

Conference report

Quality assessment in interdisciplinary research and education

Veronica Boix Mansilla, Irwin Feller and Howard Gardner

Growth of interdisciplinary research and education is accompanied by uncertainty about how to evaluate interdisciplinary work. What constitute indicators of quality that distinguish the exemplary from the mundane? What research evaluation processes are most appropriate when disciplinary standards do not suffice? To advance the discussion of quality assessment, researchers from Harvard University and the American Association for the Advancement of Science convened a select group of leading research administration (eg National Institutes of Health, Duke University), science journal editors (eg Physics Review) and social scientists, to share innovative practices and empirical results. Core insights are here summarized.

Veronica Boix Mansilla is Principal Investigator at the Harvard Interdisciplinary Studies Project, Project Zero, 124 Mt Auburn St, 5th Floor, Cambridge MA 02138, USA; Tel: 617 496 6949; Fax: 617 496 9709; Project site: <www.pz.harvard.edu/interdisciplinary>. Irwin Feller is Senior Visiting Scientist at the American Association for the Advancement of Science; Mailing address: 1426 Ridge Master Drive, State College, PA 16803, USA; Tel: 814-237-6276; Fax: 814-237-0295; Email: iqf@ems.psu.edu. Howard Gardner is Hobbs Professor of Education and Cognition at the Harvard Graduate School of Education, USA.

The meeting was held at the American Association for the Advancement of Science Washington, DC, USA, on 8 February 2006.

BEING A MANTRA IN THE TREATMENT of contemporary knowledge production, ‘interdisciplinarity’ is ubiquitously invoked across federal funding agencies, journal editorial boards, university strategic plans and research centers as a sign of dynamism and creativity. Yet awareness of the vitality of interdisciplinarity is frequently accompanied by skepticism about the quality of research it yields and the profile of scholars it attracts. Central to the debates encountered in many of these settings is the ongoing problem of assessing interdisciplinary work.

For better or worse, evaluation and accountability are expected in every sphere of public sector activities. So the question for those who celebrate or fear the proliferation of interdisciplinary initiatives is not: should we evaluate? But rather, what are the most helpful and least damaging ways to do so? Evaluation parameters and processes developed within tight disciplinary communities are deemed insufficient to assess interdisciplinary work. Considerations of quality — how it is defined and measured, who is conducting the assessment, and in what context — constitute a formidable challenge to the institutionalization of recent interdisciplinary initiatives as well as to the launching of new ones. In the United States the challenge is heightened by increasing competition for federal funding and the spread of accountability requirements that emphasize quantitative outcomes.

To advance the discussion of quality assessment we convened a select group of research administrators, journal editors and social scientists with interest in, and knowledge of, the topic (see Appendix for list of participants). With support from the Atlantic Philanthropies, the meeting was held at the American

Association for the Advancement of Sciences in Washington, DC, on 8 February 2006. Two questions guided our deliberations:

1. What constitutes quality interdisciplinary work — that is, what are *qualities or indicators* that distinguish the exemplary from the mundane?
2. What *assessment processes* seem most appropriate to discern the quality of interdisciplinary work?

In preparation for the meeting, participants produced position papers outlining their current thinking. Federal agency officials and university administrators, such as Lawrence Tabak, and Peter Lange, shared innovative assessment practices in review panel composition and the nurturing of an institutional culture of quality interdisciplinary work. Judson King and Steven Hyman highlighted key challenges and new initiatives in quality assessment of interdisciplinary centers in the University of California system and at Harvard University respectively. Scholars, such as Michèle Lamont, and Lee Shulman, offered empirical findings on the construction of notions of ‘quality’ in grant review panels and across disciplinary and professional spheres. The perspectives of science policy, sociology, psychology, economics, and pedagogy were brought together to shed light on the assessment of interdisciplinary grant proposals, manuscripts for publication, research programs and educational outcomes.

While perspectives clearly reflected the diverse backgrounds of participants, there was reassuring convergence on a number of fundamental issues across the papers preceding the meeting and the unfolding onsite discussions. We recognized the need to move beyond what have now become less necessary definitional debates about interdisciplinarity. Thus, for the purpose of our discussion, we viewed ‘interdisciplinary research’ as a form of inquiry that integrates knowledge and modes of thinking from two or more disciplines or established fields of study to produce a cognitive or practical advancement (eg explain a phenomenon, create a product, develop a method, find a solution, raise a question) that would have been unlikely through single disciplinary means.

