

This is a repository copy of *Evaluating interdisciplinary research: The elephant in the peer-reviewers' room*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/136955/>

Version: Published Version

Article:

McLeish, Tom orcid.org/0000-0002-2025-0299 and Strang, Veronica (2016) Evaluating interdisciplinary research: The elephant in the peer-reviewers' room. Palgrave Communications. 16055.

<https://doi.org/10.1057/palcomms.2016.55>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



REVIEW ARTICLE

Received 11 Mar 2016 | Accepted 18 Jul 2016 | Published 25 Aug 2016

DOI: 10.1057/palcomms.2016.55

OPEN

Evaluating interdisciplinary research: the elephant in the peer-reviewers' room

Tom McLeish¹ and Veronica Strang¹

ABSTRACT We review a selection of published reports on the evaluation and wider peer-review of interdisciplinary research (IDR), drawing on an in-depth examination of a range of interdisciplinary projects and the work of a UK-based working group of funders and researchers. Our aim is to elucidate best practice. We focus the study on integrative, interdisciplinary projects, rather than those at the level of “multidisciplinary dialogue”. Five areas of evaluation (publishing, research grants, careers, IDR centres, institutions) demonstrate both commonality and difference in the task of measuring added value in IDR collaborations. We find that, although single-discipline peer review processes are poorly suited to address IDR, a framework that starts with the assumption that IDR is a fundamental academic research practice is effective for single-discipline evaluation as well. This article is published as part of a collection on interdisciplinarity.

Introduction

Interdisciplinary research (IDR) is widely praised for its capacity to address “wicked problems”, transform and reshape the academic landscape in imaginative ways, re-integrate a fragmented world of learning, and even transform disciplines themselves. In the following discussion we adopt Giddens’ helpful definition:

Interdisciplinary research (IDR) is 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. (Land, 2011: 7, citing Giddens, 1991)

Several collections of essays theorize IDR, and explore non-summative and modal concepts of interdisciplinarity (Barry and Born, 2013; Callard and Fitzgerald, 2015). In this instance we will focus on the most transformed (and arguably transformative) mode of IDR—the deeply integrative category (whether this is motivated by external problems as Giddens suggests, or by an internal academic imperative) in which disciplines move beyond a polyphonic discourse or empirical comparison to create new ways of approaching research questions (National Academy of Sciences, 2010). But how do we connect this exciting world of integrated research and learning with the now-essential academic *sine qua non* of evaluation and assessment of quality? There is little clear consensus about how to undertake objective evaluations of IDR (National Academies of Science, 2010:166; Frodeman *et al.*, 2012: 309). The challenges of evaluating IDR have been

¹ Durham University, Durham, UK Correspondence: (e-mail: t.c.b.mcleish@durham.ac.uk)

cited as a barrier to undertaking it. The imperative of finding solutions to this problem is another driver of our choice of focus on integrative IDR—for this mode of conducting research is the least responsive to single-disciplinary views, and the most challenging to evaluate.

In this article we consider funders' perspectives on this barrier, rather than the views of researchers, but draw together evidence from both, suggesting that the insufficiency of current peer review procedures constitutes an impediment in realising the potential of IDR. A recent review of the (limited) available literature (using the terminology of “transdisciplinary research”—TDR)¹ concludes:

The lack of a standard and broadly applicable framework for the evaluation of quality in TDR is perceived to cause an implicit or explicit devaluation of high-quality TDR or may prevent quality TDR from being done. (Belcher *et al.*, 2015: 14)

We then identify published work that has suggested constructive ways forward, and review recommendations in the light of work by a group of funders and researchers (Strang and McLeish, 2015), and experience of integrative IDR projects. Although some aspects of IDR evaluation appear ubiquitously, it is important to differentiate the divergent evaluative tasks that a comprehensive review needs to recognize. There are five key levels—with major differences in scale—at which effective methods of evaluation are needed:

1. Research outputs (concerning journal and book publishers).
2. Research grant proposals (concerning funding organizations).
3. Individual career progression (concerning higher education institutions and other employers).
4. HEI-based institutes and centres in support of IDR (concerning higher education institutions).
5. Institutional research (concerning national funding councils).

Although the necessary cohesive and emergent properties of high-quality IDR apply at the level of individual outputs as much as that of entire institutions, methodological approaches to evaluation clearly need to differentiate these key areas. We note that the most promising proposed routes to IDR evaluation construct frameworks of questions based on best practice in interdisciplinary methodologies, and exemplify these by proposing examples that address strands (2) and (5) in more detail.

