proposals include a wider range of theories and methods, involve a more diverse team, pose a higher level of uncertainty, and their evaluation requires expertise from multiple disciplines. In this study, we investigate the possibility to support the evaluation of interdisciplinary research proposals with measures of interdisciplinary research quality. Based on the literature, we curated a set of qualitative criteria and bibliometric indicators. Subsequently, we examined their feasibility using interviews with interdisciplinary researchers and a re-assessment session of a grant-allocation procedure. In the re-assessment session members of an original evaluation panel assessed four original research proposals again, but now supported with our measures. This study confirmed the potential of qualitative criteria to assess the interdisciplinarity or research proposals. These indicators helped to make explicit what different people mean with interdisciplinary research, which improved the quality of the discussions and decision-making. The utility of bibliometric indicators turned out to be limited, due to technical limitations and concerns about unintended side effects.

Keywords: interdisciplinarity; peer review; research funding; indicators.

1. Introduction

In times in which public science, technology and innovation are increasingly asked to address complex societal challenges, the potential of interdisciplinary research collaborations is recognized more than ever (OECD 2020, AWTI 2022). Many scholars realize that the societal challenges we are facing span disciplinary boundaries (Leydesdorff and Ivanova 2021). Not surprisingly, scholars have demonstrated a positive association between interdisciplinary research and the societal visibility of scientific results (D'Este and Robinson-García 2023).

Despite these claims about the potential of interdisciplinarity, there is a longstanding debate about both the conceptual definition of interdisciplinary research (IDR), and its empirical operationalization (Huutoniemi et al. 2010). For pragmatic reasons, in this paper we follow the broad definition of IDR provided by McLeish and Strang (2016) and the National Academies of Sciences, Engineering, and Medicine (2005: 26):

[...] a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are

beyond the scope of a single discipline or field of research practice.

To realize the full societal and scientific potential of IDR, consistent and fair assessment of IDR proposals by research funders is essential (Lyll et al. 2013). However, the evaluation of IDR proposals is a greater challenge than the evaluation of monodisciplinary proposals (McLeish and Strang 2016). IDR proposals are generally more complex (Klein 2008) and their evaluation requires input from multiple experts from different disciplines, each with their own implicit epistemic standards of research quality and diverging concepts and expectations of interdisciplinarity (Huutoniemi 2016). Hence, IDR proposals, especially from early career researchers, are less likely to be funded (Lamont, Mallard and Guetzkow 2006; Sun et al. 2021).

A potential solution to counteract these problems is to support the assessment process with a set of measures¹ to evaluate different aspects of interdisciplinarity. Supporting information can help to clarify definitions and achieve more uniformity in the evaluation. Moreover, it can also help to reward interdisciplinary aspects in the proposal more explicitly, in programs where this is appropriate (Lyll and King 2013; Belcher et al. 2015; Strang and McLeish 2015). For applicants, an established set of measures of interdisciplinarity, in addition to established assessment measures such as scientific

excellence that tend to be more monodisciplinary-focused, will enable them to highlight these interdisciplinary aspects of their planned research in the research proposal.

The aim of this study is to explore the potential of qualitative and bibliometric measures to support the evaluation of interdisciplinary research proposals, by developing a set of relevant measures and testing their feasibility in practice. We first curated a set of relevant measures, building on available literature on measuring and evaluating interdisciplinarity in science. Second, to explore the validity of bibliometric indicators of interdisciplinarity further, we have investigated their correlation with the interdisciplinarity of the publications produced by interdisciplinary projects that received funding from the Dutch science funder NWO. Third, despite a long-standing academic tradition analyzing grant review processes of IDR, few studies have adopted a more hands-on approach in which the practicality of a new set of assessment measures is tested in a real-life setting (Huutoniemi *et al.* 2010; Belcher *et al.* 2015). To contribute to this lack of knowledge, we explored the implementation of these measures in practice, by investigating the feasibility of the measure using the case of a Dutch funding program with individual grants without thematic restrictions. In this study, we define the feasibility of the measure as the degree in which it can be implemented in real-life assessment procedures in the future. In this study, the goal was never to express a value judgment suggesting that more interdisciplinarity is always better: the desired degree of interdisciplinarity depends, among other things, on the proposed research question. We consider a high degree of interdisciplinarity not as an end in itself, but as an important discussion point during assessment meetings.

Our study focused on a funding instrument providing individual, bottom-up, research funding. These are calls with no thematic demarcation in which individual researchers (i.e. no consortia) submit proposals. In these research funding instruments, the necessity and feasibility to perform an interdisciplinary research project successfully has to be made plausible by a single applicant rather than a team of experts with varying backgrounds. Still, the grant aims to allow researchers to hire and collaborate with a team of fellow colleagues. This makes it an interesting case as it further complicates the assessment of interdisciplinary research proposals.

2. Assessing the interdisciplinarity of research proposals

Over the past few decades, the relative importance of project-based research funding has increased in relation to block grant funding in Europe, including the Netherlands (respectively Wang, Lee and Walsh 2018 and Rathenau Instituut 2023). In many countries, external grants or research contracts have become a key source of funding for academic researchers aspiring to develop a research line or to build up a research group, hence affecting researchers' careers to a great extent (Ebadi and Schiffauerova 2015). We focus in particular on competitive grants, typically allocated based on a process of peer review. Selection committees evaluate a number of proposals using a fixed set of measures, often assisted by assessment reports from academic peers on the individual proposals. Competition for grants is often fierce and success rates are low, leading to widespread dissatisfaction with the peer-review selection system in academia (Philipps 2022).

Aiming to reduce variability across reviewers to avoid mistrust among grant applicants, national research councils that provide competitive grants place a strong emphasis on procedural fairness and transparency of peer review processes (Abdoul *et al.* 2012). Yet, following the ethnographic work of amongst others Lamont (2009) and Roumbanis (2022), we regard the work of selection committees as a social practice, too. This paper builds on the holistic perspective of Oxley (*under review*), who regards the output of selection process of grant proposals as the product of individual panelist characteristics, organizational characteristics and how these interact and are moderated by the group interaction taking place in selection committee meetings.

One may argue that panelist, panel and call characteristics as well as group interactions taking place in panel meetings affect *all* peer reviewed proposal selection processes. Yet, there are indications that these dynamics are exacerbated when reviewing IDR grant proposals. First, IDR proposals are generally more complex in nature, as they often include a broader range of aims and a mixed methods research design (Klein 2008). Additionally, to perform the IDR successfully, the project team needs to have an adequate governance structure and the right personalities to create effective partnerships (*ibid*). Because of the integrative ambitions and the involvement of people with varying backgrounds, IDR often deals with greater levels of uncertainty, which require more extensive planning for contingencies and risk-mitigation. Assessing these aspects fairly requires assessors to keep an eye on the balance between the risks of IDR and its opportunities, whilst research has shown assessors to be risk averse (Lamont, Mallard and Guetzkow 2006) and they tend to favor well-established research disciplines rather than interdisciplinary research of a more experimental nature (Langfeldt 2001).

Second, the assessment of IDR requires experts from different disciplines, each with their own implicit epistemic standards of research quality (Huutoniemi 2016). Panelists tend to favor work that is similar to their own (cognitive particularism), causing a prejudice against IDR proposals (Porter and Rossini 1985; Travis and Collins 1991; Sandström and Hällsten 2008; Li 2017; Guthrie, Ghiga and Wooding 2018). A diverse group of panelists has been shown to be beneficial to the evaluation of IDR proposals (Abma-Schouten *et al.* 2023) and there are also indications that the use of diverse reviewers could reduce bias associated with cognitive particularism (Teplitskiy *et al.* 2018). Yet, ensuring appropriate diversity by complementing panels with *ad hoc* external assessments is ineffective, as they appear to have little demonstrable effect on the result of panel discussions (Thorngate, Faregh and Young 2002; Mayo *et al.* 2006; van Arensbergen, van der Weijden and van den Besselaar 2014).

