



Learning Together: Proceedings, Evaluation, and Applying Lessons Learned

National Park Service
Interpretation and Education
Evaluation Summit
Denver, Colorado
October, 2006





“Creating a culture of evaluation will be a key piece of taking the NPS from good to great.”

—DAN RITCHIE, CHAIR

NATIONAL PARK SYSTEM ADVISORY BOARD EDUCATION COMMITTEE

1916



2016



Learning Together: Proceedings, Evaluation, and Applying Lessons Learned

National Park Service
Interpretation and Education
Evaluation Summit
Denver, Colorado
October, 2006

Prepared by:

The Education Evaluation Coordination Team of the
National Education Council, National Park Service,
in collaboration with
Michael Duffin, Ph.D. and Catey Iacuzzi
through a Cooperative Agreement between
the NPS Conservation Study Institute and Shelburne Farms

Online: www.nps.gov/interp/evaluation

Opposite: An interpretive program at Delaware
Water Gap National Recreation Area engages
young visitors.

Cover: Podcasts enhance interpretation at
Richmond National Battlefield Park.

Acknowledgements

The Interpretation and Education Evaluation Summit would not have been possible without the generous, visionary, and multifaceted contributions of the following supporters:

Dan Ritchie and the **National Park System Advisory Board**; the **National Park Foundation** (special mention: **Vin Cipolla, John Reynolds**) the **NPS National Leadership Council** (special mention: **Mary Bomar**); and the **NPS National Education Council**.

Thanks also goes to **Julia Washburn, Sheri Forbes, and Loran Fraser** for the big picture leadership and nuts and bolts detail support they provided for the broader effort of which this Summit was one piece.

The bulk of the design, implementation, and evaluation of the Summit were conducted by the Education Evaluation Coordination Team, a working group of the NPS National Education Council.

This group includes the following individuals: **Deanne Adams, Neil DeJong, Michael Duffin, Lakita Edwards, Sheri Forbes, Elizabeth Hoermann, Cynthia MacLeod, Megan McBride, Nora Mitchell, Anne**

O'Neill, Amy Powers, Patti Reilly, Sam Vaughn, and Julia Washburn.

Nowhere does it become clearer, however, that this event was the result of the work of an extensive team of talented specialists, than in reviewing the following list of additional friends who provided invaluable content and technical support:

Craig Ackerman, Les Baxter, Marcia Blaszak, Brad Bennett, Tracy Bowen, Jim Boyd, Beth Boland, Lyn Carranza, Diane Chalfant, Judy Chetwin, Delia Clark, Theresa Coble, Lynn Dierking, Allison Druin, Eric Epstein, Robert Fudge, Jim Gasser, Jim Gramann, Flip Hagood, Gayle Hazelwood, Pat Hooks, Rodney Hopson, Catey Iacuzzi, Chris Jarvi, Jon Jarvis, Doug Knapp, Richard Kohen, David Larsen, Jennifer Lee, Paul Lee, Diane Liggett, Linda Lutz-Ryan, Jan Matthews, Alden Miller, Martha Monroe, Richard Nichols, Polly Nordstrand, Emma Norland, Michael Quinn Patton, Megan Richotte, Tom Richter, Shirley S. Smith, Mike Snyder, David Sobel, Antonio Solorio, Carol Stapp, Hazel Symonette, Veronica Thomas, the University of Denver, Gay Vietzke, U.S. Media, Jon Wergin, and Don Wollenhaupt.

Contents

EXECUTIVE SUMMARY 1

PART I: PROCEEDINGS OF THE SUMMIT 5

Designing the Summit 6

Setting the Context 7

Comments from Summit Host, Mr. Dan Ritchie: 7

Comments from NPS Director, Honorable Mary Bomar: 7

Comments from Summit Facilitator, Dr. Michael Quinn Patton 9

Summaries of Presentations and Discussions 10

Day One Welcome and Introductions 10

Panel One Presentations: Why Create a Culture of Evaluation? 11

First Large Group Discussion with Panelists and Participants 13

Panel Two Presentations: Evaluation and Place-Based Learning 13

Second Large Group Discussion with Panelists and Participants 15

Day One Wrap-up 16

Day Two Opening Remarks 16

Introducing the Servicewide I&E Evaluation Strategy 16

Criteria for Evaluation Pilot Projects: Fishbowl Discussion 17

Criteria for Evaluation Pilot Projects: Small Group Discussions 19

Summit Closing Remarks 19

Interactive Teleconference 20

Participants 21

Additional Materials 21

PART II: EVALUATION OF THE SUMMIT 23

Methods 24

Findings and Discussion 24

Section A: Two Major Themes, and Two Sub-themes 24

Section B: Accomplishment of Short-term Summit Outcomes 29

Section C: Tracking a Culture of Evaluation Over Time 31

PART III: APPLYING LESSONS LEARNED 37

Recommended Immediate Actions 39

Recommended Short-term Actions 39

Recommended Long-term, More Comprehensive Actions 40

Closing Reflection on Recommended Action Steps 42

REFERENCES 43

APPENDICES A-1

APPENDIX A: Complete Summit Participant List A-2

APPENDIX B: Summit Reaction Form (including selected results) A-5

APPENDIX C: Summary of Pilot Project Criteria Discussions A-10

APPENDIX D: "Chiefs" Survey Executive Summary A-12

APPENDIX E: Pre-Summit Statements From Panelists A-16

APPENDIX F: Evaluation Strategy Implementation Matrix A-48



“We need a ‘toolbox’ . . . and training and information about how to do evaluation.”

—SUMMIT PARTICIPANT

Summit panelists present experience and research about the benefits and challenges of evaluation.

Opposite: Members of the NPS National Education Council and National Interpretive Advisory Council

Executive Summary

The overarching goal of the Summit was to generate useful dialogue about “creating a culture of evaluation” within Interpretation and Education characterized by continuous learning and decision-making based on audience analysis and outcome data.

THE SUMMIT

The National Park System Advisory Board, the National Park Foundation, and the National Park Service (NPS) convened an “Interpretation and Education Evaluation Summit” at the University of Denver, Colorado, on October 25 and 26, 2006. This event brought together education, evaluation, and organizational development experts from across the country with a wide range of NPS stakeholders. Participants included members of the National Park System Advisory Board, the NPS National Leadership Council (NLC), NPS deputy regional directors, the current and three former NPS directors, several NPS partners, NPS regional chiefs of interpretation and education, and other NPS field staff from across the country. Collectively, more than 130 people worked together to better understand how to use evaluation to create a vital and relevant future for the Interpretation and Education Program in achieving the mission of the National Park Service.



The overarching goal of the Summit was to generate useful dialogue about “creating a culture of evaluation” within the Interpretation and Education Program characterized by continuous learning and decision-making based on audience analysis and outcome data.

HISTORICAL CONTEXT

The Evaluation Summit was one link in a series of actions that the National Park Service is taking to reinvigorate itself as it heads into its second century of service. During a historic general conference seven years ago (*Discovery 2000*, held in St. Louis, Missouri), the NPS reaffirmed the critical role of interpretation and education in conservation, particularly in the context of globalization and America’s changing demographics. Shortly thereafter, the National Park System Advisory Board issued its defining report: *Rethinking the National Parks for the 21st Century*, urging the NPS to embrace its role as a national education institution. The NPS National Leadership Council responded by conducting a series of six education seminars, resulting in publication of *Renewing Our Education Mission*. This led to the formation of the NPS National Education Council (NEC) and a call to establish a comprehensive

Interpretation and Education Program Business Plan, which was released in early 2007. Additionally, a *Scholar’s Forum on Civic Engagement* was held in January 2006.

The critical role of education was reinforced at each step along the way. Standards, goals, and priorities were clarified, and evaluation increasingly became viewed as an essential component of the overall effort. In October 2006 the National Leadership Council unanimously endorsed the *Interpretation and Education*

Renaissance Action Plan that was developed by the National Education Council to realize the tactics described in the evolving business plan. This true “Renaissance” has five important pillars: Standards, Access, Technology, Partnerships, and Evaluation. In concert with the action plan, a subcommittee of the NEC has drafted a *Servicewide Interpretation and Education Evaluation Strategy*. The Evaluation Summit was a first step in implementing this evaluation strategy.

Collectively, these steps aim to move the NPS from good to great in its ability to engage the public with their national parks in new, dynamic, and relevant ways.

SUMMARY OF PART I: PROCEEDINGS OF THE SUMMIT

This part of the report provides a historical record of what happened at the Evaluation Summit and gives readers a vicarious sense of how the event unfolded. It is also written as an invitation to readers beginning to contemplate what evaluation might mean for them in their NPS context.

Dan Ritchie, Chairman of the National Park System Advisory Board Education Committee, hosted the event. Mr. Ritchie claimed that: “The survival of the National Park System in the twenty-first century depends on how it interacts with society and how much society values it.” Further: “Creating a culture of evaluation will be a key piece of taking the NPS from good to great.”

Newly appointed NPS Director, Mary Bomar, delivered her support in the keynote address, stating that “this Evaluation



“Too often [evaluation] work is at least perceived and received as standing in judgment of rather than working in deliberative collaboration with . . .”

—HAZEL SYMONETTE, SUMMIT PANELIST



“We need to care about the ‘invisible’ people.”

—POLLY NORDSTRAND, SUMMIT PANELIST



“Why should we do [evaluation]? It’s good business. In very practical terms, it prepares you for opportunity.”

—FLIP HAGOOD, SUMMIT PANELIST



“Ranger-led programs far surpassed any other type of programs as the . . . number one most meaningful program [in our study].”

—THERESA COBLE, SUMMIT PANELIST



“The most exciting thing for me is that evaluation means continual learning.”

—LYN CARRANZA, SUMMIT PARTICIPANT



“We need a culture of evaluative thinking as a way of doing business, not only in interpretation, but throughout the . . . National Park Service.”

—MARY BOMAR, NPS DIRECTOR

Summit is the beginning of our Interpretation and Education Renaissance . . . and an important first step in looking ahead to our Centennial.” Additionally, she claimed, “We also need a culture of evaluative thinking as a way of doing business, not only in interpretation, but throughout the disciplines within the National Park Service.”

Renowned evaluation expert, Dr. Michael Quinn Patton, facilitated the Summit. He noted that “evaluation findings and processes are more likely to be useful when there is strong leadership support for evaluation, when the organizational culture supports inquiry, reality-testing, and learning, and when people throughout the organization value and demonstrate evaluative thinking.”

Day One of the Summit was organized around two panels, during which 14 guest experts presented experience and research about the benefits and challenges of evaluation. Most of the agenda was reserved for dialogue among and between panelists and participants in response to panelist presentations. Topics emerging from these discussions included the following:

- Holding people accountable for learning rather than results
- Practical concerns about implementing evaluation (e.g., flexible planning; involving field staff, partners, and other stakeholders; risk and innovation)
- The role of technology in place-based learning
- Cultural competence
- Evaluating visitor experiences

Day Two of the Summit targeted more tactical discussions. This included introduction of the draft *Servicewide Interpretation and Education Evaluation Strategy*, and beginning to define the selection criteria for the evaluation pilot projects called for in the *I&E Renaissance Action Plan*. The most common sentiment in reports from small group discussions was the importance of including diverse parks and audiences in pilot evaluations.

Immediately after the formal close of the Summit, NPS leaders and partners conducted an interactive teleconference. One hundred twenty individual NPS staff from around the country logged in to view this Tel, making it the second most watched interpretation and education Tel in FY 07.

SUMMARY OF PART II: EVALUATION OF THE SUMMIT

A participatory and highly collaborative approach was used for evaluating the Summit in order to model organizational learning and a user-focused approach. The evaluation was accomplished through a public-private partnership that combined knowledge of NPS interpretation and education programs with professional evaluation expertise. Data were obtained from a Summit reaction form, small group notes, participant question cards, lunchtime “scribbles” of questions and ideas, a previous survey of NPS evaluation practices, field notes, transcripts, and observations.

Themes

The following themes emerged from analysis of Summit evaluation data:

Major Theme 1 - Participants seemed enthusiastically engaged in the concept of creating a culture of evaluation.

Major Theme 2 - Participants voiced concern about how such a change will be implemented.

Additional sub-themes included discussions about terminology and language and the importance of building a culture of evaluation around existing NPS values of inclusion.

Intended Short-Term Outcomes

Before the Summit, the planning team prioritized five short-term outcomes to guide the design of both the event agenda and evaluation of the Summit itself. All five short-term outcomes were accomplished, though to varying degrees. These outcomes include (paraphrased, and presented in rank order from strongest to weakest levels of evidence):

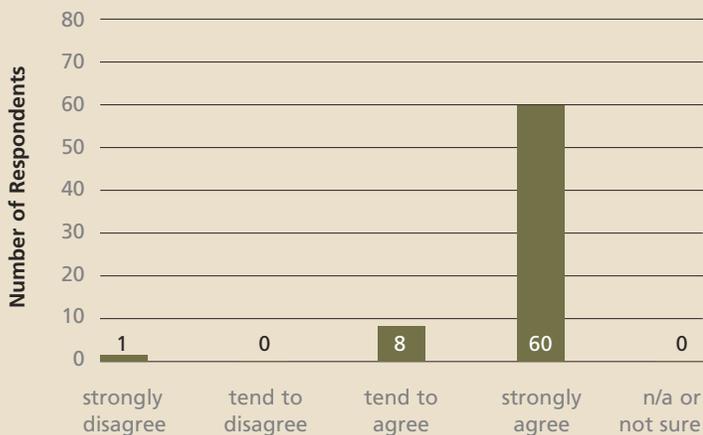
1. Enhance leadership support for evaluation.
2. Increase confidence in cost-benefit of evaluation.
3. Commence work on pilot evaluation projects.
4. Promote making decisions based on outcome data.
5. Develop action steps.