We also shared a sense of urgency in devising processes and criteria that are better suited for the assessment of interdisciplinary work than sole disciplinary parameters currently used by default. We recognized the tremendous variation in interdisciplinary work across domains of research (eg natural sciences, social sciences, humanities), institutions (eg research centers, universities), and purpose (eg advancing knowledge, curing disease, training scientists). Considerations of quality, we agreed, must respect such substantial differences while advancing a form of discourse about quality that supports meaningful dialogue across instantiations.

Our perspectives also differed. Science administrators were keen to propose and gather effective innovations in quality assessment processes. They put a premium on the identification of ‘good evaluators’ and ‘good procedures.’ If the right people were convened, the argument went, the characterization of quality would take care of itself. These participants called for empirical work that would compare the evaluation outcomes of more traditional review panels and those in which innovations had been introduced. Others agreed but put a premium on understanding the social and epistemic forces that govern peer review contexts. They called for further explanatory accounts of the mechanisms by which quality is construed in various domains. They proposed a closer study of the tacit expertise embodied in ‘good evaluators’ and called for an analysis of the qualities embodied in work deemed excellent in disciplinary research as well as in interdisciplinary work.

In what follows, we summarize the most important points and issues that emerged regarding adequate processes and criteria for assessing interdisciplinary work. Insights have been reviewed by meeting participants. Collectively, they capture the group’s effort productively to advance the state of the art in quality assessment of interdisciplinary work.

Getting the process right

‘Getting the process right’ was viewed by all participants as essential for the proper evaluation of interdisciplinary work. In Peter Lange’s words,

To the extent that interdisciplinarity is an experiment or a craft, there are really two kinds of risk. There is the risk that because you establish the process inappropriately, you will quash the good research ... but there’s also a risk that you establish a process in such a way that it really promotes work of questionable quality.

A central challenge in the evaluation process, meeting participants convened, is the selection of adequate review panelists and the effective management of their concurred expertise during the review.

Strategic breadth in panel membership

To perform their function, panels must be *broad* and embody multiple disciplinary perspectives. Yet such perspectives must also fit the specificities of the work assessed in sensible and effective ways. As Lawrence Tabak suggested, one needs to be aware of the point at which adding one more physicist to a panel yields diminishing returns. Identifying such a cadre of experts presents special difficulty in emerging interdisciplinary fields, where standards of excellence and communities of scholars are not yet

established. Two approaches were proposed for creating panels of strategic breadth.

1. *Creating fit and agile, technologically supported review groups* Nina Fedoroff proposed the use of smaller, carefully crafted ‘on-the-fly’ review groups that are brought together electronically or via videoconferencing. For example, three or four individuals whose various forms of expertise are tailored to the substance of a set of proposals can complement their individual reviews with a mediated discussion facilitated by the program officer. Applicants may be invited to address criticisms in real time to ensure reviewers understand the nature of the work fully before making their final judgments. While adding demands on program officers, the approach could overcome the limitations of traditional disciplinary or large multidisciplinary review panels by ensuring an appropriate fit between reviewers and the work to be assessed.
2. *Identifying experts and administrators who are close to the substance of the work* When the scholarly community for a given line of work is not discernible, individuals who know the problem space well can be identified by asking the applicants themselves to suggest reviewer names, and then taking such recommendations seriously. In addition, opinions may be solicited from the heads of the interdisciplinary units where the work is carried out. Peter Lange highlighted the importance of assuring interdisciplinary scholars that “their work will be appropriately evaluated by those truly able to do so”.

Interpreters

Including adequate experts in a multidisciplinary review panel is necessary but not sufficient. Capitalizing on expertise during the review demands that such expertise be carefully managed. Multidisciplinary discussions are counterintuitive — often resulting in misinterpretations or in individuals talking past one another. To address this problem Lawrence Tabak described the experimental inclusion of “interpreters” in panels at the National Institutes of Health Roadmap initiative. Interpreters are individuals able to bridge the epistemic gap among experts in multidisciplinary panels.