This bias against IDR calls for improvements to the peer-review selection system that intervene at multiple levels. Improvements should not just be procedural in nature, but also tested for their implementation in practice. As studies suggest that dedicated review measures for assessing interdisciplinary proposals may have a moderating effect on the bias against IDR grant proposals, numerous evaluation frameworks have been suggested to capture and assess the unique contributions of an interdisciplinary approach (Lyall and King 2013; Belcher *et al.* 2015; Strang and McLeish 2015). We did not just design such an instrument, but also examined its feasibility together with exactly those people who are

meant to use it, namely the members of review panels of IDR grant proposals.

In the following section, we delve into the literature about quality measures, distinguishing between bibliometric indicators and qualitative criteria of interdisciplinarity. Both types of indicators are characteristics of a research proposal that can help to predict the type or degree of interdisciplinarity of the research proposed. With qualitative indicators we refer to non-numerical characteristics which require a subjective judgement from panelists and/or the panel as a group. With bibliometric indicators we refer to factual quantitative data about the scientific publications of the applicants or the literature cited in the proposal.

2.1 Qualitative criteria of interdisciplinarity

In qualitative criteria, reviewers or applicants are asked to indicate whether, or the extent to which, certain criteria are met in a proposal. We build on the work by Laursen, Motzer and Anderson (2022), who categorized 40 criteria in literature into three key aspects of interdisciplinarity that can be used in the assessment of IDR proposals, namely: (1) interdisciplinary rigor, which we compress to integration, (2) feasibility, and (3) impact, which we interpret as scientific impact. The *integration* of knowledge can be considered a defining feature of interdisciplinarity as opposed to multidisciplinary research, and many scholars therefore list this as a criterion that should be explicitly assessed (Bruun et al. 2005; Klein 2008; Lyall and King 2013; Strang and McLeish 2015). This criterion consists of multiple underlying dimensions, of which diversity (e.g. the breadth, variety, and balance of contributions) and interaction (e.g. collaboration, communication, and stakeholder engagement) are the two main subcomponents (Lyall and King 2013; Strang and McLeish 2015).²

The *feasibility* of a proposal is, amongst others, dependent on the applicants' leadership and team readiness (Lyall and King 2013; Strang and McLeish 2015; Tate, Decker and Just 2021). Other components regarding the feasibility of an interdisciplinary proposal include fundability, goals, and institutional support (Laursen, Motzer and Anderson 2022). Interdisciplinary research projects generally require researchers with diverse, compatible fields of expertise to collaborate. Moreover, these projects are typically more dynamic and unpredictable than mono-disciplinary research projects. This requires teams to have reached common ground, to be open and hospitable to each other's expertise, and to be able to develop research projects iteratively (Klein 2008; McLeish and Strang 2016).

Third, the potential *scientific impact* of a research proposal is defined more broadly in interdisciplinary research proposals than in monodisciplinary proposals. This criterion includes the transfer of knowledge across disciplines, the transformation of disciplinary standards, and the effectiveness of integrating work across disciplinary boundaries (Bruun et al. 2005; Klein 2008; Lyall and King 2013; Strang and McLeish 2015).

2.2 Bibliometric indicators of interdisciplinarity

In addition to the qualitative assessment of a research proposal's interdisciplinary merits, the degree of interdisciplinarity of a research proposal can be assessed quantitatively ex ante using bibliometric elements such as articles published by the applicant and the list of references associated with each proposal. Ex post, the scientific output of a funded research

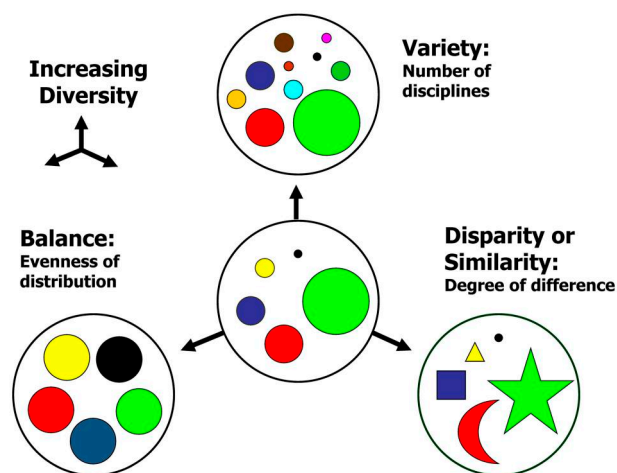


Figure 1. The three attributes of diversity according to Stirling (1998, 2007): variety, balance and disparity. Representation by Rafols and Meyer (2010) by permission of I. Rafols.

project and the citations to these articles can provide an indication of the degree of interdisciplinarity of the research project.

In bibliometric research, interdisciplinarity is often understood and evaluated in terms of three components of scientific diversity: variety, balance and disparity (Stirling 1998, 2007). These components can be operationalized as the number of distinct scientific disciplines cited (*variety*), the evenness of the distribution of citations among disciplines (*balance*), and how dissimilar these disciplinary categories are (*disparity*). Figure 1 provides a graphic representation of these attributes of diversity.

The different dimensions of diversity can be visualized using science overlay maps, which provide the user with 'a perspective of the disciplinary diversity of any given output, yet without the need to rely on combined or composite indices' (Rafols, Porter and Leydesdorff 2010). Science overlay maps allow to evaluate different components of diversity independently. Such maps offer a dynamic and bottom-up representation of the cognitive interrelationships between (sub) disciplines. Based on an overlay map, reviewers can assess proposals for their interdisciplinarity along the dimensions of variety, balance and disparity.

Figure 2 serves as an example to illustrate how a science overlay map can help evaluate the interdisciplinarity of an applicant. This example uses a base map of scientific journals that is spatially organized with respect to cognitive proximity, and uses color as an index of scientific disciplines. Firstly, the degree of variety can be inferred from the variation in the color of the dots—the more different colors, the more disciplines the author has published in. Second, the degree of disciplinary balance can be assessed by looking at how densely the superimposed journal data concentrates on certain locations vis-a-vis others. Higher balance is achieved by a more even distribution of publications to scientific disciplines. For instance, a researcher who published mainly in one discipline, and once in another discipline, scores much lower on balance than a researcher who publishes equally across different disciplines. Third, maps can convey the disparity (i.e. the cognitive distances) among disciplines by placing these units closer or more distant on the map. The further apart or widely