Tracking a Culture of Evaluation Over Time

This report provides a rough, concise snapshot of the current culture of evaluation within the interpretation and education community as a baseline for future comparison. Three relatively replicable metrics are used: average responses to several survey items; documentation of some behaviors of a few key groups of stakeholders; and application of two theoretical frameworks from the research literature.

This evaluation concludes that, as of the end of 2006, the interpretation and education function within the NPS is poised at a threshold of potential cultural change, but has not yet demonstrated systemic changes.

Endorsements for Increased I&E Program Support



Responses to item from Summit Reaction Form: "I enthusiastically endorse increased support for evaluation-related activities within I&E."

SUMMARY OF PART III: APPLYING LESSONS LEARNED

Evaluation findings from the Summit generally reinforce both the *I&E Renaissance Action Plan* and the draft *Servicewide Interpretation and Education Evaluation Strategy*, while also adding some new insights. An overall implication for practice emerging from evaluation of the Summit is that next steps for cultivating a culture of evaluation within the NPS should NOT assume a "one size fits all" approach. Evaluation activities and strategies should be segmented and designed specifically to meet the needs of stakeholders in different stages of change and innovation adopter categories.

In an effort to demonstrate the kind of learning that can emerge from a culture of evaluation, and to meet the needs of personnel who are responsible for authorizing and implementing the evaluation strategy, this section of the report provides 13 specific recommended actions:

Immediate Actions (next three months):

- A. Solicit feedback on recommended next steps from the National Leadership Council;
- B. Complete, distribute to the field, and solicit feedback on proceedings (including evaluation) and DVD of Summit; and
- C. Solicit feedback from partners on Summit proceedings (including evaluation), the DVD of the Summit, and on the evaluation strategy.

Short-term Actions (up to twelve months):

- D. Build a Web portal/evaluation resources library with *practical* tools for parks and partners;
- E. Enhance communication about evaluation within NPS and with partners;
- F. Establish selection criteria and the process to identify pilot evaluation projects to be considered as funds become available; and
- G. Fill vacant GS-13 Evaluation and Visitor Studies Coordinator position.

Long-term, More Comprehensive Actions (one to five years):

- H. *Systematically* share lessons learned from existing evaluation projects;
- I. Incorporate evaluation more tangibly into existing professional development opportunities, training programs, and reward systems;
- J. Require an evaluation component for all funding sources in the Servicewide Consolidated Call (SCC) (selection criteria to include low-cost options);
- K. Create a mini-grants program to promote small scale evaluation into questions of local interest;
- L. Provide resources to systematically involve historically underserved audiences and communities in evaluation work; and
- M. Make available individualized technical assistance and support for evaluation "champions" in the field and at the national level.



“ Our parks are not just special places for Americans, but they are special places for the entire world . . . ”

—MARY BOMAR, NPS DIRECTOR

Part I: Proceedings of the Summit

This part of the report provides a historical record of what happened at the Evaluation Summit. It opens with a brief explanation of the factors that went into crafting the agenda and outcomes, and then sets the stage with introductory and summary comments from three central participants. The majority of Part I provides concise summaries of the various presentations and discussions that occurred over the course of the day-and-a-half event, and closes with a summary of participant demographics and a brief note on how to access additional Summit-related materials.

The primary purpose of this section is to give readers who did not attend the Summit a vicarious sense of how the event unfolded. It can also serve as a basic orientation to evaluation concepts for readers who are beginning to contemplate what a culture of evaluation might look like in their part of the National Park Service world.

Opposite: Interpreters at Zion National Park conduct tours on alternative-fuel shuttle buses.

Proceedings of the Summit

This Summit was an integral step toward the long-term goal of creating a culture of evaluative thinking throughout the Interpretation and Education staff of the National Park Service.

DESIGNING THE SUMMIT

The Evaluation Summit agenda was crafted to allow maximum space for dialogue. Presentations by outside experts were kept very short, with most of the time reserved for plenary discussion. Participants were carefully chosen to represent a wide range of stakeholder roles to ensure diverse representation in the discussions. The Summit itself was evaluated by a collaborative team charged with capturing the rich variety and depth of the dialogue.



Participants share innovative ideas in breakout sessions.

Another key criterion used for designing the Summit agenda was to *do* rather than just *talk about* utilization-focused evaluation. Thus, the general flow of the Summit went from vision-focused on the first day toward a more tactical emphasis on the second day.

This Summit was an integral step toward the long-term goal of creating a culture of evaluative thinking throughout the Interpretation and Education staff of the National Park Service. Such a culture would be characterized by continuous inquiry and learning and decision-making based on using various types of outcome data. In order to demonstrate effective evaluation practices, the planners of the Summit prioritized five short-

term outcomes to guide the construction of the Summit agenda, and to serve as a basis for evaluation of the Summit itself. These included:

1. Develop clear, realistic, prioritized action steps for rolling out an evaluation strategy within the context of existing resources for Fiscal Year 2007;
2. Generate enthusiastic National Leadership Council (NLC) endorsement to move toward increased support for evaluation and social science research (e.g., pilot projects, incorporating

evaluation components into existing activities, and hiring a national evaluation coordinator);

3. Begin implementing pilot evaluation projects as specified in the *I&E Renaissance Action Plan*, in support of action steps outlined in the *Servicewide I&E Evaluation Strategy*;
4. Document evidence of park-level managers seeking and using existing and new outcome data to inform decisions (e.g., budget, program, staffing);
5. Inspire increasing confidence at all levels within the NPS that evaluation can help and support continuous learning and improvement of programs and decisions in a way that is not overly burdensome.

EVALUATION SUMMIT AGENDA

Wednesday, October 25, 2006

8:00–8:30	Welcome and Introductions
8:30–9:45	Panel One Presentations: Why Create a Culture of Evaluation?
10:05–Noon	Large Group Discussion
1:00–2:10	Panel Two Presentations: Evaluation and Place-Based Learning
2:35–4:00	Large Group Discussion
4:25–5:00	Day One Wrap-up
5:00–6:00	Reception

Thursday, October 26, 2006

8:00–8:40	Opening Remarks
8:40–9:10	Servicewide Interpretation and Education Evaluation Strategy
9:10–10:40	Fishbowl Discussion: Criteria for Pilot Evaluation Projects
10:40–11:45	Small Group Discussions: Criteria for Pilot Evaluation Projects
11:45–Noon	Closing Remarks
1:30–3:30	Interactive Teleconference

SETTING THE CONTEXT

The comments of three key participants help to frame some of the personal, political, and professional dimensions of this historic event. First, the Summit was hosted by Mr. Dan Ritchie, Chairman of the Education Committee of the National Park System Advisory Board, and Chairman of the Board of the University of Denver where he served as Chancellor for 16 years. His text is synthesized from his written comments distributed in the briefing book and transcripts of his various comments during the event. Second, Honorable Mary A. Bomar, seventeenth Director of the National Park Service, welcomed participants on the first day of the Summit. Her comments included here are distilled from the transcript of the opening remarks she delivered in person. The third key figure whose comments set the context for these proceedings is the overall facilitator of the event, Dr. Michael Quinn Patton. He is Founder and Director of Utilization-Focused Evaluation, an independent organizational and evaluation consulting firm. The text included here represents his summary reflections provided in writing after the Summit was complete.

Comments from Summit Host, Mr. Dan Ritchie

The National Park System is more than the places and objects central to the heritage of the United States. The National Park System embodies intangible meanings—beauty, health, wonder, freedom, democracy, and struggle—that are central to our collective identity. The survival of the National Park System in the twenty-first century depends on how it interacts with society and how much society values it. The Interpretation and Education Program is the primary means by which the National Park Service engages diverse publics with their national parks, provides access to meanings, establishes relevance, and connects people and communities to national heritage.

It is critical that the NPS Interpretation and Education Program be strong, vital, flexible, and effective. To that end, it is exciting that the NPS is in the midst of an Interpretation and Education Renaissance—a commitment to build on existing success, and to learn, grow, and respond to our changing society. The National Park System Advisory Board Education Committee is honored to be a catalyst for this Renaissance and is pleased to see the Director and the National Leadership Council forming a strong partnership with regional, park-level, and field leaders within Interpretation and Education.

This National Interpretation and Education Evaluation Summit was a historic step toward creating a culture of evaluation within NPS, which is one of the central pillars of the Renaissance. It was clear to me that Summit participants found the experience of this day and a half of dialogue to be fulfilling and thought provoking. The contributions of our panelists and outside experts provided fresh and useful insight, and the responsive discussion from NPS staff and partners demonstrated the depth, creativity, and commitment that we bring to this challenge and opportunity.

The National Park Service must work to find the resources that will help to create this culture of inquiry and ongoing learning. This is not something that can be accomplished overnight. This will be a long journey, but ultimately the effort will be fulfilling,



“The National Park System embodies intangible meanings—beauty, health, wonder, freedom, democracy, and struggle—that are central to our collective identity. The survival of the National Park System in the twenty-first century depends on how it interacts with society and how much society values it.”

—DAN RITCHIE, SUMMIT HOST

worthwhile, and enjoyable. Creating a culture of evaluation will be a key piece of taking the NPS from good to great.

Comments from NPS Director, Honorable Mary Bomar

This Evaluation Summit is the beginning of our Interpretation and Education Renaissance. The National Leadership Council endorsed the *I&E Renaissance Action Plan* at their August 2006 meeting and this event is an important first step in implementing the plan—an important first step in looking ahead to our Centennial. As we look ahead, I think it is important to also look back for a moment and see from whence we came. So I looked through the history of education in the National Park Service, and I went back to the beginning.

In 1918 the objectives drafted by National Parks Education Committee were clear, bold, and expansive: to educate the public in respect to the nature and the quality of the national parks; to further the view of the national parks as classrooms and museums of nature; to use existing publicity and educational systems as to produce a wide result; to combine in one interest the sympathy and activity of schools, colleges, and citizen organizations in all parts of this country; to study the history and science of each national park and collect data for future use. These objectives are among the earliest expressions of the National Park Service’s founding fathers regarding the educational aspects of park management.

A resolution adopted by park superintendents in 1922 made it clear: “The mission of the national parks is to provide not cheap amusement, but healthful recreation and to supplement the work of schools by opening the doors of nature’s laboratory to awaken an interest in natural science as an adjunct to the commercial and industrial work of the world.”

If there were any doubt about what Congress thought about the Service’s education program, it was put to rest by the Historic



“The public looks upon the national parks almost as a metaphor for America itself. But there is another image emerging here, a picture of the National Park Service as a

sleeping giant. Beloved and respected, yes. But perhaps too cautious, too resistant to change, too reluctant to engage the challenges that must be addressed in the twenty-first century.”

—MARY BOMAR, NPS DIRECTOR, QUOTING FROM
RETHINKING THE NATIONAL PARKS FOR THE 21ST CENTURY

Sites Act of 1935. While the Act placed the National Park Service squarely in the middle of a maturing historic preservation movement in this country, it also charged the Service with developing an educational program for its newly acquired cultural parks. “The Secretary of Interior shall develop,” it declared, “an educational program and service for the purpose of making available to the public facts and information pertaining to the American historic and archaeological sites, buildings and properties of national significance.” The Act also formalized the National Park System Advisory Board to advise the Secretary of the administration of parks.

To mark the twentieth anniversary of the National Park Service, the Department of Interior published “Research and Education in the National Parks.” It was divided into two parts: the educational program in the national parks and the history of the educational movement. This publication was clearly designed to praise the accomplishments of the Service’s educational program. It itemized various ways the Service delivered educational information to the public from auto caravans, nature and historic trails, exhibits, lectures and campfire talks, to museums, libraries, college and university field classes and the Yosemite School of Field Natural History, all built upon a foundation of solid research.

Let us now fast forward to the future, 20 years from now. What will the historians write about us in 2026? What will they write about education in the National Park Service in the years surrounding the Centennial? And if you agree with Emerson that “there is probably no history, only biography,” what will it say about us?

I am by nature an optimist, and I see the glass as already half full. There has been much work by the National Education Council and the National Interpretive Advisory Council in the past two years, including a business plan and an action plan, endorsed by

the National Leadership Council. Together, with the National Park System Advisory Board, many of us here attended the Scholars’ Forum last January, in Philadelphia. Many parks, regions, and our partners already have a commitment to evaluation. For example, the Northeast Region has an ongoing evaluation of its educational programs. So in many respects we are building on the good work of the past, from the era of Stephen Mather as well as good work of more recent vintage.

