BUILDING PROGRAM ASSESSMENT AND INSTITUTIONAL ACCOUNTABILITY IN NOT-FOR-PROFIT CENTERS FOR ADVANCED RESEARCH

SAR

SCHOOL FOR ADVANCED RESEARCH
SEPTEMBER 27, 2011

INTRODUCTION[1]

Over the past decade, issues of institutional accountability and program assessment have permeated higher education and the not-for-profit world, driven largely by heightened government scrutiny and increasing calls for accountability among philanthropies in the face of several notorious examples of failed oversight in the for-profit and not-for-profit worlds. This has occurred in the context of decreased public funding for higher education, rapidly growing accountability requirements from philanthropic sources, and a maturing "Baby Boomer" population with considerable assets accompanied by a desire to maximize social impact. Commentators on the not-for profit sector, both in print and in numerous blogs[2], debate how to measure social impact and guide social investing through "outcome-based grant-making" and "strategic philanthropy," while Congress considers the efficacy of venerable legislation that regulates not-for-profit organizations.

Centers for advanced research that are fully or largely independent of universities have generally avoided much of this scrutiny, perhaps because they represent a tiny sector of the not-for-profit world. Many centers, however, have felt the pressure to develop internal accountability policies and procedures due to requirements placed upon them by funders. Some have responded with token statements of program evaluation and financial accountability procedures, while a few have seen the new demands as an opportunity to gain useful information about how well their organizations are working.

In 2008, the School for Advanced Research (SAR) in Santa Fe, NM began to develop procedures for program evaluation, and in 2009, with support from the Dobkin Family Foundation, assembled the like-minded leadership of several other centers for advanced research to begin discussions of program assessment and institutional accountability, ranging from the sharing of best practices to the exploration of how to develop a common set of evaluation metrics. Subsequent funding from the William and Flora Hewlett Foundation facilitated a follow-up workshop in 2010, as well as additional program evaluation work at SAR in 2011.

This report summarizes the efforts of SAR and the other centers to explore issues of accountability and assessment for independent, not-for-profit centers for advanced research. It includes a history and discussion of accountability challenges, a summary of the workshops held in 2009 and 2010 and their outcomes, and a description of the specific assessment work of SAR from 2008 to 2011.

---

[1] This report was assembled and edited by Dr. John Kantner, Vice President for Academic & Institutional Advancement at SAR. The Accountability Challenges section was researched and written by Dr. Glen W. Davidson, Professor Emeritus of Psychiatry and Medical Humanities at Southern Illinois University, Senior Consultant/Evaluator for the Higher Learning Commission, and member of the SAR Board of Managers. Drs. Kantner and Davidson contributed to the summaries of the accountability workshops held at SAR and Dumbarton Oaks. Dr. Kantner conducted the research and wrote the summaries on SAR's program evaluation initiatives.
[2] For example, *The Chronicle of Philanthropy* regularly features articles on the topic, including in its daily email newsletter, while blogs such as *Tactical Philanthropy* discuss the merits of social investment, including contributions by professional consultants who manage philanthropic giving not unlike stock portfolios. See also Brest *et al.* 2009, Gertner 2008, Kramer 2007, Kramer *et al.* 2009, Tuan 2008.

ACCOUNTABILITY CHALLENGES FOR CENTERS FOR ADVANCED RESEARCH

When discussing what actions the Board and officers of the School for Advanced Research should be taking to strengthen our strategic plan, one board member asked, seemingly in frustration, "Why are we being hit with all of these demands for accountability?" Her question is timely. Accountability is the "buzz" word of the day. How her question is answered will make a difference between regressive efforts of compliance or opportunities to become stronger in our missions as centers for advanced research.

Like all buzz words, "accountability" has taken on multiple nuances. Here, the word is used to convey the various means by which personnel of an organization document claims of transactions, particularly accomplishments, for those who have oversight of the organization's mission and operations. The argument is that accountability methods and measures, while basic for building trust with various stakeholders, are crucial for attaining and maintaining quality.

Schools, centers, and institutes for advanced research ("centers") have yet to be publicly exposed to legislative and media issues of accountability, transparency, or accreditation in the ways that certificate-issuing, degree-granting, and status-conferring organizations are being challenged. While some business officers may disagree, the centers have been held accountable to the most benign of requirements external to the institutions' own bylaws only having to meet tax laws and regulations of 501(c)3 organizations, and then only since 1960.

In fact, most of the centers have been established, in part, to avoid the accountability measures required by their often-associated agencies, colleges, or universities. Like their associated organizations, however, evidence is emerging that the residual deference accorded to such centers and their leaders by policymakers and the general public has eroded significantly in recent years. As Peter Ewell argues:

> A parallel loss of respect is occurring for all of the professions, most notably medicine and the law. Meanwhile, the Enron scandal and the subsequent requirements of Sarbanes-Oxley in the corporate community symbolize a similar loss of trust in the nation's business enterprises. This trend signals a growing public perception—especially strong among elected officials at this point—that the traditional assumption of collective responsibility for quality by the members of any profession provides no guarantee that they will act with integrity.[3]

The demands for accountability, however, are not new[4]. While such structures as ordination for clergy, licensure for physicians, and admission to the Bar for lawyers extend back as early as the

---

[3] Peter Ewell. *U.S. Accreditation and The Future of Quality Assurance,* a report he prepared for The Council for Higher Education Accreditation (CHEA), p. 5. Ewell notes that Dan Yankelovich calls this "Wave III" of public mistrust in social institutions, following "Wave I," which emerged following the Great Depression, and "Wave II," which emerged following Watergate and the social unrest of the 1960s. [Yankelovich, Dan (2006). "Accreditation in a More Demanding World." Presentation at the 2006 CHEA Annual Conference.] Were Yankelovich a more comprehensive historian, he would have identified the "waves" of the 19th century as well, issues that remain largely unresolved, with "trust" being the most basic.

[4] They're not even new for research organizations, although previous discussions were largely academic in nature compared with the public discussions taking place now. See Altschuld and Zheng 1995.

16th century in Europe, and back as far as 12th-century China when bureaucrats had to "stand" for examinations, most accountability criteria we know have their origins in 19th-century issues about what can be trusted in practices of learning.

Following the Civil War, when there was a proliferation of "colleges" and "truth claims," a select group of educators began to struggle with definitions such as "What is a college?" When the first accreditation agency, the New England Association of Schools and Colleges, was organized in 1885, issues focused on a series of core questions—Who can be an instructor at the "advanced" level? Who can be admitted as a student? Which institution's degrees can be recognized as demonstrating "higher" education? Should students be allowed to transfer between institutions and have previous courses recognized?—all of these questions answerable based on the definition of a college. Interestingly, challenged by today's global markets of academic credentialing, the Lumina Foundation for Education has commissioned a study with three prominent American universities to pilot a project to identify what specific knowledge and skills must be taught at the college level. The project is an attempt to meet expectations of the Bologna Process whereby colleges in dozens of countries can award degrees based on comparable standards.

While some of the 19th-century issues that dominated education are still with us, "the feel" of accountability, in Ewell's words, of "self-study verified by peer review on a periodic schedule leading to formal recognition of 'accredited' status by an established commission" didn't take shape until the years immediately following World War II.[5] Today, approximately 6,800 postsecondary institutions and more than 9,000 academic programs are accredited by regional or specialty organizations in the United States. Centers for advanced research that are part of universities know well the accreditation drill even though they do not generate degrees. Most non-university affiliated centers are unaccredited and lack the kind of accountability measures that public agencies are accustomed to reviewing, and, increasingly, philanthropists and foundations are demanding.

Whatever the debated issues in the history of higher education and higher learning in the United States and Canada, trust and mistrust are the age-old and lingering reasons that any organization in the "knowledge business" is beginning to be hit with all of the demands for accountability facing other not-for-profit organizations in the United States. Organizations in other parts of the globe are also being challenged to demonstrate the will and ability to maintain, in Ewell's words, "credible self-regulatory responsibility for quality."[6] The central challenge facing all centers for advanced research, then, is whether, in partnership with our funding partners, we can establish accountability criteria and measures that reflect our unique missions and avoid destructive and resource-consuming efforts of imposed standards whether created through legislative or regulatory fiat. Further, we are challenged to construct means by which we use accountability efforts for our betterment.

---

[5] Ewell 2008, p. 18.
[6] *Ibid.*, p. 5.

Historical Background

Following World War II, W. Edwards Deming redefined 19th-century assumptions of institutional "trust" into questions of "quality." He is widely credited with being the father of the quality movement, first in Japan, then in emerging industrial economies of Asia, and finally in the United States. A statistician, Deming argued that American industries' failure to use quantitative measures to analyze production led to shoddy production, out-of-control expenses, unreliable assumptions for making production decisions, and incompetent management. Further, he argued that accountability measures had to be conducted in the context of systems—that no change in one part of the production process was without consequences in every other part of the production process, a phenomenon, he argued, most CEOs of the time failed to notice.

Deming's influence can be traced back as early as the 1940 U.S. Census, when he introduced sampling techniques for the first time as one means for determining consumers' expectations for quality, or, how an organization's stakeholders define trust. Deming argued that a system cannot understand itself. Transformation requires an independent view from outside the organization, otherwise, vested and self-interests rule rather than objectivity. Data for making decisions, and for determining (trusting) reliability, must be comparative and reflect trends, in Deming's assessment.

By the time Deming came along, accreditation in higher education had already adopted an industrial model. Assigning credit to courses, particularly for meeting degree requirements and transferability between accredited colleges and universities, was based on industrial models developed to improve production goals during World War II. "Credits" based on production metrics were assigned to every course and other learning events. Accountability was based on the availability of adequate resources for a college, such as favorable student-faculty ratios, sufficient number of books in the library for each major, and functioning facilities that set an environment for learning.

Deming and his disciples made "continuous quality improvement" an expectation for higher education. His idea that "trust" and "quality" were intertwined could be tested by market survey techniques. "Long-term commitment to new learning and new philosophy is required of any management that seeks transformation," he argued. "The timid and the fainthearted, and the people that expect quick results, are doomed to disappointment." But, he cautioned, "when information is obtained, or data is measured, the method, or process used to gather information, affects the results." For example, to base judgments on customer complaints alone ignored the general population of other opinions, or a statistical sample of the whole. Change the method and you change the results, Deming argued. "Aim and method are essential. An aim without a method is useless. A method without an aim is dangerous. It leads to action without direction and without constancy of purpose."[7]

In the present idiom, an organization's mission and the methods by which we test to see if we are fulfilling the mission must be in alignment. An impact of Deming's thought was university

---

[7] Quoted in Wikipedia, "W. Edwards Deming." Also see Demings' major publications: *Out of the Crisis* (1982, 1986) and *The New Economics for Industry, Government, Education* (1993). Also see W. Edwards Deming Institute, webpage: http://deming.org/index.cfm?content=653, accessed 2008-10-15.

trustees increasingly turning to men (and a few women) with experience in business to lead colleges and universities. By the 1990s, their mantra had become, "we can trust institutions which produce quality products (students, publications, patents, etc.)."

