8

Evaluating Outcomes of Interdisciplinary Research and Teaching

Funding organizations and academic institutions need effective ways to evaluate the outcomes and effects of their investments in interdisciplinary research and teaching, just as they do for disciplines, to determine whether their goals are being achieved.

Appropriate evaluation is critical not only to assess the outputs of research but also to view more general outcomes in terms of organizational goals. For example, in evaluating a federal program to reduce unemployment, policy makers are less concerned about the *output*: how many people participate in the program, and more so by the *outcome*: how many participants obtain employment as a consequence of participating in the program within a particular period. The same is true of research: there is less interest in the number of publications than in the impact of the publications in terms of their quality, relevance, and stature.

THE CHALLENGE OF EVALUATING RESEARCH

As discussed in the Committee on Science, Engineering, and Public Policy (COSEPUP) report *Evaluating Federal Research Programs*,[1] the useful outcome of interdisciplinary or disciplinary basic research cannot be measured directly on an annual basis because of its inherent unpredictability.

[1]National Research Council. 1999. *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*. Washington, D.C.: National Academy Press.

But that does not mean that no meaningful measures exist. COSEPUP found that measures do exist: measures of *quality*, in terms of research advancement; *relevance*, in terms of application development; and *leadership*, in terms of the ability to take advantage of opportunities when they arise, as evaluated by experts and users of research. In addition, COSEPUP concluded that human-resource development is also a key outcome of an effective research program.

A remaining challenge is to determine what additional measures, if any, are needed to evaluate interdisciplinary research and teaching beyond those shown to be effective for disciplinary activities. Successful outcomes of an interdisciplinary research (IDR) program differ in several ways from those of a disciplinary program. First, a successful IDR program will have an impact on multiple fields or disciplines and produce results that feed back into and enhance disciplinary research. It will also create researchers and students with an expanded research vocabulary and abilities in more than one discipline and with an enhanced understanding of the interconnectedness inherent in complex problems.

"I ADMIRE THE INQUIRING MIND AND THE PRAGMATIC MIND, BUT I ALSO ADMIRE SOMEONE WHO CAN HIT."

The remainder of this chapter examines the challenges of evaluating interdisciplinary research and teaching and provides examples of innovative techniques (see Box 8-1). The report does not presume to prescribe specific evaluation measures; that is best done by each institution involved in IDR on the basis of its own objectives and culture. The examples cited here are intended to demonstrate approaches that may be useful.

TOOLKIT

BOX 8-1 Measures to Evaluate Interdisciplinary Work

In a recent study by the Harvard Interdisciplinary Studies Project, Veronica Boix Mansilla and Howard Gardner looked at research and teaching practices at several interdisciplinary institutes and programs over the last 2 years. They focused on the appropriate ways to evaluate IDR by interviewing over 60 researchers. They found that researchers typically were judged on indirect or field-based measures of quality, such as numbers of patents, publications, and citations; the prestige of universities, funding agencies, and journals; and approval of their peers.

Interdisciplinary research varies broadly in specific goals and validation criteria. Researchers at the cutting edge find that they also have to develop criteria with which to gauge their progress. On the basis of these interviews, Mansilla and Gardner suggest that measures of acceptability directly addressing the substance of interdisciplinary work be considered together:

1. The degree to which new interdisciplinary work relates to antecedent disciplinary knowledge. Even though engaged in interdisciplinary work, researchers still evaluated the credibility of new findings on the basis of consistency with the "disciplinary canon"—often in more than one field. High-quality understanding required more than a sum of disciplinary rules—it required a "unique coordination of disciplinary insights."

2. The sensible balance reached in weaving perspectives together. The interviewees appreciated interdisciplinary work that thoughtfully balanced perspectives of the disciplines represented, even though disciplinary standards could conflict with regard to worthwhile topics of inquiry or measures of proof.

3. The effectiveness with which a particular piece of work advances understanding and inquiry. Among interviewees, contributions oriented toward pragmatic problem solving and product development placed a premium on standards of viability. Algorithmic models of complex phenomena were associated with measures of simplicity and predictive power. Multidimensional phenomena were evaluated on the basis of comprehensiveness and empirical grounding.

---

[a]Mansilla, V.B. and Gardner, H. "Assessing Interdisciplinary Work at the Frontier: An empirical explanation of symptoms of quality." *http://www.interdisciplines.org/interdisciplinary/papers/6/2/printable/paper.*

EVALUATING RELATIVE TO THE DRIVERS OF IDR

Chapter 2 discusses the four driving forces of IDR:

- The inherent complexity of nature and society
- The drive to explore the interfaces of disciplines
- The need to solve societal problems
- The stimulus of generative technologies

One way to evaluate IDR is to consider it in light of those forces. For example, does a given program deal with the inherent complexity of nature and society that must be addressed by multiple disciplines? How well does the program do that? Each funding organization, depending on its own mission goals, would expect to use a different combination of drivers for its evaluation.