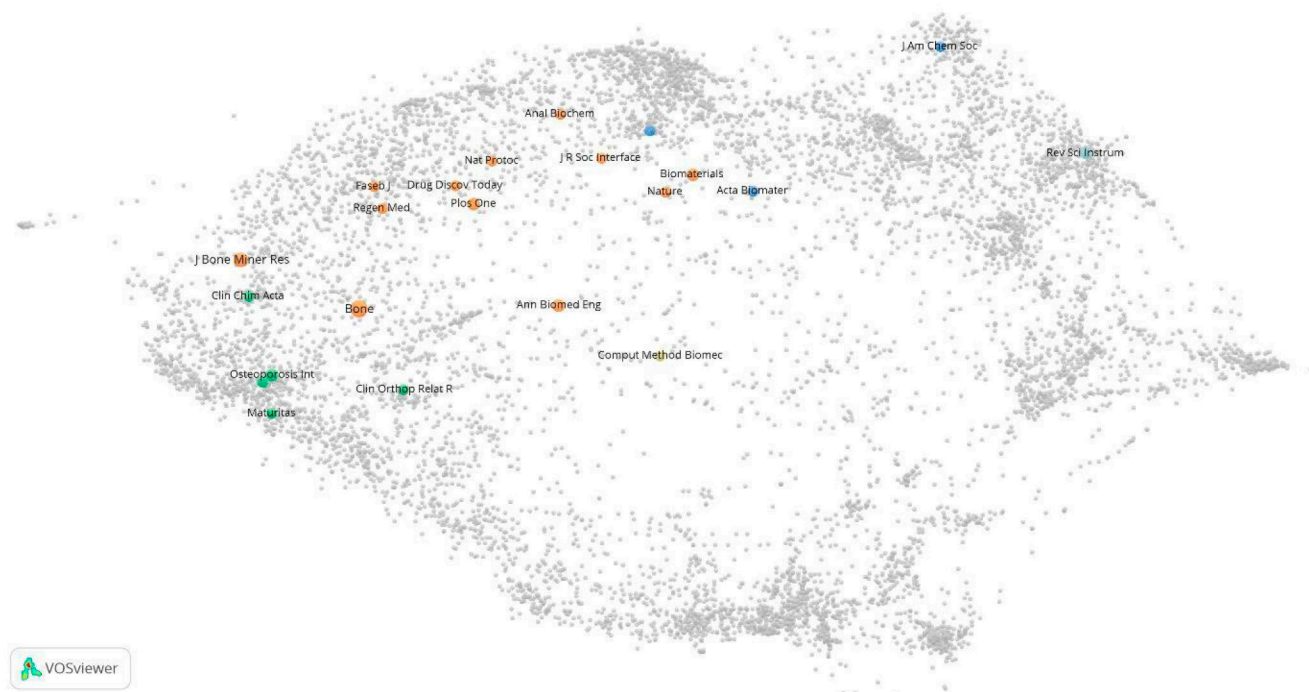


Figure 2. Exemplary science overlay map of one of the reviewed grant applications. The colors indicate disciplines the author has published in (variety); clusters of dots indicate number of articles published in similar journals (balance); disparity is visualized by the distance between the dots on the map.

scattered the journals in which a researcher publishes, the higher a researcher would score on disparity.

In addition to this graphical presentation, the components of interdisciplinarity can also be expressed in a numerical form. The Rao-Stirling index of diversity is a widely used composite measure integrating the three diversity components: variety, balance, and disparity (Leydesdorff 2018). The fields of publication need to be defined, for example by selecting the names of scientific journals, in which applicants have published or in which the cited articles have been published, as the target element. The subject category to which these journals are assigned by the Web of Science classification system can then be used to calculate the Rao-Stirling index of the applicant's list of publications or the proposal, respectively.

3. Methodology

In order to address the aim of this study, we embedded quantitative measures in a qualitative analysis. The main steps of our study are introduced in the following sections. First, we curated a set of qualitative assessment criteria and bibliometric metrics evaluating different aspects of interdisciplinarity, using literature review, bibliometric validation, and interviews with past awardees (see Sections 3.2 and 3.3 for details). Second, we examined the feasibility of these metrics in a reassessment session as described in Section 3.4.

3.1 Retrospective approach

Evaluating the assessment of research proposals on their interdisciplinarity and linking this assessment to the research process that follows once the grant has been funded can be performed either longitudinally or retrospectively. In a longitudinal study, one would have to observe the proposal evaluation and selection process, wait for the funded research

projects to start, analyze the research process of the granted projects in real time, and wait for the outputs of the project to appear in scientific journals. In this study however, we have chosen a retrospective approach in collaboration with the Dutch research council NWO. We focused on successful (funded) interdisciplinary research proposals and their subsequent projects in the NWO Talent Programme for researchers with several years of research experience (interdisciplinary Vidi and Vici grants, funded in 2016 to 2018). **Textbox 1** provides background information regarding these grants. This retrospective methodology had the pragmatic advantage that it greatly reduced the throughput time of this study, as no waiting time existed between the assessment procedure, the start of the research project once the grant had been awarded, and its execution and output phase. In addition, the stakes are lower in a reassessment session when the proposals have already been evaluated positively (see Section 3.4). This allows panelists to focus and advise freely on the feasibility of the assessment measures, rather than on the judgements themselves.

3.2 Qualitative assessment criteria

To qualitatively assess the three key main criteria for successful IDR as set out in Section 2.1, we iteratively curated a set of questions to be provided to the assessors of interdisciplinary research projects as described in Section 3.4. First, we built upon Laursen, Motzer and Anderson (2022), selecting items from literature related to the three main criteria (1) integration, (2) feasibility and (3) impact, while preventing redundancy. Items that were phrased differently in literature but aimed at assessing the same element were grouped and (wherever possible) reformulated into a single evaluative question. Second, we invited all laureates of the interdisciplinary NWO Vidi and Vici grants in the period 2016–18 to

Textbox 1 Information regarding the NWO Talent programme.

The NWO Talent programme offers individual grants to talented, creative researchers. This enables them to conduct the research of their choice. The Talent programme has three funding instruments tailored to various phases in researchers' scientific careers.

- Veni [*not included in this study*]
- Vidi
 - maximum funding €800 000 lasting 5 years
 - aimed at experienced researchers who have conducted 3 to 8 years of research after obtaining their PhD degree (excluding extension clauses)
- Vici
 - maximum funding €1 500 000 lasting 5 years
 - aimed at senior researchers who have obtained their PhD degree max. 15 years ago (excluding extension clauses), and who have already demonstrated the ability to develop their own line of research

Applicants submit their proposals to one of the four NWO disciplinary domains (Science, Social Science and Humanities, Applied and Engineering Sciences, or Health Care Research). Up until 2018, there was a fifth option for applicants to submit their proposal to an interdisciplinary office. The selection processes and criteria were identical for all proposals submitted to the various offices. Hence, at the IDR office, the interdisciplinarity of a proposal was not a quality criterion: proposals awarded at this office usually had an interdisciplinary character, but it was not what they were assessed on.

This study focuses on the successful (granted) proposals that were submitted to this interdisciplinary office between 2016 and 2018. The success rate for interdisciplinary Vidi-proposals was 20.4% on average, for interdisciplinary Vici-proposals the average success rate was 12.9%. In total, 24 interdisciplinary Vidi and Vici proposals were funded between 2016 and 2018.

(Information obtained from NWO website, accessed on 26 April 2023: <https://www.nwo.nl/en/researchprogrammes/nwo-talent-programme>).

Table 1. Characteristics of interviewees and of all laureates eligible for an interview.

Characteristics		No. of interviewees (total = 9)	No. of laureates (total = 24)
Type of grant	Vidi	8	21
	Vici	1	3
Grant awarded in	2016	2	7
	2017	3	9
	2018	4 ^a	8

^a The proposals from these four laureates were selected for re-evaluation regarding their interdisciplinarity (see Section 3.4).

an online semi-structured interview (see [Supplementary Appendix A](#) for the interview protocol). We asked interviewees to reflect on the development of their research proposal, evaluate opportunities and challenges in the research process, and take stock of the scientific and societal outputs thus far, specifically in relation to the interdisciplinary nature of their research project. We based our interview questions on the qualitative interdisciplinarity assessment criteria found in the literature. After this list was discussed, we asked the interviewees about other any success factors or barriers regarding the interdisciplinary nature of their research project that were not yet discussed. We discussed the extent to which the laureates felt these success factors and barriers could be predicted *ex ante*, in their original research proposal. These interviews refined the set of qualitative assessment criteria.

The interviews lasted approximately 1 h and were conducted in Dutch or English. From the 24 laureates, nine responded to our invitation (response rate of 37.5%, see [Table 1](#) for details of interviewees compared to all laureates eligible for an interview). All interviews were performed by two researchers (author AS and colleagues AH or MD). The interviews were audio-recorded, based on which we wrote a report that we sent to the interviewees to check for accuracy within two weeks after the interview. Author AS analyzed these reports using a thematic coding scheme based on the three main criteria in the literature,

building upon [Laursen, Motzer and Anderson \(2022\)](#): integration (diversity/interaction), feasibility and scientific impact. Data analysis of this part of the project was discussed extensively amongst authors AS and LH.

3.3 Bibliometric interdisciplinarity measures

For all 24 successful interdisciplinary Vidi and Vici proposals funded between 2016 and 2018, bibliometric data were collected regarding the list of articles authored by the applicant at the time of application, citations to the applicant's articles, the references in the research proposal, the output of the funded research project to date, and citations to the output to date. Most of the data were collected manually via Web of Science (WoS); proposal references were provided by NWO.