Deming's ideas gained great traction when, in the 1970s and 1980s, Americans began to favor Asian over American-produced goods. Malcolm Baldrige, Secretary of Commerce in the Reagan Administration, began to urge American business executives to adopt Deming's "quality principles." Baldrige's premature death prodded the Congress to pass The Malcolm Baldrige National Quality Improvement Act of 1987. Now 40 states have "continuous quality improvement" structures.[8] "Strategic planning for quality and quality-improvement programs, through a commitment to excellence in manufacturing and services, are becoming more and more essential to the well-being of our Nation's economy and our ability to compete effectively in the global marketplace," the Act reads. The Act focused on manufacturing, service, and small business. In 1999, education and health care were added as separate foci for quality improvement. In 2007, nonprofit organizations became an additional focus. "Trust" and "quality" now became identified as necessary components of the health of the nation. The private-public coordination of continuous quality improvement have led to expectations of accountability more transparent and more sophisticated than ever before.

In order to continuously improve the "product"—and for centers of advanced learning it is results packaged for dissemination—reliable systems are necessary. By 1990, systems theory became the context for accountability, and learning was by objective. More and more, legislative mandates and private sector organizations such as philanthropic foundations envision their expectations for accountability from the Baldrige model. In nursing education, and by extension in clinical practice and research, the "Magnet journey" is an example. The American Association of Medical Colleges has quality criteria for medical education. In regional higher education accreditation, the Academic Quality Improvement Program of the North Central Association is another. These, and other accountability models, apply the principles of Deming and the practices of Baldrige to specific professions. "Such practices are the norm for any organization/institution today," explained a senior member of the Senate Finance Committee staff. "For organizations, particularly not-for-profits, not to have such practices in place needs to be explained, if that is possible! Even if it were possible, it would open the organization to suspicion."[9]

---

[8] In Illinois, the federal and state recognized organization is The Lincoln Foundation for Performance Excellence. New Mexico's is Quality New Mexico.

[9] Interviews: November 14, 2008 and February 10, 2009. The interviewee is a senior staff member of The Senate Finance Committee but must remain anonymous because he is not authorized to speak either for Senators Baucus or Grassley, or on behalf of the Committee. Review of numerous online bulletins of the Senate Finance Committee was also conducted.

Analyses of the Challenges

One of the more credible answers to the question with which this section began is that perceptions are highly significant in setting people's expectations and is the foundation of their trust. A survey by the Ellison Research organization, completed in 2008, found that "most Americans believe non-profits are not financially efficient":

> We've spoken with tens of thousands of donors over the years, and one thing that is consistent is that most people really don't know much about how non-profits operate…But even when people are misinformed, their perceptions still influence how they make giving decisions. That's why it's so important to understand how people perceive charities in general, as well as why individual non-profits really need to learn how their own donors or potential donors see them…It's almost as if organizations are automatically under suspicion until they prove themselves innocent…Only 28 percent of Americans believe the typical charitable organization is spending a reasonable proportion on overhead and only 10 percent believe non-profits typically spend less on overhead than what they would consider to be a reasonable standard.[10]

The study noted that if people do not know much about a specific organization that they might consider supporting, their perception of the charity will be based on how they see charities in general.

Senator Charles E. Grassley (R-Iowa), then chair of the Senate Finance Committee, began hearings in 2006 regarding the gap between public expectations of not-for-profit organizations and the integrity of grantees. He raised accountability questions with university boards of trustees, specifically research universities, about alleged abuse of their tax-exempt status and misuse of university funds. The first major finding was "violation of the expectation of 501(c)3 organizations to meet the test of being apolitical and contributing to the common good with more efficiency than governmental agencies can provide." Committee investigations, including those of a number of "prestige" 501(c)3 organizations, found that they were, wittingly or unwittingly, fronting for 501(c)4 organizations that do have defined lobbying and advocacy permission not granted to the 501(c)3 organizations. The problems were captured in the indictments of lobbyist Jack Abramoff's problematic transactions with Native American clients. The issues identified included the following:

- Disguising the source of funds by accepting payments and passing them through to other groups, sometimes after subtracting a substantial fee.
- Accepting payment for writing op-eds and press releases favorable to Abramoff clients.
- Facilitating introductions between Abramoff clients and government officials, and accepting payment from Abramoff clients to act as front organizations sponsoring trips by members of Congress and their staff.

---

[10] Audrey Barrick. "Most Americans Believe Non-Profits are Not Financially Efficient." *Christian Post Reporter*. Feb. 14, 2008.

As Amanda Adams summarized the issue for *OMB Watch*, "The central problem with the activities is that they were unrelated to the organizations' tax-exempt purposes and benefited organizational insiders or individuals associated with Abramoff, rather than the general public."[11] An immediate implication for centers for advanced research is the need for a carefully stated by-law statement on "advocacy rights" within the limits of 501(c)3 status as well as defined limitations on how to avoid allowing the centers to be used for political advocacy or an individual's (institutional board member, officer or donor) gain.

Grassley's hearings have continued under the chairship of Senator Max Baucus (D-Montana). A summary of findings, thus far, from the Senate Finance Committee's investigations and hearings are that some not-for-profit organizations are more self-serving than engaging in activities that meet public interest and needs. Other findings include over-inflated executive salaries, heavy administrative costs for organizational promotion, fraudulent financial statements, haphazard or inadequate internal controls, deviation from mission focus, self-appointing boards that select members with conflicts of interest, and practices which are ethically and legally dubious.

In addition, research universities have been identified that have a pattern of practice designed to mislead both governmental agencies and foundations in grant accountability. In one review, a majority of the multi-million dollar federal grants were not being administered in ways agreed to in the original funding contracts. Worse, a majority of the principal investigators (PIs) of those grants not only did not comply with accountability requirements but had difficulty identifying what accountability requirements were part of their grants. As a consequence, several universities have now put in place required orientations for future PIs before applications can be initiated through university application control procedures. "At least the PIs can't claim ignorance about their accountability responsibilities," one Vice President for Research asserted. "Now let's see if they will comply."

Government agencies, most of which are under legislative scrutiny for failing to supervise expenditure of appropriated funds, fail to act on either fraud or accountability failures. One explanation is that to do so would slow down distribution and utilization of appropriated funds, for which agency directors are also being criticized. Supporting evidence for this finding comes from one professional accreditation agency that has declared that administration of many of the minority support grants are "open fraud."

Many foundations find themselves in situations similar to governmental agencies. Their mission involves the granting of funds for specific projects, but they have inadequate means for guaranteeing that grantees are using monies as contracted with program officers.

Until recently, there has been little serious research of suspected fraud and "worst practices" of not-for-profits. Until the 2008–2009 economic meltdown, the not-for-profit sector had grown every year for over a decade with revenues in excess of $665 billion. In 2007, however, the Association for Research on Nonprofit Organizations and Voluntary Action published a study based on data acquired and reported by the Association of Certified Fraud Examiners. The study found that, contrary to many not-for-profit CEOs who claim that headline scandals were just about a "few 'bad apples,'" the challenges were far more widespread within the 501(c)3 sector.

---

[11] Amanda Adams. *OMB Watch*. 10/24/06.

The authors of the study highlighted why the scandals not only threaten specific public needs and societal values, but the trustworthiness of the nonprofit sector in general:

> First, every dollar lost to fraud represents a lost ability to provide needed public services. Second, the sector is facing increased public scrutiny primarily as a result of the widespread availability of detailed financial information, and finally, a "Gresham's Law" may be at work, where publicized fraud cases may result in an unwillingness of donors to give to *any* nonprofit."[12]

There is some evidence to support the notion of a "Gresham's Law" when one institution's unethical/illegal practices are exposed. Contributions to American University (AU) declined precipitously when the trustees fired and then awarded President Ben Ladner a $3.75 million severance package. Other universities in the Middle States region, as a consequence of the AU investigations, also reported that they were under scrutiny by their alumni and donors like they had never experienced before. Further, AU applications for federal and foundation grants reportedly were far more closely scrutinized than competing applicants after the scandal. Other investigations, and their consequences, reveal that all institutions of higher learning are perceived with suspicion when one of their members violates public expectations of trust. Can the same phenomenon also apply to centers for advanced research?

A survey conducted by the Association of Certified Fraud Examiners in 2005 estimates "that all organizations lose on average 6 percent of their revenue to fraud every year. Applying this percentage to the nonprofit sector would suggest that the fraud loss would be approximately $40 billion each year."[13] In times of economic turmoil, the loss is projected to be much higher. Some of the findings apply to centers for advanced research:

- The likelihood of financial statement fraud increases as the number of outside members on a firm's board of directors decreases.[14]
- Fraud is easiest in organizations in which there is an atmosphere of trust, difficulty in verifying certain revenue streams, weaker internal controls, lack of business and financial expertise, and reliance on a volunteer board.[15]
- Financial statement fraud tends to generate the largest losses for victim organizations. Expense misreporting may be the most prevalent type of financial statement fraud.[16]
- From a 2003 sample of 38 nonprofits, 66 percent reported an average of $7 million less fundraising on their Form 990 than on their audited financial statements. Since donors and regulators often rely on program expense ratios from the publicly available Form 990, there is a high probability that the discrepancies are deliberate misreporting.[17]

---

[12] Janet Greenlee, et al., "An Investigation of Fraud in Nonprofit Organizations: Occurrences and Deterrents." p. 677.
[13] *Ibid.*
[14] *Ibid.,* p. 679.
[15] *Ibid.*
[16] *Ibid.,* p. 681.
[17] *Ibid.*

In a 2005 analysis, J. T. Wells found "three major types of frauds. The first is misappropriation of assets and occurs when an organization's assets are stolen or misused…The second is referred to as corruption and occurs when influence is inappropriately used in an economic transaction. Third, financial statement fraud is the deliberate falsification of an entity's financial statements."[18] Asset misappropriations comprise more than 97 percent of all reported frauds.[19] Playing "shell games"—using grant funds to cover unapproved expenditures, misdirecting grant funds for unapproved projects, and misleading reports to donors and granting agencies—are included in this category.

Many not-for-profit organizations have boards that rely on the annual audit process and assume that it is an adequate control procedure. But prior to 2002, "auditing standards did not encourage fraud detection procedures." After SAS No. 999 Consideration of Fraud in a Financial Statement Audit (AICPA, 2002b), there is a better opportunity for the annual audit process to detect at least major fraud activity, "but it is not a guarantee."[20] For example, of the 58 cases examined in the Association for Research on Nonprofit Organizations and Voluntary Action study, only 10 percent were discovered during the annual audit.[21]

The Senate Finance Committee and the regional accrediting agencies also have found that colleges and universities are very lax about personnel background checks. In one accreditation region, from six to ten percent of faculty are found to be without the credentials, degrees, or certifications claimed by either the individual faculty member or the college/university. In preparing for this essay, interviews with provosts, deans, and directors found that they give high to very high credibility to an individual who has been a fellow at a center for advanced research. Consequently, one provost declared that more extensive investigation of the faculty member was not necessary. "Do you think that I'm going to challenge someone who has spent a year at the _____ Institute? I've got more important things to do!"