The drive to explore the *interfaces of disciplines* could be evaluated by examining the extent to which researchers truly collaborate with other researchers in adjacent or complementary fields or stimulate the development of a new field. Especially relevant to earlier COSEPUP reports is the driving force of the need to *solve societal problems*, which usually involve at least some applied research. One measurable outcome of research generated by societal problems would be a practical answer to the original question. For example, an IDR effort to reduce hunger could measure practical progress toward that goal. The same program, of course, might produce additional outcomes of value, including basic research, that were not anticipated.

The *stimulus of generative technologies* could be evaluated by examining the degree to which new technologies are developed that enhance research capabilities in many fields through the development of new instrumentation or informational analysis.

EVALUATING THE DIRECT AND INDIRECT IMPACTS OF IDR

Many of the standard means for evaluating disciplinary research and teaching can also be applied to interdisciplinary research and teaching: the use of metrics, such as number of publications, citations of publications, and successful research-grant proposals; teaching evaluations by students; benchmarking with other programs (when comparable programs exist); and national or international awards for and recognition of researchers or teachers. However, IDR can be expected to have measurable outcomes in multiple elements of technique, theory, and application. Taking account of that expectation will require new evaluation criteria that match the crosscutting nature of IDR. Both direct and indirect outcomes could be amenable to evaluation.

Evaluating IDR productivity can also be complicated because, although in some situations IDR may take more time than disciplinary research, it may have a high degree of depth and importance of achievement. The contribution achieved by a research team may be more than the sum of the individual accomplishments.

Direct Contributions of IDR to Knowledge

One way to evaluate IDR programs is to look for direct contributions in the form of new knowledge (see Box 8-2). Some IDR programs are so large that they stimulate new understanding in multiple fields. Examples are the Human Genome Project, the Manhattan Project, the broad effort to prove the theory of plate tectonics, global-climate modeling, and the development of fiber optic cable. A current example is the study of extremophiles—microorganisms that thrive under extreme chemical and physical conditions—as part of the emerging field of geomicrobiology. Their existence has influenced both biology and geology by expanding our notions about the origin of life on Earth (including the possibility of an extraterrestrial origin of life related to meteoric bombardment) and the limits of life on Earth (studied at deep-sea hydrothermal vents sustained by chemical synthesis). The existence of extremophiles has also altered traditional geochemical ideas about the formation and mediation of processes that lead to deposits of such ores as golds and sulfides.

Sometimes, the direct contribution of IDR is the creation of a new field or discipline as a result of the interactions between researchers who have a common interest. That was the case many years ago with biochemistry, and it is happening now in the formation of cognitive science, computational biology, nanoscience, and other fields (see Box 6-9 and Appendix D).

IDR may also add *value* to many traditional fields of research. For example, people studying nanoscience must bridge several disciplines seamlessly. Chemists are required for synthesis of nanostructures, materials scientists for characterization of structures, physicists for establishing new principles that relate quantum-like molecular states to new physical behavior on the nanoscale, and engineers for designing and building new devices and systems. At the same time, people use the richness of their nanoscience research experience to open up new *disciplinary* research directions and applications, such as the incorporation of nanostructures into bulk materials (see Box 6-6).

IDR may lead directly to the development of *new technologies* or *products*. Mathematical techniques developed for radiology now provide tools for oil companies to image the earth's upper crust. Researchers using prin-

INNOVATIVE PRACTICE

BOX 8-2 Evaluating IDR Center Proposals and Programs: The National Science Foundation Engineering Research Centers

The engineering research centers (ERCs) sponsored by the National Science Foundation (NSF) are systems-focused, interdisciplinary centers at universities all across the United States, each in close partnership with industry.[a] Primary goals of ERCs are to integrate engineering education and research, build competence in engineering practice, and produce engineering graduates with the depth and breadth of education needed for success in technological innovation and leadership of interdisciplinary teams.

NSF views ERCs as change agents for academic engineering programs and the engineering community at large. The mechanism of centers was chosen because centers can bring disciplines together. Since the ERC program was founded in 1985,[b] the ERCs collectively have brought substantial changes in the culture of academic engineering research and education.[c]

Proposal Review

ERC proposals are generated by program announcements.[d] Proposals are reviewed by a panel of peers selected by the NSF program manager. Panel members have scientific and technical expertise and experience in cross-disciplinary research, engineering education, industrial R&D, technology transfer, and research management. Panels recommend whether a prospective ERC should submit a full proposal. A second panel narrows the field and determines which sites to visit. The site-visit team consists of evaluators and two or three members of the panel. In 2002, there were 77 pre-proposals, 16 proposals, seven site visits, and four awards.

Program Review

An ERC begins operation with a 5-year award under a cooperative agreement with NSF. The agreement has the potential to extend NSF support to 10 years. After that, the ERC is expected to be self-sustaining. The progress and plans of each ERC are assessed annually through merit review by outside experts; review in the third year of operation can lead to extension of the cooperative agreement for 3 years to year 8, and a second review can take place in year 6. A period of phased-down support is provided to an ERC that is not renewed.