The strategy [at NIH] has been to not only assemble content experts but also to assemble what we like to call “interpreters,” people who do have that unique gift of being able to understand more than one of the disciplines. And the dynamic of the study section is quite interesting and informative. You literally have a collaboration among reviewers, with the interpreter playing as the intermediary to try and come up with some common sense of the value or not of the particular application.

Entrepreneurial leadership

To overcome the markedly conservative disposition of review panels that — as Steven Hyman put it — tend to “favor feasibility over significance”, a new kind of science leadership is needed. The group agreed with Nina Fedoroff that successful university officials and program officers function as science *novelty detectors* and *entrepreneurs*. These individuals are familiar with earlier work in a field. They are also sufficiently critical of mediocrity or stagnation that they can delineate which lines of work are novel in ideas, may lead to novel conclusions, or feature novel methods able to nurture areas of inquiry that will, in turn, advance science generatively. They are able to detect novelty in productive ways and are also willing to take risks.

What counts as ‘quality’?

The problem of defining quality

“So, how do we know an interdisciplinary approach is productive? [Is] there a set of questions we can ask peer reviewers and research participants, the answers to which will help us discriminate among more and less promising efforts?” asked Charlotte Kuh. Our discussion suggested that, broadly speaking, there is continuity between expectations about quality in disciplinary and interdisciplinary work. Whether interdisciplinary or not, we seek work of *relevance* — that is, work which can offer a solution to an important societal or theoretical problem. We seek results that can *impact* scholarship and teaching, proposing novel paradigms and as a result generating new lines of work in multiple fields. We seek works of *scientific merit*, in which conceptual frameworks, methods, and analyses are adequately developed to fit the aims of the project; potential problem areas are considered and alternative strategies spelled out. We value *original* work, where novel concepts, approaches, or methods are employed in innovative ways, perhaps challenging existing paradigms.

In well-established interdisciplinary fields (eg biochemistry, history of science), scholars have negotiated standards of quality over time and their views are, as Peter Lange suggested, “sufficiently shared to warrant thinking of them as a scholarly ‘community’”. In this sense such fields operate very much as “disciplines” holding more or less common beliefs about what counts as quality work. However, when interdisciplinary work in incipient fields brings together disciplinary perspectives in less precedented ways, constructs such as ‘scientific merit’ or ‘originality’ prove more problematic. Panelists must calibrate individual beliefs about the meaning of quality in more fundamental ways. In her empirical study of multidisciplinary panels in the social sciences and the humanities, Michèle Lamont found

important differences in the way such markers of quality were defined. For example, she noticed that while the literature in the sociology of science depicts 'originality' as the production of new findings and new theories, peer reviewers in social sciences and humanities panels construed originality more broadly as 'a new approach', 'a novel theory, method, or data', 'studying a new topic', or 'researching an understudied area.'

If the diversity of disciplinary perspectives called upon by interdisciplinary panels militates against simple definitions of quality, the sheer variety of fields, institutions, and goals of interdisciplinary research itself makes defining exacting universal metrics for quality even more elusive. Interdisciplinary research is used to cure disease as well as to understand fundamental natural laws. It takes place within and outside of academia. It is conducted by faculty and students alike. Not surprisingly, we concluded that there is no single quantifiable formula to measure quality in interdisciplinary research. However, understanding the unique demands of interdisciplinary research can illuminate particular aspects of the work as 'hot spots' worth special attention in quality assessment. Such a pragmatic focus was perhaps not surprising given the panel's composition and the meeting's espoused goal of productively "moving the discussion forward." In what follows, we outline four of the *hot spots* addressed during our deliberations:

1. Focusing on "the right shared problem";
2. Establishing social conditions for good work;
3. Meeting multiple disciplinary standards; and
4. Reaching effective syntheses.