Using the free software provided by Leydesdorff in association with [Leydesdorff and Rafols \(2012\)](#), Rao-Stirling indices on the basis of WoS analysis results organized by publication titles were calculated regarding the;

- 1) articles authored by the applicant at the time of application (Rao_PI),
- 2) citations to the applicant's authored articles (Rao_PI_cit),
- 3) references in the research proposal (Rao_Prop),
- 4) output of the funded research project to date (Rao_Out), and
- 5) citations to the output to date (Rao_Out_Cit)

Subsequently, we calculated the correlations between the Rao-Stirling indices of diversity among the five sets of bibliometric data, to investigate how the interdisciplinarity of a proposal or applicant's list of articles correlated to the interdisciplinarity of the project's output.

Last, we selected four proposals from the successful Vidi proposals submitted to the NWO interdisciplinary review committee in 2018. This selection was based on whether the proposal would be re-evaluated regarding its interdisciplinary nature (see Section 3.4). For these four proposals we created science overlay maps of the articles published by the

applicant at the time of application and the references in the research proposal that the evaluation committee could use in their re-evaluation.

3.4 Reassessment session

We invited the original NWO review committee from the 2018 VIDI allocation procedure to re-evaluate the interdisciplinary nature of four proposals using a new set of qualitative assessment criteria (as developed in Section 3.2) and the bibliometric interdisciplinarity information regarding the proposals (as calculated in Section 3.3). These four proposals were selected based on the availability of the laureate for an interview (as described in Section 3.2), and the fact that the grants were all awarded in the same year, thus reviewed by the same members in the review committee.

Out of 12 members of the review committee, seven accepted this invitation. The chair of the original NWO review committee was present and took on her role as (technical) chair of the meeting, similar to the original review committee's meetings. A NWO programme officer was present to assist the chair in this role. All participants were reimbursed for their time following the standard NWO reimbursement policy.

Similar to the original procedure, in preparation of the session the participants were asked to review either one proposal as a first reader, or two proposals as a second reader. The set of qualitative assessment criteria was provided to the committee members through email prior to the meeting. In addition, for each proposal under scrutiny we provided the:

- science overlay map of the articles published by the applicant at the time of application,
- science overlay map of the references in the proposal, and
- Rao-Stirling index of the references in the proposal (Rao_prop score).

We provided a short explanatory note on the bibliometrical background of science overlay maps and the Rao-Stirling index.

The online session lasted 1.5 h. Author AS welcomed the participants, briefly introduced the purpose of the session and provided some essential information on the qualitative assessment criteria and the bibliometric metrics provided for each proposal. Following this introduction, the technical chair of the meeting took over. In the second part of the meeting the members of the review committee were asked to verbalize all their questions, doubts, and considerations, as inspired by a thinking aloud protocol (Ericsson and Simon 1993; Alhadreti and Mayhew 2017). The members of the review committee discussed the interdisciplinary nature of the four proposals, taking ~12 min per proposal. In the final section of the meeting, author AS invited the committee members to reflect more generally on the qualitative assessment criteria and the bibliometric indicators. Discussion topics were the usability of the bibliometric indicators, and the applicability, comprehensibility and weighing of the qualitative assessment criteria. Moreover, the committee members were invited to discuss the possible side effects of assessing the interdisciplinarity of research proposals in a bibliometric and qualitative way.

Three researchers involved in this study were present at the session—author AS facilitating and introducing the session, and two colleagues (author DL and colleague JZ) observing the interactions. Based on these observations, we wrote a report that we sent to the participants in the session for member check three weeks after the event took place. Quotes

from interviews and the re-assessment session have been translated to English by the authors.

3.5 Ethical considerations

All interviewees invited for this study took part in the research voluntarily and we obtained their formal (verbal or written) informed consent prior to the interview. Their anonymity was guaranteed, both towards NWO as their funding organization and externally. Prior to the reassessment session all participating members of the original NWO review committee signed a confidentiality agreement stating that the contents of the original research proposals under scrutiny were still to be treated strictly confidentially and that any personal interests, had they arisen in the meantime, should be declared. We informed the applicants of the proposals that were to be re-evaluated about this session by email; none of them raised objections. We used the NWO data in this research in accordance with European GDPR legislation: its use for academic or statistical research to improve NWO's processes is compatible with the original processing purpose.

4. Results

The aim of this study was to curate a set of qualitative criteria of interdisciplinarity and bibliometric metrics regarding the interdisciplinarity of proposals: see Sections 4.1 and 4.2 respectively for the results regarding this aim. We discussed the practical feasibility of both these types of measures in the reassessment session and the interviews with laureates (qualitative criteria only), presented in Section 4.3 of this section.

4.1 Qualitative criteria of interdisciplinarity

A set of qualitative criteria of interdisciplinarity was developed. The three main criteria, building upon Laursen, Motzer and Anderson (2022): Integration (subsections diversity and interaction), Feasibility, and Impact- were used as the main framework. Within these categories, evaluative questions were formulated based on suggestions from literature, while preventing redundancy.

Refinement of these questions took place based on the interviews with interdisciplinary researchers. The interviewees reflected on the criteria, which led to adjustments to the original set of evaluative questions;

- 1) Laureates indicated that departments where interdisciplinary working is common, or valued, contribute to the smooth running of the project. Therefore, the institutional embedding of the prospective IDR project was added as an evaluative criterion (criterion 2.4 in Table 2). Interestingly, this issue has rarely been identified in the literature regarding the assessment of IDR (Porter *et al.* 2006; Nair 2011).
- 2) In contrast, the flexibility of the research plans was eliminated as a criterion, as interviewees indicated that their research projects indeed pursued new unforeseen avenues compared to their proposals, but this was (according to them) not in particular attributable to the interdisciplinary nature of their work.
- 3) Team composition refers to a diverse, balanced background of team members, but especially to their attitude towards colleagues with different kinds of expertise, and towards interdisciplinarity in general (criterion 2.2 in Table 2). Laureates regard an open attitude, flexibility

Table 2. Qualitative criteria of interdisciplinarity (categorized along integration, feasibility, and impact).

1.1 Integration: diversity	
1.1.1	Is the emergent whole of the IDR greater than or different from the sum of its parts? Do the ingredient disciplines do more than work in parallel but interact, communicate, recombine? Is there synthesis of knowledge?
1.1.2	Has a convincing case been made that only interdisciplinary or transdisciplinary research can make the promised contribution to solving the problems indicated/answering the questions addressed? Is there a clear justification for the choice of disciplines based on the needs of the research questions?
1.1.3	Does the proposal draw on literature/material/methods from the various disciplines involved (e.g. via citations)?
1.1.4	How diverse are the disciplines, methods and researchers and how suitable is the combination of disciplines?
1.2 Integration: interaction	
1.2.1	How is the collaboration organized—is there an understanding of the challenges of interdisciplinary integration, including methodological integration, and the ‘human’ side of fostering interactions and communication, and an effective strategy to achieve this?
1.2.2	Are additional resources and time planned for dialogue, co-learning and integration between the contributing disciplines?
2. Feasibility	
2.1	Is the necessary experience with IDR represented by the team and the leadership as well as training and development in place? Is there evidence of interdisciplinary leadership?
2.2	How do you assess the applicant’s collaborative skills (open mindedness, self-reflection, dealing with changing hierarchies, ability to bear and manage tensions)?
2.3	Does the proposal describe how the disciplines involved will be integrated (in the design and conduct of the research as well as in subsequent publications) and how this relates to the type of interdisciplinarity involved; does it demonstrate how the quality of integration will be assured?
2.4	Is the applicant (prospectively) embedded in an ID institutional environment?
3. Scientific impact	
3.1	Do the outputs have the potential to provide transformative feedback into specific disciplinary areas?
3.2	Do the outputs create novelty by the integration of established knowledge within disciplines?

and pleasure in learning about new disciplinary perspectives as important traits of researchers for making an IDR project successful.