---

[a]Engineering Research Centers Association home page *http://www.erc-assoc.org/*; NSF, Division of Engineering Education and Centers, Engineering Research Center Program *http://www.eng.nsf.gov/eec/funding/pgm_display.cfm?pub_id=9971&div=eec.*

[b]The New Engineering Research Centers: Purposes, Goals, and Expectations. 1986. Washington D.C.: The National Academy Press.

[c]Suh, Nam P. The ERCs: What Have We Learned. *Engineering Education.* October 1987, p. 15-17. Engineering Research Centers Best Practices Manual *http://www.erc-assoc.org/manual/bp_index.htm.*

[d]Engineering Research Centers (ERC) Program Solicitation NSF 04-570 *http://www.nsf.gov/pubs/2004/nsf04570/nsf04570.htm.*

A team of peers similar to that constituted for proposal review conducts the third-year renewal review. Before the visit, the members of the site-visit team review the center's renewal proposal, using the ERC program performance-review criteria[e] available to all program participants. The first generation of ERC awards were evaluated on how well research needs were met for the industrial partner, 70 percent reported that participation in an ERC favorably affected their competitive position. Industry partners reported that the most important benefit was working with students.[f]

Renewal reviews are divided into six categories: systems vision and value added, strategic research plan, research program, education and educational outreach, industrial-practitioner collaboration and technology transfer, and strategic resource and management plan. Criteria in each category change in the three review periods—years 1-3, 4-6, and 7-10. For example:

Evaluation Category	Years 1-3		Years 4-6	Years 7-10
	High Quality	Low Quality	High Quality	High Quality
Research program	Thrust team is appropriately cross-disciplinary, projects are becoming interdependent within the thrust and contributing to other thrusts	Thrusts and projects are single-disciplinary and isolated from one another, project results have no role in other projects and thrusts	Thrust team is appropriately cross-disciplinary, projects are interdependent within the thrust and contributing to other thrusts	Thrust team is appropriately cross-disciplinary, projects are interdependent within the thrust and contributing to other thrusts
Education & educational outreach	Cross-disciplinary research culture is developing, students work in teams, ratio of graduate to undergraduate is at most 2:1	Little inter-dependence between faculty and students, involving few students in teams, ratio of graduate to undergraduate students is over 2:1	Cross-disciplinary research culture has been developed, students work in teams, ratio of graduate to undergraduate students is at most 2:1	Cross-disciplinary, team culture for students flourishes and impacts other parts of the university

continues

[e]Third-Year Renewal Review Process and Review Criteria for Engineering Research Centers in the ERC Class of 2000. Engineering Research Centers Program, Division of Engineering Education and Centers, National Science Foundation. August 2002.

[f]Lynn Preston, Deputy Division Director (Centers), Division of Engineering Education and Centers, National Science Foundation. Staff interview conducted November 17, 2003.

BOX 8-2 Continued

Outcomes

The 41 ERCs have made substantial contributions to US industry, are leaders in developing interdisciplinary cultures in academe, and produce a wide array of knowledge and technological advances. Innovations in research management, education, precollege outreach, and technology transfer are documented by NSF and the Engineering Research Centers Association.[g]

[g]National Science Foundation, Division of Engineering Education and Centers, Engineering Research Center Program Achievement Showcase *http://www.erc-assoc.org/showcase/index.htm.*

ciples of molecular biology, nanofluidics, materials science, and engineering have been able to address well-identified medical and clinical needs, and this has led to progress in developing artificial tissues and drug-delivery systems.

Indirect Contributions of IDR to Knowledge

Some contributions of IDR are less direct but substantial, and some institutions have begun attempts to evaluate them. For example, developing the engineering technologies necessary to achieve space flight has led to advances in the computer control of engineering processes, which have resulted in improvements in the reliability of industrial products and processes.

Information-Sharing Networks

Researchers who divide their time between traditional disciplinary departments and interdisciplinary programs or centers often form "networks of practice"[2] through which they share information that does not always appear in immediate or traditional forms, such as publications in academic journals (see Box 8-3). Such information-sharing networks may yield other important outputs, such as congressional testimony, public-policy initiatives, mass-media placements, alternative-journal publications, and long-term product development.[3]

[2]Brown, J. S. and Duguid, P. *The Social Life of Information*, Cambridge, MA: Harvard Business School Press, 2002.

[3]Rhoten, D. 2004. "Interdisciplinary Research: Trend or Transition." *Items and Issues* 5, no. (1-2):6-11.

EVOLUTION

BOX 8-3 Social Network Evaluation of IDR Centers

The Hybrid Vigor Institute[a] conducted a social network analysis[b] of six interdisciplinary research centers. The object was to model the structure, relations, and positions of the research network in each of the centers, assess the relationships of the researchers in a given center, and identify "hot spots" of interdisciplinary activity.