The past, it is said, is the key to the future. When we look back over the National Park Service of 15 years ago, how far have we come? In 1991 we had the Vail Agenda—looking for ways to diversify our workforce, broaden our stories, and reach new groups of visitors. And in a published version of the report, there were some interesting predictions: Everyone will belong to a minority group. Whites will no longer be a majority group in several states, such as California. Asian and Hispanic populations will dramatically increase, with Hispanics outnumbering African Americans by 2010. Politics will be altered by 2000. Many mayors in the nation’s great cities will be people of color. Racial crossover voting will be common. The Vail Agenda also recommended that the Service should revise its list of cultural themes to more accurately reflect the breadth of American culture; that individual units publicize their unique purpose to their employees and the local population of visitors; and that new studies by the Service include the need for cultural diversity throughout the National Park System.

In 2001 the National Park System Advisory Board prepared *Rethinking the National Parks for the 21st Century*. At the time, the Board was chaired by one John Hope Franklin, a great thinker and a most humble man despite his many accomplishments. And while the report was the work of many, I sometimes like to think that it was he who penned these words: “The public looks upon the national parks almost as a metaphor for America itself. But there is another image emerging here, a picture of the National Park Service as a sleeping giant. Beloved and respected, yes. But perhaps too cautious, too resistant to change, too reluctant to engage the challenges that must be addressed in the twenty-first century.” Later that thread continues... “The Park Service must ensure that the American story is told faithfully, completely, and accurately. The story is often noble but sometimes, as we all know, shameful and sad. In an age of growing cultural diversity, the Service must continually ask whether the way in which it tells these stories has meaning for all our citizens.”

The world is, indeed, different from the time the original National Parks Education Committee was established. The U.S. population was 110 million in 1922 and it is 300 million today. It is expected to double yet again in this century, and the demographic forecast in the Vail Agenda Report was pretty much on the mark. With changing population, demographics, and technology, it is clear that our approach to interpretation and education must also change if we are to continue engaging the American public with their natural and cultural heritage. We can certainly use the newest in technology to reach our visitors in many ways, both those who physically visit a park and those who do it in the virtual realm. And we must embrace partners who can help in this effort.

We sometimes need that outside shot in the arm to help us change. We also need a culture of evaluative thinking as a way of doing business, not only in interpretation, but throughout the disciplines within the National Park Service. When people ask me for my vision of an ideal park, my mind's eye takes me to a very special day at Independence when I was the Superintendent. I left the office after a very long day, and I walked through the park. As I arrived at Independence Hall, I saw my perfect vision of what a park could be. Our staff was busy keeping the grounds and buildings looking good. A group of school children was there listening to one of our rangers give an Underground Railroad tour. And all around me I heard languages from all over the world from our visitors who had traveled from far corners of the globe. Our parks are not just special places for Americans, but they are special places for the entire world, and that is my vision for our national parks, not just for one day, but every day, not just for one park, but for all our parks. That is the true challenge for our Centennial, to make the best idea America ever had the best it can possibly be. With the vision outlined by President Bush, with the leadership of Secretary Kempthorne, and with your help, that perfect vision can become a reality.

I congratulate you. But we must demonstrate results if we wish to garner the resources we need to move forward. You have my support. You've had my support over the last two years with my involvement with the National Education Council. Now it's up to all of you to roll up your sleeves and make it work.

Comments from Summit Facilitator, Dr. Michael Quinn Patton

I came to this Summit from two paths, one as a professional evaluator and the other as a long-time supporter of and visitor to parks. Indeed, just a week before the Summit I had hiked the Grand Canyon for a week, revisiting the route I had written about in a book about the Canyon that became my own venture into interpretation and education.¹ As an evaluator I brought to the Summit a commitment to making evaluation useful and meaningful, the long-time focus of my consulting, training, and writing.² Indeed, it was because the background, preparation, and planning for this Summit demonstrated a commitment to evaluation as learning-oriented and useful that I agreed to serve as a facilitator and resource. After 30 years as an evaluator and a founding member of the American Evaluation Association, I'm at the stage in my career when I no longer waste time with organizations that are simply going through the motions of pretending to take evaluation seriously. I was convinced by preparations for the Summit that the NPS was ready to take evaluative thinking and action to a new level. I was not disappointed. Here, then, are some of the things that still stand out to me as I reflect on the extraordinary Summit experience.

The dominant theme of the NPS Evaluation Summit was *creating a culture of evaluation* within the National Park Service. This

¹ *Grand Canyon Celebration: A Father-Son Journey of Discovery* by Michael Quinn Patton. Prometheus Books, 1999.

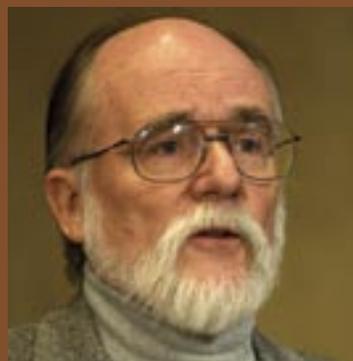
² *Utilization-Focused Evaluation, 3rd. ed.*, by Michael Quinn Patton, Sage Publications, 1997.

³ *Ibid.*

theme is consistent with research on the utilization of evaluation showing that evaluation findings and processes are more likely to be useful when there is strong leadership support for evaluation, when the organizational culture supports inquiry, reality-testing, and learning, and when people throughout the organization value and demonstrate evaluative thinking.³ This shift in emphasis is critical, it seems to me. It means looking beyond the use of discrete evaluation findings reported at a moment in time, as significant as that can be. Rather, the conditions for evaluation use are understood to be embedded in the values, attitudes, and behavioral patterns that are manifest day-to-day as people throughout the organization interpret what is important, pay attention to what gets rewarded, and notice what priorities get attention.

I experienced participants in the Summit as taking seriously the challenges of building an evaluation-friendly culture within the National Parks Service. Some talked about the need to move from an audit and compliance culture to a culture of inquiry. Panel members discussed the difference between just focusing on evaluation reports versus bringing evaluative thinking into all aspects of organizational planning, implementation, and decision-making. Those present acknowledged the tensions between accountability-driven evaluation and learning-focused evaluation. There was also dialogue about the importance of making evaluation meaningful and useful at the frontline level—in the field where interpretation and education take place, as well as having a system of evaluation that allows findings to be synthesized and shared at regional and national levels to inform strategic decision-making.

To assure the relevance of evaluation findings and processes throughout the National Park Service, evaluations will need to focus on significant issues that affect both practice and policy. To this end, participants devoted significant time to examining how the NPS *Interpretation and Education Logic Model* can inform the selection of evaluation pilot projects and determine evaluation



“... the conditions for evaluation use are understood to be embedded in the values, attitudes, and behavioral patterns that are manifest day-to-day as people throughout the organization interpret what is important, pay attention to what gets rewarded, and notice what priorities get attention.”

—DR. MICHAEL QUINN PATTON, SUMMIT FACILITATOR

“... the Summit discussions were consistent with the standards for evaluation adopted by the American Evaluation Association, namely that evaluations should be useful, practical, ethical, and accurate.”

—DR. MICHAEL QUINN PATTON, SUMMIT FACILITATOR

priorities.⁴ Indeed, I was impressed by the serious engagement with the logic model, not always an easy thing for non-evaluators to get into. The logic model and priorities discussion affirmed that there will need to be alignment between strategic NPS priorities and evaluation priorities. For example, as the National Park Service strives to increase the diversity of visitors to parks to better reflect the changing demographics of the country, evaluation approaches will need to provide timely and relevant feedback from visitors with diverse backgrounds to effectively capture what they experience, value, understand, and retain from their visits. As the NPS incorporates new technologies in support of interpretation and education, evaluation will need to provide real-time data about the reactions to and consequences of various technological approaches. While specific evaluations contribute findings about the effectiveness of particular programs, attention to cumulative and synthesized findings across evaluations should increase overall understanding about and effectiveness in efforts to enhance the power of place-based learning.

I was impressed and pleased that participants in the Summit emphasized the importance of developing a full range of evaluation approaches and methods to assure that evaluations are appropriate for and adapted to different kinds of parks—historical sites, geographical and wilderness sites, small parks, large parks, urban parks, isolated parks, parks connected to local communities, and parks in the different regions throughout the country. Participants also discussed using a variety of methods—quantitative, qualitative, mixed methods, rapid feedback studies, and longer-term inquiries. Differences between research and evaluation were discussed. The Summit included NPS field staff and leaders from all levels of the organization as well as representatives of many NPS partners. Thus, the dialogs and discussions included ways of partnering around evaluation. Personally, it was a pleasure to meet and work with so many energetic and thoughtful people all coming together in a shared commitment to preserving our national parks and improving interpretation and education programs.

At times the discussion was quite specific and technical. One example centered on how to pose genuinely learning-oriented questions. An evaluation that asks, “Did students retain what they learned at the park?” implies by its grammatical construction that a yes/no answer is appropriate and informative. But things are seldom that simple. To improve interpretation and education programs, it can be more helpful to ask: “Under what circumstances, in what ways, and to what extent, if at all, do various students retain what they learned at the park?” This reframing of the question guides the inquiry to look at variations in learning so as to better understand what works for some

students and doesn’t work for other students, and how to adjust and adapt an interpretation and education program to better meet the diverse needs of diverse students.

At other times the discussion was more general, focusing, for example, on examining what kinds of resources are needed to conduct evaluations. And there was regular and astute recognition that the National Park Service operates in a political environment and evaluation is therefore subject to political entanglements. Different stakeholders have different evaluation interests and varying information needs, so evaluation designs and measures need to be developed taking into account diverse perspectives and interests while also maintaining the integrity and credibility of the evaluation.

The tenor of the discussions was that evaluations should be done with people not to them. Participants recognized that new directions and pilot evaluations should build on the extensive evaluation work that has already gone on in many different places and at many different levels in the National Park Service. Further, evaluations themselves should be evaluated to assure that scarce resources are well used. Toward that end, I would note that the Summit discussions were consistent with the standards for evaluation adopted by the American Evaluation Association, namely that evaluations should be useful, practical, ethical, and accurate.⁵

Returning to the same personal tone with which I began these reflections, and as a longtime evaluation professional, let me conclude by affirming how impressed I was. The discussions were serious and sophisticated. The participants were forthright in identifying and addressing complex and challenging issues. There was an overall commitment to building a strong culture of evaluative inquiry throughout the National Park Service in support of the NPS’s important historic mission and future vision. It was both a pleasure and honor to be part of this historic gathering, and I look forward to following the implementation and realization of the commitments made in Denver.

SUMMARIES OF PRESENTATIONS AND DISCUSSIONS

The following pages provide a chronological summary of what happened during each scheduled segment of the Summit agenda. The purpose is to provide a vicarious experience of the event for a reader was not in attendance. For further detail, DVDs of the Summit are available in an edited summary version and complete record from sources at www.nps.gov/interp/evaluation.

Day One Welcome and Introductions

The Summit opened with remarks from three key people. Mr. Dan Ritchie (Chairman, Education Committee, National Park System Advisory Board; Chairman, Board of Trustees, University of Denver) welcomed everyone as host of the Summit. Mr. Ritchie highlighted the importance of evaluation for organizations and

⁴ See page 18 of this report for a copy of the draft *Interpretation and Education Logic Model*.

⁵ For the Program Evaluation Standards see: www.eval.org/EvaluationDocuments/progeval.html

stated that the goal of this summit was to “create a culture of evaluative thinking” throughout interpretation and education staff. Next, NPS Director Mary A. Bomar shared an inspiring set of insights about how the history of the National Park Service connects to the current need for an “Interpretation and Education Renaissance.”

Finally, the facilitator for the event, Dr. Michael Quinn Patton, set the stage by sharing his insights about the factors that make evaluation effective and useful. Dr. Patton began by stating that support from leadership is the single most important factor for effective evaluation. As leadership support is also one of the rarest factors, Dr. Patton praised the NPS for the level of commitment to evaluation demonstrated by the range of participants in attendance at this Summit. A related factor in successful evaluation is an organizational culture that integrates it into decision-making. Dr. Patton concluded by highlighting the importance of engaging all levels of an organization—a willingness to ask “How do we know [rather than assume] that what we are doing is good?”

Panel One Presentations: Why Should We Do This?—Creating a Culture of Evaluation within NPS: Vision and Rationale

The first panel of the Summit brought together experts from various fields to talk about what evaluation is, what a culture of evaluation looks like, and to provide their ideas about how to make evaluation successful.

Martha Monroe, Panel One Moderator
Associate Professor, Natural Resources Education and Extension, University of Florida

“A culture of evaluation depends on being able to use the results of evaluation. If we’re afraid of evaluation or not sure what’s going to happen, then we tend to ask ‘DUH’ questions. We ask the questions that we know the answers to and we know the answers aren’t going to hurt us any. But they’re also not going to help us very much either.”

Dr. Monroe began Panel One by highlighting the importance of allowing for risk-taking in a culture of evaluation. She provided the example of Ben and Jerry’s Ice Cream who have a “Flavor Graveyard” at their factory in Vermont celebrating flavors that have been unsuccessful or otherwise discontinued. Dr. Monroe encouraged the NPS to engage field staff and build the capacity for evaluation starting with the field. She pointed out that the more people are involved in the whole process of evaluation, the more engaged and excited they become. By allowing for risk-taking and mistakes, the NPS can engage staff at all levels in growth and learning. She noted that good questions develop in environments where there is confidence in the usefulness of evaluation. That happens when the questions being asked are important to the people asking.