More than one foundation executive reported a similar trust in applications where the applicant has been a center fellow. But how closely have the centers vetted the background of their fellows? How many centers have organized peer-reviewed accountability measures, including standards for reviewers? What centers have processes of accountability in place? How many centers share their data in the spirit of transparency as opposed to self-promotion? How many centers take the time to validate publication and award claims on curriculum vitas? Answers to all of these questions could lead to voluntary rather than imposed standards and regulations. Voluntary compliance with accountability demands have been the dominant model for not-for-profit operations since the 19th century. But, as one of the Senate Finance staff concluded, "Why should we trust voluntary compliance when presented with poorly constructed or monitored accountability claims, particularly from research centers?"

---

[18] *Ibid.,* p. 684.

[19] *Ibid.,* p. 686.

[20] *Ibid.,* p.690.

[21] The Association of Certified Fraud Examiners (ACFE) makes available a "fraud prevention check-up" to help identify and fix problems before it is too late.

Possible Responses from Centers for Advanced Research

Let us assume that no "credible" center for advanced research would ever intentionally practice or experience any of the misrepresentations and frauds uncovered in these analytical studies. But how do we know? As importantly, how do our stakeholders know? What measures do we have in place to practice and protect the integrity of our institutions?

Many lawyers and consultants are now advising not-for-profit organizations to adhere to a number of stringent requirements in the Sarbanes-Oxley Act that was enacted following the Enron, Worldcom, and other public company scandals. While some interpreters of the Act distance its applicability to not-for-profit organization, others note that two specific provisions of the law apply directly to not-for-profit organizations: (1) the whistleblower provision, which says that an organization can't fire an employee for reporting illegal activities involving a federal issue, and (2) the document retention provision, which requires organizations to keep and maintain documents after they become aware of an investigation.

Interviews with Senate Finance Committee staff confirm the expectation that a new act will soon be brought before Congress that makes application of increasingly strict Sarbanes-Oxley requirements specifically to 501(c)3 organizations. One proposal even holds private foundations accountable for vetting grantees' compliance with the proposed standards. Some state legislatures have not waited. By 2006, California, New Jersey, Rhode Island, and especially Illinois pushed through proposals to hold tax-exempt organizations to stricter standards and regulations. As Birchard observed, "Outsiders no longer take for granted the good intentions of nonprofit organizations. The general public expects nonprofit organizations to establish sound internal controls, distribute external audit documents, and provide more performance information to both the internal and external stakeholders."[22]

The challenges uncovered in the analyses could appear to be overwhelming. Based on extensive research and experience in higher education, however, scholars are in agreement that how we respond to these challenges will be the difference between a dying organization and a vibrant institution. Knee-jerk reactions, or even thoughtful attention to findings like those of the Senate Finance Committee, could lead centers for advanced research into paranoid reactions of noncompliance, on the one hand, or an overemphasis on control, on the other. The findings of violation of public trust; the extensive literature questioning validity of traditional, single methods of inquiry; and the legion of findings on how the inertia of traditional ways of operation turn an institution's mission into self-serving operations should be used to motivate us to find more creative approaches for better implementing our respective missions. Indeed, we do need to take challenges about veracity and verifiability very seriously. But we also need to recognize that the challenges are more than demands to tidy our workshops; we need to address these challenges by defining what kind of "trust" we want to establish between our centers and those with whom we need to connect.

We need to wrestle with the questions: For what and to whom should centers for advanced research be accountable? How do we define "accountability?" How should it be measured? What are valid methods? Who are peers with whom to compare "best practices?" If our centers were

---

[22] B. Birchard (2005) "Nonprofits by the Numbers." *CFO*. Quoted in Greenlee, et al., p. 691.

for-profit organizations, how would we operate differently and what functions/services would we be unlikely to provide?

Perhaps we can better address accountability expectations by distinguishing between simple, complicated, and complex challenges. The following categorization may help:[23]

> **Simple challenges.** Most of the identified challenges about organizational roles and responsibilities, fiscal management, and accounting procedures are matters of compliance with institutional, state, and federal regulations. Board members, CEOs, and staff cannot beg ignorance like the grant PIs in the example given above. Accountability expectations to simple challenges require competence. Measures can be summative, outcomes focused, and sequential. Centers can use the guidelines set forth by certified accounting organizations. Results can be validated though audits and comparisons with results from peer institutions.

> **Complicated challenges.** Some of the identified challenges revealed above are challenges of accountability to our stakeholders, particularly donors and funders. Most of these are about systems failures, flawed processes, poor communication, strained relationships, and misuse of accountability measures (e.g., for image-building rather than trust-building). As Deming promoted 50 years ago, we must find the means and measures for finding, analyzing, and applying the results of the systems in which we operate. Some contradictions and competitions, say between institution and community, between institution and funders, or between institutional mission and regulatory requirements, cannot be resolved, but we can have processes for addressing them. Both the models and measures of accountability need to be formative and summative. The Baldrige criteria provide us with guidelines.

> **Complex challenges.** Summative and regulatory accountability methods and measures do not give us help in facing the unpredictable and changeable realities faced by our centers. Nor are they likely to foster creative thinking, continuous analyses of our work, styles of negotiated accountability that encourage "buy in" of all participants, or promotion of staffs', scholars', and donors' development. Complex challenges require integrating accountability thinking from vision to product. Unlike responses to simple and complicated challenges, in which stakeholders tend to bring their preconceived notions of "trust" to the challenge, complex challenges solicit recommendations or guidance from "experts" about what stakeholders should expect and trust in largely new or unusual situations.

For more than 50 years, business leaders and others have provided leadership for challenges that are simple, or simple and complicated. Political and educational leaders have tended to adopt their methods and measures. But what about complexity issues? Business leaders have failed, some would say, miserably failed! Are centers for advanced research in a unique position to address the complex accountability challenges of our day and thereby influence what legislators, regulators, funders, and administrators of not-for-profit organizations seek in an environment without precedent?

---

[23] These distinctions are borrowed from Michael Quinn Patton's Evaluators' Institute course.

Options for Complex Accountability

Centers need to be prepared to address the demands for accountability outcomes from a range of stakeholders, regulatory agencies, and chains of command. But whatever we do, we need to acknowledge the complexities that growing demand for data brings with it and the calls for better ways of interpreting the data. These demands bring increasing (often exponentially) uncertainties to address. Following are several different perspectives that may help us look at our options.

Joshua Cooper Ramo, a managing director at Kissinger Associates, one of the world's leading geostrategic advisory firms, co-chaired the Santa Fe Institute's first working group on complexity theory and its application to advancing knowledge in the social sciences. In his book, *The Age of the Unthinkable*, he challenges:

> What we need now, both for our world and in each of our lives, is a way of living that resembles nothing so much as a global immune system: always ready, capable of dealing with the unexpected, as dynamic as the world itself. An immune system can't prevent the existence of a disease, but without one even the slightest of germs have deadly implications. The idea of deep security as an immune system is useful also because the stakes here could not be higher. The problems we are failing to confront now, from nuclear proliferation to global climate change to the rise of new and angry powers, are on a historic scale, and their cost will ultimately be weighed in the lives and deaths of tens of millions of people.[24]

He continues, "Can we find a way of understanding this revolutionary age that doesn't require us to do all the rounding and footnoting that doomed the old models [of investigation]?"[25] Applying his challenge: As centers for advanced research, how do we create methods and measures that gain validity in an environment of continuous change? Few centers have provided a response to the challenge. But accreditation organizations and some progressive foundations have.

Mark Kramer, a founder of FSG Social Impact Advisors, investigating foundation practices, writes:

> Our Advisory Board agreed unanimously that no single methodology represents the optimum approach to evaluation. In the words of Les Baxter, Chief Evaluation Officer of the Pew Charitable Trusts: 'There is a growing recognition that there are lots of different approaches to evaluation; it's not a monolithic one-size-fits-all endeavor.' Victor Kuo, Program Officer of the Bill & Melinda Gates Foundation, adds: 'There are a variety of purposes and a variety of audiences for evaluation, which all lead to the need for different approaches and methodologies. The challenge foundations face in evaluation, therefore, is to understand the full range of choices available, the different purpose they serve, and the circumstances in which they are relevant, in order to choose the approach that best captures the information needed.[26]

---

[24] Joshua Cooper Ramo. *The Age of the Unthinkable*. p. 18f.
[25] *Ibid.,* p. 46.
[26] Mark Kramer. *From Insight to Action: New Directions in Foundation Evaluation.* p. 11.

Michael Quinn Patton, a nationally recognized consultant/evaluator, observes, "Every model, no matter how detailed or how well conceived, designed, and implemented, is a vastly simplified representation of the world, with all of the intricacies we experience on a day-to-day basis."[27]

Centers for advanced research have the advantage of tapping from the experiences of accreditation organizations and foundations upon which to build our own approaches. Both quantitative and qualitative models have provided organizations far more objective and valid data upon which to investigate challenges and to base decisions than the era of management up to World War II. Both models of investigation have fostered highly sophisticated measures. Some historians argue, however, that they can date not only accountability documents, but academic publications, by the exclusive method in vogue at the time of writing. And, in most cases, the documents also reveal an "either/or" approach, a rejection of "both/and," and a distrust of the model (quantitative or qualitative) not chosen.

Evidence-based decision-making, and the environment for accountability represented by the Baldrige quality movements, use measures of both models. They also assume theoretical models for making corrections and changes in research discoveries. At The Evaluators' Institute in 2009, Michael Quinn Patton noted, "We are moving from studies to streams; we are relying on systems, not individual evaluators, to produce evaluative knowledge; we need multiple evaluative streams; we find most evaluative information now produced by non-evaluators; monitoring and evaluation are starting to merge." Accountability has become both a theoretical as well as a logistical matter. "Emphasis has shifted to continuous organizational adaptation and improvement."[28] Accountability thinking amply demonstrates that decision-tree thinking and fixed model management are inadequate to meet our challenges, and in some instances have failed us.

Patton proposes a developmental model of inquiry. He defines developmental evaluation as:

> Evaluation processes, including asking evaluative questions and applying evaluation logic, to support program, product, staff, and/or organizational development. The evaluator is part of a team whose members collaborate to conceptualize, design, and test new approaches in a long-term, on-going process of continuous improvement, adaptation, and intentional change. The evaluator's primary function in the team is to elucidate team discussions with evaluative questions, data, and logic, and facilitate data-based decision-making in the developmental process.[29]

The table on the following page illustrates the distinctions that Patton makes between traditional and developmental evaluations:[30]

---

[27] Francis Westley, et al., *Getting to Maybe: How the World Is Changed?*
[28] Michael Quinn Patton. Evaluators' Institute course.
[29] *Ibid.*
[30] *Ibid.*

Traditional Evaluation	Developmental Evaluation
(1) Testing models.	Supporting innovation and adaptation.
(2) Render definitive judgments of success or failure	Provide feedback, generate learnings, support direction or affirm changes in direction in real time.
(3) Measure success against predetermined goals.	Develop new measures and mechanisms as goals emerge and evolve.
(4) Evaluator external, independent, objective.	Evaluator part of a team, a facilitator and learning coach bringing evaluative thinking to the table, supportive of the organization's goals.
(5) Evaluator determines the design based on the evaluator's perspective about what is important. Evaluator controls the evaluation.	Evaluator collaborates with those engaged in the effort to design an evaluation process that matches philosophically and organizationally.
(6) Design the evaluation based on linear, cause-effect logic models.	Design the evaluation to capture system dynamics, interdependencies, and emergent interconnections.
(7) Aim to produce generalized findings across time and space.	Aim to produce context-specific understandings that inform ongoing innovation.
(8) Evaluation often a compliance function delegated down in the organization.	Evaluation a leadership function: reality-testing, results-focused, learning-oriented.
(9) Evaluation engenders fear of failure.	Evaluation supports hunger for learning.
(10) Accountability focused on and directed to external authorities and funders.	Accountability centered on the innovators' deep sense of fundamental values and commitments to learning.
(11) Accountability to control and locate blame for failures.	Learning to respond to lack of control and stay in touch with what is unfolding, and thereby respond strategically.