Finally we show that this kind of reflection on the evaluation of IDR radically illuminates the structural position it holds in academic practice. As Callard and Fitzgerald (2015: 23) put it, “[moments of peer review] stage the complexities, tensions, and excitements of ‘interdisciplinarity’, precisely at the moment in which interdisciplinarity inveigles itself into the strictures and assumptions of (to use a flat-footed term) ‘normal science’”.

The evidence—three bases and five levels of practice

In this study, evidence for the challenges in evaluating IDR on the one hand, and experience in meeting them on the other, come from three sources: Previously published work on the evaluation of IDR (sometimes denoted by other terms), selected for a focus on integrative IDR rather than “in-dialogue” MDR.

Experience of working groups of funders, researchers and funding bodies. For example, such a group, convened by Durham University's Institute of Advanced Study, published recommendations on best practice in evaluation at levels (1)–(5) above (detailed in Strang and McLeish, 2015). Other current projects² are drawn on in their early stages.

Personal experience of a range of interdisciplinary projects involving intense integration of widely separated disciplines and methodologies, analysed in the context and evidence of (1) and (2) (McLeish and Strang, 2014; also Strang, 2009; Gasper *et al.*, 2016). Lyaal and King (2013) point to several funding organizations that have seen the need to adopt special review procedures for IDR evaluation. The Swiss National Science Foundation (SNSF), for example, is not alone in identifying the evaluation of IDR as one of the most intransigent obstacles to its adoption and resourcing (Defila and DiGiulio, 1999). After a trial period of using a special interdisciplinary panel to assess IDR proposals, SNSF concluded that there was a need both to maintain such a specific evaluation body for interdisciplinary proposals, and to create specific funding instruments for IDR. Other funders have fallen at the fence, discontinuing IDR funding calls because of the seemingly insurmountable challenges of evaluating IDR proposals and their consequent outputs.

A further barrier presented to the peer review of IDR is the increasingly narrow remit of most new academic journals (the present journal is a notable exception), as well as the preference of high-profile “general interest” journals for publishing a range of single-discipline research outputs rather than IDR. This arises, in part, because they too experience difficulties in finding peer reviewers, or adequate processes of evaluation, for the assessment of the latter.

Throughout the levels of academic evaluation identified above—grant proposals, journal articles, individual career progression and institutional research evaluations (for example, research excellence framework (REF))—the evidence reflects a lack of capacity within current review processes to address IDR. Summarising by level of evaluation:

- *Research outputs*: specifically, at the level of journals and academic books, several academic respondents to a recent consultation by the British Academy (forthcoming) commented that the readership, and thus the impact factor, of interdisciplinary journals are frequently small, creating a disincentive to publish with them.
- *Research grant proposals*: reports have consistently reflected a conservative, discipline-based process at the level of research funding. Researchers describe experiences, with many funders, of being required to “pigeon-hole” their research into pre-determined and narrow disciplinary categories, rather than being able to let the breadth of disciplines and methodologies speak for itself.
- *Research element of career progression*: noting pressures to establish a distinct disciplinary identity, individual researchers often express concerns that their career progression will not be propelled by IDR as effectively as by single-discipline work.
- *IDR centres*: both institutional and individual evidence points to the increasing, and successful, deployment of IDR centres and institutes. Some of the difficulties associated with establishing and maintaining them are, however, associated with the complexities inherent in evaluating their effectiveness.
- *Institutional research assessment*: at the level of institutional assessment of research quality there is evidence that the disciplinary structure of the REF has disincentivised IDR, or at least the submission of IDR outputs. A citation-based quantitative analysis performed for the Higher Education Funding Council of England (HEFCE), of the proportion of IDR submitted to the³ REF compared with the proportion of IDR in United Kingdom research as a whole, lent partial support to these views (Elsevier, 2015). From the entire array of outputs produced through UK research, a lower

proportion of IDR was submitted to the REF. On the other hand, when it came to evaluation, the quality of submitted IDR outputs did not differ from that of the entire output distribution. Various measures were taken in the REF2014 exercise to encourage the submission of IDR outputs and to evaluate them effectively. These included the encouragement of joint submissions from more than one institution, and the multiple use of interdisciplinary outputs by more than one unit of assessment in an institution.

There was a similar arrangement for the submission of impact case studies,⁴ and this has proved to be an important area of strength for IDR. A subsequent analysis of the research cited indicated that interdisciplinary projects constituted a very high percentage (the report cites 87%) of the endeavours underpinning universities' impact case studies in the period 2008–2014.