Laureate 2: *It’s definitely challenging, but in a good way. [...] For both PhD’s it has been very advantageous to see each other’s work because one is very not philosophical and the other not very technical. And I think that works very well.*

Interviewer AS: *How have you fostered or facilitated these interactions?*

Laureate 2: *Meetings. A lot of meetings with the whole group. People talking about their research and asking questions to each other. And I was also lucky socially, they interacted very well. They were very willing to make an extra step to understand each other. [English original]*

- 4) According to the laureates, pleasant interactions within a research team are very important for the success of an interdisciplinary research project. Also, a lot of time and patience is needed for conceptual and/or methodological translations amongst team-members, in order to interweave their different (disciplinary) frames of thinking (criterion 2.3).

Laureate 3: *It just takes a lot more time. You have to keep talking to each other all the time, about what do we mean when we talk about [example from proposal]. The terms that mean one thing to me as a [one discipline] and something else to a [other discipline]. Yes, you have to keep discussing that, and what is relevant.*

- 5) All laureates believe that interdisciplinary research questions have added value compared to a monodisciplinary approach (criteria 3.1 and 3.2)—which is unsurprising given that they submitted an IDR proposal. Nevertheless,

Table 3. Correlation matrix of Rao-Stirling indices of diversity across five sets of bibliometric data, for all proposals.

	1.	2.	3.	4.	5.
1. Rao_PI	–				
2. Rao_PI_cit	0.89*	–			
3. Rao_Prop	0.52*	0.54*	–		
4. Rao_Out	0.51*	0.42	0.75*	–	
5. Rao_Out_Cit	0.43	0.37	0.55*	0.73*	–

* $P < 0.05$.

for the sake of scientific impact, some laureates indicate that they strategically choose to publish their results in monodisciplinary journals, for example by focusing on distinct subprojects.

Laureate 3: *We are just now finishing a paper. [We are discussing] how much do you put into a manuscript about data from the brain, using a method from physics. To what do you pay attention and what do you emphasize. [...]. It is almost never possible to do that in a balanced way, just because you choose a target journal, which often puts either one or the other at the top so to say.*

Table 2 provides the complete set of qualitative criteria of interdisciplinarity as presented to the review committee for their re-evaluation session.

4.2 Bibliometric analyses

We calculated the Rao-Stirling indices of diversity for five sets of bibliometric data, for all 24 proposals in the dataset. We used this information to calculate how strongly these Rao-Stirling indices correlate. As one can observe in Table 3, there is a relatively strong significant correlation (0.75) between the Rao-Stirling scores based on the research proposal (Rao_Prop) with the output of the funded research project (Rao_out). This suggests that the interdisciplinarity score of the references in the proposal could be a reasonable predictor of the

interdisciplinarity of a project's output. This means that a quantification of the interdisciplinarity of the references in a proposal could be used as a metric to predict how interdisciplinary the output of a proposed research project will be.

For each of the proposals under scrutiny in the reassessment session, we therefore provided the Rao-Stirling index of the references in the proposal (Rao_prop score). Additionally, we provided the science overlay map of the articles published by the applicant (at the time of application), and the science overlay map of the references in the proposal.

4.3 Feasibility of qualitative and bibliometric measures for interdisciplinarity

Several themes regarding the assessment of interdisciplinary proposals were discussed in the reassessment session. The practice of conducting IDR and to what extent its success can be predicted *ex ante* came up in the reassessment session. Moreover, laureates' discussions regarding the scientific and societal outputs of their research projects thus far, and the choices they make in publishing their research results helped the assessment committee reflect on the impact of a proposal in terms of publications (criterion 3.2). Last, the assessment committee reflected extensively on the use of bibliometric metrics and qualitative assessment criteria for interdisciplinarity.

4.3.1 Assessing practice of interdisciplinary work

In an interdisciplinary proposal, cooperation with the right people is essential, according to the assessment committee. An applicant does not have to master everything perfectly himself; it is about the right composition of the team.

Reviewer 1: *In a collaboration, a candidate might not need to have perfect command of everything himself either. You could also take that aspect of cooperation with the right people as a good criterion. This could include cooperation with people outside those directly involved in a project.*

The committee considers that the criteria assessing the practice of conducting IDR (criteria 1.2) should weigh heavily. Yet, according to the assessment committee, these criteria are difficult to assess both on paper and in an interview. For example, if asked about handling conflicts in an application form or as an assessment criterion, applicants are going to write down what they think an assessment committee wants to hear. Alternatively, this aspect could be assessed using a narrative about the collaboration the applicant envisions. The explicit inclusion of knowledge-sharing moments in the proposal design could also contribute to the practice of interdisciplinary working.

4.3.2 Tension between content and strategy

Laureates indicate there is a tension between the choices they make regarding their interdisciplinary research questions and strategically playing the scientific publishing game (see Section 4.1 point 5). The assessment committee recommends thinking about the impact of a proposal on the contributing disciplines, both in terms of content and in terms of publications (criterion 3.2). In theory, a proposal does not even have to be interdisciplinary to have impact on different disciplines, for example because the work is cited by researchers from a wide range of disciplines.

4.3.3 Reflections on use of qualitative assessment criteria

The members of the assessment committee are generally positive about the use of the qualitative assessment criteria to evaluate the interdisciplinarity of a proposal. Most importantly, the use of qualitative criteria without scoring them avoids sham precision and it shows respect for the quality of applicants and proposals, the evaluation committee believes. One question, however, is how these criteria can be put to use without weighing them; the committee recognizes that this does not remove the thorny issue of selection, of funding and rejecting proposals.

Reviewer 3: *We have to make a decision somehow. With qualitative criteria, we show more respect for applications but how can we translate this into who gets funding or not?*

Reviewer 5: *Yes, you describe a kind of tension between being competitive and recognizing those values.*

Furthermore, the members of the assessment committee think the list of qualitative assessment criteria on interdisciplinarity is rather long. They question whether it is feasible to assess them all per proposal and also weigh the criteria against each other. The committee suggests that the list could also be used as a starting point for a substantive discussion. Assessors are then presented with many criteria that they do not all have to score individually but that serve as a guide for an evaluative discussion between committee members. This strategy may seem to undermine the feasibility of qualitative indicators. In the discussion we will offer some suggestions for overcoming this difficulty.

4.3.4 Reflections on use of bibliometric information

In the reassessment session, the bibliometric information that is provided about the proposals confirmed the committee members' intuitive assessment of the interdisciplinarity of the proposals. Applications that the assessment committee rated highly in terms of interdisciplinarity generally also had a higher Rao-Stirling index score and vice versa.

However, the assessment committee had some critical reflections on the use of bibliometric information on the interdisciplinarity of proposals. The criticism concerned: (1) interpreting the information without comparative material; (2) the aspects of interdisciplinarity that do not emerge from the index or science overlay maps; (3) possible side effects of bibliometric information on assessment practice; and (4) on a more technical level, the assessment committee recommends improvements to the calculation of the Rao-Stirling index and the science overlay maps based on more complete data.

First, it is difficult for reviewers to interpret the bibliometric information when they lack any context or comparative material. For example, it is not yet clear to reviewers what a high or low score on the Rao-Stirling index is, and what it actually means.

Reviewer 2: *I do find it difficult to interpret that number of that index, because what does it mean? For example, when it is half as high as the highest score of 0.31. Then what does that say?*

Therefore, the reviewers would find it useful to work with a confidence interval of the index score. The reviewers also raise the question of what it means if the science overlay maps of an applicant and of a proposal are much, or little, alike. This could be an indication that an applicant wants to develop more interdisciplinary or that, on the contrary, he already publishes a lot in different fields. More experience with the index can help assessment committees interpret the score, although a high score is not necessarily better if interdisciplinarity has to be primarily fit for purpose. The assessment committee thinks that a discussion on the interpretation of bibliometric information can serve as a kind of calibration in which the validity of bibliometrics is tested against a qualitative assessment of interdisciplinarity (and vice versa).

Second, the assessment committee indicates that the bibliometric information does not provide information about how the interdisciplinarity of a proposal is constructed. For example, the index and maps cannot indicate whether a proposal is interdisciplinary in conceptual or methodological terms, nor whether the design of a project proposal moves from more to less diversity or vice versa.

