Evaluation and Program Planning 35 (2012) 390–397

Contents lists available at SciVerse ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Expert panel reviews of research centers: The site visit process

Frances Lawrenz [a], Mao Thao [a,*], Kelli Johnson [a,b]

[a] University of Minnesota, Twin Cities, College of Education and Human Development, Department of Educational Psychology, 250 Educational Sciences Building, 56 East River Road, Minneapolis, MN 55455, USA

[b] University of Minnesota, Twin Cities, College of Education and Human Development, Department of Organizational Leadership, Policy, and Development, 330 Wulling Hall, 86 Pleasant Street SE, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 28 February 2011
Received in revised form 6 January 2012
Accepted 8 January 2012
Available online 15 January 2012

Keywords:
Site visit
Expert panel review/evaluation
Research center

ABSTRACT

Site visits are used extensively in a variety of settings within the evaluation community. They are especially common in making summative value decisions about the quality and worth of research programs/centers. However, there has been little empirical research and guidance about how to appropriately conduct evaluative site visits of research centers. We review the processes of two site visit examples using an expert panel review: (1) a process to evaluate four university research centers and (2) a process to review a federally sponsored research center. A set of 14 categories describing the expert panel review process was obtained through content analysis and participant observation. Most categories were addressed differently through the two processes highlighting the need for more research about the most effective processes to use within different contexts. Decisions about how to structure site visits appear to depend on the research context, practical considerations, the level at which the review is being conducted and the intended impact of the report. Future research pertaining to the selection of site visitors, the autonomy of the visitors in data collection and report writing, and the amount and type of information provided would be particularly valuable.

1. Introduction

There is consensus that interdisciplinary approaches to research are necessary to address today's societal challenges (De Jong, Van Arensbergen, Daemen, Van der Meulen, & Van den Besselaar, 2011; Gray, 2008; Hansson, 2010). Interdisciplinarity involves the confluence of several different research streams to the comprehensive study of important issues. Within interdisciplinarity, societal problems are investigated from a variety of angles through the lenses of diverse disciplines (De Jong et al., 2011). In order for this type of research to thrive, however, the researchers from the different areas need to come together, learn each other's languages and research methods, and creatively move in new directions (Hansson, 2010; Toker & Gray, 2008). This integrative space is generally a research center and the development of research centers is bourgeoning both through federal funding requirements, such as the National Science Foundation's Science of Learning Centers or the National Institutes for Health's Clinical Translational Science awards, and within universities (Bulter & McAllister, 2011; Toker & Gray, 2008; Youtie & Corley, 2011). Unfortunately the funds presently available for research are severely limited (Rosbash, 2011). Limited resources result in pressure to carefully evaluate existing centers and to only provide funding for the most effective and productive ones (Hansson, 2010). The question remains, however, as to what is the best way to provide this evaluation.

In their edited *New Directions for Evaluation* volume, Coryn and Scriven (2008) point out that peer review and bibliometrics have dominated the evaluation of scientific research and that peer review is the predominant method used. They make a distinction between two types of peer review: one being a process used to evaluate individual researchers, research products, or research proposals (such as promotion and tenure reviews or the review of journal articles); the other being an expert panel evaluation. Hansson (2010) points out that expert panel review of research centers is common and expanding. This use of expert panels to review research centers mirrors the use of expert panels by educational accreditation associations and regulated industries which routinely conduct site visits for data and information gathering, and the use of expert panels in program evaluations in both the United States and internationally which often include site visits as part of the data collection strategy (Borum & Hanssen, 2000; Kaiser & Brass, 2010; Love, 1996; Segerholm, 2001). The use of site visits by expert review panels is particularly common in program evaluations conducted by United States government agencies. Since passage of the 1993 Government Performance and Results Act (GPRA), federal agencies have been required to measure performance and evaluate their progress. GPRA implementation guides and related materials have

* Corresponding author. Tel.: +1 612 625 2046; fax: +1 612 624 8241.
E-mail address: thao0181@umn.edu (M. Thao).

0149-7189/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.evalprogplan.2012.01.003

discussed expert review panels and site visits as part of a comprehensive performance measurement and evaluation strategy (Jordan, 2003; United States Department of Energy, 2006). As an example, the Department of Energy's Energy Efficiency and Renewable Energy program offers a guide to conducting peer review of projects in progress. Studies by the Government Accountability Office (GAO) found that numerous evaluation studies conducted by federal agencies used "judgmental assessment" or "experts' judgment of program effects" and that expert panel review and site visits were frequently used in evaluating federal program performance (GAO, 1998, 1999, 2003). Although internet searches identify documents with the search terms in any order within it, as a quick indication of frequency of the mention of site visits, a Google search (December 2011) was conducted for 'site visit review panel research program evaluation' which resulted in 4.79 million hits, with evaluation references from the Centers for Disease Control and Prevention and the American Association for the Advancement of Science topping the list. In terms of its validity, expert panel review is viewed as the gold standard for ascertaining research quality and other similar methods are often compared to it (Bulter & McAllister, 2011; Rons, DeBruyn, & Cornelis, 2011). The Committee on Science, Engineering and Public Policy suggested that expert review is the optimal method for evaluating applied research (COSEPUP, 2001). Furthermore, expert panel review has been shown to be an effective procedure for improving research behavior and providing summative information about research quality (Barbosa & Grayson, 2009; Lane et al., 2011).