There is limited reported work on the development of general peer review processes in the evaluation of IDR at all levels, but some significant recent studies have made experience- and theory-based suggestions. This body of work also provides an additional evidence-base for the crucial issues identified above:

- Research by Lyall *et al.* (2013) proposed a much closer working relationship between researchers and funders in the case of (typically large) IDR projects.
- A report commissioned by the RCUK Research Group identified aspects of international best practice in peer review of IDR (Lyll and King *op. cit.*).
- A Canadian group reviewed literature on evaluation of IDR (they refer to it as TDR) and suggested an evaluation framework (Belcher *et al.*, *op. cit.*)
- Klein (2008) draws together literature on IDR evaluation, proposing a seven-point categorization of “principles” that rehearse but go beyond those applied to disciplinary research.
- Callard and Fitzgerald (2015) provide a detailed textual analysis of experiences of peer review in IDR that clarify the epistemic crevasses into which IDR can fall within discipline-base peer review.

The problem—encompassing complexity

The wider evidence collected by the Durham IAS working group above (facilitated by the present authors); the selected literature explored; and our in-depth experience with a number of IDR projects, together indicate a complex collection of challenges associated with the evaluation of IDR. The academy already possesses a highly evolved set of mechanisms, procedures and criteria for the peer-review of single-discipline work at all of the Levels (1)–(5) identified above. The difficulties arise from the special character of IDR when compared to the single-disciplinary research, which has shaped our current peer review frameworks. The additional evaluatory challenges are multiple, presenting at first a bewildering set of ideas to negotiate. It is initially a challenge to get a clear sense of the shape of the problem, but a good starting point for an emergent understanding of the challenge is to list as comprehensively as possible the major elements that we have identified from our three sources of experience, focus-group discussion and literature review:

- The need to evaluate multiple disciplinary expertise in linked and concurrent form
- The deployment of mixed methodologies arising from disparate disciplines
- The requirement to evaluate the extra dimensions of team-building and management that IDR calls upon (examples

include the notion of “disciplinary hospitality”) (Strang and McLeish, *op. cit.*: 3)

- The more extended timeframe, size and cost typical of IDR projects, and the differing temporal needs of disciplinary areas
- The requirement to evaluate the role of participants external to the academic team (especially important in TDS and “challenge-driven” IDR).
- The achievement of equality and productive partnerships between disciplines in institutional contexts in which they often experience inequalities in roles, status and access to resources
- The limited ability to frame the outputs and outcomes of IDR within the existing evaluation frameworks of the participating single disciplines
- The requirement to frame IDR projects or outputs in language that may be unfamiliar to potential evaluators
- The occasional involvement of “token” disciplines in response to inadequately framed IDR funding calls or misconceived responses to them
- The bewildering multiplicity of criteria that arise in IDR proposals
- The danger of double, or multiple, jeopardy in the sequential evaluation of IDR through single-disciplinary lenses
- The perception that the journals in which the IDR outputs are published are less prestigious or of lower profile than those focused on single-discipline research
- The increased need for openness and flexibility in the less-familiar territory of IDR, and the consequent complexity of planning contingencies and risk-mitigation
- The increased range of criteria required in the evaluation of IDR
- The occasional simultaneous presence of more than one evaluating body (for example, funding council) with incompatible criteria.
- The risk that disciplinary components of a project may simply proceed in parallel with one another, without intensive interaction and mutual dependency

The items on this list might suggest unmanageable complexities implicit within this challenge. However, there is a summative core issue that underlies these differentiated factors and reveals them as multiple aspects of the same structure. This summative issue arises from the motivation and rationale for IDR in the first place. The question of evaluating the emergent whole of IDR is not expressible in terms of its (disciplinary) parts. Rather it tests the extent to which the disciplinary participants have communicated and engaged to such a degree that new knowledge and understanding can no longer be expressed as a sum of their separate contributions. We might articulate the core question in these terms: Does the project, centre or output successfully and effectively integrate its disciplinary components, so that it generates an emergent whole, addressing an interdisciplinary research question, or programme of questions, and producing outcomes that are demonstrably greater than the sum of its (disciplinary) parts?

Klein reports the same high-level aspiration from a Harvard study: “More primary or epistemic measures of ‘good’ work [other than discipline-based proxy measures such as citations] are needed that address the substance and constitution of the research” (Klein, 2008: S118). This is a central question, and as it emerges from an apparently bewildering set of special requirements of IDR, it needs to be broken down into a detailed yet connected set of evaluative measures. For example, as most large interdisciplinary projects are conducted by teams of researchers, such integration represents a successful process of knowledge exchange and collaboration. Thus evaluations also need to consider how this process has been managed and

supported in the creation of “epistemic communities”, as well as the examination of research outputs. However, many individual scholars conduct IDR too, and in this instance the question addresses the extent to which they have, individually, integrated knowledge and expertise from different disciplinary areas.