Reviewer 3: [...] *With these maps it's hard to see, for example: does the project start from more or less diversity, or does it move the other way around?*

Third, there are possible side effects when using the new bibliometric information. The evaluation committee cautions against drawing hasty conclusions based on the visual image of the science overlay maps. While these images may seem to depict the interdisciplinarity of a proposal or applicant at a glance, they are in fact a proxy for this aspect, as they are based on the extent to which an applicant has used citations from different disciplines for the proposal, or has published in a range of disciplines.

In line with this point about the reliability of the bibliometric data, gaming is also a concern. If interdisciplinarity (formal or informal) were a criterion to score high on, what would applicants do to boost their index? And in what other ways might this criterion influence research practice? The assessment committee believes it is important not to replicate the problems with the use of other indices in assessment practice (such as the journal impact factor and the H-index), to the assessment of interdisciplinarity.

Reviewer 4: *Do we want IDR to become a criterion that people score high on? Because then indeed it gets problematic, like the H-index, if people want to make their proposal as interdisciplinary as possible.*

Lastly, a number of technical improvements to the Rao-Stirling index and science overlay maps are needed, according to the assessment committee. For instance, publications in books and conference proceedings are not yet included in the current beta version of the science overlay maps, even though these are important and common publication types in some disciplines. This limits the value of the current method to devise the maps for the humanities, for example.

5. Discussion

To realize the full societal and scientific potential of IDR, the complexity of evaluating IDR proposals needs to be resolved

to achieve consistent and fair assessment of IDR proposals by research funders. In this study, we aimed to address this issue by developing a set of practically applicable measures to evaluate the interdisciplinarity of research proposals.

Building on the work by Laursen, Motzer and Anderson (2022) we have collected a relatively comprehensive set of qualitative assessment criteria reflecting the integration, feasibility and scientific impact of the research proposed. In addition to the criteria taken from literature, refinement of the assessment criteria took place based on interviews with laureates. For example, the laureates stressed the importance of having team members with the right mindset to conduct IDR; the practice of conducting IDR as an assessment criterion should therefore weigh heavily. The applicants' leadership and the team readiness are considered paramount in the success of IDR, even when the grant provides *individual* research funding to talented researchers (see [Textbox 1](#)). This underlines the importance of somehow evaluating these aspects even in individual grant assessment procedures, given that the grant provides enough funding to establish a research team. In addition, institutional embedding is a success factor for IDR projects according to laureates. These arguments from our interviewees all point to the value of including hands-on experience from IDR researchers in research proposal evaluation processes. Yet, we should reflect on the possible side effects of converting these success factors into selection criteria for IDR; adding institutional embedding as a selection criterion for example may favor (or even limit) IDR conducted in dedicated IDR working environments, jeopardizing the IDR efforts of scholars working in more traditional, monodisciplinary-oriented, institutional settings.

Building on the work of Stirling (1998, 2007), Porter and Rafols (2009), Rafols, Porter and Leydesdorff (2010), and Leydesdorff and Rafols (2012), we have chosen bibliometric indicators that reflected the variety, balance and disparity of the research proposals and the articles authored by the applicant at the time of application. These were the Rao-Stirling indices of diversity for five sets of bibliometric data, and science overlay maps visualizing the interdisciplinarity of the articles authored by the applicant and the references in the proposals.

In a small sample, we detected a correlation between the interdisciplinarity of a proposal and its scientific outputs. This correlation speaks in favor of the use of bibliometric indicators to assess the interdisciplinarity of research proposals; in the current sample they predict the level of interdisciplinarity of the subsequent scientific outputs. Yet from a qualitative perspective, this relationship was problematized as laureates sometimes strategically publish their IDR work in monodisciplinary journals.

To assess the practical feasibility of qualitative assessment criteria and bibliometric metrics, we derived valuable insights during a reassessment session (see Section 5.1 for our methodological reflections on this innovative approach). Despite their predictive value, objections can be raised against the implementation of bibliometric indicators in the assessment of the interdisciplinarity of research proposals. Indeed, many panel members hesitated to use the bibliometric indicators provided. Reviewers especially feared undesired side effects of these indicators, similar to the concerns regarding the mindless use of the Hirsch's citation index and Journal Impact Factors to evaluate applicants' quality (van Raan 2006; Barnes 2017). Without further development of these

bibliometric indicators that resolve these issues, their best use might be as an optional tool that researchers with strongly interdisciplinary careers can use to communicate this aspect of their resume (e.g. through a science overlay map). Overall, we conclude the use of qualitative assessment criteria is preferable. This is indeed more in line with the San Francisco Declaration on Research Assessment (DORA) and commitment 2 of the CoARA agreement, which have both been signed and are actively supported by NWO (SF DORA n.d.; CoARA 2022). Also, introducing qualitative assessment criteria resonates with pilots within academia, including NWO, to introduce narrative curriculum vitae's (Bordignon, Chaignon and Egret 2023). We advise to use qualitative assessment criteria as a starting point for an evaluative discussion between committee members rather than to close a discussion and form a decision. This approach is in line with the assessment pathway which Laursen, Motzer and Anderson (2023) have called *Critical Questions to evaluate research integration*. We believe the reviewers' substantiation of their evaluations should be considered more insightful than the evaluation scores themselves. In any reviewer's assessment meeting in which these qualitative assessment criteria are used, the chair of the meeting (and/or the secretary on the funder's behalf) should first encourage a discussion on how the assessment criteria were used by the reviewers. This discussion can establish a shared understanding of different quality aspects of IDR. As shown by Vienni-Baptista and Pohl (2023), IDR is a multidimensional phenomenon of which people hold very divergent conceptions, which often remain implicit in committee discussions. Our proposed list of criteria possibly complemented with the heuristic tool of Vienni-Baptista and Pohl (2023), can be helpful to make different conceptions and disagreements explicit. A committee can decide for each proposal which criteria should have the largest weight in the scoring process, given the goals of the individual proposal and the aims of the overall funding program. At NWO, the so-called 'calibration sessions' that take place prior to the actual assessment meeting provide a good time to have this discussion. These are separate meetings in which a committee discusses their understanding of the provided assessment criteria, compares their preliminary scores and decides whether any adjustments are needed in the scoring system in order to arrive at a fair judgement.

It should be noted that future application forms will need to ask applicants to provide relevant information about these interdisciplinary measures.

To conclude this section, a reflection is needed regarding the performative nature of indicators that has been shown convincingly in science policy literature (de Rijcke et al. 2016). Any quality measure that affects money or reputation influences research practice, either consciously or unconsciously, in two ways (Butler 2007). The first is goal displacement: scoring high on the assessment measure becomes the goal rather than a means of evaluating if certain objectives have been met (Colwell et al. 2012). The second is related to more structural, yet less obvious, changes to research priorities, publication activities, and research capacity-building and organization. While de Rijcke et al. (2016) warn against the potential bias of disciplinary assessments against IDR, one might also argue that the very principle of assessment affecting practice also applies to qualitative and bibliometric measures assessing the interdisciplinarity of a proposal. This calls for the responsible implementation of IDR assessment

measures, as proposed in principle nr. 9 in the Leiden Manifesto on responsible metrics: 'Recognize the systemic effects of assessment and indicators' (Hicks et al. 2015).

5.1 Methodological considerations and further research

In this study, we embedded quantitative and qualitative measures in a qualitative analysis of data collected by various methods such as interviews and a reassessment session of successful research proposals. In general, our retrospective approach was appealing since it gave us, within a reasonable time span, access to original IDR grant proposals, insights from the laureates regarding the execution of the research process, and the outputs of the projects as they were close to conclusion. A downside of this approach was that it inevitably included self-reporting of laureates, which may have colored our findings regarding the research process. We attempted to stimulate the laureates to be as honest as possible, also about the hardships of conducting IDR, by guaranteeing anonymity both towards NWO as their funding organization and externally.