Expert panel evaluation of entities, such as research centers, usually involves a site visit by the panel to the research center being reviewed. Commonly, these site visits are conducted by teams of experts who provide a connoisseurship approach to evaluation (Eisner, 1998). The selection of the expert site visitors is an important process. Site visitors not only need to have the expertise, but also the ability to decide what data and methods are appropriate, as well as the ability to gather such data (Lawrenz, Keiser, & Lavoie, 2003). Some common methods used in site visits include observations, presentations, document review, and interviews. Additionally, site visitors need to be prepared to conduct the visit. The site team traditionally prepares and provides extensive documentation of its activity both in written form and/or in presentations to the site visitors. Site visitors may ask questions of the site team members about the documentation, presentations, and other issues relevant to the evaluation of the quality of the site. Typically, the site visit is arranged by someone outside of the site who is responsible for funding or evaluating the site.

Expert panel evaluation sponsored by the funder is distinct from the evaluation provided for a center to track its progress. Quinlan, Kane, and Trochim (2008) provide a review of four large-scale federally funded research programs and suggest that the challenges in conducting evaluations of such centers might be addressed through participatory planning using concept maps, triangulation of evidence, and a systems approach. Youtie and Corley (2011) provide a detailed account of the effect of evaluation on one large research center in relation to the center's growth in understanding. They describe how centers might overcome the challenge of dealing with evaluation information and avoiding a 'vicious circle' of continuous restructuring. The focus of this paper, however, is on the process of implementing expert panel site visits, not on how the centers make use of the information or how decisions about the quality of the research should be made.

Evaluative site visits rely heavily on the expertise of the visitors, essentially providing the site visitors' personal answers to the evaluation questions based on the information they obtain during the visit, as filtered through their backgrounds and other experiences (Rons et al., 2011). The purposes of site visits include observing facilities, interacting with staff and students, accessing documentation, scrutinizing tangible evidence of student achievements, and checking the veracity of self-assessment statements (Harvey, 2009). In a methodological review of evaluative site visits, Lawrenz et al. (2003) define them as occurring when "...persons with specific expertise and preparation go to a site for a limited period for time and gather information about an evaluation object either through their own experience or through the reported experiences of others in order to prepare testimony addressing the purpose of the site visit" (p. 341).

Peer review is ubiquitous in scientific research (Foltz, 2000). In fact, recently a review panel was commissioned to review another review panel (Spotts, 2010). Review panel site visits are used routinely to make important decisions about the quality and worth of continuing programs or centers. Although there is research about the peer review process (Foltz, 2000; Langfeldt, 2006), it tends to focus on the criteria for and quality of the reviews or reviewers, not on the processes used to obtain them. As Foltz (2000) points out, there is no single form of the peer review process and the process should be evaluated. Especially little empirical research has been provided about the most appropriate methods for conducting review panel site visits (Coryn & Scriven, 2008; Lawrenz et al., 2003). Although there are various manuals about conducting site visits in accreditation processes, these do not provide research evidence for the structural choices in the processes (Commission on Accreditation, 2001; Commission on Accreditation for Respiratory Care, 2010).

Given the lack of attention to this aspect of evaluative site visits, the purpose of this study is to describe and analyze the processes used in two types of expert panel reviews of research centers, to compare and contrast the different approaches toward the goal of determining optimal site visit structures, and to identify areas where future research is needed to inform the improvement of the review panel site visit process.

2. Methods

Two different site visit processes were reviewed. The first was a process used by a university to evaluate four of its research centers and the second was a process used by a federal agency to evaluate its research center. The methods used to compare and contrast the different approaches were content analysis and reflective review by a participant observer. In the first example, the observer perspective is represented by the person commissioning the review; in the second, it is an expert panelist. The content analysis was facilitated by the creation of a cross-process comparison rubric based on the activities and products of the two review processes. This rubric provided the basis for Table 2 which compares the activities and products of the two review processes and highlights their similarities and differences.

3. Site visit examples

Two types of site visit processes are compared in the sections below. Table 1 displays the similarities and differences between two different agenda styles.

3.1. Expert panel review of four university research centers

The research centers developed detailed self-studies that included in-depth information about each center. The self-studies were provided to the site visit team well in advance of the visit. Additional information was provided as requested by the site visit teams. The site visit teams were moderately sized groups of three to five people with complementary expertise representing different institutions. Site visitors were chosen by administrators

Table 1
Typical site visit agenda for the two processes.

Day	Agenda for university site visits	Agenda for federal site visits
Day 1	Team dinner with one or two administrators	Team dinner with federal representatives Make writing assignments
Day 2	Team breakfast with university administrators Meet with the director of the center Meet with professional faculty Meet with graduate faculty Lunch with external advisory board Meet with center professional staff Meet with relevant collegiate deans Meet with heads and directors of related departments Meet with professional students Meet with graduate students and post docs Team dinner and evening work on draft report	Presentation by center director Presentations of initiatives by center staff Private meeting with team and funder Presentations Poster session Meet with students as a group Private meeting with team and funder
Day 3	Meet with center directors for any questions Preparation of final recommendations Lunch exit session with administrators and center directors Lunch exit session with administrators	Presentations Tour of facilities Presentations Private meeting with team and funder Meet with appropriate local administrators Team and funder determine any 'overnight' questions Team and funder private dinner and SWOT process
Day 4		Presentations of answers to overnight questions Continue SWOT process Write sections of report Share sections with other team members Final report put together Team reads the report aloud and agrees to wording Individual team members vote on continuing the center Team members sign off on final draft

sponsoring the review and were asked to recuse themselves if they believed they had a potential conflict of interest.