Patton's developmental model undoubtedly has its limitations, some obvious and some which will come to light in the push and shove of real world challenges. Nevertheless, he provides an advanced model for addressing issues of complexity.

Some commentators on the challenges of accountability argue that at this time, it is enough that not-for-profit organizations demonstrate that they have some processes in place. For some institutions in some settings, this "it's better to have something rather than nothing" approach to compliance may work, at least for a while longer. Whatever methods are used, however, require validity components that establish and maintain trust in the minds of stakeholders. It is not a matter of just finding the limitations to Patton's or other's initiatives; it is imperative that, in finding the limitations, we design improved models of knowledge acquisition and accountability models.

Mark Kramer, speaking of his study of foundations, argues that the "traditional approach" is linked to a specific theory of change: that foundations discover new solutions to the root causes of social problems, test them on a small scale, demonstrate their efficacy, then leave it to government or other funders to replicate and expand their efforts. This theory is widely accepted but rarely plays out in practice.[31] He continues:

> Newer ways of working require a different approach to evaluation. Accordingly, we have seen a pronounced shift away from the use of evaluation to measure the impact of past grants and toward a more timely and pragmatic process of gathering forward-looking information that will enable both grantors and grantees to make ongoing improvements in their work. The question driving evaluation has broadened from 'what was the impact of our grants?' To 'what do we need to know to increase our effectiveness?'…Foundation leaders have not, however, abandoned their desire to understand the impact directly attributable to their foundations' own interventions. The pursuit of these two different approaches—trying to isolate past effects through long-term outcome studies while at the same time seeking more flexible and timely sources of data to improve performance—has caused considerable tension in the field. No single methodology can meet both requirements and the resulting discord has undermined the credibility of both approaches.[32]

Where trust in the 19th century arose from definitions of "college" or "higher" education, and became intertwined with "quality" products in the 20th century, so today, validated knowledge and accountability have combined and are intertwined with the expectations of trust. But what about the 21st century? How do we help prepare people to have trust when our world is so rapidly changing and increasingly inconsistent? What are ways we can be accountable, yet address the noncausal and unpredictable factors of complexity? What methods and measures can we develop to account both for integrity and radical change?

---

[31] Mark Kramer. *From Insight to Action: New Directions in Foundation Evaluation*. p. 5.
[32] *Ibid.*

Summary and Conclusion

We began this essay with the question, "Why are we being faced with all of these accountability challenges?" Historical analysis reveals that the underlying issues of "trust vs. mistrust" not only dominated in the latter half of the 19th century, but we know through numerous polls that every profession and every institution is susceptible to a loss of trust today, particularly since the 1990s. "Quality" and contributions to "societal and national well-being" also continue to be integral to peoples' notions of trust. The overwhelming majority of Americans expect to see these values demonstrated, particularly in institutions of higher and advanced learning.

In this section, several well-received studies that have analyzed the witting and unwitting failures of not-for-profit organizations have been summarized to help locate and understand stakeholders' expectations for credible financial accountings, consistent mission-driven behaviors, and renewing and transparent "best practices." The analyses have also helped us to understand why failure to meet expectations of legislators, regulators, and the media put centers at risk for imposition of rigid standards and regulations. The most prominent accountability challenges to centers for advanced research have been listed.

The key lesson is that leaders of centers, and their supporters and funders, should endeavor to avoid a regressive and knee-jerk reaction to the challenges with overly conservative, restrictive, and single-method approaches to accountability. Rather, all not-for-profit organizations, but particularly centers for advanced research, should move quickly and decisively to get organizational principles and practices in compliance with expectations for 501(c)3 organizations and find trust-building methods and measures for addressing the challenges of complexity and chaos.

While balancing the tensions generated by competing perspectives of "trust," we still have the opportunity to create valid accountability measures that identify "best practices," trends, and comparative data in order to better understand and manage our own institutions. We can better manage our challenges by distinguishing between the simple, complicated, and complex. But to adequately address the complexities of our time and environment, we will have to lower our competitive and proprietary defenses and find the will to cooperate and coordinate efforts and resources. If we do, we will turn challenges of accountability into opportunities for one of the most creative periods in our respective histories.

Finally, while our underlying primary challenge is to operate our centers with the integrity that builds and rebuilds trust in our organizations, our unique opportunity may well be bringing the insights of accountability—accountability thinking—to disciplined inquiry; acquisition, analysis, and dissemination of new knowledge; as well as "best practices" of advanced research.

COLLABORATIVE EFFORT: ACCOUNTABILITY WORKSHOPS

To begin exploring "accountability thinking" in independent centers for advanced research, SAR launched an initiative to identify peer institutions and to bring them together to discuss each center's approach to accountability, consider challenges facing each of our institutions, and share "best practices" on program evaluation. The initiative was in part inspired by internal SAR efforts to identify meaningful quantitative and qualitative measures of the social impact of the School's programs, and in part by a recognition that, like many not-for-profit organizations, centers for advanced research are undergoing increasing legislative, media, and foundation scrutiny of their financial and programmatic activities. With funding from the Dobkin Family Foundation, SAR hosted a workshop in Santa Fe, NM for representatives of these peer institutions in 2009, which was followed up a year later by a second workshop at Dumbarton Oaks in Washington, DC, funded with the support of the William and Flora Hewlett Foundation. The participants, activities, and outcomes of these workshops are described below.

Our collaborative efforts recognized that numerous other organizations are working on assessment and accountability policies and procedures, and their efforts certainly informed our own approaches. This includes the work of not-for-profit and for-profit organizations such as the National Research Council and Academic Analytics, which evaluate and rank academic programs, as well as initiatives by foundations such as the Gates and Hewlett Foundations, which focus on the social impact of philanthropic activities and have supported research on assessment metrics[33]. However, at the time that SAR first assembled these peer institutions, we were aware of no similar collective efforts by independent centers of advanced research to develop shared approaches for accountability and program evaluation[34].

June 2009 Workshop

In June 2009, directors and key staff members from seven centers for advanced research joined representatives from the Wenner-Gren Foundation and Council on Foundations, as well as Paul Brest of the Hewlett Foundation, for a two-day workshop on "Performance Excellence and Accountability" held at SAR's meeting facilities in Santa Fe, NM. The June 2009 workshop had four primary objectives:

- to help centers for advanced research improve organizational performance practices, capabilities, and results;
- to identify and analyze the differing perspectives on accountability among the participating institutions and among the foundations that support our work;
- to identify means for addressing accountability expectations, particularly those from government agencies and private foundations; and
- to establish cooperation among centers through shared data and performance indicators.

---

[33] For example, the Gates Foundation assessed eight different cost-oriented approaches for evaluating the "social value creation" potential of grants. See Tuan 2008.

[34] There is, however, an extensive literature touting the benefits of shared measurement across institutions, with Kramer *et al.* 2009 advocating for more systematic collaborative measurement systems.

Participating centers were selected based on their shared missions to promote research in the humanities and social sciences, similar programming that includes fellowships and seminars, and comparable responsibilities towards maintaining collections of research materials. Most of the centers are autonomous of larger institutions and share similar infrastructural demands and funding models. The final list of participants included:

- Amerind Foundation (Dragoon, AZ)
- Dumbarton Oaks (Washington, DC)
- Omohundro Institute for Early American History & Culture (Williamsburg, VA)
- Santa Fe Institute (Santa Fe, NM)
- School for Advanced Research (Santa Fe, NM)
- Southwest Museum of the American Indian (Los Angeles, CA)
- The Newberry Library (Chicago, IL)
- Wenner-Gren Foundation (New York, NY)

Prior to the workshop, each participating center shared information on institutional history and mission, organizational structure, and accountability practices and challenges. The workshop agenda focused on these topics, with participants first comparing and contrasting the institutions and their stakeholders and constituents. Later discussions focused on fellowship programs and considered issues of assessment and accountability, with particular attention paid to the process of peer review, the challenges of taking risks and the lessons of failure, and the difficulties of evaluating program impact on scholarship. Especially valuable was the sharing of best practices for implementation and evaluation of fellowship programs.

Participants in the 2009 "Performance Excellence and Accountability" workshop at SAR.

As the two days in June 2009 came to an end, the participants realized that while much progress had been made towards achieving the workshop's objectives, more remained to be done. With the encouragement of Paul Brest, the participants expressed enthusiasm for a second workshop,

and Dr. Jan Ziolkowski, the director of Dumbarton Oaks in Washington, DC, offered his institution's meeting facilities for such an event.

June 2010 Workshop

With the support of the Hewlett Foundation, the second workshop was scheduled for June 2010 at Dumbarton Oaks. This brought together representatives from most of the centers represented in the first workshop, along with some new participants, to grapple with defining outcomes and sharing metrics with the end goal of increasing the institutions' collective ability to evaluate the social impact of our efforts. In early 2010, a core group of four participants from the original workshop met at SAR for a planning meeting to further develop the specific objectives and create a detailed agenda for the second workshop. Topics that were refined during the planning meeting included the following interrelated questions:

- What methods can centers for advanced research develop and/or use to analyze social impact of the organizations' activities?
- How do centers deal with questions of measurable outcomes and dissemination of scholarship?
- Are there internal evaluation methods that might prove applicable as foundations and government agencies move toward new standards for accountability?
- What shared quantitative and qualitative measures might allow the participating centers to mutually assess and support each other's mission fulfillment?
- How are principles of validity and integrity in advanced research—such as is assumed in the peer review process—protected in the process of defining accountability methods and measures?
- How do we address "failure"? Is it important to have a certain rate of failure to know that we are pushing conventional boundaries of academic and/or artistic production rather than just "playing it safe" with conventional projects?
- How can the productive engagement created through the workshop be sustained in the future, and how can the results be shared among a wider group of constituents and peer organizations?

The final list of participating organizations at the three-day June 2010 workshop at Dumbarton Oaks included the following:

- Amerind Foundation (Dragoon, AZ)
- Center for Advanced Study in the Visual Arts (Washington, DC)
- Council of American Overseas Research Centers (Washington, DC)
- Dumbarton Oaks (Washington, DC)
- Folger Shakespeare Library (Washington, DC)
- The John Carter Brown Library (Providence, RI)
- Omohundro Institute for Early American History & Culture (Williamsburg, VA)
- School for Advanced Research (Santa Fe, NM)
- The Newberry Library (Chicago, IL)
- Stanford Humanities Center (Stanford, CA)
- Wenner-Gren Foundation (New York, NY)

In addition, Paul Brest of the Hewlett Foundation was again in attendance, as was Dr. Ruth Bowman, Vice President of Evaluation for the Margaret A. Cargill Philanthropies.