We next examine the integrative question, and the problems it poses, in a little more detail.

The elephant in the room: mono-disciplinary engagement

A classic metaphor is provided by the ancient Indian parable about a group of blind people attempting to identify an elephant. One can feel a smooth tusk, another a gnarled and tree-like leg and a third the tasselled tail, but none can perceive the whole animal. Yet the evaluation of the emergent whole is precisely the core task that differentiates the evaluation of IDR from that dealing with single-discipline research. It is vital, quite simply because the difference between high-quality and poor IDR is most often *not* to be found in the quality of its disciplinary ingredients, in its individual researchers or in their knowledge sources, but rather in how these are combined to generate the whole research project and its findings. The same metaphor was invoked by Repko and Szostak (2012), who advocated a seventh blind examiner of the elephant, with a role to question the other six and lead a process of integration producing a theory of the elephant as a whole.

A core question aimed at “more than the sum of the parts” naturally draws the evaluation towards key issues within IDR projects: the co-generation of research questions and project design (Belcher *et al.*, 2015); the compatibility of epistemologies (Klein, 2008); mutual learning and language-acquisition within teams (Marzano *et al.*, 2006); high-level responsibilities for managing and nurturing internal communication (Marzano *et al.*, 2006); development of interdisciplinary skills (Strang and McLeish, 2015); shared methodologies and interpretations (Callard and Fitzgerald, 2015); the creation of common ground (Repko and Szostak, 2012); combination of research results at high levels (Somerville and Rapport, 2000); and so on.

The inability of current frameworks of peer-review, career advancement or institutional evaluation to assess these core issues stems from the absence of prior incentives to develop measures and methodologies that address emergent structures, knowledge, understanding, wisdoms, whose articulation cannot be framed within a single discipline. A mono-disciplinary view may even lead to a rejection of the validity of IDR on the part of a reviewer with no experience or grasp of the benefits of working. (For example, in a project bringing medieval scholars and scientists together in a re-examination of thirteenth century work on light, colour, motion and sound, a reviewer opined that the very notion of employing scientists to examine historical work would be anachronistic Gasper *et al.*, 2016).

Seen this way, the “problem” tells us why it has also proved ineffective to address its challenges by simply *adding* to existing frameworks. The “sticking-plaster” approach to IDR evaluation assumes an additive, linear structure, rather than the non-linear processes that drive the emergence of qualitatively new results (Newell, 2001). Merely assembling experts corresponding to the constituent disciplines within a single-IDR proposal does *not* guarantee the effective evaluation of its whole. A Finnish study of panel evaluation of IDR proposals confirms wider qualitative evidence suggesting that, without effective coaching, or without the inclusion of people whose expertise lies in the identification of good IDR, such panels are the peer-review equivalent of the elephant in the room (Bruun *et al.*, 2005).

This way of describing the difficulties of evaluating IDR, and the other findings of this report, point to the fundamental

importance of such work. Although the challenge is severe, the rewards are great: there is much to be gained, not by building a superstructure upon the disciplines into which we currently divide academia, but achieving renewed access to a foundational level of learning that we are in danger of losing. This suggests that the most productive way forward is not to create additional evaluatory criteria, but to recognize a more fundamental framework for evaluation from the beginning (Strang and McLeish, *op cit.*; McLeish and Strang, 2014).

Towards solutions—creating an evaluative framework for Levels (2) and (5)

In this section, we illustrate how the joint summative and detailed evaluatory approach to IDR may be developed in practice, by focussing on two of the five levels identified in Section I: the interdisciplinary grant proposal (Level 2) and the national research evaluation exercise (Level 5).

The few suggested frameworks for evaluation of IDR proposals have an interesting commonality in *form*: they have generally attempted to identify the holistic structures of good IDR through the formulation of *questions* to address in relation to the proposal, project or output. For example, both Lyall and King (2013), and Strang and McLeish (2015) condense their findings into a “checklist”. We have indicated **in bold** in the box below the questions proposed by Lyall and King that appear to be specific to the evaluation of the fundamentally integrative nature of IDR.