The site visit lasted about two days total, beginning with an evening meeting on the first day and ending around noon on the third day. Site visitors listened to presentations and met with various groups and individuals to conduct their assessments. At times, certain site visitors would meet independently with people connected to a particular aspect of the center. The site visitors were somewhat autonomous and could ask to meet with groups or individuals in addition to those scheduled.

Site visitors prepared a draft of issues that would be included in the site visit report before leaving the site and shared the draft with the administrators and center staff which allowed the opportunity for questions, clarifications and justifications. The final reports were prepared off-site by the visitors after the conclusion of the site visit. The final site visit reports were shared with the center staff who then prepared a formal response. Both the site visit report and the center's formal response were then shared with the administrators.

3.2. Expert panel review of a federally sponsored research center

The federally sponsored research center provided a variety of existing information and prepared new documentation and presentations for the site visit team. The site visit team was fairly large, consisting of 10 members with a variety of expertise and representing a variety of institutions. Site visitors were chosen by the federal agency and extensive checks for potential conflicts of interest with anyone working with the center were conducted. Approximately one week before the site visit, materials were made available electronically to the site visit team. These materials included reports from prior site visits with the center's responses, specific yearly implementation plans, examples of work produced by the center, the evaluation report from an external evaluator, and a proposal for continued funding.

The site visit continued over a span of four days, beginning with an evening meeting on the first day and ending just before supper on the fourth day. During dinner the first evening, writing assignments for the final site visit report were finalized. The first two full days of the visit were spent with the entire team and all of the center participants listening to formal presentations about each of the different aspects of the center followed by questions from the site team. The team also toured the facilities and observed the staff during the presentations. There were no small group or individual interactions except casual conversation during breaks. The site team met in executive session several times during the visit. These sessions included the site team and the federal agency representatives and any questions that arose during the sessions were posed to the center participants to provide answers. The center director attended one such session and responded to questions from the site team directly.

At the end of the second full day, the site team initiated a SWOT (strengths, weaknesses, opportunities, and threats) analysis based on their understanding of the center's work. This summary was used as the basis for report preparation and to identify any remaining questions about center operations or accomplishments. The next morning, the center participants made a final presentation responding to questions previously raised by the site visit team. Afterward, in an executive session including expert team members and federal administrators, the SWOT analysis was completed and the team members drafted their assigned sections. Each section had a lead and secondary author. Once all the parts were written, the entire report was projected and read aloud so that the team members could hear and reach consensus on the written material. This part of the process involved a great deal of debate and discussion. Once the process was completed, the site visit team departed. The federal agency kept the draft report which they edited and much later returned to the center along with other evaluative materials including a statement about their decision regarding continued funding.

4. Content analysis

Table 2 presents the processes described below following the 14-category rubric developed to facilitate the cross-process comparison. This tabular format highlights the similarities and differences in the two processes. The rubric shows that only three of the categories were identical across the two processes. The purpose of these site visits and the criteria imposed were the same—to recommend whether funding should be continued based on assessment of research productivity/focus, professional reputation, efficient use of funds, and widespread involvement. Additionally the preparation of the site visitors was the same in terms of providing written information before the visit. Despite having the same purpose, the way the site visits proceeded was different.

The period of time in which the site visits were conducted was very similar; two to four days, although the federal process was slightly longer in order to incorporate the finalization of the draft report. The selection and number of site visitors varied. The university center site visit teams included small groups of three to five experts, while the federal center site visit panel included 10 experts. The federal selection process placed much more attention on potential conflicts of interest than the university center selection process. In fact, the university center site visitors were often quite familiar with both the center being evaluated and with its staff.

The data provided were similar although the university center reviews were much more self-study oriented, whereas the federal center utilized prior reports that had been required for other purposes and from different sources. In both cases the amount of information was very large, requiring extensive reading. There were different data gathering formats during the site visits. The federal center site visit prioritized ensuring that all members of the panel heard the same information throughout the visit. Data presentations and question and answer sessions were very formal with all members of the center leadership, the federal representatives and the site visit team present. The only time the center staff members were not at a presentation was when the team and the federal representatives talked with the students who were connected with the center. In contrast, the university visits emphasized breaking up the site visit teams and having meetings with smaller groups of different types of people separately from the center leadership. The university review structure did not require uniformity in the information provided to review panel members, allowing variation in the agenda and smaller group discussions.