Aiming to assess the practical feasibility of qualitative and bibliometric measures, we employed a methodologically innovative strategy of organizing a reassessment session, which was a simulated meeting of a selection committee with original panel members and original research proposals. We are not aware of any comparable experiments in science policy studies. The value of this approach lies in its simulation of real-world assessment settings, providing insights into the discussions that reviewers (should) have on the interdisciplinarity of the proposals. In this study's set-up, there was little at stake since the panel members were re-evaluating successful proposals. This allowed the evaluation committee to advise freely on the feasibility of the assessment measures. One should take note that the context in which the original assessment meeting (in 2018) and its rerun (in 2023) took place has changed considerably in the Netherlands. NWO has adopted and promoted a more reserved attitude towards the use of bibliometric indicators in the assessment of research proposals. This has become an important topic of discussion in the academic community in The Netherlands after the introduction in 2020 of a narrative resume, which has the goal to discourage reviewers to base their assessment on quantitative indicators not directly related to individual performance (e.g. journal impact factors). This may have influenced the discussions at the reassessment session, especially considering the reviewers' hesitation to adopt the Rao-Stirling index as a measure for interdisciplinarity. Following up on this study, we therefore advise to employ an action-oriented research approach in which the new qualitative assessment measures are included in an on-going grant assessment procedure.

To conclude, this study focused on two specific funding instruments (Vidi and Vici) which were both individual and bottom-up grants, with no thematic demarcation. The calls were not specifically aimed at IDR, and neither were the IDR proposals in direct competition with monodisciplinary proposals. The relatively limited scope of our study raises the question to what extent our findings are applicable to other funding instruments, for example grants aimed at IDR consortia or grants with a scientific or societal thematic focus. One can imagine that especially the evaluation processes of research proposals from IDR consortia

encounters its own challenges, for example assessing the proposed team's readiness and leadership qualities. Rather than speculating about these particulars, we propose that aforementioned action-oriented future research plans should include a comparative element to test the feasibility of the qualitative assessment measures in various funding instruments.

5.2 Conclusion

This study shows that bibliometric indicators that reflect the variety, balance and disparity of research proposals correlate with the interdisciplinarity of the research output, suggesting a predictive value. Distinguishing between different dimensions of diversity also provided conceptual clarity to the assessment discussions. Yet, based on an innovative simulated assessment meeting of a selection committee, we conclude that qualitative criteria focusing on integration (diversity and interaction), feasibility, and impact, are currently more feasible and useful to support selection decisions to allocate funding for interdisciplinary research. Such qualitative assessment criteria should serve as a starting point for an evaluative discussion between committee members rather than close a discussion and form a decision. The set of criteria that we have collected in this paper can serve as a draft to be refined into a balanced and workable set for similar assessment situations in real-life. In a next iteration, perhaps the three dimensions of diversity can help to further structure the set.

Acknowledgements

The authors thank all participants in this study (interviewees and members of the assessment committee) for their time and valuable inputs. The authors express their gratitude to Annemarie Horn (Radboud University) and Michael Dale (TU Eindhoven) for partaking in the interviews with laureates as co-interviewers, and Jochem Zuiderwijk (Radboud University) for observing the reassessment session. We gratefully acknowledge Kristin Oxley (University of Oslo), Philippe Larrue (OECD), Marianna Marra (University of Sussex), Paul Diederer (Rathenau Instituut) and Jasper Deuten (Rathenau Instituut) for their helpful feedback on earlier drafts of this paper.

Supplementary data

Supplementary data is available at *Research Evaluation Journal* online.

Funding

This work was supported by the Dutch research council NWO. Design of the study, data collection, analysis and interpretation were all part of the pilot phase of the national Knowledge platform for interdisciplinary and transdisciplinary research. More information can be found at: <https://www.nwo.nl/en/knowledge-platform-for-interdisciplinary-and-transdisciplinary-research>.

Notes

1. We largely follow Laursen, Motzer and Anderson (2022) in the use of our terminology regarding measures, criteria, metrics and indicators assessing IDR. We use the word 'measure' as a covering term for all

methodologies throughout this article to simplify sentences. When we refer to quantitative measures only, we use 'metrics' or 'indicators'. Qualitative measures are signaled by 'criteria'.

2. Some authors consider 'emergence' as another dimension of integration, but we have chosen not to include this because of a lack of clear definition in the literature.

References

- Abdoul, H. et al. (2012) 'Peer Review of Grant Applications: Criteria Used and Qualitative Study of Reviewer Practices', *PLoS One*, 7: e46054. <https://doi.org/10.1371/journal.pone.0046054>
- Abma-Schouten, R. et al. (2023) 'Evaluation of Research Proposals by Peer Review Panels: broader Panels for Broader Assessments?', *Science and Public Policy*, 50: 619–32. <https://doi.org/10.1093/scipol/scad009>
- Alhadreti, O., and Mayhew, P. (2017) 'To Intervene or Not to Intervene: An Investigation of Three Think-Aloud Protocols in Usability Testing', *Journal of Usability Studies*, 12: 111–32.
- AWTI (2022) *Grenzeloos onderzoeken: Stimuleer interdisciplinariteit met twee onderscheidende overheidsrollen*. Den Haag: AWTI.
- Barnes, C. (2017) 'The h-Index Debate: An Introduction for Librarians', *The Journal of Academic Librarianship*, 43: 487–94.
- Belcher, B. M. et al. (2015) 'Defining and Assessing Research Quality in a Transdisciplinary Context', *Research Evaluation*, 25: 1–17. <https://doi.org/10.1093/reseval/rvv025>
- Bordignon, F., Chaignon, L., and Egret, D. (2023) 'Promoting Narrative CVs to Improve Research Evaluation? A Review of Opinion Pieces and Experiments', *Research Evaluation*, 32: 313–20. <https://doi.org/10.1093/reseval/rvad013>
- Bruun, H. et al. (2005) *Promoting Interdisciplinary Research: The Case of the Academy of Finland*. Helsinki: Academy of Finland.
- Butler, L. (2007) 'Assessing University Research: A Plea for a Balanced Approach', *Science and Public Policy*, 34: 565–74.
- CoARA (2022) *Agreement on Reforming Research Assessment*. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf, accessed 24 Jan. 2024.
- Colwell, R. et al. (2012) *Informing Research Choices: Indicators and Judgment. The Expert Panel on Science Performance and Research Funding*. Ottawa: Council of Canadian Academics.
- D'Este, P., and Robinson-García, N. (2023) 'Interdisciplinary Research and the Societal Visibility of Science: The Advantages of Spanning Multiple and Distant Scientific Fields', *Research Policy*, 52: 104609. <https://doi.org/10.1016/j.respol.2022.104609>
- de Rijcke, S. D. et al. (2016) 'Evaluation Practices and Effects of Indicator Use—A Literature Review', *Research Evaluation*, 25: 161–9.
- Ebadi, A., and Schiffauerova, A. (2015) 'How to Receive More Funding for Your Research? Get Connected to the Right People!', *PLoS One*, 10: e0133061. <https://doi.org/10.1371/journal.pone.0133061>
- Ericsson, K. A., and Simon, H. A. (1993) *Protocol Analysis: Verbal Reports as Data*, Revised ed. Cambridge: MIT Press.
- Guthrie, S., Ghiga, I., and Wooding, S. (2018) 'What Do We Know about Grant Peer Review in the Health Sciences?', *F1000Research*, 6: 1335. <https://doi.org/10.12688/f1000research.11917.2>
- Hicks, D. M. et al. (2015) 'Bibliometrics: The Leiden Manifesto for Research Metrics', *Nature*, 520: 429–31.
- Huutoniemi, K. (2016) 'Interdisciplinarity as Academic Accountability: Prospects for Quality Control across Disciplinary Boundaries', *Social Epistemology*, 30: 163–85.
- Huutoniemi, K. et al. (2010) 'Analyzing Interdisciplinarity: Typology and Indicators', *Research Policy*, 39: 79–88.
- Klein, J. T. (2008) 'Evaluation of Interdisciplinary and Transdisciplinary Research: A Literature Review', *American Journal of Preventive Medicine*, 35: S116–23. <https://doi.org/10.1016/j.amepre.2008.05.010>
- Lamont, M. (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Harvard: Harvard University Press.
- Lamont, M., Mallard, G., and Guetzkow, J. (2006) 'Beyond Blind Faith: overcoming the Obstacles to Interdisciplinary Evaluation',