Panel One experts discuss evaluation with the audience.

Jon Wergin, Professor, Antioch University
Ph.D. in Leadership and Change

“Bottom line, what we found confirmed our worst fears, that there was a vast sense of the futility of evaluation in most of the institutions we visited. Compliance mentality was alive and well . . . but evaluation works when people on the ground are addressing questions of interest to them . . . good evaluation flourishes in places that encourage risk-taking . . .”

Dr. Wergin continued the conversation by presenting his research on evaluation within higher education. In this research, it was discovered that often people go along with evaluation without any real commitment to the process—what Dr. Wergin called a “compliance mentality.” While this is common, he also offered hope that it can be overcome. Dr. Wergin noted that when both leadership and staff within departments are open, self-reflective, and communicative, evaluation is more likely to be successful. He further noted that good evaluation happens in environments that hold people accountable for learning and not necessarily for results. Finally, he highlighted the importance of evaluation coming from and being useful to the front lines.

Reginald “Flip” Hagood, Senior VP for Strategic Initiatives,
Student Conservation Association

“So when we’re asked the question ‘Why should we do [evaluation]?’ [The answer is that] it’s good business. In very practical terms, it prepares you for opportunity.”

Mr. Hagood pointed out the importance of accountability and the use of evaluation for survival as an organization. In addition, he highlighted the partnership between the SCA and the NPS and spoke about several types of evaluations and the value that they have added to SCA. He noted that evaluations might run from very informal through to much more formal, including audit type evaluations. Evaluation has helped the SCA to diversify funding, move to meet the needs of the market, and continue to thrive as a business. Mr. Hagood's central message was that evaluation helps an organization meet the goals it sets for itself.

**Carol Stapp, Director, Museum Education Program,
George Washington University**

“In order to do this seamlessly, you have to take on a completely different approach in your thinking, because the entrance narrative that we bring to evaluation generally is one that’s been imposed on us . . . as something where we’re trying to prove that we’ve done something right as opposed to looking to see how we can become more effective.”

Dr. Stapp shared her experience as an educator within the museum world and the role that evaluation has played in her work. She noted that shifting to a culture of evaluation is both valuable and challenging. Dr. Stapp provided the example of working with her students and their reactions to evaluation. As she noted, her students often take several semesters to adjust to the idea that evaluation does not have to be punitive. It can be difficult to promote positive attitudes about evaluation. In order to accomplish this, she encouraged focusing on competencies and building capacity for evaluation. Overall, Dr. Stapp highlighted the fact that shifting to a culture of evaluation in the NPS will take time, will have challenges, and will require a new understanding of what evaluation means and how it is used.

**Les Baxter, Deputy Director for Evaluation,
Pew Charitable Trusts**

“So why do evaluation? It’s not easy. It takes time and effort and resources. And you will make mistakes. But [to quote Michael Patton]: ‘If you want to believe that the universe is unfolding as it should, avoid evaluation for it tests reality.’”

Dr. Baxter described ways that evaluation can inform program planning and design. He noted that without evaluation, you do not know where you are going, if you are accomplishing what you set out to accomplish, or if you are making the best use of limited resources. Dr. Baxter shared two examples of the benefits

of evaluation that he has witnessed at Pew. In the first example, evaluation played a central role in a major internal reorganization of Pew. The second example highlighted the fact that although an evaluation may indicate changes, the process of organizational change still takes time. Dr. Baxter pointed out that evaluation is an important reality-testing mechanism and important for organizational survival.

Emma Norland, Ohio State University

“The PARKS Project, Parks As Resources for Knowledge in Science, included 36 parks . . . and it was a wonderful project. It went on for three years . . . we learned stuff . . . But here’s the thing. . . this was an episodic evaluation, a one-time evaluation. So you know what it was? This was a sacrifice fly . . . It was a wonderful evaluation . . . we advanced the runner and here we are today . . . But because there wasn’t a system in place to keep it going . . . it went nowhere.”

Dr. Norland spoke about the importance of evaluation occurring within a broader context and a more robust system. She provided the example of the PARKS Project, a large 36-park evaluation. She described some of the learning that was garnered, but emphasized that the episodic nature of the evaluation meant that many of the best learning opportunities were lost. She noted that evaluation should be part of a larger picture. She also commented that the best evaluations focus on the use of the evaluation by the stakeholders involved. Finally, she encouraged the NPS to create a large database in which all evaluation data could be gathered so that, over time, larger questions could be answered.

**Hazel Symonette, Senior Policy and Program Development
Specialist, University of Wisconsin–Madison**

“There are voices from the future, and they’re calling all of our names. All of our names . . . Too often this [evaluation] work is at least perceived and received as standing in judgment of rather than working in deliberative collaboration with . . .”

Dr. Symonette highlighted for participants the “who” of evaluation. She spoke about working collaboratively with all stakeholders rather than standing in judgment when conducting evaluations. Dr. Symonette talked about identifying our goals, both personal and organizational, and using evaluation as a learning tool to reach these goals. In particular, she highlighted the importance of including all stakeholders—frontline staff, partners, leadership—in a collaborative process of evaluation.

FIRST LARGE GROUP DISCUSSION WITH PANELISTS AND PARTICIPANTS

Following the panelists' remarks, participants in the audience were asked to jot down on index cards questions, comments, and thoughts for discussion. These were collected, organized during the break, and then used to guide a large group discussion that lasted approximately two hours and addressed several topics. Below are quotes and notes that capture some of the key themes from this discussion.

"I really do encourage you to be very mindful about how you structure this so that people can fail fast, fail frequently, and fail forward...[as a way to] give folks a space for innovation, and for creativity to live and breathe."

—Hazel Symonette, Summit Panelist

Accountability for learning. Dr. Wergin clarified his earlier comments by explaining that people should be held accountable, but that the accountability is for learning and utilizing what they have learned. This discussion also brought up issues about what data is being collected and for what purpose.

"What are the questions you want to be able to answer in five years? Instead of, what data do you want to have to share?"

—Terri Behrens, Kellogg Foundation

Planning must be flexible. Dr. Patton commented that you cannot set a course in genuine innovation because you do not know what will happen yet. Not every activity is or should be innovative, and innovation ought to be based on a solid theoretical foundation.

"I agree 100 percent that we need to develop this evaluation culture in the organization. How do you do that with a very distributed organization?"

—Jon Jarvis, NPS Staff

Practical concerns. Issues such as available resources for evaluation, the reality of needing to be accountable to funders, staff movement within the NPS, and others were earnestly discussed. Comments included the idea of doing both the "accountability" type evaluation and the evaluation that is oriented to learning and growth.

"What's the responsibility at different levels of the organization in terms of evaluation?"

—David Larsen, NPS Staff

Involving the field. The more involved frontline staff is in this process, the more embedded evaluative thinking will become in the culture. The culture of interpretation and education is intrinsically oriented to curiosity and learning, which provides a good foundation for building a culture of evaluation. Ultimately, the message was that all levels of the organization and its partners need to be involved and feel responsible for evaluation, and that this change will take time and patience.



Panel Two discusses how evaluation enhances cultural competency.

Panel Two Presentations: The Role of Evaluation in Enhancing the Power of Place-Based Learning

The second panel brought together experts to discuss the use of evaluation in documenting the impact of services and improving programs. In addition, the use of evaluation in enhancing cultural competency was highlighted.

Lynn Dierking, Panel Two Moderator, Vice President for Special Initiatives, Institute for Learning Innovation

"[In] an institution I've been working with . . . [at first] we did not realize [our evaluation] was using such a top-down approach. . . so we began to work with the people that actually managed the day-to-day activities . . . This process had . . . wonderful consequences."

Dr. Dierking began the second panel of the day by sharing an example of an evaluation in which she has been involved. In this example, the evaluation was initially imposed from leadership with limited buy-in at other levels of the organization. Dr. Dierking explained that the focus of the evaluation was shifted to include managers responsible for the day-to-day operations of the program. By engaging these individuals in the process, the attitude toward evaluation shifted. People in this organization are now excited about evaluation and are engaged in a wonderful learning experience. Dr. Dierking offered this example as a way of encouraging the NPS to engage a range of stakeholders in the process of evaluation.

Theresa Coble, Assistant Professor of Forest Recreation and Interpretation, Arthur Temple College of Forestry & Agriculture, Stephen F. Austin State University

“I think this Summit is being held in part because we all recognize the need to be more specific about the ‘and then a miracle happens’ aspect of visitor on-site experience . . . I recognize the need for humility if one seeks to conduct outcome-based assessment of interpretive programs at national parks. To a large extent we’re trying to measure and quantify and understand experiences that vary from the mundane all the way up to the miraculous.”

Dr. Coble continued the conversation by sharing the results of an outcome evaluation called “Visitor Voices” that she worked on with the NPS Intermountain Region. She explained that this evaluation was an outcome-based assessment and showed where it fit in the Interpretation and Education Logic Model. Dr. Coble also stated that the outcomes needed to be contextualized in order to be meaningful. By placing the results in context, this evaluation was able to provide information about the factors that contributed to the outcomes found. For instance, different types of experiences (e.g., park film, interpretive exhibit, and guided tours) evoked different levels and types of meaning, making connections in visitors. One finding was that “ranger-led programs far surpassed any other type of programs as the . . . number one most meaningful program.” Dr. Coble offered some thoughts on how this particular evaluation approach that focuses on visitor context could be expanded for even greater utility.

Polly Nordstrand, Curator, Native American Collection, Denver Art Museum

“We need to care about the ‘invisible’ people.”

Ms. Nordstrand began her presentation with a story about her mother’s experience of being “invisible” as a result of her brown skin in a predominantly white society. She continued by sharing stories of her own experience of discrimination as a Native American when she was formerly an employee of the National Park Service. Ms. Nordstrand described how it is often more difficult to collect data from non-white groups and emphasized that these groups have valuable information to share. She noted that most NPS sites have a Native American story, but that these (and the stories of other marginalized groups) are not always shared with visitors. Ultimately, Ms. Nordstrand urged the NPS to utilize evaluation as a means of making those people who have been invisible visible.

Doug Knapp, Associate Professor, Indiana University

“. . . the fun thing is that parks have different powers. You don’t know what they are . . . and it’s been a fascinating voyage to be able to find out what those powers are.”

Dr. Knapp shared his findings regarding the powers of the national parks. He offered examples from Clingman’s Dome, George Washington Carver National Monument, and Denali National Park. In each of these examples, Dr. Knapp explained that the experiential components of the interpretive and education programs had long lasting impacts on participants. At Clingman’s Dome, children in the 5th grade retained information they learned one year later. At George Washington Carver National Monument, children continued to experience empathy for George Washington Carver a year after the program. Similar information was found at Denali National Park. In each of these cases, the data from the evaluation have been utilized to continue to improve already effective programs. Dr. Knapp closed by sharing that each park has unique powers that are communicated through interpretive programs, but we may not know what each park’s power is yet. Evaluation can help with discovering these powers.

Veronica Thomas, Professor, Department of Human Development and Psychoeducational Studies, Howard University

“We had to hear their voices of what they want and what they needed rather than us imposing what we thought based on the literature or based on what we thought from our expertise . . . So it’s really important to . . . engage the audience that you are trying to reach.”

Dr. Thomas began her presentation with a discussion of cultural competency relative to the NPS. She shared data indicating that the NPS is not representative of the U.S. population, in either staffing or visitors. As such, she emphasized the importance of asking those populations that the NPS hopes to reach what they want and need from their national parks. Dr. Thomas noted that place-based learning can play a very important role in educating children “placed at risk”⁶ by offering programming that relies on local culture and local geography. Dr. Thomas distinguished between co-construction of a program and its evaluation versus a less rigorous process of collecting feedback from stakeholders. Finally, Dr. Thomas explained that when working with culturally diverse populations it is important to address their needs and interests, not just one’s own agenda.

⁶ This terminology highlights that “at-risk” situations derive from external, social causes, and contexts.

Alison Druin, Director, Human-Computer Interaction Lab,
University of Maryland

“... what roles do your stakeholders play in this evaluation experience? Are they partners in brainstorming before anything has been designed? ... It’s about giving people a voice in changing their world.”

Building on the idea of co-construction offered by the previous panelist, Dr. Druin offered examples of her work with children and technology. She shared a range of examples highlighting the ways that children can and should be included in the design, implementation, and evaluation of education programs. Dr. Druin also provided a range of examples of creative evaluation techniques (e.g., using sticky notes, paper, and glue). By offering participants the opportunity to co-construct their education program, their experience in parks becomes highly individualized and more meaningful. Ultimately, the more involved all the stakeholders (including visitors of all ages) are in this process of program design and evaluation, the more effective it will become.

David Sobel, Co-Director, Center for Place-based Education,
Antioch University New England

“A series of evaluations over the course of about five years has led to a variety of changes in what the [Forest for Every Classroom] project is doing and how Marsh -Billings -Rockefeller and the Conservation Studies Institute think about their programs.”