The preparation and delivery of reports were somewhat similar in that for both examples the teams had ideas of the final recommendations in mind before they left the site. In the federal process, the actual report was drafted before the site visitors left and the visitors had no opportunity to later reconsider what they said or how best to say it. Furthermore, the report was not shared with the center management prior to the visitors leaving, so there was no opportunity for clarification of misperceptions. In the university process, the report was not written but the main elements of the report were presented orally to both the funders and the center management simultaneously. This presentation allowed for further clarification of ideas and for correction of any misperceptions. The team members left the site and had the next few weeks to consider what they had learned and how best to convey that knowledge in a final report.

Table 2
Content analysis comparing the two expert panel processes.

Expert panel consideration	University process	Federal process
Purpose of site visit	Should support be continued	Should support be continued
Criteria for assessment of quality	Research productivity/focus, professional reputation, efficient use of funds, widespread involvement	Research productivity/focus, professional reputation, efficient use of funds, widespread involvement
Time frame for visit	2 days	4 days
Data gathering techniques during visit	Listen to presentations, ask questions of presenters, team breaks up to gather different data, interview individuals or small groups, observe center staff during presentations, tour facilities, funder not present	Listen to presentations, ask questions of presenters, team stays together, observe center staff during presentations, tour facilities, funder present
Selection of site visitors	3–5 experts in the research areas of the center	7–10 experts in the research areas of the center
Consideration of conflict of interest	Some conflict of interest considerations, panelists knowledgeable about the center and its work, could have worked with the center staff	Very formal conflict of interest considerations, mix of reviewers knowledgeable and not knowledgeable about the center
Preparation of site visitors	Written charge to the team, meet with funder before on site data gathering to frame needs	Written charge to the team, meet with funder before on site data gathering to frame needs
Autonomy of site visitors during the visit	Site visitors were in charge of the process and could change the agenda, time frames and activities	The site team followed the prearranged agenda
Roles of site visitors in preparation of final report	Site visitors determined their own roles in the report preparation	Roles in writing the final report were assigned by the funder with input from site visitors
Data provided to site visitors before and during visit	Materials from a self-study by the center staff	Materials from proposals or strategic plans by the center staff, written answers to specific questions from the funder, other reports prepared by the center or external entities (progress reports, evaluation reports, previous visit reports, etc.)
Participation of the funder in the site visit	Administrators met with team at beginning and end of visit and read reports	Federal program officers were present in all aspects of the visit
Preparation of reports	Rough draft in mind before leaving site with a verbal presentation of main points, final report 2–4 weeks later	Writing assignments for report sections, SWOT analysis after first day, report writing at the site, final draft before leaving
Delivery of report to site	Verbal report to site and funder before leaving the site, site gets change to respond to verbal comments in front of funder, written report to site for comments and any changes	Site does not receive report until additional information and considerations
Delivery of report to funder	Verbal report to site and funder and then to funder independently, written report to funder after site has a chance to respond to it	Funder is on site when draft report is finalized

5. Participant observer reflections

5.1. Reflections on the university process

Strengths of the university approach included conducting an extensive self-study; completion of the site visit in a short time frame; advice from experts in the field; meetings with the sponsoring administrators; individual, small group and large group interview sessions; site visitor input to the selection of interviewees; drafting and presentation of main ideas of the site visit report before leaving; and the opportunity to discuss potential findings with the center staff and administrators. Weaknesses included balancing prior opinions about the center, synthesizing information obtained by different site visitors in different settings, and a long time interval between the end of the site visit and the filing of the final reports.

After conducting the four university research center site visits, opinions of the strengths and weaknesses of the approach were obtained from site visitors, center staff, and administrators through debriefing conversations both in person and by telephone or participants from the centers felt that the self-study required to provide the information specified by the self-study outline was valuable and had helped them to better understand their programs. However, they also felt that it took a great deal of time that might have been more productively spent conducting the research of the center. Additionally, after the extensive self-reflection, much of the input that was ultimately provided in the site visit by the site team had already been identified by the center staff themselves. The site visitors felt that the amount of material provided was overwhelming and that it was impossible to read and understand it all in the required time frame. They suggested providing more limited information directly related to specific questions about the value judgments that were to be made as a result of the site visit. The administrators also felt that the process was cumbersome and required too many different meetings. Table 3 presents the self-study outline used during the site visits and the revised outline based on the feedback. All agreed the allowing time for the center participants to hear the potential main components of the report and to present their own opinions in the presence of the administrators were strengths of the process. This process helped to resolve any misperceptions. However, it also allowed for push-back and may have resulted in less critical reviews. Having site visitors that were familiar with the work of the center was also viewed as valuable and potential conflicts of interest or preconceived ideas were not seen as problems by the individuals interviewed. It should be noted, however, that the pre-existing relationships may have contributed to feedback that was more related to these relationships than to the material covered in the site visit. Additionally, prior familiarity with the center and the small number of experts on the team could have resulted in less comprehensive feedback in the final report. Finally, once the reviewers were off site, it was difficult for them to find time to write and revise the final report resulting in fairly long time-to-completion rates.

Table 3
Comparison of the initial and revised self-study report outlines.