Data were collected in two phases. First, social networks were evaluated using field survey and bibliometric methods. The survey determined a person's professional background with regard to disciplinary and interdisciplinary exposure, relationship with every other person in the center, and the center's organizational practices and processes. Second, researchers visited sites to collect observational data and perform interviews. The data were compiled and analyzed with a social network analysis.[c] Social network analysis provides useful insights into how well researchers in an interdisciplinary center interact with one another, and it can determine critical personnel for fostering collaboration. However, it does not match performance results with interactions.

Hybrid Vigor found that center networks were shaped by the diversity of and functional distance between the disciplines. There was a greater rate of connectivity among researchers of different disciplines than like disciplines; this suggests that researchers do seek interdisciplinary connections in the centers. In fact, on the average, 84 percent of the current connections were formed after the researchers joined the interdisciplinary center. Regardless of group size, researchers did not tend to interact with more than 15 other researchers. Position in the network is affected by professional rank and status; center directors act as nodes. Graduate students and postdoctoral scholars were connected with more people outside their discipline than were senior faculty. That indicates that, although center directors may act as the organizing force, it is graduate students and postdoctoral scholars that weave the web.

---

[a]Hybrid Vigor Institute home page *http://www.hybridvigor.org/.*

[b]Rhoten, D. Final Report, National Science Foundation BCS-0129573: A Multi-Method Analysis of the Social and Technical Conditions for Interdisciplinary Collaboration. September 29, 2003. Available at: *http://www.hybridvigor.net/interdis/pubs/hv_pub_interdis-2003.09.29.pdf.*

[c]For a discussion of social network analysis, see National Research Council. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. 2002. Washington, D.C.: National Academy Press.

Quality of Educational Experience

One indirect impact of interdisciplinary programs is enrichment of the quality of undergraduate and graduate education. Interdisciplinary education programs have increased enrollments of undergraduate majors in IDR fields and enhanced non-majors' understanding of science and engineering.

- One example is an increase in earth-science enrollments at Stanford University that followed a shift in curriculum. After a substantial decline in the number of geology majors beginning in 1984, an interdisciplinary earth-systems degree was initiated in 1991-1992, and it led to a substantial increase in degrees awarded by the School of Earth Sciences (see Figure 8-1).
- Many universities have noted the popularity of science and engineering academic programs that are integrated with social-science issues. Two such programs are the Global Change Program of the University of Michigan[4] and the Program in Human Biology at Stanford University.[5]
- Experience at the University of Colorado at Boulder, a member institution of the National Center for Atmospheric Research, has shown that interdisciplinary programs that feature problems of greater breadth, societal relevance, or public policy (such as global change) are attracting more of the general student population to science courses. That is evidenced by the higher percentage of undergraduates who are taking more than the typical single required science course.

Enhancing an Institution's Reputation

Another indirect impact of interdisciplinary research efforts and curricula is enhancement of an institution's reputation by establishing programs of high quality in cutting-edge, niche fields. That, in turn, can strengthen an institution's ability to attract outstanding graduate students, faculty, and postdoctoral scholars, as happened at the Joint Institute for Neutron Sciences at the University of Tennessee and Oak Ridge National Laboratory,[6] the Keck Graduate Institute,[7] and the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara.[8]

Demonstrating the Value of Instrumentation

IDR may demonstrate the value of a major tool or instrumentation that has multiple applications. For example, synchrotron radiation, which provides an ultrabright photon source, has had a major impact on many fields

---

[4]School of Natural Resources and Environment, Global Change Program home page *http://www.snre.umich.edu/faculty-staff-directory/list.php?unit_id=35*.

[5]Program in Human Biology home page, *http://www.stanford.edu/dept/humbio/*.

[6]Magid, Lee. Comments at Convocation on Facilitating Interdisciplinary Research, Washington, D.C., January 29, 2004. *http://www7.nationalacademies.org/interdisciplinary/Convocation_Agenda.html*.

[7]Keck Institute home page *http://www.kgi.edu/index_flash.shtml*.

[8]Kavli Institute home page *http://www.itp.ucsb.edu/*.

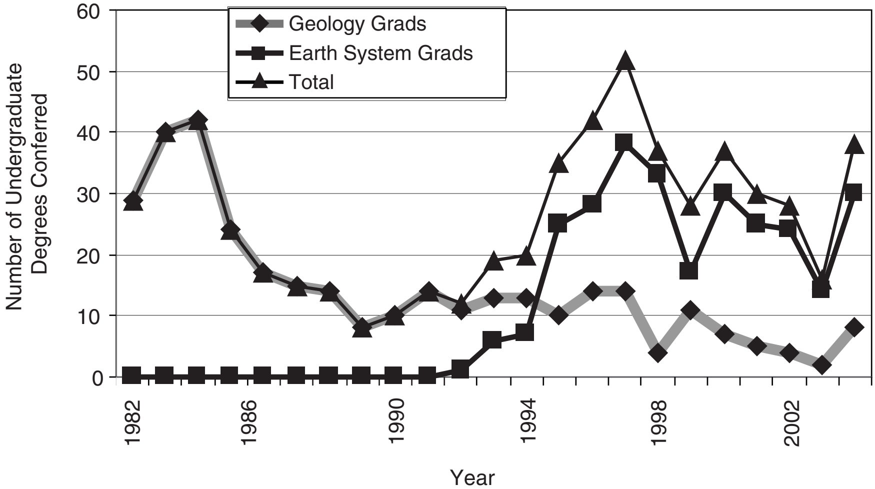

FIGURE 8-1 Degrees awarded by Stanford School of Earth Sciences.
NOTES: The undergraduate program at Stanford University, in an effort to offset a steady decline in student interest, created a new interdisciplinary earth sciences program. Student interest is reflected in the increased number of graduates declaring earth science as a major.