Mr. Sobel shared some of the results from an evaluation that is ongoing involving the national parks. He described how the impetus for launching a collaborative evaluation several years ago was, in part, a response to a foundation’s reluctance to fund environmental education programs for youth because the outcomes were not considered measurable. Several organizations conducting place-based education programs decided to create an evaluation collaborative and work together to find ways to measure and describe the outcomes of their place-based educational programming. As a result of a series of evaluations over several years, the programs have each undergone a variety of changes to better meet their goals. In particular, the NPS program “A Forest for Every Classroom” was able to more effectively and deeply engage local schools and the local communities. The results of these evaluations have also fueled replication of this place-based education program at other areas, such as the Appalachian Trail through “A Trail to Every Classroom.” Ultimately, the evaluation conducted by Mr. Sobel’s group has demonstrated that outcomes from place-based education are measurable and that the results of evaluation can be used to improve programs and enhance their effectiveness.

SECOND LARGE GROUP DISCUSSION WITH PANELISTS AND PARTICIPANTS

Index cards were used to capture the range of participant questions and comments as a way to seed the plenary discussion. This afternoon discussion was more wide-ranging than the earlier discussion, mirroring the diversity of topics highlighted by presenters. These quotes and notes provide a flavor of some of the key points that emerged.

“I think you need to be looking to your partners as providing continuity, connection to the communities, and ideas about where you’re going with evaluation.”

—Ken Voorhis, Great Smoky Mountains Institute

Involve stakeholders early. Stakeholders can offer a great deal of assistance in developing projects and maintaining continuity, especially when included from the beginning.

“Technology is a tool for social interaction [and] inquiry. It’s not sitting down at a computer and just reading. It’s really using it as an opportunity to provoke people and send them off doing something that is tangible and more real.”

—Lynn Dierking, Panel Moderator

The role of technology in place-based learning. Can technology foster emotional connections? Comments ranged from the democratic nature of Web-based tools to the shift in culture in which technology is a common method of building community. Concerns were also raised about the use of technology, such as disappointing visitors who have done their research on the Internet before coming to the park.

“You have to go into the communities, go into the schools, go into to the churches. You have to go to where they are... [They may not] share the same values.”

—Veronica Thomas, Summit Panelist

Cultural competence. In creating culturally competent evaluations, the key is dialogue, learning, and listening. Shifting from a deficit model (What can we do to help people?) to an asset model (How can people help us?) is one method of engaging diverse populations. There is a need to reach out to populations that are not being reached and tell all the diverse stories of a park. It was also suggested that in order to conduct evaluations with diverse populations, the NPS may need to adopt alternative evaluation methods.

“I do think we do need to look at behavior based outcomes... stewardship, civic engagement, these are the pay dirt, the hoped for long term outcome... But we also have to be very mindful of the complicated [research] process, the zigs and zags along the way.”

—Theresa Coble, Summit Panelist

Evaluating visitor experiences. Discussions highlighted the need to evaluate more than just the formal education programs as well as the need to use a wide range of evaluation methods to conduct such evaluations. Behavior-based outcomes are important to evaluate, but this can be difficult.

Day One Wrap-up

Dr. Patton closed the first day of the Summit with a genesis story from the Maori tradition in New Zealand. The story highlighted the reciprocal relationship between evaluation and learning and showed how the attitude of inquiry and questioning can exist at the foundation of an entire culture, as well as being part of the culture of a large contemporary organization.

Day Two Opening Remarks

Mr. Dan Ritchie welcomed everyone back to the University of Denver on a very snowy day. After recognizing three people, Julia Washburn, Loran Fraser, and Sheri Forbes, for their role in organizing this Summit, he turned the podium over to Dr. Michael Patton to offer some reflections on Day One and to set the stage for Day Two.

Dr. Patton reiterated that one part of a culture is the stories that are shared and that part of changing a culture will include storytelling. Other factors that influence a culture include the things that people get rewarded and punished for doing. Dr. Patton explained how organizations that are successful at building a culture of evaluation and inquiry are open to honest discussion of what is really happening. These organizations have developed a culture with stories about both what has worked and what has not worked.

Dr. Patton offered a comparison of the nonprofit and government sectors with the business sector's approach to new ideas, success, and failure. He noted that, in the business world, failure is more of a rule than success. Businesses risk failure in order to try a new idea and evaluate what has happened. In contrast, the nonprofit and government sectors tend to punish failure as someone's error. Overall, Dr. Patton highlighted the need for balanced reality testing in creating a culture of evaluation—a willingness to ask tough questions and obtain honest answers.

Presentation Introducing the Servicewide Interpretation and Education Evaluation Strategy

Sheri Forbes, Chief of Interpretation, Mount Rainier National Park⁷, and chair of the Education Evaluation Coordination Team, introduced the draft *Servicewide Interpretation and Education Evaluation Strategy* (also known simply as the evaluation strategy) with a brief discussion of the importance of tangibles and intangibles in interpretation. By conducting evaluations and identifying the tangibles of interpretation (i.e., evaluation data), the Interpretation and Education Program can be strengthened. Ms. Forbes commented on the possible benefits of engaging in evaluation, including greater focus on outcomes and results, the ability to answer questions about programs, identification of the longer-term impacts of programs, and justification for financial decisions. (See *Figure 1 below.*) As an example, Ms. Forbes outlined the results of an informal survey of parks that found park staff members are eager for information about visitors, program outcomes, input for planning, and how to conduct and use evaluations.

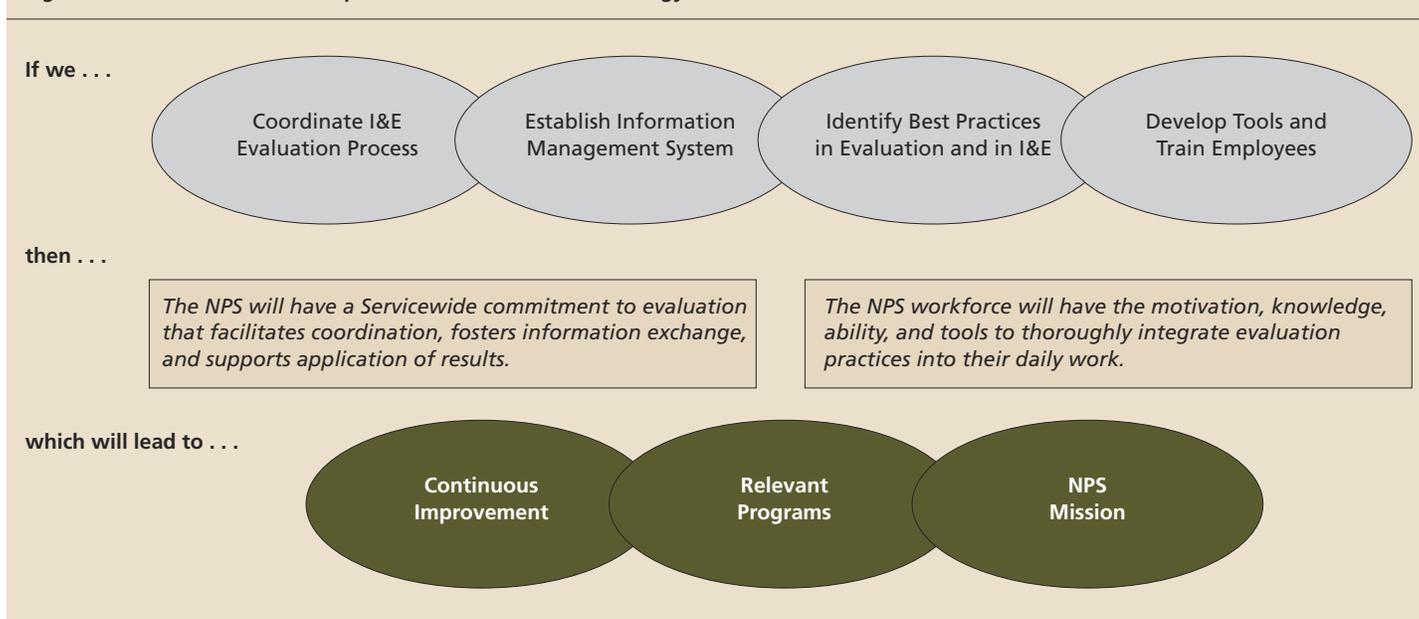
Ms. Forbes then outlined the primary goals of the draft evaluation strategy:

1. Foster a Servicewide commitment to evaluation; and
2. Support all NPS stakeholders in implementing useful evaluation.

Effective implementation of the evaluation strategy will lead to continuous improvement, relevant programs, and accomplishment of the NPS mission. All participants were invited to review the draft strategy and offer feedback. Invoking Freeman

⁷ Between the Summit and the publishing of this report, Ms. Forbes became the new Chief of Interpretation for Glacier National Park.

Figure 1. NPS Servicewide Interpretation and Evaluation Strategy





... key leaders were straight-forward in their support for ... creating a culture of evaluation ...

“Fishbowl Discussion” panelists address aspects of the draft *Interpretation and Education Logic Model*.

Tilden, she concluded with reflections on how the science of evaluation can validate and improve the art of interpretation.

At the conclusion of Ms. Forbes’ presentation, Dr. Patton invited her and the audience to engage in an exercise about reframing evaluation questions. For example, instead of asking “Do visitors make intellectual and emotional connections?” ask “In what ways, under what conditions, and to what extent do visitors make intellectual and emotional connections?” The first phrasing seems to require a “yes” or “no” answer and implies judgment, whereas the second phrasing invites learning that is probably more useful for practitioners.

⁸ Seated from left to right in the following order: Jon Jarvis, *Regional Director, Pacific West Region*; Tom Richter, *Chief of Interpretation and Education, Midwest Region*; Alden Miller, *Chief of Interpretation and Education, Sitka National Historic Park*; Gay Vietzke, *Superintendent, Fort McHenry National Monument and Historic Shrine*; Martha Monroe, *University of Florida, School of Forest Resources and Conservation*; Diane Chalfont, *Chief of Interpretation, Yellowstone National Park*; Antonio Solorio, *Education Interpretive Park Ranger, Santa Monica Mountains National Recreation Area*; Gayle Hazelwood, *Superintendent, National Capital Parks East*; Tracy Bowen, *Executive Director, Alice Ferguson Foundation*.

Defining Selection Criteria and Process for Evaluation Pilot Projects: Fishbowl Discussion

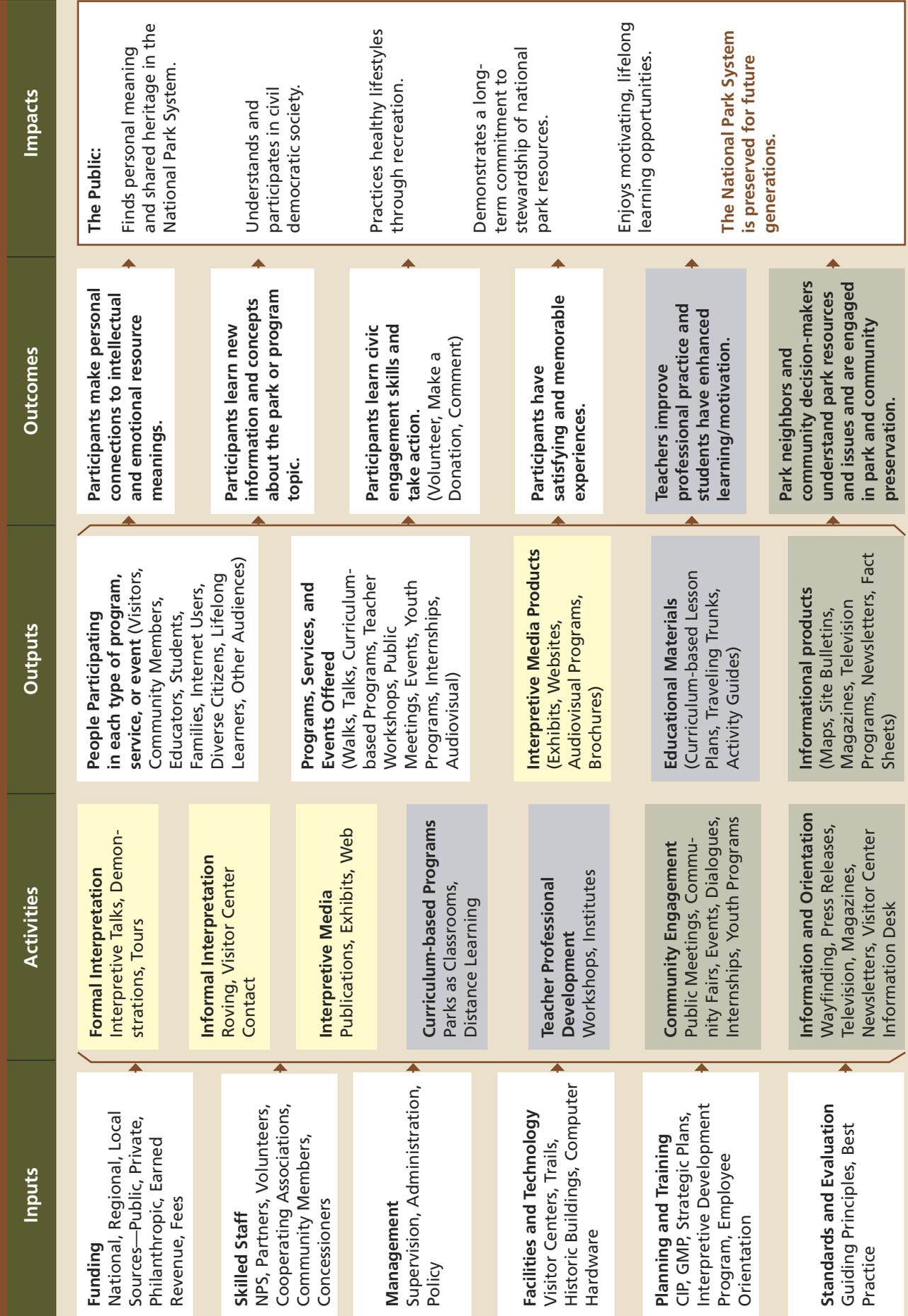
Dr. Patton began the “Fishbowl Discussion” by asking panelists⁸ to identify some key decisions or challenges that they are facing. A sampling of the issues offered included the following:

- the role of the orientation function in relation to the knowledge and expectations that people have when they visit parks;
- the challenge of most effectively allocating limited resources to maximize the returns on that investment;
- the use of Web-based technology; and
- questions about using park programs to help children and schools.