- Research Evaluation*, 15: 43–55. <https://doi.org/10.3152/147154406781776002>
- Langfeldt, L. (2001) ‘The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome’, *Social Studies of Science*, 31: 820–41.
- Laursen, B. K., Motzer, N., and Anderson, K. J. (2022) ‘Pathways for Assessing Interdisciplinarity: A Systematic Review’, *Research Evaluation*, 31: 326–43. <https://doi.org/10.1093/reseval/rvac013>
- Laursen, B. K., Motzer, N., and Anderson, K. J. (2023) ‘Pathway Profiles: Learning from Five Main Approaches to Assessing Interdisciplinarity’, *Research Evaluation*, 32: 213–27. <https://doi.org/10.1093/reseval/rvac036>
- Leydesdorff, L. (2018) ‘Diversity and Interdisciplinarity: How Can One Distinguish and Recombine Disparity, Variety, and Balance?’, *Scientometrics*, 116: 2113–21. <https://doi.org/10.1007/s11192-018-2810-y>
- Leydesdorff, L., and 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: 387–402. <https://doi.org/10.1002/asi.24416>
- Leydesdorff, L., and Rafols, I. (2012) ‘Interactive Overlays: A New Method for Generating Global Journal Maps from Web-of-Science Data’, *Journal of Informetrics*, 6: 318–332.
- Li, D. (2017) ‘Expertise versus Bias in Evaluation: Evidence from the NIH’, *American Economic Journal: Applied Economics*, 9: 60–92.
- Lyll, C. et al. (2013) ‘The Role of Funding Agencies in Creating Interdisciplinary Knowledge’, *Science and Public Policy*, 40: 62–71.
- Lyll, C., and King, E. (2013) *International Good Practice in the Peer Review of Interdisciplinary Research*. Edinburgh: University of Edinburgh.
- Mayo, N. E. et al. (2006) ‘Peering at Peer Review Revealed High Degree of Chance Associated with Funding of Grant Applications’, *Journal of Clinical Epidemiology*, 59: 842–8.
- McLeish, T., and Strang, V. (2016) ‘Evaluating Interdisciplinary Research: The Elephant in the Peer-Reviewers’ Room’, *Palgrave Communications*, 2: 1–8.
- Nair, K. M. (2011). ‘Understanding and Assessing Interdisciplinarity in Health Research Teams’, Doctoral dissertation. McMaster University. <http://hdl.handle.net/11375/10680>
- National Academies of Sciences, Engineering, and Medicine (2005) *Facilitating Interdisciplinary Research. Chapter 2 the Drivers of Interdisciplinary Research*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/11153>.
- OECD (2020) *Addressing Societal Challenges Using Transdisciplinary Research*. Paris: OECD.
- Oxley, K. (under review) ‘Conservatism and Variability: A Systems View on Panel Peer Review’, PhD thesis. University of Oslo.
- Philipps, A. (2022) ‘Research Funding Randomly Allocated? A Survey of Scientists’ Views on Peer Review and Lottery’, *Science and Public Policy*, 49: 365–77. <https://doi.org/10.1093/scipol/scab084>
- Porter, A. L. et al. (2006) ‘Interdisciplinary Research: meaning, Metrics and Nurture’, *Research Evaluation*, 15: 187–96.
- Porter, A. L., and Rafols, I. (2009) ‘Is Science Becoming More Interdisciplinary? Measuring and Mapping Six Research Fields over Time’, *Scientometrics*, 81: 719–45. <https://doi.org/10.1007/s11192-008-2197-2>
- Porter, A. L., and Rossini, F. A. (1985) ‘Peer Review of Interdisciplinary Research Proposals’, *Science, Technology, & Human Values*, 10: 33–8.
- Rafols, I., and Meyer, M. (2010) ‘Diversity and Network Coherence as Indicators of Interdisciplinarity: case Studies in Bionanoscience’, *Scientometrics*, 82: 263–87. <https://doi.org/10.1007/s11192-009-0041-y>
- Rafols, I., Porter, A. L., and Leydesdorff, L. (2010) ‘Science Overlay Maps: A New Tool for Research Policy and Library Management’, *Journal of the American Society for Information Science and Technology*, 61: 1871–87.
- Rathenau Instituut (2023) *Nederlandse overheidsbijdrage aan R&D—naar type financiering*. <https://www.rathenau.nl/nl/wetenschap-cijfers/geld/wat-geeft-nederland-uit-aan-rd/nederlandse-overheidsbijdrage-aan-rd-naar>, accessed 23 Jul. 2023.
- Roumbanis, L. (2022) ‘Disagreement and Agonistic Chance in Peer Review’, *Science, Technology, & Human Values*, 47: 1302–33. <https://doi.org/10.1177/01622439211026016>
- Sandström, U., and Hällsten, M. (2008) ‘Persistent Nepotism in Peer-Review’, *Scientometrics*, 74: 175–89.
- SF DORA (n.d.) *The Dutch Research Council Inclusive Assessment Toolkit*. <https://sfdora.org/resource/the-dutch-research-council-nwo-inclusive-assessment-toolkit/>, accessed 27 Jul. 2023.
- Stirling, A. (1998) ‘On the Economics and Analysis of Diversity’, Brighton: SPRU Electronic Working Papers No. 28.
- Stirling, A. (2007) ‘A General Framework for Analysing Diversity in Science, Technology and Society’, *Journal of the Royal Society Interface*, 4: 707–19.
- Strang, V., and McLeish, T. (2015) *Evaluating Interdisciplinary Research: A Practical Guide*. Durham: Durham University.
- Sun, Ye. et al. (2021) ‘Interdisciplinary Researchers Attain Better Long-Term Funding Performance’, *Communication Physics*, 4: 263. <https://doi.org/10.1038/s42005-021-00769-z>
- Tate, E., Decker, V., and Just, C. (2021) ‘Evaluating Collaborative Readiness for Interdisciplinary Flood Research’, *Risk Analysis*, 41: 1187–94. <https://doi.org/10.1111/risa.13249>
- Teplitskiy, M. et al. (2018) ‘The Sociology of Scientific Validity: How Professional Networks Shape Judgement in Peer Review’, *Research Policy*, 47: 1825–41. <https://doi.org/10.1016/j.respol.2018.06.014>
- Thorngate, W., Faregh, N., and Young, M. (2002) *Mining the Archives: Analyses of CIHR Research Grant Adjudications*. Ottawa, Ontario: Carleton University.
- Travis, G. D. L., and Collins, H. M. (1991) ‘New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System’, *Science, Technology, & Human Values*, 16: 322–41.
- van Arensbergen, P., van der Weijden, I., and van den Besselaar, P. (2014) ‘The Selection of Talent as a Group Process. A Literature Review on the Social Dynamics of Decision Making in Grant Panels’, *Research Evaluation*, 23: 298–311. <https://doi.org/10.1093/reseval/rvu017>
- van Raan, A. F. J. (2006) ‘Comparison of the Hirsch-Index with Standard Bibliometric Indicators and with Peer Judgment for 147 Chemistry Research Groups’, *Scientometrics*, 67: 491–502. <https://doi.org/10.1556/Scient.67.2006.3.10>
- Vienni-Baptista, B., and Pohl, C. E. (2023) ‘Exploring Interdisciplinarity and Transdisciplinarity as Knowledge Regimes: A Heuristic Tool for Disentangling Understandings in Academia and Policy’, *Science, Technology & Human Values*. <https://doi.org/10.1177/01622439231216789>
- Wang, J., Lee, Y. N., and Walsh, J. P. (2018) ‘Funding Model and Creativity in Science: Competitive versus Block Funding and Status Contingency Effects’, *Research Policy*, 47: 1070–83. <https://doi.org/10.1016/j.respol.2018.03.014>