of research and on the development of industrial applications. It has also brought together multiple disciplines and groups of researchers, even though it was originally developed for the study of electronic and structural properties of materials. For example, in molecular biology, researchers use a synchrotron to obtain structures of proteins rapidly, and this enables pharmaceutical companies to develop new drugs (see Box 2-5).

EVALUATING THE PEOPLE WHO PERFORM IDR

Many organizations would like to develop more effective ways to evaluate students and faculty who engage in IDR (see Box 8-4). One approach to such evaluation is to measure the degree to which steps suggested in Chapters 4 and 5 are implemented. When applied, those steps should have detectable effects on the success of students and faculty.

Undergraduate and Graduate Students

Undergraduate and graduate students who work in more than one department might be expected to have experiences that they might not otherwise have; these experiences can provide starting points for evalua-

TOOLKIT

BOX 8-4 Evaluating the NSF Integrative Graduate Education and Research Trainee (IGERT) Program

The National Science Foundation (NSF) created the Integrative Graduate Education and Research Trainee (IGERT) program in 1997 in response to a growing recognition that graduate students in science and engineering needed to be better prepared for research that involves two or more disciplines. Since the first year of funding in 1998, NSF has added about 20 new programs each year. An institution that receives an IGERT grant currently receives up to $640,000 per year, the bulk of which is distributed as graduate student traineeships.

Cross-Site Program Evaluation

NSF has commissioned two cross-site reports to evaluate the impact of the IGERT program. The more recent focuses on the first two cohorts of IGERT projects in their third year of program implementation.[a] Data collection centered on interviews with students, faculty, and associated department chairs and university administrators. The key components evaluated were: project management, impact on students, impact on faculty, impact on institutions, and institutionalization.

Self Assessment

As described in the 2004 program solicitation, "IGERT projects are expected to incorporate and integrate . . . strategy and methodology for formative assessments of the project's effectiveness by individuals internal and external to the institution and program improvements based on these assessments." NSF used an

tion. The following examples may convey a sense of useful questions to ask of students in interdisciplinary programs:

- Are they working with and learning from students in other disciplines?
- Are they developing mastery of more than one discipline?
- Are they developing a sense of what it means to integrate more than one discipline in addressing a complex research question?
- Are they learning to use instrumentation or techniques that their own discipline might not provide?

Postdoctoral Scholars

In a similar spirit, one might ask many of the same questions of postdoctoral scholars who are involved in IDR:

external evaluator to appraise assessment methods of 79 IGERT programs that were initiated between 1998 and 2001.[b] A direct survey of principal investigators (PIs) found that 71 percent were responsible for program assessment. In about half the programs, a committee of IGERT participants performed an assessment, whereas less than one-third of the programs relied on an external person or group to evaluate the program.[c]

The methods of assessment reported are listed below in order of prevalence:

- Informal feedback to PI or committee responsible for evaluation (84 percent).
- Annual or more frequent surveys of trainees (68 percent).
- Annual or more frequent meetings of project faculty members serving as an assessment committee (58 percent).
- Annual or more frequent meetings of project participants to discuss project management and problems in program implementation or function (57 percent).
- Survey of faculty for concerns (33 percent).
- Interview of participants and observation of classes, seminars, and laboratories by external evaluator (30 percent).
- Continuing observation by external evaluator (14 percent).

---

[a]IGERT Annual Cross-Site Report: 1998 and 1999 Cohorts. Fall 2003. Prepared by Abt Associates, Inc for NSF.

[b]IGERT Implementation and Early Outcomes: 2002. June 2003. Prepared by Abt Associates, Inc. for NSF.

[c]Kusmierek, K. and Pionte, M. “Content, Consciousness, and Colleagues: Emerging Themes from a Program Evaluation of Graduate Student Progress Toward Multidisciplinary Science.” 42nd Annual Association of Institutional Research Forum. June 2002.

- Are they applying their own expertise in ways that add new value to a project and to their own grasp of one or more fields?
- Are they able to interact with specialists in other disciplines?
- Are they able to learn the language, content, and culture of another discipline?

Faculty

Faculty who work in more than one department or discipline would be expected to receive many of the benefits sought by students and postdoctoral scholars, such as extending the range of their understanding, working on exciting topics at the frontiers of their field, and learning new disciplinary languages and cultures.