Focusing on the right shared problem

Promising interdisciplinary collaborations devise effective forms of self-assembly around "the right shared problem" — a problem that clearly calls for complementary and integrative disciplinary perspectives. The group emphasized the importance of research and collaborative arrangements that are genuinely driven by a problem that admits no single disciplinary solution and is of intrinsic interest to all of the participants in a multidisciplinary group. Timing is important, argued Steven Hyman. "Things really have to be right. There have to be problems. There have to be tools." Several years ago, Hyman recalled, the National Institute of Mental Health called for proposals in the emerging field of social neuroscience, a move that was perhaps premature given the level of interest among many psychologists and neuroscientists:

Social psychologists were skeptical that the brain was really important. They were "brainless," so to speak. A number of neuroscientists on the other hand thought that social behavior was just too complicated a phenomenon.

They wanted to remain "mindless." It really took the development of tools. Some social psychologists began wandering into imaging laboratories. It actually took the family groups who were advocating for research on autism finally to create incentives for this field that was really ripe in terms of problems and tools.

"The key to success is the right shared problem," Hyman concluded, "which has to be audacious but still tractable and important. Of course there has to be money there, but it's [about] the problem." He pointed out that in the social sciences and the humanities, the sizeable centripetal problems that we see in the sciences (eg curing diabetes, understanding the human genome) have not yet clearly emerged.

Establishing social conditions for good work

Does the scientific environment in which the work will be done contribute to the probability of success? In assessing interdisciplinary proposals the conditions needed for effective collaborative research merit close attention. The group saw both the bottom-up emergence of a self-assembled research team and the top-down commitments of a leader with a track record of quality work as playing a central role in the potential success of the work. As Diana Rhoten suggested, leaders must be individuals who see themselves in a role akin to that of a film producer, able to channel team member's talents, rather than that of a director closely managing their actions.

Thus, assessing the conditions for good interdisciplinary work involves asking several questions:

- What are the complementary contributions of members of a team?
- What patterns of interaction and levels of physical proximity have been established?
- In the best-case scenario, is this work taking place within a larger institutional "culture of quality interdisciplinary work"?

Close analysis of past and present collaborations may enable reviewers to outline the likelihood of continuous collaboration beyond the funded period of a project. In this regard, social network analysis was suggested by Lawrence Tabak and Diana Rhoten

Leaders must be individuals who see themselves in a role akin to that of a film producer, able to channel team member's talents, rather than that of a director closely managing their actions

as a powerful tool to assess conditions for quality interdisciplinary work.

Meeting multiple disciplinary standards

A basic premise of quality interdisciplinary work, the group agreed, is that it satisfies quality standards arising from the disciplines involved. "Interdisciplinary work needs to speak to multiple disciplinary communities. In fact if interdisciplinary work is not recognizable as good research by a variety of peers perhaps it isn't good research after all," claimed Charlotte Kuh. "Good interdisciplinary research expands the disciplines but someone in the disciplines should resonate with what is being found." According to Martin Blume, interdisciplinary manuscripts submitted for *Physical Review* journals and *Reviews of Modern Physics*, tend to involve theoretical and experimental techniques used by physicists as applied to problems in other disciplines. Manuscripts are assessed by physicists who determine the correctness of the use of physical methods and by economists or biologists who comment on the significance of the problem and the proposed solutions in their discipline.

In assessing interdisciplinary research we must recognize the commitment of interdisciplinary work to more broadly defined problems of study and multiple disciplinary standards, if we are to avoid rejecting good interdisciplinary research because it is viewed as "overly ambitious and unfocused" or "lacking a hypothesis" warned Lawrence Tabak.

Judson King and Lee Shulman pointed out that professions are inherently interdisciplinary and differ from academic research. In the case of the professions, creativity is also exercised through *proficient activity* rather than publications alone, thus having to withstand standards of effectiveness and transformative power in the realities (medical, educational, legal, pastoral) they inform. Drawing on examples from engineering, nursing, and architecture, Lee Shulman described professionals as individuals who must know how various systems of knowledge interact. The point was illustrated by an engineer's self-description as someone who: "uses math and science to mess with the world by designing and making things people will buy and use. And once you mess with the world, you are responsible for what you have made."