The 2010 workshop again focused a great deal of attention on sharing best practices, especially regarding fellowship programs and their evaluation. With many new institutions represented, many topics similar to those discussed in June 2009 were covered, albeit with additional insights from the new participants. Additional discussion focused on the role of outreach to professional and general audiences, and the degree to which this is or should be an important part of the missions of centers for advanced research; for some, it is an explicit part of the mission, but not for others. In this context, some discussion centered on the use of emerging digital media for communicating outside of the centers.

Key Findings

The productive conversations that took place during the 2009 and 2010 workshops allowed the participating centers to share practices that they considered to be especially effective—or potentially promising if not yet implemented—for achieving their missions. A selection of suggestions on fellowship programs that participants found useful for consideration are listed below:

- Peer Review Process
  - To ensure that review panels provide the best selections, it was suggested that panel deliberations occur on one day, but that final fellowship selections not occur until the following day so that panelists have time to consider their discussions.
  - To maximize the outcomes of review panels, it was suggested that centers endeavor to provide feedback on unsuccessful applications, especially to graduate students and their advisors and departments.
  - While most centers use blind peer review of submitted applications to select their fellows, at least one institution felt that this is costly and yields results that are not always satisfactory.[35] A nomination process was proposed as a potentially better method for selecting fellows.
- Program Evaluation
  - Insofar as the goal of fellowships is to positively influence fellows' work beyond what they could have achieved without the fellowship, exit interviews need to ask how their experiences at the center changed their project.
  - Insofar as a goal of fellowships is to encourage trans-/inter-disciplinarity, evaluation needs to examine fellows' engagements beyond their disciplines before and after their time as fellows.
  - To measure the impact of fellowship programs beyond the fellows themselves, longitudinal network analysis of alumni and their impact through professional connectivity would be helpful, if not complex and expensive to implement.

---

[35] A growing body of literature is pointing out that while the flaws of peer review are well documented, its benefits are only assumed and have yet to be empirically demonstrated (see Smith 2010). The National Science Foundation recently launched efforts to assess the use of peer review in their granting decisions.

- To distinguish a center's success in selecting exceptional applicants from the center's positive impact on fellows' scholarship, it might be useful to evaluate a selection of fellowship alumni vs. alternates whose applications were considered high-quality but who did not receive a fellowship.
- Similarly, for graduate student fellows, interviews with their graduate program advisors can provide an evaluation of the fellowships' impact on the fellows' research, degree completion, and/or career success.
- A focus on the shared characteristics, selection criteria, and/or fellowship experiences of the most successful and/or most disappointing fellows might be a cost-effective approach to evaluation.

In addition to the sharing of best practices, numerous challenging issues for managing and evaluating fellowship programs were considered at the workshops. Evaluation especially invoked much discussion, with its goals and costs more of a concern than its methods. Some participants expressed concern over whether the costs of evaluating fellowship programs in the face of tight budgets justified the potential benefits, while the issue of evaluation to promote a center's profile vs. evaluation to strengthen the fellowship program was also a subject of much discussion—mirroring the larger discussion of assessment described in the previous section of this report.[36]

Participants in the two workshops felt that the key questions to be raised in developing fellowship program evaluation strategies include: Who are the intended users of the evaluation results? What do we want to learn from evaluations? What data are reasonably collectible, and what data can be used for improving programs? Is it cost-effective to collect both qualitative and quantitative data? And, how do we evaluate the impact of fellowship programs on graduate students vs. junior scholars vs. senior scholars? An especially challenging issue in causation was how to distinguish between a center's success at *predicting* which applicants would be successful from a center's *actual impact* on fellows.

Best practices on outreach programs also generated much discussion at the workshops, especially considering that most of the participating centers were independent of larger institutions. Several centers explicitly include outreach to public and professional audiences in their missions. But some participants argued that, for their institutions at least, the support of research was the core element of their mission, and dedicating resources towards outreach diluted this element while potentially jeopardizing scholarship with too much concern regarding public perception of the center's work. Suggestions of useful practices in outreach programs included the following:

- Communicating with the General Public
  - To advance critical thinking and to share the outcomes of center programs, the inclusion of journalists in these programs can be especially effective.
  - Similarly, some centers have established media relations positions to coordinate and prepare external communications.

---

[36] Many if not most of the workshop participants regarded the subject of evaluation with considerable trepidation, perhaps the result of over-exposure in large university bureaucracies to punitive forms of "traditional evaluation" as opposed to the arguably more productive "developmental evaluation" approach (as defined in Michael Quinn Patton 2006 and others).

- To encourage outreach, Wenner-Gren offers a supplement to their research grants to those recipients who present plans to demonstrate the social value of their work to a larger audience. Few applicants, however, have pursued these supplements, arguably mirroring disinterest in public outreach and/or the absence of training in how to do it.
- Communicating with Professionals
  - To encourage and recognize scholarly potential while promoting the center, it was suggested that centers establish graduate student "associate" programs in which exceptional students from nearby universities are given the honor of center affiliation.
  - To encourage international participation, centers should consider supporting seminars in non-U.S. settings, perhaps in collaboration with foreign centers for advanced research.
  - Several centers have concluded, based on direct and indirect experience, that New Media projects such as webinars are inadequate as direct substitutes for traditional programs such as face-to-face seminars.
  - On the other hand, the Folger's experiment in providing an online review process of already-accepted manuscripts for its print journal was considered to be a success.

Related to outreach was evaluation of publication programs, although discussion of this was brief since few of the participating centers had their own presses. Issues that came up in this context were concerns about the interference of economic concerns in scholarly publication, the role of peer review in publication, and the measurement of publication impact. This is an area that merits further attention in any future workshops.

Summary and Conclusions

By the end of the 2010 workshop at Dumbarton Oaks, participants generally agreed that the efforts were heading in two complementary directions. First, a periodic gathering of representatives from centers for advanced research to share best practices, explore issues in assessment and evaluation, and discuss possible collaborations was determined to be useful. At the time, the Stanford Humanities Center was offered as the hosting institution for a 2011 workshop. Subsequently, a change in leadership there contributed to the decision to delay the third workshop by a year or more, which would also have the benefit of allowing participating organizations time to review and selectively adjust their own practices in light of the first two gatherings.

The second outcome of the 2010 workshop was the decision to have a smaller group of centers pursue a pilot fellowship evaluation project. With the encouragement of Paul Brest, it was thought that three or four centers—rather than all participating institutions—could effectively conduct a multi-year analysis of their fellowship programs, the results of which could be shared with the larger group. Changes in the leadership of some of these institutions have delayed pursuing this joint evaluation project, but it is anticipated that this will be picked up again in the near future. As noted by Brest and others, if independent, not-for-profit centers for advanced research have value in the increasingly competitive field of higher education and research, this case has to be made with systematically collected data used to evaluate explicit and detailed claims of impact on scholarship.

INDIVIDUAL EFFORT: SCHOOL FOR ADVANCED RESEARCH

Beginning in 2008, SAR began to dedicate time and resources towards institutional accountability and program evaluation, which in turn inspired the collaborative efforts described earlier. SAR's individual efforts were the result of some key personnel changes, combined with growing demands by potential funders that applicants specify their evaluation procedures. At the same time, Dr. Glen W. Davidson, senior consultant/evaluator for the Higher Learning Commission and Professor Emeritus of Psychiatry and Medical Humanities at Southern Illinois University, joined the SAR Board of Managers and began educating the board and staff as to the value of evaluation efforts.

Over the past few years, Dr. John Kantner, SAR's Vice President for Academic & Institutional Advancement, has been exploring evaluation strategies for select programs with the goal of identifying approaches that SAR could routinely use to identify programmatic shortcomings and successes, with the goal of improving the School while also demonstrating to stakeholders SAR's intention to maximize the impact of its programs. The general approach has been to take each of SAR's programs, determine what part of the School's mission statement each was created to fulfill, and then identify quantitative measures to evaluate how well any individual program is achieving this goal. The emphasis on quantitative measures is not meant to reflect a preference for that approach over qualitative evaluation. Rather, SAR has a long history of collecting qualitative reviews from program participants, a practice that has been very helpful for making improvements to our operations. In contrast, quantitative measures have not been systematically collected in a way that ties program evaluation to mission fulfillment. The three evaluation implementations described below reflect SAR's attempts to provide quantitative data to complement the qualitative evaluations.

Scholarly Productivity

The first evaluation attempt at SAR focused on the productivity of SAR's permanent community of researchers, rather than on one of the programs designed for scholars from outside the institution. The reason for this was threefold. First, the academic staff, senior scholars, and research associates with multi-year appointments are key members of the School community who contribute to the mission's goal to create "a dynamic environment for the advanced study and communication of knowledge about human culture, evolution, history, and creative expression." At any given time, these researchers comprise half of the scholarly community at SAR, and thus their work influences the efforts of visiting fellows and seminarians. Second, at the time when evaluation work was beginning in earnest at SAR in 2008, much attention in the academic world was focused on larger assessment tools and efforts, ranging from the for-profit work of Academic Analytics to the oft-delayed doctoral program rankings from the National Research Council.[37] Finally, SAR staff believed that starting evaluation efforts with a resident population of scholars would be an efficient way to begin these efforts while also providing insights for later work with the alumni of the School's other programs.

---

[37] See Ostriker *et al.* 2010.

For this evaluation of the productivity of SAR's scholarly community, an attempt was made to replicate the methods of Academic Analytics, a private consulting firm that emerged out of scholarly evaluation work conducted by Lawrence B. Martin, a primatology professor and graduate dean at SUNY-Stonybrook. The company's best-known product is the Faculty Scholarly Productivity Index (FSPI), which is based on a proprietary algorithm for ranking academic program productivity. The claimed purpose of the FSPI is to assist decision-makers at universities and colleges as they decide how to allocate increasingly limited resources among programs. Unlike other ranking systems, such as that of the National Research Council,[38] the FSPI does not take peer assessment of program reputation into account, instead using quantitative data on grants, journal and book publication, citation rates, and honorific awards.[39] Academic Analytics also uses a moving window to assess productivity, which ensures that a scholar's past successes are discounted over time. To accommodate disciplinary differences in what is considered to be scholarly productivity (e.g., books in humanistic disciplines vs. journal publications in the hard sciences), FSPI comparisons are meaningful only within rather than among fields.

In 2008, Academic Analytics released disciplinary rankings along with the data used to generate FSPIs for the majority of programs in U.S. higher education. The algorithms used to create the FSPIs were not made available, but since both the data and the resulting indices were provided, it was possible to use regression analysis to approximate the method used to create FSPIs for programs in any given discipline. These regression formulas could in turn be used to generate an index for SAR's scholarly community, as if the School were a university department or program. The results could then be compared with the published rankings from Academic Analytics to provide a quantitative assessment of SAR's scholarly productivity.