Evaluating the work or contribution of faculty who are participating in

IDR is not simple, however. For example, how does an academic department evaluate the radiochemist who has carried out substantial portions of a positron-emission tomography study but will not be the study's senior author? Each institution or funding organization that supports IDR is encouraged to devise ways of answering such questions.

On a more general level, the following questions might help to frame an evaluation:

- Are faculty members doing interdisciplinary work of high quality and reporting on it in leading journals or conferences?
- Are they working on topics that they might not otherwise be able to address in their original discipline?
- Have they extended their expertise in new directions?
- Have they participated in establishing new subfields?
- Do they include students or faculty from other disciplines in their own research work?
- Are their students successfully merging disciplines?
- Do they take part in multidisciplinary advisory or review groups?
- Have they been evaluated at their own institution by a multidisciplinary review group?
- Have they achieved recognition, such as awards and lectureships, for IDR or from another professional society outside their own field?
- Have they been invited to present work in venues outside their discipline (an interdisciplinary mathematician, for example, invited to give a presentation to a biology department or at a biological professional society)?

It is reasonable to assume that a series of such questions will point to a framework for evaluating faculty who are engaged in IDR.

EVALUATING PROGRAMS, INSTITUTES, AND CENTERS THAT ENGAGE IN IDR

Many universities—motivated by the desire to organize work efficiently and to attract funds, students, and necessary infrastructure—have set up formal centers, programs, and institutes for IDR. Such structures are customary in industry and government, but their effectiveness in academe has not been thoroughly studied.

Indeed, the difficulty of developing effective review criteria is illustrated by a recent evaluation of NSF programs in IDR. Despite NSF's long-standing leadership, the evaluation urged the agency to "establish supplementary review criteria that will help to assess the quality of interdisciplinary effort in those programs where both single and multiple discipline proposals compete for a common pool of funds."[9] The report also stated that "no effective mechanism is in place to track or set performance goals for interdisciplinary research that can be used for planning, budget, and management decision-making."[10]

Convocation Quote

We need to do a better job of measuring results. . . . There is not a lack of data. There is a profusion of invisible data that needs to be better collected and disseminated.

Julie Thompson Klein, professor of humanities,
Wayne State University

At the heart of any evaluation process must be not only stringent peer review but also site visits that include personal interviews and objective observations. For most IDR programs, both internal and external reviews are essential to combine familiarity with institutional processes and objectivity of independent observation (see Box 8-5).

External review groups should represent all appropriate sectors; for example, in evaluating university centers, review groups should include the "users" of research outputs, such as industry, government, and policy representatives. To address the complexity of IDR, reviews should include mechanisms with two key qualities: depth of expertise in the core disciplines and related disciplines, and experience in carrying out IDR.

Recommendation of future directions for interdisciplinary centers should also include a "sunset" option. Initiatives will not be equally productive or equally long-lived. Reviewers should consider how much relevant new knowledge and understanding an IDR effort is generating and whether it should be terminated or moved in a new direction if the field itself changes.

For example, the discovery in the late 1980s of the fascinating C_{60} molecule, with its icosahedral symmetry, attracted many researchers to study it and its associated cage molecules, collectively called fullerenes. After 5-8 years, much of the basic fullerene research had been accomplished; research priorities moved elsewhere, and centers closed or evolved. A good example of the latter situation occurred at Rice University, where the fullerene center successfully changed to a nanotechnology center (see Box 6-6). Addressing the complex strategic questions involved in identifying new directions and finding sufficient support can be aided by advisory committees that have the interdisciplinary expertise mentioned above.

[9]NAPA, ibid., 2004, p. ix.
[10]Ibid, p. 96.

TOOLKIT

BOX 8-5 Assessment of Disciplinary and Interdisciplinary Research in the Netherlands

Assessment of scientific research at Dutch universities was started in the early 1980s. External review committees, consisting mainly of non-Dutch members, carry out evaluations on the basis of an evaluation protocol. The recently renewed assessment strategy was based partly on a model developed by the European Federation for Quality Management.[a]

Self-Assessment

Self-assessment reports are to be written every three years and reviewed externally every 6 years. The self-assessment reports should contain several elements, of which the main ones are:

- Characterization of the institute: mission, formal collaborations, and affiliations.
- Leadership: organizational structure, list of research programs, and program leaders.
- Research strategy: organizational context, plans for short and long term.
- Researchers: personnel policy—selection, training, career planning, and mobility.
- Resources and funding: financial situation, research contracts, future funding prospects.
- Processes to support research: teamwork, supervision of PhDs, quality assurance.
- Reputation: expressed in, for example, citation scores, prizes, and awards.
- Internal assessment: monitoring of research management.
- External appreciation: dissemination of research outcomes.