Dr. Patton then shifted the discussion to locating these questions in the draft *Interpretation and Education Logic Model* (see Figure 2, next page) that is a foundation of the draft evaluation strategy. As a group, the panelists and audience members examined several of these questions and discussed where they fit in the logic model. From this discussion, participants learned that more detailed logic models already exist for the three main activity areas in interpretation and education, and that it could be useful to generate additional logic models at even finer program levels. It was also noted that any one question may address multiple areas within a logic model. Finally, Dr. Patton explained that a

Figure 2. National Park Service Interpretation and Education Program Logic Model

Premise: If the NPS offers high quality interpretive, curriculum-based, and informational programs to a diverse public, the public will have better quality of life and will be better equipped to help preserve and protect the National Park System for future generations.



“... the results of an informal survey of parks ... found park staff members are eager for information about visitors, program outcomes, input for planning, and how to conduct and use evaluations.”

—SHERI FORBES, CHAIR
NPS EDUCATION EVALUATION COORDINATION TEAM

“... a logic model should not be a static document, but rather should be evaluated and modified as needed.”

—DR. MICHAEL QUINN PATTON, SUMMIT FACILITATOR

“... in the business world, failure is more of a rule than success. Businesses risk failure in order to try a new idea and evaluate what has happened. In contrast, the nonprofit and government sectors tend to punish failure as someone’s error.”

“... it is possible for evaluation to “fall flat” because organizations attempt to do too much too soon without adequately addressing the natural learning curve.”

—SUMMIT PARTICIPANT, DURING GROUP DISCUSSION

“... more detailed logic models already exist for the three main activity areas in interpretation and education ...”

—GROUP DISCUSSION OF SUMMIT PARTICIPANTS

“... include diverse parks and audiences for pilot evaluations.”

—GROUP DISCUSSION OF SUMMIT PARTICIPANTS

logic model should not be a static document, but rather should be evaluated and modified as needed, and as insights from evaluation are gathered.

Dr. Patton ended the fishbowl discussion by asking panelists to identify explicit criteria that they would recommend for selecting pilot projects. Dr. Patton began by offering two such criteria:

- A. Pilot project questions and answers should be able to inform action; and
- B. Pilot projects should locate within the logic model.

A sampling of other criteria identified by panelists included:

- address diverse projects (e.g., geography, park and budget size, and park type);
- the ability of the pilot to be a model;
- projects designed by or with field staff; and
- evaluations that can be done in a reasonable timeframe.

Defining Selection Criteria and Process for Evaluation Pilot Projects: Small Group Discussions

Participants were assigned to small groups and asked to spend forty minutes addressing four questions related to potential pilot evaluation projects. These questions were:

1. What did you hear in the fishbowl discussion that you want to reinforce?
2. What was missing or underrepresented in the fishbowl conversation?
3. What projects would you like to nominate for consideration for pilot evaluation?
4. Are there specific selection criteria or process items that you would add?

After convening again as a large group, participants were invited to share “Ah-ha” insights from their small working groups. The most common response was the following:

- the importance of including diverse parks and audiences for pilot evaluations.

Other comments were more strategically oriented, such as:

- the importance and benefits of smaller scale, informal self-assessment;
- balancing Servicewide and park-specific questions;
- choosing pilots with potential political traction; and
- modeling innovative risk taking.

Others were more tactical in nature, including:

- advocating for sharing results of new pilots and existing evaluations;
- developing systems to support evaluation implementation at the field level; and
- considering evaluations by clusters of parks or programs.

Summit Closing Remarks

Dr. Patton’s closing comments included an observation, a recommendation, an inspiration, and a story.

He observed that this Summit had generated a lot of energy and excitement about evaluation, but that this presents a risk for



Several participants commented that ethnic diversity was higher than usual for NPS events . . .

The second most watched Tel in FY 2007 was viewed by NPS employees at 120 individual log in stations.

implementation. Dr. Patton noted that it is possible for evaluation to “fall flat” because organizations attempt to do too much too soon without adequately addressing the natural learning curve. Thus, he recommended that culture change in the NPS should begin with asking focused questions that are answerable in a brief period and that offer some rapid and usable feedback.

For inspiration, he reminded participants of the importance of the vision and values of the NPS, as named by Director Bomar in the opening remarks, and as exemplified in the discussions throughout the Summit. Such values become the lens through which evaluation data needs to be interpreted. Dr. Patton noted that as interpreters by trade, participants at the Summit were well positioned to create a culture of evaluation within the NPS that truly honors the vision and values of the organization.

Dr. Patton concluded his comments with a recitation of the soliloquy that Don Quixote offers at the end of his comic and profound adventure in trying to make the world a better place. In evaluation terms, the Don Quixote story is an excellent example of the joys and dangers of “reality testing” in the light of values and vision: “Maddest of all, is to only see the world as it is and not also as it should be and could be.” A culture of evaluation can help connect what is with what could be.

Mr. Ritchie closed the Summit by noting that building this culture of evaluation will be a long journey, but a valuable one. He also shared that evaluation will be an important component of bringing interpretation and education—and the NPS as a whole—from good to great. With that thought, he thanked everyone for their enthusiastic participation and wished everyone safe travels home.

Interactive Teleconference: NPS Leaders and Partners Address NPS Staff Nationwide on Interpretation and Education Evaluation

After a busy and stimulating one and a half days of discussion about evaluation, NPS leaders conducted an interactive teleconference to invite others across the NPS to join in the conversation that started during the Summit. One hundred twenty individual NPS staff logged in to view this Tel, making it the second most watched interpretation and education Tel in FY 07.

Cindy MacLeod, Superintendent, Richmond National Battlefield Historic Park, opened the Tel and served as overall moderator. Deputy Director Steve Martin also offered a few opening remarks, expressing a commitment to building this culture of evaluation over the long term. Julia Washburn shared some highlights of the Summit and Sheri Forbes offered a brief explanation of the evaluation strategy. Following these opening comments,

participants on the Tel shared their thoughts about “take-away” messages from the Summit, which included thoughts about evaluation being collaborative and nonjudgmental, practical and useful, and an opportunity for continual learning.

At this point, the Tel was opened for comments or questions from staff members in the field. Dr. Michael Patton facilitated discussion among the Tel panelists⁹ and several callers. Similar to the questions that were raised during the Summit, people in the field asked about:

- the practical aspects of implementation
- ways of including diverse populations
- concerns about funding evaluation
- the usefulness of evaluation, and
- concerns about evaluation becoming punitive and imposed from above.

Participants on the Tel shared what they learned during the Summit in response to these comments and questions. The Tel ended with an invitation to all to remain engaged in the conversation about evaluation that was started at the Summit.

SUMMIT PARTICIPANTS

One of the more notable features of the Summit was the diversity of the participant list in terms of role and region, coupled with thorough representation from the senior leadership of the National Park Service. Several participants commented that ethnic diversity was higher than usual for NPS events as well. The accompanying figures provide a snapshot of participant characteristics. A complete participant list (including name, role, and affiliation) is provided in the Appendix.

ADDITIONAL MATERIALS

In the days immediately preceding the Summit, all participants received a three-ring binder with supporting materials. Contents included: Summit agenda and intended outcomes, participant list and contact information, three-page pre-Summit statements from each of the panelists, drafts of evaluation strategy documents (Executive Summary, Logic Model, Volume One: The Strategy, Volume Two: The Foundations, informal list of current on-going evaluations within NPS), and proceedings from the Scholar’s Forum on Civic Engagement. The most current versions of these documents are available online at www.nps.gov/interp/evaluation.

Figure 3. Summit Attendance by Role (N=132)

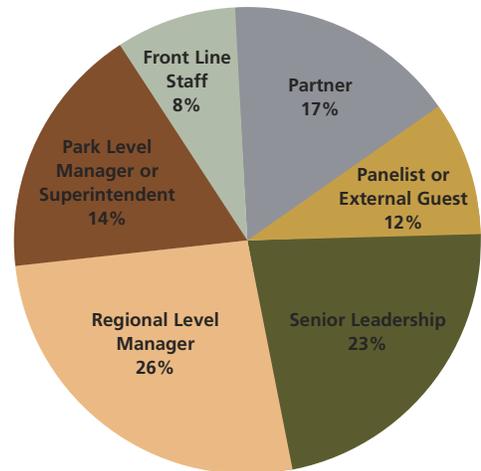


Figure 4. Summit Attendance by Years of NPS Service (N=130)

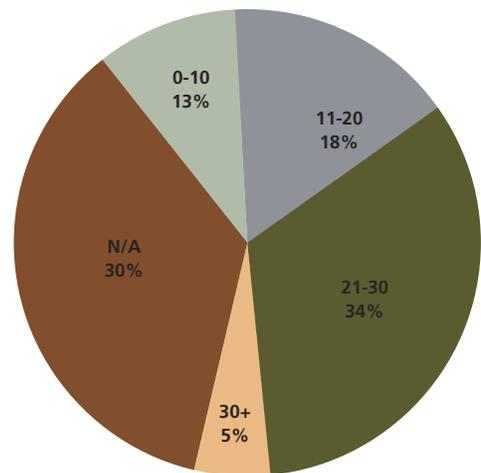
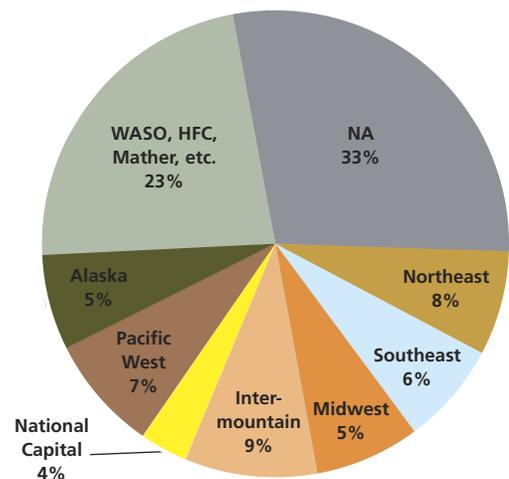


Figure 5. Summit Attendance by Region (N=133)



⁹ Tel panelists included: Steve Martin, *NPS Deputy Director*; Julia Washburn, *Co-chair, National Education Council*; Sheri Forbes, *Chief of Interpretation, Mount Rainier National Park*; Pat Hooks, *Regional Director, Southeast Region*; Lynn Dierking, *Associate Director Institute for Learning Innovation*; Jon Jarvis, *Regional Director, Pacific West Region*; Marcia Blaszak, *Regional Director, Alaska Region*; Mike Snyder, *Regional Director, Intermountain Region*; Antonio Solario, *Park Ranger, Santa Monica Mountains National Recreational Area*; Chris Jarvi, *Associate Director Partnerships, Interpretation and Education, Volunteers, and Outdoor Recreation*; Lyn Carranza, *Chief of Interpretation, Petrified Forest National Park*; Sam Vaughn, *Associate Manager, Interpretive Planning Center, Harpers Ferry Center*; and Tracy Bowen, *Executive Director, Alice Ferguson Foundation*.



“In assessment, in learning, and in improvement, it is critical that we engage our partners and our constituents.”

—MIKE SNYDER, INTERMOUNTAIN REGIONAL DIRECTOR

Part II: Evaluation of the Summit

This evaluation of the Interpretation and Education Evaluation Summit was conducted to “walk the talk” and model a culture of evaluation, thus demonstrating how what is learned through evaluation can be used to inform important next steps. Thus, this part of the report is about learning. Whereas Part I focused on the “What happened?” of the Summit, Part II provides a more in-depth consideration of the “So what?” of creating a culture of evaluation moving forward from the Summit. Evaluation methods are summarized first, followed by an extended presentation of the findings and discussion.