Data comparable to those used by Academic Analytics were collected from scholarly productivity reports that SAR's permanent scholars are asked to submit each year. Although collected in 2008, like Academic Analytics, only the productivity data through 2006 were used—this lag is necessary when using citation data, since these are only dependably available a few years after publication. All the SAR productivity data were then fed into the regression formula calculated from Academic Analytics's published information for U.S. anthropology departments. The results, shown on the next page, demonstrate that SAR's permanent scholarly community is performing quite well, comparable to the productivity of the most prestigious anthropology programs.

---

[38] In the last few years, many other approaches to measuring scholarly productivity have emerged, including proprietary for-profit initiatives such as those from Scopus, SciTech Strategies, and Web of Science, as well as free tools such as that offered by Scholarometer.

[39] Other variables that have been considered in assessments of scholarly productivity include peer assessment and median expenditure per journal article. None of them have been immune to extensive criticism of data accuracy and statistical procedures (e.g., Brainard 2011, Fogg 2007).

Ranking of SAR Faculty Productivity
Among Top 10 Anthropology Programs, 2002–2006

UCSF Penn State Harvard SAR Chicago ASU UCLA Rutgers Stanford Michigan
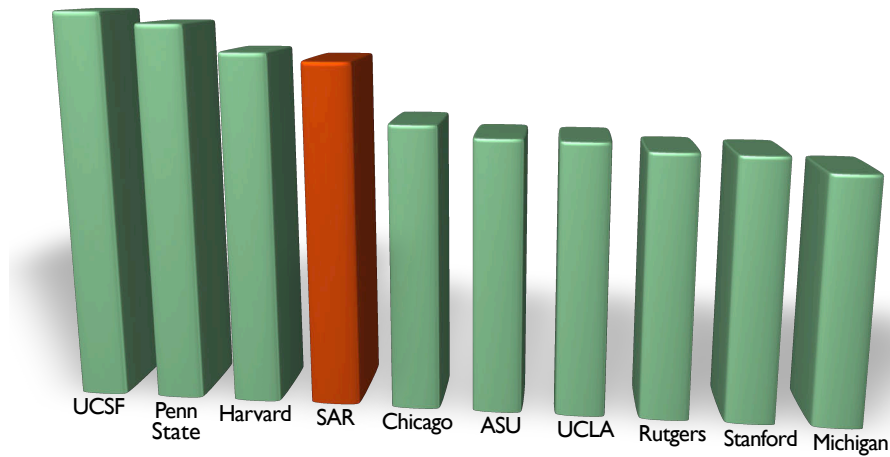
Academic Analytics no longer makes public its scholarly productivity data or FSPI rankings, and thus SAR can no longer be compared with U.S. anthropology programs.[40] However, an annual FSPI for the institution continues to be produced to allow for a year-to-year comparison of the productivity of the School's permanent scholarly community, as illustrated below.

Annual Faculty Scholarly Productivity Index

2006   2007   2008   2009

---
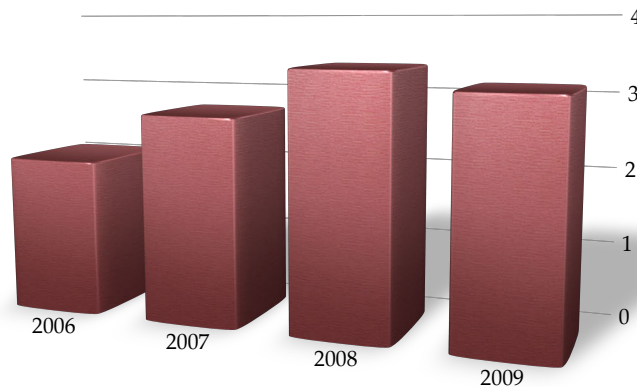
[40] The National Research Council (NRC) rankings and final methodology (Ostriker *et al.* 2010) were revealed only recently and therefore were not included as comparative data for the SAR scholarly productivity analysis. However, a review of the NRC data and rankings for anthropology programs confirms the high level of productivity described here for SAR. For example, Pennsylvania State University's anthropology program, the top-ranked for research productivity, is highlighted by 0.772 publications/faculty, 61% of faculty with grants, and 1.038 awards/faculty, compared with SAR's figures of 1.000, 25%, and 1.38 respectively. Anthropology at Duke University, tied for number one in some of the NRC measures with Penn State's program, came in with figures of 1.043, 56%, and 0.192 for those same variables. Note that some of the NRC variables, such as citations per *total career* publications, were not measured by Academic Analytics nor in the SAR analysis.

This shows a generally positive trend in scholarly productivity among the permanent SAR scholarly community. The slight drop in 2009 reflects the discounting over time of some older books and awards. If such declines continue, SAR leadership may want to consider possible changes in research policies for administrative faculty and/or changes in personnel.

Of course, the type of evaluation provided by Academic Analytics's FSPI is largely summative. An assessment of the factors that contributed to SAR's score, however, can be quite informative and help School leadership consider ways to improve the experience and thus productivity of its scholarly community. As illustrated in the following table, for example, SAR's score is positively influenced by honorific awards, which are heavily weighted in the FSPI. Books and citation rates also provide a significant boost to SAR's ranking. The high values of these measures are likely because the SAR community includes several senior scholars whose work continues to be routinely cited, who are the recipients of several career awards, and who are producing books summarizing their years of research. Journal articles[41] and research grants, in contrast, are underrepresented in SAR's FSPI, and in general the more junior members of the SAR community are underperforming, most likely because of the heavy administrative load they carry.

	2006	**2007**	**2008**	**2009**
Books	43.6%	39.1%	32.2%	32.5%
Articles	11.6	12.9	9.3	11.2
Citations	5.1	16.5	14.8	19.4
Grants	15.4	3.3	0.8	2.1
Awards	51.7	49.2	59.8	53.0

SAR Press

The second evaluation tool developed at SAR was created to assess the impact of scholarly books produced by SAR Press. The Press serves a key function at SAR by fulfilling the component of the mission to "[communicate] knowledge about human culture, evolution, history, and creative expression." Initiated in 2009, this evaluation concentrated on the Press's flagship Advanced Seminar series, which is composed of almost 90 edited volumes dating back to 1972, as well as the more recent Resident Scholar series, which features 18 titles dating back to 1992. The goal of the evaluation was to determine the degree to which these scholarly books are cited in the professional literature, assuming that the level of peer citation would be a reasonable proxy for communication and impact of the knowledge represented by the books. A related assumption was that book sales, an otherwise easy comparative to use, are more likely to

---

[41] The Academic Analytics approach, like the NRC's analyses, does not consider the impact of a journal in which faculty articles and citations appear. Thomson Reuters produces annual impact factors for many journals, and competing journal rankings are appearing, such as SCImago Journal Rank and Eigenfactor. All are subject to the same kinds of criticisms that are being leveled at other scholarly evaluation metrics. Nonetheless, scholars intuitively know that certain journals are more "prestigious" than others, even if they're reluctant or unable to articulate why, and including journal impact factors might begin to assess quality vs. quantity in scholarly productivity.

reflect marketing efforts, course adoptions, and editor status than they are the specific impact of the work on scholarship.

For the analysis, the journal citation rates for an initial sample of 20 Advanced Seminar and Resident Scholar books were compared with 20 similar titles from peer presses in the U.S. and U.K. published since 2000. Peer presses were identified as those most often cited in SAR Press books, regardless of nonprofit or for-profit status and size compared to SAR Press: University of California Press, Oxford University Press, Routledge, University of Chicago Press, Duke University Press, and Cambridge University Press. For each SAR Press title, a book published in approximately the same year, on the same topic and discipline, and in the same format (i.e., single-authored vs. edited collection) was selected from one of the peer presses, such that every SAR Press book was paired with a peer book. The next step was to use Thomson Reuters's Web of Science citation database to search for year-by-year citation rates for all of the books.[42] Because Web of Science only includes journal citations, and not those from books, it necessarily better represents disciplines that favor journal publications. To mitigate this, the pairwise comparisons were required to be between books from similar disciplinary backgrounds, which ensured that, for example, a book in biological anthropology was not being compared with a book in history. Citation rates were compared both by absolute calendar year and by number of years after publication.

The 2009 results, which are illustrated below, demonstrate that SAR Press books are heavily cited in professional journals at a rate roughly equal to or above that of peer presses, with only the University of California Press significantly outperforming SAR Press (67% more citations) in this specific measure of intellectual impact. On average, the performance of the six peer presses is only 83% of SAR's citation rates. This suggests that SAR Press is playing an important role in fulfilling the institution's mission to communicate scholarly knowledge with academics.
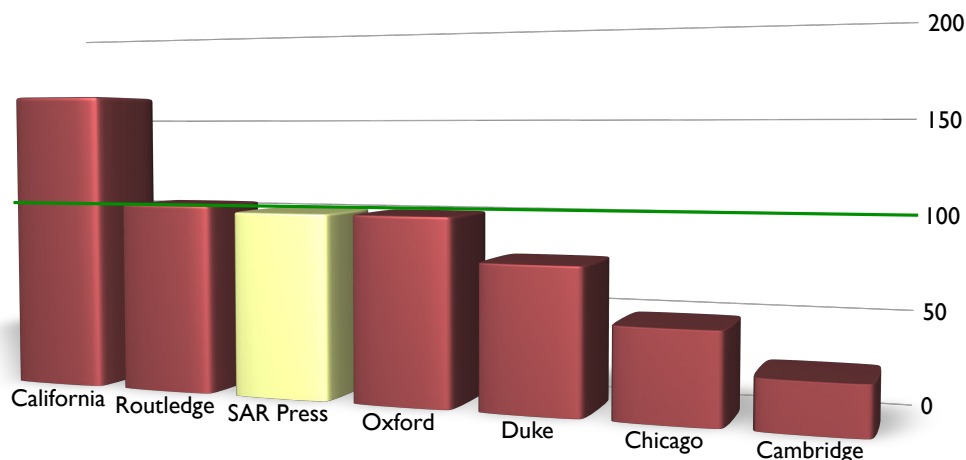

Peer Press Citation Rates as % of SAR Press Rate

---
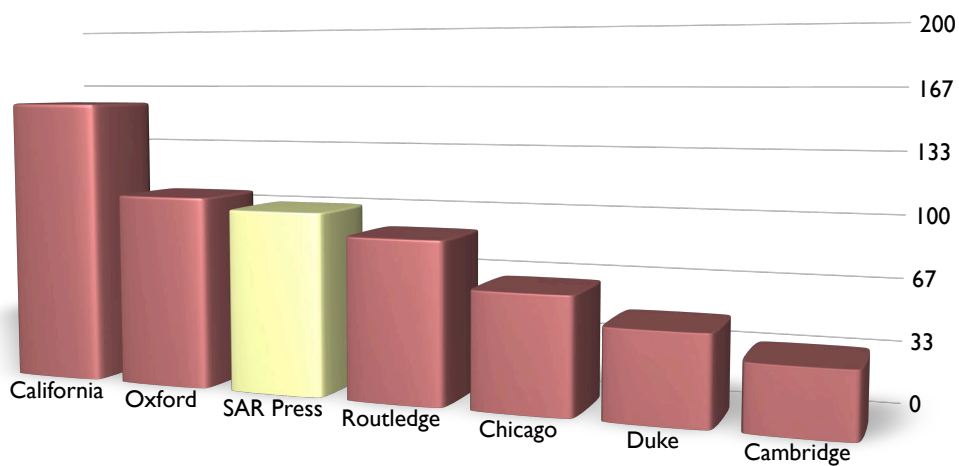
[42] One issue to consider in future analyses is whether to include some assessment of the impact of the journal in which each citation appears. See previous note on journal impact factors.