Finally, academic institutions have begun to implement interdisciplinary courses and minors, and even a few majors, but development of course and program evaluations is in its early stages. The beginning questions to ask in framing an evaluation mechanism might include the following:

- Is interdisciplinary teaching attracting more of the general student population to science courses?
- Are interdisciplinary courses and programs attracting a new or different mix of students to careers in science?
- Are interdisciplinary courses effective vehicles for instilling science literacy and awareness of the roles of science and technology in modern life?

- Research outcomes: publications in refereed journals, and so on; patents.
- Future perspectives.

External Review

In the external assessment, four aspects are to be considered: the quality of the scientific research, the productivity of the scientific output (such as refereed publications), the relevance of the research for academe and society, and the future perspective, feasibility, and vitality of the research. In addition, it has proved useful to assess the research according to a five-point qualitative scale with scores, which are given for all four aspects separately:

5. *Excellent:* research that is internationally at the forefront and has a high impact.
4. *Very good:* research that is internationally competitive and nationally at the forefront.
3. *Good:* research that is nationally competitive and internationally visible.
2. *Satisfactory:* research that is solid but not exciting; nationally visible.
1. *Unsatisfactory:* research that is not solid or exciting; not worth pursuing.

Interdisciplinary Themes

For large interdisciplinary themes, the contributions by the different disciplines can be rated separately, for example, in written reports from experts in specific disciplines. The reports are offered to the assessment committee for final assessment. In a typical field, such as the biomedical and health sciences, in which multiple disciplines contribute (for example, physics, chemistry, biology, informatics, clinical medicine, and epidemiology), all research themes were assessed by a multidisciplinary committee, and the contributions of the different disciplines were assessed beforehand in writing by experts in those disciplines.

[a] *http://www.pgmm.org/efqm.htm.*

- Are students demonstrating a grasp of the complex interconnectedness of real-world problems?

COMPARATIVE EVALUATIONS AND RANKINGS OF RESEARCH INSTITUTIONS

Comparative evaluations of research institutions, such as the National Academies' assessment of doctoral programs (see Box 5-1) and similar activities that rank university departments,[11] should include the contributions of both interdisciplinary activities and single-discipline contributions.

[11] For example, *U.S. News and World Report* issues annual department and program rankings.

For example, organizations evaluating such institutions can experiment with new forms of "matrix evaluation" to capture the activities and accomplishments of interdisciplinary researchers. A matrix approach would consider IDR as an integral part of the disciplines in which the IDR is "embedded" and make visible the cross-departmental efforts of people who form interdisciplinary teams.

A matrix-based evaluation might include in its criteria the comentoring of doctoral students, the contributions of people to multiple departments, and some publication criteria. Among their publication criteria might be the nature of the journal audiences for whom the work is published; citation analysis that reveals a broad interdisciplinary interest in the work being cited; double counting of publications, by which credit for a given paper is awarded to all coauthors; multiple authorship and coauthorship patterns that would reveal the disciplinary backgrounds of coauthors; and other measures that are still being developed.

CONCLUSIONS

Better methods to evaluate IDR are needed to help funding organizations to assess the results of their investments better, to help sustain America's preeminence in higher education and research, and to enhance the contribution of IDR to the general advancement of science and engineering.

There has been little systematic study of the people, institutions, or funding organizations taking part in interdisciplinary activities. A few studies have begun, including the study by the Transdisciplinary Tobacco Use Research Centers (see Box 8-6), and their results will begin to add much-needed information to this nascent skill.

Despite the complexities of evaluating activities that span multiple disciplines, long experience in peer review and other assessment methods suggest that useful assessment techniques can be developed for IDR. This chapter has attempted to outline some of the topics to be studied and questions to be asked in constructing frameworks for evaluation. Given the inherent difficulty and expense of most interdisciplinary activities and the need to balance investments in research, it is essential to measure and maintain its value to the research enterprise.

FINDING

Reliable methods for prospective and retrospective evaluation of interdisciplinary research and education programs will require modification of the peer-review process to include researchers with interdisciplinary expertise in addition to researchers with expertise in the relevant disciplines.

TOOLKIT

BOX 8-6 Determining How to Assess a Program: The Case of the Transdisciplinary Tobacco Use Research Centers

The Transdisciplinary Tobacco Use Research Centers (TTURCs) were created in 1999 with funding from the National Cancer Institute, the National Institute on Drug Abuse, and the Robert Wood Johnson Foundation.[a] As with the NSF IGERT program (see Box 8-4), the funding agencies required that centers include a core program on evaluation. TTURC evaluation researchers have developed the following outcome metrics for measuring and evaluating science that bridges two or more disciplines:[b]

- How well is the *collaborative transdisciplinary work* of the centers (including *training*) accomplished?
- Does the collaborative transdisciplinary research of the centers lead to the development of new *or improved research methods* and/or *new or improved scientific models and theories*?
- Does research result in scientific publications that are *recognized* as high quality?
- Does research get *communicated* effectively?
- Are models and methods translated into *improved interventions*?
- Does research influence *health practices, health policy, or health outcomes*?