Effective synthesis

Being able to wear multiple disciplinary 'hats' properly is necessary but not sufficient for the production of interdisciplinary work. The capacity to synthesize is key to advancing knowledge in ways that would have been unlikely through parallel disciplinary means. Successful syntheses advance researchers' understanding of their problems of study, yield comprehensive explanations, reveal productive dimensions, or grant novel solutions. Reviewers of

interdisciplinary research often find it difficult to articulate what makes a synthesis exemplary. In such case, the group convened, understanding how syntheses might go wrong may prove as illuminating as characterizing excellence. For example, grant proposals and research papers may succumb to the temptation of including too many disciplinary perspectives, surpassing a point of diminishing returns. In other cases a research design may bring disciplines together but do so in perfunctory ways that fail to advance the overall purpose of the work.

In sum, faced with the diversity of forms and contexts for interdisciplinary work, understanding whether or not that work is effective involves understanding the core intellectual problem addressed by potential grantees or manuscript authors and the unique socio-institutional and epistemic challenges that they confront in their efforts to bridge disciplinary traditions.

Conclusion

A discussion of the future of quality assessment of interdisciplinary work yielded overall recommendations. If we are to ensure the sustainability of quality interdisciplinary research, Julie Thompson Klein observed, we must institutionalize interdisciplinary work (and its evaluation), routinizing it in the cultural fabric of higher education and research institutions. Michael McPherson emphasized that if we are to advance interdisciplinary work of quality, we must embrace the multiplicity of forms and contexts for interdisciplinary work *without* side-stepping the problem of articulating workable quality criteria. "We need to develop the ability to be more articulate about what we mean by quality and, to this end, we have to think about the purpose of a work of research. 'We know it when we see it'," he claimed, "does not constitute an acceptable standard for evaluation." Others agreed.

We concluded the meeting with a call that further research be directed at designing and evaluating experimental approaches to quality assessment of interdisciplinary work. Most of the new initiatives in quality assessment, such as the inclusion of interpreters in peer review panels or science entrepreneurs in funding agencies, are indeed experiments to be examined closely for the fruits they might bear. Understanding the forms of expertise necessary for quality assessment of interdisciplinary work is essential in order to train future assessors and to educate current and future researchers. There is an enculturation process of "internalizing values, the set of habits of mind" that makes interdisciplinary work "feel natural and not unnatural" to individuals working between disciplinary boundaries, Lee Shulman explained. Without these values and habits, he argued, for "people [who] are trained in deeply disciplinary ways, interdisciplinary work becomes an unnatural act ... and difficult to sustain". In sum,

understanding interdisciplinary research as an emerging form of knowledge production demands that we take learning how to do interdisciplinary work seriously. It demands that we distill desirable

qualities of mind to be nurtured, whether the 'learners' be graduate students, members of an expert interdisciplinary team, review panel *interpreters*, or program officers.

Appendix: Meeting participants

Martin Blume
Editor-in-Chief
American Physical Society

Veronica Boix Mansilla
Principal Investigator, Interdisciplinary Studies Project
Harvard Graduate School of Education

Nina Fedoroff
Willaman Professor of Life Sciences
The Pennsylvania State University
Chair, Transformative Research Taskforce,
National Science Board

Irwin Feller
Senior Visiting Scientist
American Association for the Advancement of Science

Howard Gardner
Hobbs Professor of Education and Cognition
Harvard Graduate School of Education

Ray Handlan
Former President
Program Director
The Atlantic Philanthropies

Steven Hyman
Provost
Harvard University

C Judson King
Director, Center for Studies in Higher Education
University of California Berkeley
Former Provost University of California

Julie Thompson Klein
Professor of Humanities
Wayne State University

Charlotte Kuh
Deputy Executive Director, Policy and Global Affairs
National Academies of Science

Michèle Lamont
Professor of Sociology
Harvard University

Peter Lange
Provost
Duke University

Michael McPherson
President
Spencer Foundation

Diana Rhoten
Program Director
Social Science Research Council

Lee Shulman
President
Carnegie Foundation for the Advancement of Teaching

Lawrence Tabak
Director, National Institute of Dental and Craniofacial Research
National Institutes of Health

Albert H Teich
Director, Science and Policy Programs
American Association for the Advancement of Science

Acknowledgements

We would like to thank the Atlantic Philanthropies for supporting

this project and Ted Hullar for his creative contributions to its success. We are especially thankful for meeting participants for their invaluable contributions prior to and during the meeting. We thank Matthew Miller for his careful note-taking.

Copyright of Research Evaluation is the property of Beech Tree Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.