Initial outline	Revised outline[a]
Cover page	Cover page
Executive summary	Executive summary (1–2 pages)
Background and history	Background and history (1 page)
Current research, education and public outreach activities	Scope, mission and accomplishments (3–7 pages)
Activities in building intellectual community	Strategies for building intellectual community (1–2 pages)
Resources	Resources (1–2 pages)
Organizational structure, governance and management	Organizational structure, governance and management (2–3 pages)
Major accomplishments and value added	Evidence of national and international stature (1–2 pages)
National and international stature	Vision statement (1–3 pages)
Strategic planning and future directions	Appendices
Summary statement	
Appendices	

[a] To keep the self-study report concise, page limits were added in the revised outline.

5.2. Reflections on the federal process

Strengths of this approach included finishing in a short time frame, making use of existing documentation, checking for potential conflicts, and obtaining consensus from a large group of people with diverse expertise. Another strength of the process was that it provided an opportunity for the funding agency to engage not only with the site visitors, but also with center personnel in presenting information and discussing implications. In this way, the funding agency directly and more completely understood the relationship between the information about the center and the opinions of the site visit team members. Developing the SWOT analysis before preparing the report focused the writing process and reduced the difficulty of different people composing the various sections of the report. It was also a strength that the report was written when all of the information was fresh in the site visit team's minds.

There was a large amount of data about the center available to the site visitors both before and during the site visit. The center staff was aware of potential weaknesses in their written materials and several presentations addressed issues the site visitors were likely to have questions about. The large group, formal presentation style precluded in-depth, detailed questioning but ensured that everyone received the same information. A key weakness of this approach is that there was really no opportunity for individual conversations with participants and the large group questioning format did not provide the opportunity for nuanced questions or individualized responses. In addition, not all of the participants in the center (a very large number of people) were able to attend the site visit so the perspectives presented to the site visitors were selective. Finally, the speed with which the report was written limited the site visitors' opportunity for thoughtful reflection and complex analysis prior to drafting the report. The large number of experts on the site team allowed for direct expert opinion feedback on a variety of areas of the center's operation which substantially contributed to the entire team's understanding. However, the expertise was sometimes not directly connected to the type of research conducted by the center; some of the experts were in a sense operating 'out-of-field'. Also the careful conflict of interest considerations ruled out site visitors that might be more knowledgeable about more connected issues. The report writing mirrored the expertise so individual sections were not always as connected as they could have been. With the large number of people with varied expertise, a single consensus opinion about whether the center should be funded was difficult to obtain. It was difficult for experts within an individual area to balance strengths and weaknesses of the center across areas to make a more overarching statement. In contrast, the large number of experts allowed rich discussions with careful consideration of a variety of issues.

6. Discussion

The purpose of this paper was to examine the expert panel review processes used in two different settings in order to characterize and consider the structure of the site visit process. The

14 different aspects that were included in the comparison rubric attest to the wide variety of questions and answers involved in commissioning an expert review panel, even when the reviews have the same general purpose.

As suggested by De Jong et al. (2011) and Bulter and McAllister (2011), the actual criteria for judging the research center should be contextualized to align with the goals of the center and the type of research being conducted. Many researchers (Bulter & McAllister, 2011; Coryn & Scriven, 2008; De Jong et al., 2011; Rons et al., 2011) have provided excellent suggestions and guidelines for criteria for judging research centers. Development of these types of criteria was not the focus of this study even though the review team site visit process should be structured to complement these purposes. This study was designed to highlight considerations in the structuring of a site visit. Based on this study, conditions that affect decisions about how to structure the review team site visit include practicality, level of the review, and intended impact of the review.

Many of the decisions about structuring a site visit review process appear to be based on pragmatic considerations. These include common practices, such as careful selection of experts, provision of information to the experts prior to the visit, a short time frame for the visit, presentations during the visit, and production of a final report. An example of a pragmatic decision would be the short time frame, 2–4 days, for conducting panel reviews which appears to be based on two considerations: first, the amount of time the experts can afford to be away from their primary responsibilities while on location at the site visit and second, the consideration of the effects on the site's productivity knowing that all or some of the day-to-day work of the center will take a back seat to the demands of hosting a team of expert visitors.

The level at which the review was conducted also affected the structure of the review process. The federal process was perforce very public and needed to control any perception of bias, hence the formal and strict conflict of interest procedures. Rons et al. (2011) suggest controlling for possible positive and negative bias as a key element of effective review of research. Langfeldt (2006) said a larger team with several experts mitigates against scholarly bias, "The number of reviewers indicates the amount and breadth of scholarly opinions invited to the decision making process, and is thereby a measure of thoroughness of review" (p. 36). Rons et al. (2011) also point out that controlling for negative bias is much less common. However in this case, negative bias was also considered because the reviewers were asked to explain relationships with any of the site members that they knew in any way, as opposed to the usual conflict of interest procedures where reviewers are only asked about positive bias such as if they have worked recently with a site member or stand to benefit in some way from a positive review of the site. Rons et al. (2011) suggest that the panel process by its very nature limits bias through team discussion.