The evaluation focuses on the program as a whole and not necessarily on the individual research centers. To answer the questions, researchers analyze annual progress reports and the federal financial report and conduct a survey of each of the researchers involved. That is accomplished through survey analysis, content analysis of the progress report, peer evaluation of the progress report, bibliometric studies, peer evaluation of the publications, personnel analysis, and financial analysis. To date, data indicate progress toward intellectual integration within and between centers and changes in collaboration behaviors, and they highlight how pathways to integration are affected by environmental, organizational, and institutional factors.

---

[a] The Web Center for Social Research Methods. *http://www.socialresearchmethods.net/*. Accessed June 11, 2004.

[b] Stokols, D., et. al. "Evaluating Transdisciplinary Science," *Nicotine and Tobacco Research*. (2003) 5:1-19.

RECOMMENDATIONS

E-1: IDR programs and projects should be evaluated in such a way that there is an appropriate balance between criteria characteristic of IDR, such as contributions to creation of an emerging field and whether they lead to practical answers to societal questions, and traditional disciplinary criteria.

For example, organizations that review IDR can measure

- The degree to which IDR contributes to the creation of an emerging field or discipline; emerging fields have included nanoscience and nanotechnology and cognitive science.
- How well IDR enhances the training of students and the careers of researchers in ways that surpass the results expected from disciplinary research; these might include employment in a broader array of positions, more rapid progress in gaining tenure and other goals, and greater numbers of speaking invitations.
- Whether the research leads to practical answers to societal questions; for example, an IDR effort to reduce hunger should produce some measurable progress toward that goal. The same IDR program might produce additional outcomes of value, including basic research, that were not expected.
- Whether participants demonstrate an expanded research vocabulary and abilities to work in more than one discipline.
- The extent to which IDR activities, institutes, or centers enhance the reputation of the host institutions; reputation can be measured in research funding, external recognition of IDR leadership, awards, and recognition of participants in the research.
- The long-term productivity of a program; not all initiatives will have the same lifetime, and the use of "sunset" provisions should be considered in the planning of IDR centers and programs.
- Multiple measures of research success, as appropriate to the fields being evaluated, such as conference presentations or patents in addition to publication in peer-reviewed journals.

E-2: Interdisciplinary education and training programs should be evaluated according to criteria specifically relevant to interdisciplinary activities, such as number and mix of general student population participation and knowledge acquisition, in addition to the usual requirements of excellence in content and presentation.

For example, organizations reviewing interdisciplinary education and training programs can begin with such criteria as the following, to be supplemented with others appropriate to the organizations' missions:

- Are interdisciplinary courses attracting more of the general student population to science and engineering courses?
- Are interdisciplinary courses and programs attracting a new or different mix of students to careers in science and engineering?
- Are interdisciplinary courses effective in instilling scientific and technologic literacy and awareness of the roles of science and technology in modern life?

E-3: Funding organizations should enhance their proposal-review mechanisms so as to ensure appropriate breadth and depth of expertise in the review of proposals for interdisciplinary research, education, and training activities.

For example, organizations that fund IDR could

- Involve researchers who have experience with and are knowledgeable about interdisciplinarity and ensure representation of the most important disciplinary points of view on panels that review IDR proposals.
- Evaluate a proposal to its cell-biology research program by using researchers in cell biology and including a substantial number in chemistry, physics, computer science, the social sciences, and the humanities as appropriate; this practice would help to ensure disciplinary breadth and reduce bias.
- Review a proposed interdisciplinary program in climate change by using input not only from experts in climate change and related fields—such as oceanography, meteorology, atmospheric chemistry, and land use—but also experts in the constituent disciplines—such as physics, chemistry, and statistics—and nonresearchers for whom the research is relevant; the contributions of different disciplines might be submitted separately in written form, and these reports would be offered to a full review panel, including both disciplinary and interdisciplinary researchers, for final assessment.

E-4: Comparative evaluations of research institutions, such as the National Academies' assessment of doctoral programs and activities that rank university departments, should include the contributions of interdisciplinary activities that involve more than one department (even if it involves double-counting), as well as single-department contributions.

For example, organizations that evaluate such institutions can

- Survey emerging interdisciplinary fields to identify demographic information (e.g. numbers and characteristics of participants in various interdisciplinary fields, and/or the the kinds of activities in which they are engaged).
- Experiment with "matrix evaluation" to capture the activities and

accomplishments of interdisciplinary researchers; a matrix approach is one that would consider IDR as an integral part of the disciplines in which the researchers are "embedded" and make visible the cross-departmental efforts of the researchers who make up the interdisciplinary teams.

- Include as evaluation criteria the comentoring of doctoral students, the contributions of individuals to multiple departments, and publication criteria. The publication criteria might include the nature of the journal audiences for whom the work is published; citation analysis that reveals a broad interdisciplinary interest in the work being cited; "double counting" of publications, by which credit for a given paper is awarded to all coauthors; multiple-authorship patterns that would reveal the disciplinary backgrounds of coauthors; and others that are still being developed.
- Include the facilitation of interdisciplinarity as part of the accreditation process.