Part II of the report presents three groups of findings. Section A describes the key themes that emerged from a holistic analysis of Summit-related data. Section B considers the extent to which intended short-term outcomes of the Summit were attained. Section C discusses prospects for tracking change over time, including presentation of a baseline snapshot of the current level of a culture of evaluation within National Park Service.

Opposite: Superintendent Gay Vietzke and Maryland Governor Mike O’Malley welcome school children to participate in the Youth Defenders’ Day activities at Fort McHenry National Monument and Historic Shrine.

Evaluation of the Summit

The two major themes in the data gathered to evaluate the Summit were: (1) enthusiasm for a culture of evaluation; and (2) concern about implementing such a culture change. These themes were distinct, yet interconnected.

METHODS

A highly participatory and collaborative approach was used to evaluate the Summit in order to demonstrate organizational learning and a utilization-focused evaluation (Patton, 1997) approach. The evaluation was accomplished through a public-private partnership that combined knowledge of the NPS and its interpretation and education programs with professional evaluation expertise. Members of the Education Evaluation Coordination Team (EECT) of the NPS National Education Council worked closely with NPS cooperating partners from Shelburne Farms to guide and design the evaluation plan. The cooperating partner team played a leadership role in data collection and analysis, and included evaluators Dr. Michael Duffin and graduate research assistant, Catey Iacuzzi. The NPS and cooperating partners together discussed and interpreted the findings for presentation in this report. In particular, the EECT played a leadership role in crafting the next steps based on the evaluation results, as presented in Part III of this report.



Detailed discussions allow for a thorough analysis of the Summit.

Several different data sources were collected to evaluate the Summit (see Figure 6), and together they were analyzed for emergent themes and patterns.¹⁰ Detailed notes, descriptions, and observations from presentations and discussions informed much of the substance for the overall analysis. Data were integrated with participant opinions captured after the Summit on a reaction form. Other documents and data sources from both before and after the Summit served to put the analysis in a broader context. The analysis as a whole was drafted by evaluators from cooperating partners and further refined through the participatory processes of the full public-private partnership team described here.

FINDINGS AND DISCUSSION

Section A: Two Major Themes, and Two Sub-themes

Two major themes emerged from analysis of the data gathered for evaluating the Summit.

Major Theme 1 - Participants seemed enthusiastically engaged in the concept of creating a culture of evaluation.

Major Theme 2 - Participants voiced concern about how such a change will be implemented.

These two major themes were distinct, yet interconnected. The enthusiasm was tempered by the concern, and the concern was interpreted as evidence that the enthusiasm was strong enough for people to take the concept of evaluation seriously. Both themes are consistent with a normal response to the possibility of major impending organizational change.

The following two sub-themes were prominent in the evaluation data, but not as ubiquitous as the major themes.

Sub-Theme 1 - Discussions about terminology and language were fertile ground for dialogue and learning.

Sub-Theme 2 - NPS values of inclusion are an important factor in creating a culture of evaluation.

Major Theme 1 - Enthusiasm

“Having the NPS . . . be extremely open as we have been during the Summit . . . and put in all of our collective efforts together for evaluative thinking as a way of life in our culture was just really encouraging to me. . . . this is going to be something that’s going to be quite beneficial for the NPS since our director and our regional directors and indeed the leadership council is on board.”

—PAT HOOKS, SOUTHEAST REGIONAL DIRECTOR
COMMENT DURING TEL BROADCAST

¹⁰ Using NVivo 7 qualitative analysis software (QSR, 2006), data were coded to illuminate key emergent issues (Miles & Huberman, 1994). Quantitative data were reviewed and graphed descriptively using Survey Monkey on-line survey tool and Excel.

The Tel broadcast referenced in this quote began only 90 minutes after the formal close of the Summit, and provided perhaps the most vivid example of the enthusiasm sparked by the event. The energy was not limited to one particular sector or role, but rather came from leadership, field staff, partners, and guest experts as well.

As the Tel broadcast opened, Deputy Director Steve Martin placed the event in context:

“There has been so much going on that has led up to this. The Business Plan, the Action Plan, in January of this year the Scholars’ Forum, and now the Evaluation Summit. I think it shows that we are starting to really build something here. It is already tremendously exciting.”

A little later, Antonio Solario, a frontline interpretative ranger added the following observation:

“What was really encouraging and exciting for me to hear during the last day and a half was that there is a commitment from the leadership that they are willing to make this evaluation process an

accessible, tangible thing that we can actually use being out there in the field, and being able to apply to the programs that we actually deliver.”

A chief of interpretation, Lyn Carranza, picked up the theme when she noted:

“This has been an exciting week for me . . . I learned a lot . . . I felt that interpretation is on the verge of some very exciting times . . . The most exciting thing for me is that evaluation means continual learning.”

Much of the evaluation evidence for enthusiastic engagement as a major theme was less explicit and more inferential than the distilled sentiments expressed during the Tel broadcast. For instance, evaluators observed that despite a long day of sitting, participants stayed fairly engaged through the end of the first day, and almost everyone came back for the second day with a sense of renewed energy. Conversations during breaks often built directly upon content from the formal sessions. Body language was generally open and positive. The small group sessions on the second day generated many creative ideas and deep questions,

Figure 6. Summary of Data Sources Analyzed for Evaluating the Summit	
Data Source	Description
Reaction Form	Administered in paper at the close of the Summit (n=49) and by on-line survey with multiple email reminders over the two weeks immediately following the Summit (39 additional responses, plus another 36 respondents choosing to only view the summary results). Overall response rate = approximately 80 out of 120, or 75%. Comprised of 10 Likert scale items about Summit outcomes and evaluation readiness, plus 4 open-response items about participant insights, needs, behaviors, and advice. Instrument and complete data are presented in the Appendix.
Small Group Session Notes	The final working session of the Summit consisted of 12 small groups conducting actively facilitated discussions about potential criteria and selection of projects for pilot evaluation studies. All notes were recorded.
Participant Question Cards	For each of the panels on Day One of the Summit, participants wrote their questions on index cards after all the panelist presentations were completed. These question cards were quickly organized during a break and then used to initiate and guide the subsequent full group discussion. Panel One generated 49 cards; panel 2 generated 36 cards.
Lunch Time “Scribbles”	During lunch on Day One, participants were encouraged to use colored pens and crayons to mark up their “place mats” with reflections, questions, and comments in response to creative prompts posted at the front of the room. Forty-seven written comments were captured for analysis.
Field Notes	In addition to evaluators’ systematic field notes, two participants volunteered their own detailed summaries of Summit discussions, and the detailed summary of another participant was forwarded to evaluators after traveling through various email networks. These unsolicited summaries were used to triangulate evaluator notes.
Summary of Prior Survey	In June 2005 a brief, informal evaluation needs assessment questionnaire was sent with a memo to regional chiefs of interpretation/education. A basic content analysis of the 81 (voluntary) responses that were received was included in a targeted review of evaluation literature prepared by Dr. Nina Roberts, intern Jenni Mullins, and several others in July of 2005. Results from this report were informed the baseline snapshot of the current culture of evaluation presented below.
Transcripts	All full-group sessions of the Summit were video recorded and fully transcribed. The Tel Broadcast immediately following the Summit was also recorded and fully transcribed. Transcripts were used to verify and expand upon participant quotations, but were not systematically coded.
Other	Informal observations by participants and evaluators; notes from debrief with EECT; notes from NEC conference call discussing preliminary Summit data analysis; comparison between I&E logic model and draft national citizen survey; list of questions emailed in during Tel broadcast; article written for Inside NPS; miscellaneous email correspondence.

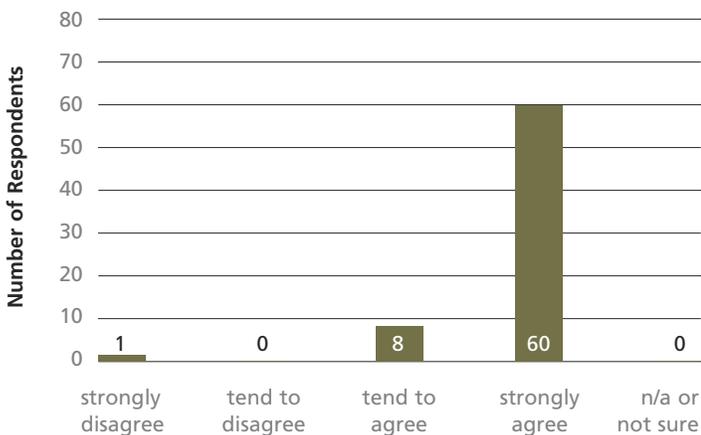
an indicator of engaged participation. The same could be said for the insights, behaviors, and advice offered by participants on the “Summit reaction form” that they filled out at the close of the Summit or later through an online survey.

The quantitative data from the Summit reaction form presented unambiguous evidence in support of this major theme as well. Figure 7 depicts the results for the item that most directly touches on this theme. It reads: “I enthusiastically endorse increased support for evaluation-related activities within I&E.” Literally 99 percent of respondents (n = 69) agreed with that statement, with 87 percent agreeing “Strongly.” The sample represented a fairly equal distribution of respondents’ roles in the NPS, and a participant’s role within NPS was not a factor in this overall result. Other Likert scale items from the reaction form that address the related issue of readiness for evaluation are discussed later in this report, and the complete reaction form results are presented in the Appendix.

Evidence of enthusiasm for evaluation came from Summit participants with diverse roles within the NPS community, and also was represented in diverse types of evaluation data collected for this Summit. Such triangulation strengthens confidence in the claim.

Another dimension of the data analysis is more subtle, but perhaps more useful when considering next steps. This is the dimension of time, or momentum. During a debrief with the planning team one week after the Summit, one person’s reflection succinctly captured the sense of enthusiasm that had occurred during the event. They described a “quiet buzz throughout the room” that pervaded the two days. As highlighted above, that energy rose to a high pitch during the Tel broadcast that immediately followed the event, as evidenced by what the panel members said as well as by observation of the Summit participants who watched the broadcast in a nearby auditorium. Over the course of the Summit, momentum grew stronger, as reflected in a planning team member’s debrief comment: “this event surpassed my wildest dreams.”

Figure 7. Item From Summit Reaction Form



Responses to item from Summit Reaction Form: “I enthusiastically endorse increased support for evaluation-related activities within I&E.”

The obvious challenge is how to keep that momentum growing in the months and years following the Summit (or at least not fading into business as usual, or, worse, dashed expectations). Much of the evaluation data for the Summit directly or indirectly anticipated this challenge; thus, we come to the second major theme in this analysis.

Major Theme 2 - Concern for Implementation

“So, once you have the data . . . How do you use it effectively? Can we afford it? How do you allow/encourage change?”

—FROM PARTICIPANT QUESTION CARD FOLLOWING PANEL ONE

The Summit format was designed to maximize dialogue and to actively encourage and model critical reflection. Thus, it is not surprising that it generated an extensive amount of evaluation data documenting questions and concerns about how to implement creating a culture of evaluation within interpretation and education. Most of these concerns fell roughly into three (interconnected) categories: resource support; technical and logistical matters; or trust and safety issues.

Concerns about finances were expressed both implicitly and explicitly. Jon Jarvis, Pacific West Regional Director, acknowledged funding constraints implicitly when summarizing the Summit: “We want to understand how we’ve been effective and how we can be more effective. We recognize that interpretation and education has been in decline . . . and we have this opportunity to rebuild it.” Several comments referenced recent budget cuts as a major factor. One example is the lunchtime “scribble” comment that said: “There has to be a sea change in funding for interpretation or there won’t be any programs to assess. 18% reduction in seasonal staff since 2000 is a tragedy.” Another comment from the reaction form boldly claimed: “The ‘evaluation strategy’ needs to be better grounded in the reality of the NPS today, so it does not become another unfunded mandate . . . SHOW ME THE MONEY!!!” The tacit link between funding, leadership, and politics seemed clear to everyone. Many participants expressed appreciation for the level of commitment the NPS leadership seems to be lending to this effort, though it is likely that concerns about the extent to which such support can or will continue remain lurking beneath the surface.

Technical and logistical concerns covered a wide range of “How to?” questions or comments, many of which were about training or communication needs and processes. The following comment captures the sentiment of several participants:

“We need a ‘toolbox’ and a comprehensive list of research and evaluations that have already taken place in parks that would apply to multiple parks. We also need training and information about how to do evaluation and how to find researchers.”



Evaluation can also be one way to begin searching for feedback from people who are not currently engaged with the NPS.

Small group discussions reveal concerns about finances, time, and technical and logistical matters.

The most common advice offered by participants through the reaction form and small group discussions was to **serve the frontline field staff** and **involve partners**. Buy-in and ownership from all levels of stakeholders within the NPS was clearly something participants at this Summit think is critical.

Several logistical concerns centered on the issue of time. Some expressed urgency, as in these examples: “How do we transfer the energy and excitement of this opportunity to improve to the frontline I&E folks as clearly and as soon as possible?” or “Spend more money right now to get more results ASAP,” or “Get the field (real field) and partners involved now—don’t wait until we’re ‘ready.’” There were also a few comments that lamented the lack of time available, such as the participant who wrote, “I have not yet read the actual Strategy document (because daily tasks are overwhelmingly consuming).”