Additionally, the open federal context required that the panel members be publically identifiable as experts. This need exacerbated the difficulty identified by Quinlan et al. (2008) of trying to find experts without conflicts of interest. Rons et al. (2011) point out that the choice of highly qualified active peers that cover all aspects of the research is the most important factor for a successful review of research. Therefore, the federal team was composed of several single discipline area experts who were not necessarily familiar with all of the content covered by the center. This provided a greater variety of expertise and perspectives; however, the visitors were less connected to the center's research overall. Rons et al. (2011) also suggest that more diverse expert backgrounds lead to more different and less expected suggestions. Reliance on scholarly, usually disciplinary, reputation may add difficulty to the assessment of interdisciplinarity within the center (Langfeldt, 2006). Quinlan et al. (2008) argued that when putting together teams to review interdisciplinary centers, reviewers with expertise in interdisciplinarity were needed as well as disciplinary experts. It has been argued that disciplinary experts have a conservative bias and are not as supportive of cutting-edge, transformative research (Foltz, 2000; Langfeldt, 2006; Scriven & Coryn, 2008). For example, Hansson (2010) states that because it uses the existing body of knowledge, the expert peer review system tends toward structural conservatism.

Because the university process was more internal and less public, it could afford to risk potential bias in the reviewers and thereby gain the advantage of using reviewers who had more prior knowledge of the center and its staff, as well as broad knowledge of all of the activities and products of the center. This breadth of knowledge coincided with the interdisciplinary recommendations of Quinlan et al. (2008). However, the risk of potential biases of the reviewers should be carefully considered, as reviewers, particularly those with an interest in the work of the center or prior connections to its staff, may be less likely to provide negative feedback.

The different sizes of the review teams used in the different processes affected the group process. In both processes reaching a consensus opinion about the quality of the center and recommendations for continuation was difficult especially in the larger federal team. Reviewers often hold different views and the outcome depends on what kind of expertise is included and the group dynamics (Langfeldt, 2006). Bolman and Deal (2008) suggest that to encourage high performance teams should be the smallest size that can get the job done, somewhere between 2 and 25. They suggest that the structural frame of a team stresses the critical link between specialization and expertise and that more members mean more structural complexity. Manners (1975) reports that smaller groups obtain consensus more easily. Bolman and Deal (2008) point out that although groups have more knowledge and diversity of perspective than individuals, they also can overrespond to social pressure or individual domination or personal agendas. Anyone planning an expert panel site visit should keep these issues in mind and more research about how to optimize group dynamics in expert panel site visits should be conducted.

The level at which the review was conducted was also related to the degree of reviewer autonomy. In the university setting the site visitors were much more in charge of the process with the administrators (funders) more content to let the external experts be their guides. The process was informal, flexible and the individual reviewers could operate somewhat autonomously. Although with such autonomy where visitors could meet with groups and individuals separately, all visitors did not have access to the same information and had to depend on each other's observations and perceptions that, again, presents the risk of potential bias. Additionally the reviewers might use this autonomy to pursue issues that are individually important to them but not critical to the operation of the center. In the federal process, there were very clear controls on how the process was to operate, a very formal feeling to the procedures and the interactions were always as a team. As Langfeldt (2006) suggests, government agencies attempt to guard against potential bias by following strict procedures. The larger size of the federal team probably also contributed somewhat to the more formal processes. This formality, however, precluded more richly nuanced discussions with researchers suggested as optimal in review of research by Hansson (2010) and Rons et al. (2011). It also mitigated against the team having interactions with stakeholders which was recommended as critical to a relevant review by De Jong et al. (2011). The university process, on the other hand, encouraged rich small group discussions with site members and interaction with stakeholders as recommended by Hansson (2010), De Jong et al. (2011), and Rons et al. (2011), although as mentioned above this type of discussion might contribute to biased results.

The desired impact of the report was the third consideration but it interacted with the level of the review. As Rons et al. (2011) suggest, goals set by university research management may be very different from those set by federal authorities. Although at both levels, the purpose of the report was a summative determination as to whether funding should be continued, the university process also specifically sought formative information.

Consequently, the university process provided for much more interaction between the site, the administrators and the site visitors over the report through the verbal reporting and the draft report reviews. This is in line with the views of Langfeldt (2006), Quinlan et al. (2008) and Hansson (2010) that letting the evaluees have input provides for a better quality review. Gray (2008) also suggests that feedback should be timely and although the final report was produced quite a while after the site visit, the oral reporting was immediate. The reporting pattern also provided the university administrators with varied views on the research center. In support of their decision, they heard a verbal report, heard verbal reactions to the report, received a written site visit report that allowed time for reflection based on verbal reactions to the initial report ideas and received a final written reaction from the site. Rons et al. (2011) recommends this type of multi-level reporting which provides public reports as well as open more confidential advice. Oral reporting is also recommended in site visit manuals for accreditation (Commission on Accreditation, 2001; Commission on Accreditation for Respiratory Care, 2010). In contrast, the federal process allowed little interaction of the site members with the site visitors during the report writing. The only input from the site to the actual report was answers to 'overnight' questions provided formally as a presentation to the site visit team as a whole. Additionally, the federal process required the report to be written before the site visit was over. While the short amount of time in which site visits occur is an issue in it of itself, Mitchell (1990) points out that the validity of site visits can be substantially affected by the lack of independence of site visitors and that preparing a report during the limited time allocated for the visit can result in less thoughtful conclusions. Rons et al. (2011) also point out the need for sufficient time to make valid decisions. Perhaps in some compensation for the interaction that was part of the university process, the federal personnel were in attendance throughout the site visit and had access to everything the expert panelists saw, heard or did. The university funders saw the center more through the eyes of the reviewers, whereas the federal funders also gathered their own information. Additional research on reporting of expert panel reviews is needed and should include better understanding of the appropriate delivery and composition of final reports and how different issues and contexts affect their quality.