1. Does the proposal describe clear goals, adequate preparation, appropriate method, significant results, effective presentation, reflective critique?
2. How was the problem formulated?
3. **How diverse are the disciplines, methods and researchers and how suitable is the combination of disciplines?**
4. **Is there a clear justification for the choice of disciplines based on the needs of the research questions?**
5. Is the study sufficiently anchored in relevant literature?
6. What is the relationship with the methodology?
7. How will communication be tackled?
8. **Does it describe how the insights of 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?**
9. **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?**
10. Is the leadership role and management strategy to deliver the desired outcomes clearly articulated?
11. **Do the researchers involved have demonstrable interdisciplinary skills and experience?**
12. **In particular, is there evidence of interdisciplinary leadership?**
13. Is there an appropriate plan for stakeholder/user engagement from the outset of the project?
14. Does the proposal budget for, and justify, the additional resources needed?
15. **Is it clear how interdisciplinarity will be reflected in the project outputs and outcomes?**

The questions **in bold** seem to lean towards the assessment of the integrative and emergent, and in the case of strongly IDR may require particular expertise. However, such questions

could equally be addressed to any research proposal or output where “disciplines” might be replaced by “integrated knowledge” or “methodologies”. Furthermore, if they are included in such an evaluation, they enhance the quality of scrutiny given to even a single-disciplinary research programme. This transferability is supported by the working group report from the Durham IAS (Strang and McLeish *op. cit.*). The report avers that:

With the recognition that IDR represents a foundation, rather than a superstructure, in the organization of knowledge (for a historical perspective see Weingart in Frodeman *et al.*, 2012), it is evident that:

- principles that guide good IDR can also serve as guidelines for good disciplinary research;
- approaches to evaluation that work well for IDR may usefully inform evaluations of single-disciplinary research (2015: 6).

The observation that such transferability does not work reciprocally is the central reason for the challenge addressed in this article. When the starting point for evaluation is that of single-discipline research, attempts to add special “bolt-on” criteria for IDR can be awkward. But if a holistic, interdisciplinary perspective is assumed from the beginning, then there is no point at which special criteria need to be inserted into an evaluatory scheme. Disciplinary and interdisciplinary evaluatory frameworks do not commute.

The Durham IAS report is also couched in the form of a checklist—a rather large one as separate frameworks of probing questions are derived for each of the levels of evaluation. But these detailed lists are generated from an overarching set of criteria, reproduced in the box below:

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? Are they sufficient?
2. Is the leadership structure characterized by inclusivity, facilitation, transparency of roles and an equality of contributing disciplines in terms of voice and status?
3. Are additional resources and time planned for dialogue, co-learning and integration between the contributing disciplines?
4. Is it clear how the individual disciplines may benefit on their own terms by engaging with the IDR, noting that this can be transformational?
5. Is there a disciplinary hospitality between the researchers, and to external participants, which avoids a hierarchical view of the contributing disciplines?
6. Are there ways of supporting the social cohesion of the collaborators (recognizing that interdisciplinary support structures may help)?
7. Have the different scales, and communication between them, been recognized in the structure of the research?
8. Are there processes for cohering the different data in the research, quantitative and qualitative, recognizing the need for translation where this is necessary?
9. Is the necessary experience with IDR represented by the team and the leadership as well as training and development in place?
10. Are research plans sufficiently open and flexible to adapt to new questions or directions that might arise unforeseen at the outset?
11. If there are “service disciplines” identified in the research, has this been driven by the project needs and not by assumed prevalence of one discipline over another?

Note that there is a strong correspondence between the key points in the two frameworks we have summarized above, correspondences identified by listed point in Table 1.

The framework proposed by Belcher *et al.* (*op. cit.*), also drawn from a wide survey of the literature, is rather different in form. These authors focus on “transdisciplinarity”, identifying such research as having “explicit goals to contribute to ‘real world’ solutions and strong emphasis on context and social engagement” (Belcher *et al.*, 2015: 1). There are some—possibly tangential—questions about a supposed division between research and a putative “real world”, and about the notion that “trans” disciplinarity possesses a more practical and applied focus than “inter” disciplinarity. But their list of criteria is useful, in that it is more general and more universally applicable to the evaluation of research quality in general. Thus it aims to assess (1) Relevance, (2) Credibility, (3) Legitimacy (this heading contains much of the special requirements of healthy IDR explicit in the “checklists”) and (4) Effectiveness (also comprising aspects such as training and development with IDR in mind). These, too, can be mapped onto the cross-corresponding classes of evaluative criteria from Lyall and King, and Strang and McLeish (Table 1).

King (2008) extracted seven perspectives, or “principles” in evaluation of IDR from her comprehensive review. These were (in her specific definitions): (1) variability of goals; (2) variability of criteria and indicators; (3) leveraging of integration; (4) interaction of social and cognitive factors in collaboration; (5) management, leadership, and coaching; (6) iteration in a comprehensive and transparent system; and (7) effectiveness and impact.