Since the 2009 evaluation, new citation rate data have been collected every year and new books added to the comparison. The 2011 results, illustrated below, demonstrate that SAR Press continues to have a positive impact on scholarship compared to peer presses, despite its much smaller size and very modest marketing budget.

## Peer Press Citation Rates as % of SAR Press Rate[*]

[*] Peer Presses = Cambridge U Press, Duke U Press, Oxford U Press, Routledge, U of California Press, U of Chicago Press

The analyses also allow for a comparison of citation rates per year after publication. This is useful because it takes time—often several years—for a book to be recognized for its scholarly importance and for citations to appear in the broader literature. The 2009 results, shown at the top of the next page, illustrate that SAR Press books are on average cited more often than their peer books in each year after publication. However, the first year after publication sees fairly modest citation rates for SAR Press books, perhaps a function of the smaller marketing apparatus. The second and third year after publication see significant growth, followed by what appears to be a leveling off of modestly higher citation compared to peer books. The latter years, however, are currently influenced by small sample size, and the patterning will likely continue to change until more years of new citation data are added to the analysis.

The SAR Press evaluation will be updated each year with new books and new citation data. Interestingly, as noted in the opening paragraph of this section, citation rates and sales numbers showed only a modest correlation—with a correlation coefficient of 0.62—suggesting that while sales might be important for making business decisions about SAR Press's productivity, citation data more accurately measure how well the press is contributing to mission fulfillment.

SAR Press Citation Rates
## as % of Rates of Peer Presses

Year After Publication Date
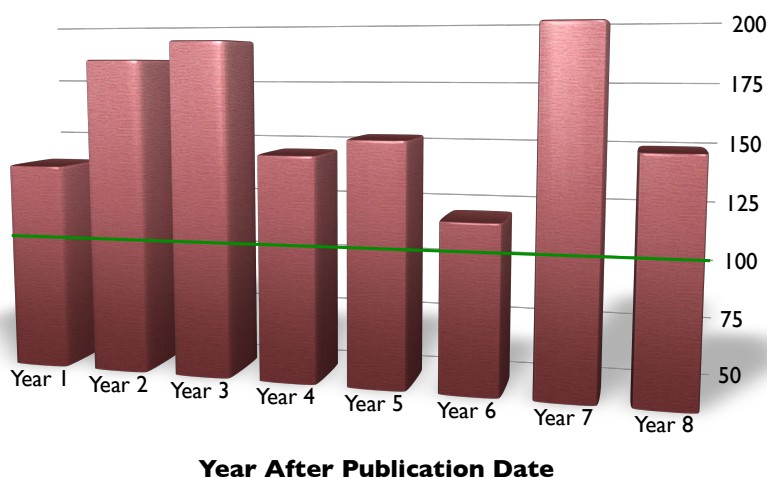
Advanced Seminar Program

Efforts began in 2010 to evaluate SAR's mission to "provide a dynamic environment for the advanced study…of knowledge about human culture, evolution, history, and creative expression." This component of the mission is most directly served by SAR's Scholar Programs, which include the well-known Resident Scholar and Advanced Seminar programs. The latter was established in 1968 and has featured almost 140 seminars, 80% of which have resulted in book publications—including the edited volumes analyzed as part of the SAR Press evaluation presented earlier. Because evaluation of the Advanced Seminar program is related to the SAR Press evaluation, and because the Resident Scholar program presents considerable assessment challenges, as discussed earlier, it was decided that the first Scholar Program to be analyzed would be the Advanced Seminars. The goals were to determine whether SAR seminars add value to the knowledge produced and, specifically, whether SAR's format—papers shared in advance, five days of seminar discussions, living and eating together during those five days, etc.—worked at least as well as other formats for collaborative engagement, such as professional conferences.

In a manner similar to the SAR Press study, it was decided that peer citation of the products of Advanced Seminars would provide the most direct quantitative data for the evaluation, to complement the qualitative information we collect from seminar participants. With funding from the Hewlett Foundation, a random selection of 20 edited volumes produced through Advanced Seminars since 1994 was compared with a pairwise sample of edited volumes of similar topic, discipline, and timeframe. The citation data in some cases were identical to those used for the SAR Press evaluation. However, because the analysis intended to compare SAR seminars with non-SAR collaborative formats rather than to compare SAR Press with peer presses, the non-SAR books were required to be the products of conferences or workshops. When possible, the events leading to the books included some of the same participants. To further control for the influence of press size and prestige, books from presses that had been identified as consistently

29

outperforming (e.g., University of California Press) or underperforming (e.g., Cambridge University Press) SAR Press books were excluded from the Advanced Seminar evaluation, but otherwise this analysis was not restricted to any particular presses.

The results show that SAR Advanced Seminars produce edited volumes that are cited in the professional literature more frequently than edited volumes produced through other collaborative engagements, such as conference sessions and workshops. In the figure below, the vertical axis represents percentage of non-SAR citations; by definition, all other total citations are at 100%. The figure accordingly shows that SAR Advanced Seminar volumes have 31% more total citations than do their peer volumes.

## Impact of Advanced Seminars vs. Other Collaborative Approaches
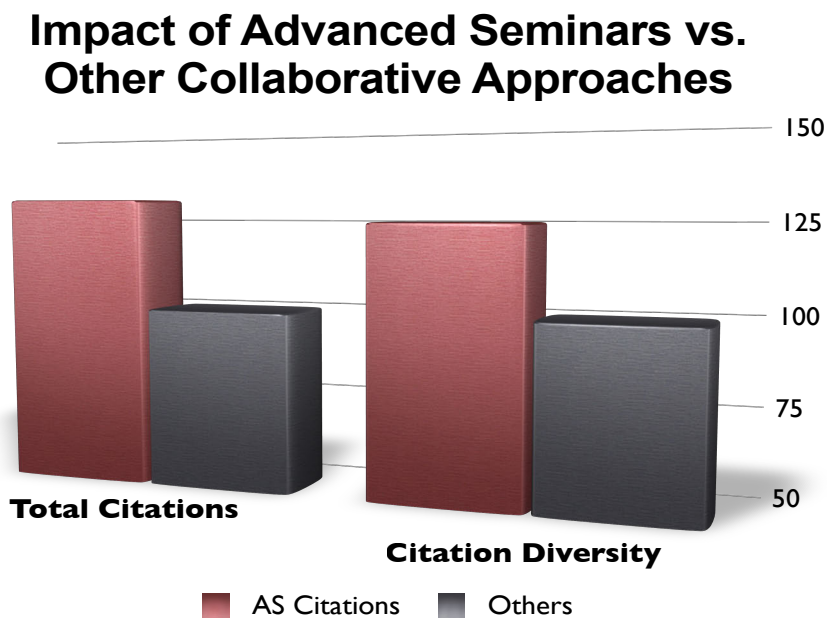
While one goal of the Advanced Seminar program is to generate important new knowledge that impacts scholarship, another important goal is to facilitate exchange among disciplines. Interdisciplinary seminars and diverse contributors are encouraged, as is international participation, as a fulfillment of the SAR mission to create a dynamic environment for the exchange of knowledge about the human experience writ large. Therefore, an additional analysis considered the number—or diversity—of disciplines citing SAR Advanced Seminar books, as measured by the disciplinary designations assigned to each citing journal in the Web of Science database. The figure above shows that the Advanced Seminar books are cited by 25% more disciplines than the peer group.

The impact of SAR Advanced Seminars on scholarship is likely underestimated because of the lag in citation rates that was noted in the previous section as a likely consequence of the modest size of SAR Press. As illustrated in the figure below, books that are the product of Advanced Seminars experience a meaningful lag in both citation rates and citation diversity. In the first year after publication, the books are cited at a rate that is less than 50% of their peers and by 50% fewer disciplines. However, the following years see a remarkable increase until, in the fifth

year after publication, the Advanced Seminar books are cited 88% more frequently by 67% more disciplines than their peer group. This suggests that the dissemination of the knowledge created in Advanced Seminars has a slow start but then realizes its full impact within a few years.

Total Citations & Citation Diversity as Percentage of Peer Group

Legend:
- Total Citations
- Citation Diversity
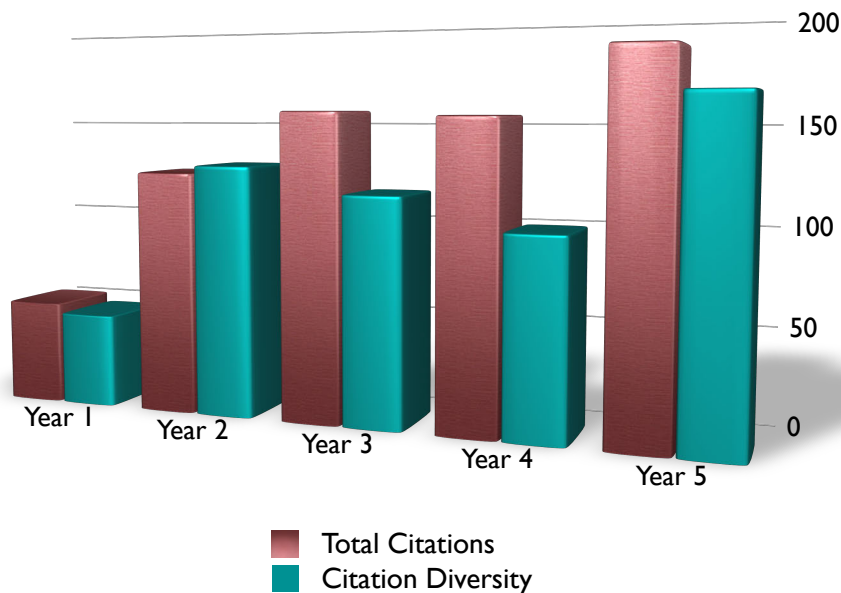
Summary and Conclusions

The three quantitative evaluations of SAR programs described in this section complement the qualitative data that the School has routinely collected through exit interviews and follow-up questionnaires. As noted earlier in the background section of this report, the assessment of the activities of centers for advanced research is most effective when both approaches are used, as they reinforce one another or reveal issues and opportunities not identifiable if only quantitative or qualitative evaluation is conducted. For example, the quantitative analysis of the productivity of SAR's permanent researchers reinforces qualitative information obtained through annual interviews with junior administrative faculty, which suggest that the administrative workload is compromising their research agendas. On the other hand, both the SAR Press and Advanced Seminar analyses reveal a pattern not identified in qualitative assessments, that the full impact of SAR programs consistently experience a delay, almost certainly due to the small size of SAR Press and its marketing budget. This and other information revealed by these analyses are useful for SAR's executive management and Board of Managers for deciding where and how to allocate resources to best fulfill the center's mission, even while providing confidence that the School's programs are performing at a very high level.