The amount of information provided in both processes was daunting. In the case of the university process, the self-study outline was completely revised and substantially shortened after the series of center reviews. This helped to ease the burden both on the visiting team and on the center. Use of a self-study to produce the information used by the expert panel was reported by center staff as valuable but it also took some of the thunder away from the findings of the panel, since the strengths and weaknesses had already been identified. Hansson (2010) points out the value of a self-study in the review of research centers. The federal process made use of several types of pre-existing information from a variety of sources but this did not decrease the overall amount of information required to be read by the panel members. Any information provided to site visit teams must be of high quality because reviewers do not have the time necessary for in-depth analysis of it (Foltz, 2000). If more than one type of evidence is provided, the reviewers must cross-check data sources or the site needs to do it for them (Quinlan et al., 2008). It appears because there is little research on which type of information (e.g., self-studies, external reports) or how much information is optimal for making decisions about research centers, any information that is available is provided to the site visit team. More research about this is necessary.

The use of pre-existing reports helped somewhat with the burden on the center for the particular visit but as so eloquently pointed out by Youtie and Corley (2011) in their study of one center, the production of all of these types of reports was a continuous burden for the center and could lead to inefficiency in research productivity. In contrast to Quinlan et al. (2008) who recommend centers be externally reviewed every five to seven years, Scriven and Coryn (2008) claim centers should be accountable after the first three and then every 2 years after that. Federal centers are often externally reviewed every year in addition to internal and external evaluation efforts, annual center progress reports and renewal proposals. Because the goal of this paper is to consider the expert review panel process, these different ideas on timing raise the issue of how often these reviews should take place. The two examples discussed here were conducted at one point in time and so do not provide data to address this question; therefore it should be investigated through future research.

In summary, evaluative site visits are exceedingly common and are used to make high-stakes decisions about research centers, as well as other entities and programs. Unfortunately, there is very little research examining the details of the processes that can guide the development of effective evaluative site visits. Decisions about how to structure site visits appear to depend on the research context, practical considerations, the level at which the review is being conducted and the intended impact of the report. Three issues emerged as both critical for individuals who commission expert panel reviews to consider and as fruitful areas for future research: the selection of the site visitors (e.g., numbers, types of expertise, objectivity, bias, openness), the autonomy of the visitors in data collection and report writing, and the amount and type of information provided.

References

Barbosa, M., & Grayson, D. (2009). Site visits: Assessing and improving the climate for women in physics. In Harline, B. K., Horton, K. R., & Kaicher, C. M. (Eds.), *AIP conference proceedings* (Vol. 1119, pp. 14–15). American Institute of Physics doi:10.1063/1.3137748.

Bolman, L., & Deal, T. (2008). *Reframing organizations: Artistry choice and leadership* (4th ed.). San Francisco: Jossey Bass.

Borum, F., & Hanssen, H. F. (2000). The local construction and enactment of standards for research evaluation. The case of the Copenhagen business school evaluation. *Evaluation, 6*, 281–299.

Bulter, L., & McAllister, I. (2011). Evaluating university research performance using metrics. *European Political Science, 10*(1), 44–58.

Commission on Accreditation. (2001). *Council on chiropractic education site visit team manual: Guide for site team chairs and members of visit teams*. Scottsdale, AZ: Council on Chiropractic Education.

Commission on Accreditation for Respiratory Care. (2010). *Site visit process*. Bedford, TX: Commission on Accreditation for Respiratory Care. Retrieved from: http://www.coarc.com/20.html.

Committee on Science, Engineering and Public Policy, National Academy of Sciences, National Academy of Engineering, Institute of Medicine Policy and Global Affairs (2001). *Implementing the government performance and results act for research: A status report*. Washington, DC: National Academy Press. Retrieved from the National Academies Press website: http://www.nap.edu/catalog/10106.html

C. L. S. Coryn, & M. Scriven (Vol. Eds.), (2008). Reforming the evaluation of research. *New directions for evaluation, No. 118*California: Jossey-Bass.

De Jong, S. P. L., Van Arensbergen, P., Daemen, F., Van der Meulen, B., & Van den Besselaar, P. (2011). Evaluation of research in context: An approach and two cases. *Research Evaluation, 20*(1), 61–72.

Eisner, E. W. (1998). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New Jersey: Merrill.

Foltz, F. (2000). The ups and downs of peer review: Making funding choices for science. *Bulletin of Science Technology & Society, 20*(6), 427–440.

General Accounting Office (GAO). (1998). Program evaluation: Agencies challenged by new demand for information on program results. GAO/GGD-98-53. Retrieved from: http://www.gao.gov/archive/1998/gg98053.pdf

General Accounting Office (GAO). (1999). Federal research: Peer review practices at federal science agencies vary. GAO/RCED-99-99. Retrieved from: http://science.energy.gov/~/media/bes/pdf/rc99099.pdf

General Accounting Office (GAO). (2003). Program evaluation: An evaluation culture and collaborative partnerships help build agency capacity. GAO-03-454. Retrieved from: http://www.gao.gov/new.items/d03454.pdf

Gray, D. O. (2008). Making team science better: Applying improvement-oriented evaluation principles to evaluation of cooperative research centers. In (Series Ed.) & C. L. S. Coryn, & M. Scriven (Vol. Eds.), *New directions for evaluation, No. 118: . Reforming the evaluation of research* (pp. 73–87). .