Meshing these four approaches (a rather comprehensive set, as they include reviewed work themselves), reveals a strong emergent classification of evaluative criteria. Together these draw on structural, epistemological and participative aspects of entire IDR projects to articulate powerful sets of guiding questions. We have labelled these criteria sets (see Table 1) as Holistic, Social, Experience, Leadership And Effectiveness. The way these break down into particular guidelines at the five levels identified in the introduction, is specific to each of those levels. We indicate how that process might develop below in the two cases of research grants (Level 2) and institutional review (Level 5).

Practical guidelines for implementation: (Level 2) research grant proposals

All of the approaches to evaluating IDR in the surveyed literature make recommendations as to how they might be implemented, but as yet there have been no serious attempts to measure the effectiveness of different implementations. For example, the Academy of Medical Science’s *Team Science* report includes in its list of recommendations (recommendation 6 of the report):

Team science grant proposals need to be appraised holistically, as well as from the perspective of the relevant disciplines.

- Funders should review policies and processes for obtaining appropriate peer review and appraisal of team science grant applications, and make changes where necessary
- Funders should induct and train peer and panel reviewers, as well as grant managers, to meet this challenge

The composition of review panels, and the process through which they address IDR funding proposals, is clearly critical. Research and consultation on this matter have regularly identified the need to employ reviewers who have experience of “translation” between the languages of different disciplines,

Table 1 | Correspondences of criteria in four recent surveys of IDR evaluation

Lyall & King	3, 4, 8	9	11	12	15
Strang & McLeish	1	1, 6	9	2, 9	1, 8
Belcher	4, 3	2	2	4	1
Klein	3, 2	4, 5	4,5	5	7
Criteria Class	Holistic	Social	Experience	Leadership	Effectiveness

and of the nature of well-integrated and high-quality IDR (Lyall *et al.*, *op. cit.*). Ideally these reviewers should have led effective IDR projects themselves, and in any case conscientious training of panels with the holistic perspective of the research in mind is an essential recommendation (Academy of Medical Sciences, 2016). The seventh blind examiner of the elephant imagined by Repko and Szostak (2012) urges the selection of reviewers who can go beyond bringing their own single-disciplinary perspectives to bear, but can also play an integrative role to the process of review itself.

While single-discipline experts have an important place in IDR evaluation, their role ought to be supportive of those chosen for their ability to judge the critical “emergent” outcomes of the research. This is the single intervention most commonly reported as effective. Other examples of good practice in implementation have been cited in support of the core IDR evaluation criteria:

- Arranging for referees to communicate in the production of a single assessment of an IDR proposal, rather than producing individual assessments.
- Ensuring that proposers are able to address, in writing, comments by individual referees before a proposal is assessed or ranked by a panel.
- Including user-community or other non-academic reviewers on panels.
- Avoiding “2-stage” review processes for IDR, in which a single-disciplinary hurdle is placed before the integrative evaluation of the proposal.
- Communicating between referees and panels to ensure that they individually and corporately understand the process and criteria of IDR.
- Considering of the track records of the researchers, especially of the research leaders, for experience in IDR.
- Avoiding reliance on quantitative publication measures, such as citation rates and impact factors.
- Probing beyond the research proposals or programmes themselves to consider the support and development structures, such as centres and institutes, of the institution (s) in which the research will be pursued.

The last point is an important one: a supportive institutional context is more important in IDR than in single-discipline research, and a strong track record at this level is a good indication of likely success. But all of these implementations are needed to improve the quality of IDR evaluation for research funders contemplating grant-giving.

Practical guidelines for implementation: (Level 5) national evaluation exercises

There are some particular challenges at the highest level of IDR evaluation—that of national research assessment exercises such as the REF in the United Kingdom. The consultations in Durham (Strang and McLeish, 2015), and in preparation for the British Academy (2016) indicated that, in the minds of researchers, the disciplinary (or “unit of assessment”) structure of these exercises

currently pre-disposes a fragmented and disciplinary approach, disincentivising submission of IDR-generated outputs to the exercise. This in turn devalues the entire lifecycle of support for IDR within universities. From this starting point, remedial measures are driven towards framing IDR as a “superstructure” or super-addition to the disciplinary landscape, rather than as foundational to it. As we have shown, no matter how effective additional checks and systems are at this point, they are not well-suited to identify the transformational and emergent value of IDR.

Current measures, such as the ability to “flag” outputs as interdisciplinary; to cross-reference between panels; to have outputs reviewed by a different panel than the one to which the researcher’s unit of assessment is submitting; and allowing multiple submissions of interdisciplinary outputs to different panels, are reported to have made IDR more acceptable and raised its profile in such exercises. However, they have not overcome (in the United Kingdom at least) a disincentive to submit IDR outputs to the exercise, that emerged not only qualitatively in consultation, but quantitatively in a citation review (Elsevier, 2015). Similarly, other evaluation criteria such as research “environment” (which has the capacity to reward structural support of IDR) and especially “impact” (on which there is strong evidence that IDR comprises a major proportion of the supporting research) have not shifted the impression that core-disciplinary research will earn higher rewards.