The evaluation and accountability efforts at SAR have identified additional approaches for evaluating the center's programs and their success at mission fulfillment. For example, other

ways to assess the value of the Advanced Seminar program in fulfilling the SAR mission would include a study of new professional networks created through the seminars,[43] as well as the impact that the seminars have on the direction of individual scholars' research.[44] Both of these approaches would be time-consuming and require the ongoing collection of longitudinal data from seminarians, but SAR staff are designing survey instruments for collecting such information.

SAR still has additional programs to evaluate using both quantitative and qualitative approaches. This includes one of the School's flagship programs, the Resident Scholar fellowships, which presents numerous evaluation challenges. Currently, qualitative information is collected from fellows during exit interviews, and while these are useful for making small improvements to the program, what is needed is the systematic collection of longitudinal data from program alumni to determine the impact of the "SAR experience" on their careers and scholarly output. Even more challenging will be evaluation of SAR's Artist Fellowship program. Again, exit interviews provide an important assessment of this program, but no formal procedures are yet in place for following the careers of artist alumni.

SAR plans to continue engagement with other like-minded organizations to advance accountability and assessment efforts. Several of the participants in the 2009 and 2010 workshops are eager to see the group assemble to continue sharing information on program evaluation and best practices, while we are still hopeful that SAR can spearhead a joint data-analysis project. In the meantime, SAR has formed another collaborative engagement with five not-for-profit research organizations in Mexico and the U.S. Southwest to begin exploring issues unique to our region. This group, which we are calling the Consortium of Southwest Research Centers,[45] is currently focusing its attention on outreach and has launched a $5,000 annual prize for an outstanding public education or outreach program that brings archaeological knowledge about the past to inform issues and problems of the present.

---

[43] Some examples of this approach have been developed over the past several years, especially within the area of science mapping. See Boyack and Börner 2002, among many others by those authors.

[44] Lavinghouze *et al.* 2007 of the Centers for Disease Control and Prevention suggest that "success stories" can be more effective at communicating an organization's effectiveness than formal quantitative measures.

[45] The Consortium includes SAR, Amerind Foundation, Center for Desert Archaeology, Crow Canyon Archaeological Center, Instituto Nacional de Antropología e Historia–Sonora, Museum of Northern Arizona, and SRI Foundation.

SUMMARY AND RECOMMENDATIONS

Since 2008, SAR and several other independent, not-for-profit centers for advanced research have launched initiatives in program evaluation and institutional accountability. This has included individual efforts by SAR to assess how well its programs are serving the center's mission, and therefore how effectively the School is using its funding. Additional collaborative work among the centers has occurred in the context of two workshops on evaluation and accountability that were held in 2009 and 2010. This report has summarized those efforts and identified many of the issues identified by the participating centers.

The centers for advanced research participating in the accountability workshops recognized the value of collaboration, especially for sharing best practices and identifying joint funding opportunities. Some of the centers' representatives, however, also identified potential challenges. "Some of us are concerned about possible conflicts between the realities of being competitors and the accountability demands to be collaborators," noted one representative. Especially when it came to discussions of possible joint assessment projects, several expressed concerns about sharing proprietary data. Others worried about "mission creep" as centers scrambled to show the successes of their programs in the face of accountability pressures. A general skepticism about evaluation was shared by a few representatives of the participating centers, partially due to unsatisfactory experiences with traditional evaluation in the past, but also because of methodological and financial challenges in implementing developmental evaluation.

Paul Brest, however, argued that centers for advanced research "must take far more responsibility than they have for shaping the questions of importance for our society, the methods which give us valid data to investigate those questions, and the distribution of research outcomes." Glen Davidson further suggested that, likely within the next five years, accountability expectations from foundations and federal regulations will indeed place greater demands for "outcome data" generated by validated metrics, and that centers need to be proactive in developing these metrics before they are imposed upon us. Recognizing complex accountability challenges and implementing developmental evaluation approaches are especially critical for centers for advanced research insofar as we are important "gatekeepers" of careers and research.

As independent, not-for-profit centers for advanced research plan and initiate new accountability projects, especially as related to program evaluation, the experiences of SAR and the other participants in the accountability workshops suggest that the following challenges and issues need to be considered:

> **The Goals of Evaluation:** Centers need to seriously consider their reasons for evaluating their programs. The temptation to consider evaluation as a fundraising tool in which only the successes are identified and touted needs to be tempered, and true developmental evaluation planned for and implemented. Fortunately, most foundations and observers of the philanthropic world are embracing this approach and expect that not all programs will be successful—but they also want us to recognize and respond to failure quickly and efficiently.

The Role of Collaboration: The efforts to develop collaborative data-collection and evaluation projects described in this report met with resistance due to concerns about sharing data and competing for donors and fellows. And yet some of the questions that our centers would like to answer can most easily be addressed through joint efforts. Collaborative projects also are easier to fund and can be more efficient through the sharing of the personnel needed for successful evaluation.

The Efficacy of Peer Review: One of the questions that might best be answered jointly is the effectiveness of peer review. Long held unassailable within the academy, and in varying degrees relied upon heavily by centers for advanced research, peer review is beginning to be challenged by a number of empirical studies. Insofar as centers are the testing grounds for the creation and dissemination of knowledge, we are in a position to experiment with peer review and its alternatives. And since we have a profound influence on individual careers, we also have a responsibility to evaluate the success of peer review.

The Challenges of Evaluation: Program evaluation at centers for advanced research presents significant theoretical and methodological challenges. Evaluation needs to be developed with our mission statements in mind, to ensure that the data we're collecting and results we're generating actually speak to our missions. For example, a fellowship program may be designed to reward the best and brightest, or it may have the goal of making fellows think and research in new ways; evaluating each of those would require very different approaches with quite distinctive sources of qualitative and quantitative data. Evaluation requires a significant commitment in time and money to do effectively.

BIBLIOGRAPHY

N a. "Assessment as a Core Strategy." The Higher Learning Commission of The North Central Association. Discussion Draft. March 1, 2009.

N a. "University of California Accountability Framework." The University of California. Discussion Draft. September 21, 2008.

N a. *Quantitative Indicators For Research Assessment—A Literature Review.* Research School of Social Sciences Research Evaluation and Policy Project. Canberra: The Australian National University, 2005.

Altschuld, James W., and Henry Y. Zheng. "Assessing the Effectiveness of Research Organizations." *Evaluation Review.* 1995, 19(2):197–216.

Boyack, Kevin W., and Katy Börner. "Indicator-Assisted Evaluation and Funding of Research: Visualizing the Influence of Grants on the Number and Citation Counts of Research Papers." *Journal of the American Society for Information Science.* 2003, 54(5):447–461.

Brainard, Jeffrey. "Johns Hopkins and Texas A&M Were the Least Productive Research Universities, Study Finds." *The Chronicle of Higher Education.* May 22, 2011.

Brest, Paul, and Hal Harvey. *Money Well Spent: A Strategic Plan for Smart Philanthropy* New York: Bloomberg Press, 2008.

Brest, Paul, Hal Harbey, and Kelvin Low. "Calculated Impact." *Stanford Social Innovation Review.* 2009, 7(1):50–56.

Ewell, Peter T., ed. U.S. Accreditation and the Future of Quality Assurance: A Tenth Anniversary Report From The Council For Higher Education Accreditation. Washington, D. C.: The Council for Higher Education Accreditation, 2008.

Fields, Robin. "Nonprofits Panel Puts Getty on Probation." *Los Angeles Times.* California/Local. December 21, 2005. P. B1.

Fogg, Piper. "A New Standard for Measuring Doctoral Programs." *The Chronicle of Higher Education.* January 12, 2007.

Gamble, Jamie A. A. *A Developmental Evaluation Primer.* Toronto: The J. W. McConnell Family Foundation, 2008.

Gertner, Jon. "For Good, Measure." *The New York Times.* March 9, 2008.

Greenlee, Janet, Mary Fischer, Teresa Gordon and Elizabeth Keating. "An Investigation of Fraud in Nonprofit Organizations: Occurrences and Deterrents." *Nonprofit and Voluntary Sector Quarterly.* 2007, 36:676–694.

Hatry, Harry, Elaine Morley and Elisa Vinson. *Outcome Measurement in Nonprofit Organizations.* Urban Institute.

Kramer, Mark (with Rebecca Graves, Jason Hirschhorn and Leigh Fiske). *From Insight to Action: New Directions in Foundation Evaluation.* Boston: FSG Social Impact Advisors, 2007.

Kramer, Mark, Marcie Parkhurst, and Lalitha Vaidyanathan. *Breakthroughs in Shared Measurement and Social Impact.* Boston: FSG Social Impact Advisors, 2009.

Lavinghouse, Rene, Ann Webb Price, and Kisha-Ann Smith. "The Program Success Story: A Valuable Tool for Program Evaluation." *Health Promotion Practice.* 2007, 8(4):323–331.

Miller, Franklin G., John P. Gluck, Jr., and David Wendler. "Debriefing and Accountability in Deceptive Research*." Kennedy Institute of Ethics Journal.* 18:3 (September 2008). The Johns Hopkins University Press.

Ostriker, Jeremiah P., Paul W. Holland, Charlotte V. Kuh, and James A. Voytuk (editors). *A Data-Based Assessment of Research-Doctorate Programs in the United States.* Washington, DC: National Academies Press. 2010.

Panel on the Nonprofit Sector. *Strengthening Transparency, Governance and Accountability of Charitable Organizations. A Report to Congress and the Nonprofit Sector*. Washington, D. C.: Independent Sector, June 2005. On line: www.NonprofitPanel.org.

--------. *Strengthening Transparency, Governance, Accountability of Charitable Organizations. A Supplement to the Final Report to Congress and the Nonprofit Sector.* Washington, D. C.: Independent Sector, April 2006. On line: www.NonprofitPanel.org.

Patton, Michael Quinn. "Developmental Evaluation: Systems Thinking and Complexity Science." Course Title: Alternative Evaluation Designs: Implications From Systems Thinking and Complexity Theory. San Francisco: The Evaluators' Institute, January 10–11, 2009.

--------. *Utilization-Focused Evaluation*, 4th edition. Thousand Oaks, CA: Sage, 2008.

Ramo, Joshua Cooper. *The Age of the Unthinkable: Why The New World Disorder Constantly Surprises Us and What We Can Do About It*. New York: Little, Brown and Company, 2009.

Smith, Richard. "Classical Peer Review: An Empty Gun." *Breast Cancer Research.* 2010, 12(Suppl. 4):S13–16.

Thompson, Roger. "Kash Rangan: …HBS Social Enterprise Initiative." *Harvard Business School Bulletin*, June 2008.

Tuan, Melinda T. "Measuring and/or Estimating Social Value Creation: Insights into Eight Integrated Cost Approaches." Report prepared for the Bill & Melinda Gates Foundation's Impact Planning and Improvement Division. December 2008.

U. S. Senate Committee on Finance. Website bulletins. http://finance.senate.gov.

Venverloh, Sally L., and Mary K. Samsa. "IRS and Congress Plan Investigation of Nonprofits Paying Excessive Compensation." Client Memorandum. Gardner Carton & Douglas. July 2004. On line: www.gcd.com.

Westley, Frances, Brenda Zimmerman, Michael Q. Patton. *Getting to Maybe: How the World Is Changed?* Toronto: Random House Canada, 2008.