Hansson, F. (2010). Dialogue in or with the peer review? Evaluating research organizations in order to promote organizational learning. *Science and Public Policy, 37*(4), 239–251.

Harvey, L. (2009). Analytic quality glossary. Retrieved from: http://www.qualityresearchinternational.com/glossary/sitevisit.htm

Jordan, G. (2003). GPRA Tutorial Section 4: Program Performance Evaluation, presentation to the Sixth Annual Performance Conference of the National Academy of Public Administration, Washington, DC. Retrieved from the Environmental Protection Agency website: http://www.epa.gov/evaluate/pdf/gpratutor.pdf

Kaiser, F., & Brass, C. (2010). *Independent evaluators of federal programs: Approaches, devices and examples*. Washington, DC: Congressional Research Service (CRS).

Lane, J., Wade, J., Down, L., Bonnington, S., Holding, P., Lennon, T., et al. (2011). A peer review intervention for monitoring and evaluating sites (PRIME) that improved randomized controlled trial conduct and performance. *Journal of Clinical Epidemiology, 64*(6), 628–636.

Langfeldt, L. (2006). The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation, 13*(1), 31–41.

Lawrenz, F., Keiser, N., & Lavoie, B. (2003). Evaluative site visits: A methodological review. *American Journal of Evaluation, 24*(3), 341–352.

Love, A. (1996). Visits to the world of practice. *Evaluation, 2*, 349–361.

Manners, G., Jr. (1975). Another look at group size, group problem solving and member consensus. *The Academy of Management Journal, 18*(4), 715–724.

Mitchell, R. (1990). Site visits in the accreditation process of the Western Association of Schools and Colleges (WASC). *Evaluation and Research in Education, 4*, 75–79.

Quinlan, K. M., Kane, M., & Trochim, W. M. K. (2008). Evaluation of large research initiatives: Outcomes, challenges, and methodological considerations. In (Series Ed.) & C. L. S. Coryn, & M. Scriven (Vol. Eds.), *New directions for evaluation, No. 118: . Reforming the evaluation of research* (pp. 61–72). .

Rons, N., DeBruyn, A., & Cornelis, J. (2011). Research evaluation per discipline: A peer-review method and its outcomes. *Research Evaluation, 17*(1), 45–57.

Rosbash, M. (2011). A threat to medical innovation. *Science, 333*, 136.

Scriven, M., & Coryn, C. (2008). The logic of research evaluation. In (Series Ed.) & C. L. S. Coryn, & M. Scriven (Vol. Eds.), *New directions for evaluation, No. 118: . Reforming the evaluation of research* (pp. 73–87). .

Segerholm, C. (2001). National evaluations as governing instruments: How do they govern? *Evaluation, 7*, 427–438.

Spotts, P. (2010). IPCC climate change panel needs transparency, review panel finds. The Christian Science Monitor. Retrieved from: http://www.csmonitor.com/Environment/2010/0830/IPCC-climate-change-panel-needs-transparency-review-panel-finds

Toker, U., & Gray, D. (2008). Innovation spaces: Workspace planning and innovation in U.S. university research centers. *Research Policy, 37*, 309–329.

United States Department of Energy (DOE) and Oak Ridge Associated Universities. (2006). The performance-based management handbook: A six-volume compilation of tools and techniques for implementing the government performance and results act of 1993. Retrieved from: http://www.orau.gov/pbm/pbmhandbook/Volume%204.pdf

Youtie, J., & Corley, E. (2011). Federally sponsored multidisciplinary research centers: Learning, evaluation, and vicious circles. *Evaluation and Program Planning, 34*(1), 13–20.

Frances Lawrenz, PhD is a professor of evaluation in the Quantitative Methods in Education Program in the Educational Psychology Department at the University of Minnesota, Twin Cities. Her specialty is in science, technology, engineering, and mathematics (STEM) program evaluation. She is also the Associate Vice President for Research at the University.

Mao Thao is a student in the Quantitative Methods in Education PhD. Program with an emphasis in evaluation within the Department of Educational Psychology at the University of Minnesota, Twin Cities. She has worked as a program evaluator for five years, primarily in mental health and youth program evaluation. Her interests in evaluation include culturally responsive evaluation, mixed methods, and instrument design. Mao received a Bachelor of Science in Sociology and a Bachelor of Arts in Communication Studies from the University of Minnesota, Twin Cities.

Kelli Johnson is a doctoral candidate in Evaluation Studies in the Department of Organizational Leadership, Policy, and Development in the College of Education and Human Development at the University of Minnesota, Twin Cities. She is also a Senior Research Fellow in the School of Public Health at the University of Minnesota where she serves as the Center Director of the State Health Access Data Assistance Center (SHADAC), a research and technical assistance center that helps states utilize data to inform health policy and the implementation of health reform at the state level.