More radical suggestions would respond to the foundational nature of IDR. The results of this work, through the emergent evaluator perspectives of Holistic, Social, Experience, Leadership And Effectiveness, suggest the following measures:

- Creating one or more explicitly interdisciplinary panels, either welcoming any IDR-flagged outputs, or focussing on integrative topics such as “energy”, “security”, “global policy”, “biophysical sciences”.
- Identifying and deploying a pool of panel members with strong interdisciplinary expertise and experience, either within a focussed IDR panel and/or with members on all subject panels.
- In the case of IDR with non-academic partners, create evaluatory structures that do not differentiate the categories of “output” and “impact” as strongly as at present, but combine them in ways that respond to the non-linear nature of IDR that involves partners external to the university.
- When evaluating the research environment of institutions, include an explicit and detailed examination of the incentives for building IDR, including the support for interdisciplinary communities and communication, the development of centres that preserve healthy reciprocal links to departments, and for career development of staff, including leadership of and mentoring in IDR.

Research evaluation exercises at national scale, apart from their direct results in terms of rankings and funding, shape and communicate at the highest volume the value structure of research. It is imperative to build strong evaluative messages into their criteria that support highly effective IDR.

Conclusion

The evaluation of IDR is central to a wider understanding and appreciation of its value. It illuminates the fundamental role that IDR can play in the acquisition of learning and understanding, rather than framing it as a super-additive or optional structure bolted on to disciplinary foundations.

The guideline question that asks about the benefits of IDR to the participating disciplines also points to the relevance of a more fundamental view of IDR. Participants in outstanding IDR research regularly comment on how its constructively disruptive context, and its broader view of shared research questions, can accelerate change, provide fresh perspectives and identify new and relevant data for their own disciplines. Callard and Fitzgerald (2015) counter the common claim that IDR is a “risky option” for a career with the notion that, in the twenty-first century, it is not particularly “safe” to remain within the confines of traditional disciplines either, in the face of rapidly changing academic opportunities.

Furthermore, all of the studies we have drawn on either explicitly or implicitly reflect a conception of IDR that identifies it as academically foundational. Invoking the common metaphor of territories, IDR does not build land-bridges across the borders between disciplines: it takes its participants into the underlying and deeper-dimensional spaces of learning that underlie, and support the current disciplinary structures of the academy. This is demonstrated by the effectiveness with which it is possible to apply evaluation criteria designed with IDR in mind to single-disciplinary research. Once a research evaluation framework is designed from scratch with a broad IDR in mind, and the question-sets we reviewed are generated, then it is immediately apparent that such a framework creates an excellent process for the evaluation of single-discipline research. In this sense, too, IDR lies at a fundamental level.

Reflecting the ultimate unity that this exercise has illuminated, we have seen that the same, emergent methodology for effective evaluations applies to each scale at which they are required, in suitably tailored form, from individual outputs to entire institutions. At each level there is a need to discern the underlying connections between disciplinary categories and structures.

The core question in any effective evaluation of IDR is the emergence of a new and integrated whole from the disciplinary ingredients. This holistic approach generates detailed frameworks that can be cast in the form of “checklists” evaluating a rich range of academic practices, including career development, continual learning, interdisciplinary translation, “disciplinary hospitality”, the formulation of shared research questions, the co-design of projects, the integration of epistemologies and data and the production of joint outputs. However, they have been articulated, the questions that constitute effective review of IDR examine its holistic and social dimensions, and explore the experience, leadership and effectiveness of its research communities.

From embarking upon a very practical task—the formulation of an evaluatory framework tuned to IDR—we have arrived, rather surprisingly, at a reappraisal of the shape of the academy itself. Anthropologist Marilyn Strathern employs an ancestral metaphor, suggesting that disciplines can differentiate themselves and “multiply their positions... precisely because they have common origins” (Strathern, 2008: 18). But if we return to our own (and Repko’s) original metaphor, we might say that a focus on interdisciplinarity revives a sense of the academy as a holistic intellectual and social organism, in which multiple flows and exchanges between all of its parts ensure its vitality. This suggests that IDR is a rather essential elephant to have in the room.

Notes

- 1 We recognize that in some academic/cultural contexts “transdisciplinary” is defined as being different from “interdisciplinary”, but in this context it is used interchangeably.
- 2 These include a major project on current practice in IDR running through 2015 and 2016 by the British Academy and another on opportunities under HEFCE (TCBM member of both working groups).
- 3 The REF is the 6-yearly evaluation of the research strength all disciplines in all higher education institutions in the United Kingdom, to which a central and significant stream of government research funding is linked.
- 4 The measure of public benefit of research defined broadly as evidenced impact outside academia, recorded and evaluated in the REF exercise.

References

- Academy of Medical Sciences. (2016) *Improving Recognition of Team Science Contributions in Biomedical Research Careers*. Academy of Medical Sciences: London.
- Barry A and Born G (eds) (2013) *Interdisciplinarity: Reconfigurations of the Social and Natural Sciences*. Routledge: London and New York.
- Belcher MB, Rasmussen KE, Kemshaw MR and Zornes DA (2015) Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*; 25, 1–17.
- British Academy. (forthcoming) *The British Academy Interdisciplinarity Report*. British Academy: London.
- Bruun H, Hukkinen J, Huutoniemi K and Klein J (2005) *Promoting Interdisciplinary Research: The Case of the Academy of Finland*. Publications of the Academy of Finland 8/05: Helsinki, Finland.
- Callard F and Fitzgerald D (2015) *Rethinking Interdisciplinarity Across the Social Sciences and Neurosciences*. Palgrave MacMillan: New York.
- Defila R and DiGiulio A (1999) Evaluating transdisciplinary research, *Panorama: Swiss National Science Foundation Newsletter*; 1 (1): 4–27.
- Elsevier. (2015) *A review of the UK’s interdisciplinary research using a citation-based approach*. Report to the UK HE funding bodies and MRC by Elsevier. Higher Education Funding Council: UK.
- Frodeman R, Klein JT and Mitcham C (eds) (2012) *The Oxford Handbook of Interdisciplinarity*. Oxford University Press: Oxford, New York.
- Gasper GEM, McLeish TCB and Smithson HE (2016) *Listening Between the Lines: Medieval and Modern Science*, this press.
- Giddens A (1991) *Self and Society in the Modern Age*. Stanford University Press: Stanford.
- Klein JT (2008) Evaluation of Interdisciplinary and Transdisciplinary Research—A Literature Review. *American Journal of Preventative Medicine*; 35 (2): S116–S123.
- Land R (2011) Crossing tribal boundaries: interdisciplinarity as a threshold concept In: Becher T, Trowler P, Bamber V and Saunders M (eds). *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*, 3rd edn, Society for Research into Higher Education and Open University Press: Buckingham, UK.
- Lyll C and King E (2013) International good practice in the peer review of interdisciplinary research. Report to the RCUK Research Group, October, www.tinyurl.com/idwiki.
- Lyll C, Bruce A, Marsden W and Meagher L (2013) The role of funding agencies in creating interdisciplinary knowledge. *Science and Public Policy*; 40 (1): 62–71.
- Marzano M, Carss D and Bell S (2006) Working to make interdisciplinarity work: investing in communication and interpersonal relationships. *Journal of Agricultural Economics*; 57 (2): 185–197.
- McLeish T and Strang V (2014) *Leading interdisciplinary research: Transforming the academic landscape*. Stimulus paper, The Leadership Foundation for Higher Education: Durham.
- National Academies of Science, USA. (2010) *Facilitating Interdisciplinary Research*. National Academies Press: Washington DC.
- Newell W (2001) A theory of interdisciplinary studies. *Issues in Integrative Studies*; 19, 2.
- Repko A and Szostak R (2012) *Interdisciplinary Research: Process and Theory*, 2nd edn. Sage: Los Angeles and London.
- Somerville M and Rapport D (eds) (2000) *Transdisciplinarity: Recreating Integrated Knowledge*. EOLSS: Oxford.
- Strang V (2009[2007]) Integrating the Social and Natural Sciences in Environmental Research: a discussion paper. *Journal of Environment, Development and Sustainability*; 11 (1): 1–18.
- Strang V and McLeish TCB (2015) *Evaluating Interdisciplinary Research: A Practical Guide*. Durham University Institute of Advanced Studies: Durham, UK.
- Strathern M (2008) Knowledge identities In: Barnett R and Di Napoli R (eds). *Changing Identities in Higher Education: Voicing Perspectives*. Routledge: London, New York, pp 12–21.

Additional information

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available at http://www.palgrave-journals.com/pal/authors/rights_and_permissions.html

How to cite this article: McLeish T and McLeish V (2016) Evaluating interdisciplinary research: the elephant in the peer-reviewers' room. *Palgrave Communications*. 2:16055 doi: 10.1057/palcomms.2016.55.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>