Participants in roles further from the field tended to encourage a softer approach and more incremental expectations in an attempt to mitigate anxieties regarding time. Guest experts said things such as “Be patient. Try different approaches . . . Focus first on your willing partners,” or “People don’t just read a report and decide ‘Oh, I’m going to do something different now.’” Facilitator Michael Patton advised in his closing comments, “Gradually build up to the more complex and longer-term questions. Don’t take it all on at once.” Host Dan Ritchie opened the Summit by claiming, “Those who have succeeded at [creating a culture of evaluation] find that it’s a joy, that it really makes life worth living and what

they do better. It’s really a transformative thing that takes time and patience.” He came back to that theme in his closing comments: “And finally [creating a culture of evaluation] is itself a long journey. You don’t do it over a year or two or three or ten. It’s a long lifetime commitment that you’ll find enjoyable. You’ll find it worthwhile, and looking back, you will be very proud.”

In any case, the Summit evaluation data showed clearly that money, logistics, and time issues intertwine to make the goal of creating a culture of evaluation within interpretation and education a substantial challenge, worthy of concern. Successful long-term implementation of the evaluation strategy will need to account for these political and practical concerns.

Perhaps the most important cluster of concerns expressed in the evaluation data could be gathered under the heading of trust and safety. Panelist Jon Wergin’s suggestion to hold people accountable for *learning* rather than *results* generated much interest and discussion. Similarly, panelist Hazel Symonette’s claim that evaluation is too often perceived and unnecessarily used as “standing in *judgment of* rather than working in *deliberative collaboration with*” was one of the most repeated phrases in participant comments. Her recipe for learning, for example, “fail fast, fail frequently, and fail forward,” was also a popular takeaway notion. Discussion about how a compliance mentality stifles risk taking and innovation resonated strongly with participants. Some participants wrote of wanting more of an “honor culture” rather than a “gotcha” or “punitive” or “audit” culture. The

following question, posed on one of the question cards during a panel discussion, seems to capture the tip of a cultural iceberg of norms experienced within the National Park Service. It also hints that much of the attention and cuing within the existing culture focuses up the chain of command:

“NPS employees have been bombarded with new approaches to evaluation (e.g., GPRA). It has been a struggle in many places to institutionalize the process in any meaningful way. What techniques should NPS leadership use to get employee buy-in and support to create a culture of evaluation?”

This question seems to address both major themes. It starts with a critique or concern, but ends with a forward-looking, perhaps even enthusiastic, request. If these two overarching and interconnected themes were to be distilled down into a single word, that word might be hope, as expressed in the following participant comment:

“I have hope rather than a concern. I hope that this continues and is effectively instilled in the field.”[emphasis added]

Another concept that links enthusiasm and concern came from facilitator Michael Patton. Between his closing comments at the Summit and his opening comments for the Tel broadcast, he noted the following:

“One of the things that I’ve observed about organizations that have a genuine culture of inquiry and a culture of evaluation parallel to a culture of caring is that there is a willingness to talk about what’s really going on, to do genuine reality testing . . . I was very much impressed with the energy that came to this event . . . And a part of what I look for as a professional evaluator in such events is a sense of balance . . . of inquiries about strengths [and] recognizing weaknesses . . . about both what works and doesn’t work, and then to have a way to engage that . . . We found that throughout the Summit.”

Sub-Theme 1 - Terminology and Language

“The Latin root of assessment is assidere, which means to sit beside. So the idea of evaluation as a process of sitting beside, talking together, deliberating with one another, helping to make meaning of what you see happening, is probably one of the most important and critical functions an evaluation can perform.”

—JON WERGIN, SUMMIT PANELIST

Words are powerful, perhaps especially so for a group of interpretation and education professionals whose essential job is to communicate ideas, knowledge, inspiration, and meaning. During the Summit there were several instances where the discussion and insights hinged on how certain words are used.

Several participants suggested that the word “assessment” was friendlier or more appropriate than the word “evaluation.” It was also noted that many practitioners and academics use “evaluation” to refer to program level outcomes and “assessment” to refer to individual level outcomes. A few people also advocated for using yet another term to describe this effort within the NPS: “a culture of care.” In a related discussion, some participants concentrated on the distinction between “research” as looking at longer-term outcomes and impacts and “evaluation” focusing on more discrete program level outcomes. As engaging and useful as many of these discussions were, most participants used the terms evaluation, research, and assessment fairly interchangeably. Facilitator Michael Patton pointed out that these are contested terms in professional academic circles, not likely to be fully sorted soon. What is important is that people in each organization figure out which terms make the most sense for their own context.

Other wording shifts that resonated strongly with Summit participants seemed to be less about the words themselves and more about the important ideas they invoke. One panelist described organizational research that revealed “strategic planning” led to less innovation and constructive change, advocating instead for “strategic visioning.”¹¹ Several panelists advocated for evaluation being framed as “sitting in deliberation with” as opposed to “standing in judgment of.” On multiple occasions, facilitator Michael Patton referenced the important difference between asking “yes” or “no” questions (which tend to imply judgment) and questions framed as “under what circumstances, and to what extent” (which tend to invite learning and exploration).

The terminology distinctions that emerged at the Summit can serve as touchstones for engaging NPS stakeholders in meaningful dialogue in the future. They can also be viewed as high priority targets for the coordinated communication efforts that are likely to be critical to the big picture goal of creating a culture of evaluation (or assessment, or continuous learning, or whichever phrase is chosen as the accepted norm within the NPS).

Sub-Theme 2 - NPS Values of Inclusion

“In assessment, in learning, and in improvement, it is critical that we engage our partners and our constituents. Almost every hour of the last day and a half, I heard about reaching out and connecting with underrepresented communities.”

—MIKE SNYDER, INTERMOUNTAIN REGIONAL DIRECTOR
COMMENT DURING TEL BROADCAST

¹¹ Smith, V. B. (1998). *The futures project: A two-year activity of 18 independent colleges and universities in California*. Final report to the James Irvine Foundation.

A highly participatory and collaborative approach was used to evaluate the Summit in order to demonstrate organizational learning and a utilization-focused evaluation approach.

A consistent thread throughout the Summit was the importance of engaging ALL Americans in the NPS mission. In her opening remarks, Director Bomar made substantive reference to the accuracy of the Vail Agenda’s predictions about the changing face of America. Several panelists spoke specifically about the challenges and rewards of reaching out to traditionally underserved populations, suggesting that evaluation can and should reinforce culturally competent practices. Evaluation can also be one way to begin searching for feedback from people who are *not* currently engaged with the NPS. During the second day’s focus on identifying criteria for pilot evaluation projects, the theme of inclusiveness was one of the most prominent. Typical comments that emerged from small group working sessions included these remarks: “How do we reach a broader audience (i.e., underserved, not traditional; urban versus rural)?;” “Be inclusive—telling whole story;” and “Reaching out to the ‘invisible’ people.”

In the last few minutes of his final closing remarks, facilitator Michael Patton addressed how important it is to connect the deeply held values of the NPS (such as inclusion and engagement of all Americans) to the practice of evaluation:

“The centerpiece . . . is a culture of inquiry that continues to honor the vision that you have about the national parks and the values that are part of that . . . In the middle of the word ‘evaluation’ is the word ‘value.’ The way that we interpret findings is through values. Data don’t speak. They have to be interpreted. You’re interpreters . . . A culture of inquiry, an evaluation culture, evaluative thinking, ought to support . . . the things that you care about, the things that get you up in the morning and have you doing this work.”

Section B: Accomplishment of Short-term Summit Outcomes

One way that the team planning the Summit practiced evaluative thinking was to prioritize a list of five short-term outcomes, and to use that list to inform the planning of both the event itself and the evaluation of the event. The goal was to focus on outcomes that could be observed immediately or within a few weeks or months of the Summit. Evaluation data show clear evidence of having succeeded in meeting all five identified short-term outcomes, though with varying levels of attainment. Each of these outcomes is discussed below, presented in rank order from strongest to weakest levels of evidence.

It should be noted that the data considered for this evaluation in general (and these outcomes in particular) are drawn from a sample of people that is highly representative of the mix of participants at the Summit, but slightly less representative of the NPS community as a whole. This is due to attendance at the Summit reflecting a relatively higher proportion of regional and national leadership as compared to field- and park-level practitioners or partners.

1. Enhance leadership support for evaluation.

Intended short-term outcome for the Summit:
Enthusiastic NLC endorsement to move toward increased support for evaluation and social science research (e.g. pilot projects, incorporating evaluation components into existing activities, and hiring a national evaluation coordinator).

This was the short-term outcome that was most clearly attained. Statements from key leaders were straightforward in their support for moving in the general direction of creating a culture of evaluation. As described earlier in this report, the overarching theme of enthusiasm included leadership as well. Shortly before the Summit, the National Leadership Council had unanimously approved the *Interpretation and Education Renaissance Action Plan*, which included multiple references to evaluation. Deputy Director Steve Martin made the following summary statement:

“I want to say that the whole National Leadership Council (a few of them are represented here, [but] speaking for all of them. . .), we’re truly committed to this, not just as a short-term initiative but as the continuation of the way of life of the Park Service.”

These comments by Director Bomar also indicated that leadership support is embedded in a context of political processes and sustained effort:

“We must demonstrate results if we wish to garner the resources we need to move forward. You have my support. You’ve had my support over the last two years with my involvement with the National Education Council. Now it’s up to all of you to roll up your sleeves and make it work.”

2. Increase confidence in cost-benefit of evaluation.

Intended short-term outcome for the Summit:
Increasing confidence at all levels within NPS that evaluation can help and support continuous learning and improvement of programs and decisions in a way that is not overly burdensome.

The notion of evaluation as continuous learning has clearly gathered strength in the minds of Summit participants. As one example among many, a participant who identified herself or himself as “frontline staff” wrote on their reaction form:

“I plan to encourage rangers/field staff to take risks in their everyday work, hold them accountable for learning from them if the risks don’t work well... [and to] re-think the methods of an upcoming formative evaluation for new exhibits.”

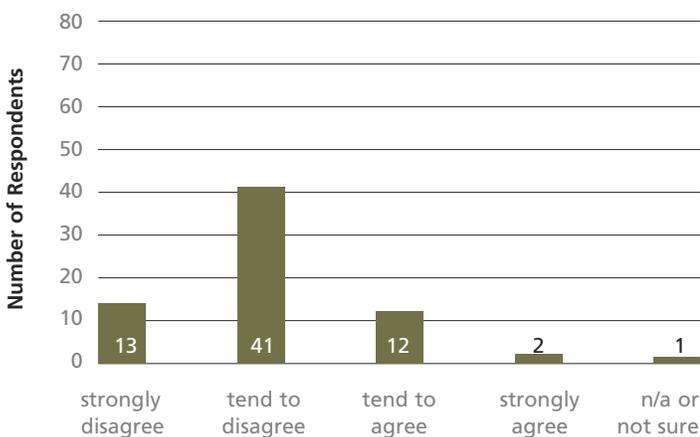
The evidence for confidence is less abundant than that for enthusiasm. For instance, Figure 8 shows that more than three-quarters of Summit participants thought that there is more to evaluation than just extra work and burden. While this is a strong result, it is noticeably less strong than the endorsement shown in Figure 7 in the previous section of this report. A participant’s role within the NPS was not a factor in this overall distribution, though frontline staff tended to be more likely to indicate stronger opinions in both directions (i.e., agreement *and* disagreement).

3. Commence work on pilot evaluation projects.

Intended short-term outcome for the Summit:
Begin implementing pilot evaluation projects as specified in the Interpretation and Education Renaissance Action Plan, in support of action steps outlined in the evaluation strategy.

Substantial work during the second day of the Summit centered on discussing potential criteria for selecting the evaluation pilot projects that are called for in the *Interpretation and Education Renaissance Action Plan*. Popular recommendations included making sure the pilots are applicable to a diversity of park units (i.e., geography, budget, type), reaching out to diverse “publics”

Figure 8. Item From Summit Reaction Form



Responses to item from Summit Reaction Form: “Program evaluation would only increase our workload.”

(i.e., historically underserved populations), and making pilots user friendly and accessible. In short, participants want pilot evaluations to be practical. While the conversation about pilot projects was started and a good foundation was laid, the time constraints and flow of the Summit discussions did not generate any definitive decisions or next steps with respect to evaluation pilots. A complete summary of suggestions for selection criteria and specific project ideas is provided in the Appendix of this report, to be used as a starting point for the body that is empowered to continue the process of selecting and implementing pilot evaluation projects.

4. Promote making decisions based on outcome data.

Intended short-term outcome for the Summit:
Finding evidence of park level managers seeking existing and/or new outcome data to inform decisions (e.g. budget, program, staffing).

This was perhaps one of the more ambitious intended short-term outcomes of the Summit since there is often a substantial time delay between the introduction of a new idea or program (e.g., creating a culture of evaluation) and measurable behavior change in target audiences. However, the responses to the open-ended question on the reaction form about intended behavior change¹² were clearly supportive of the claim that seeds have been sown. Most of the responses were fairly general, matching the somewhat abstract level of the dialogue at the Summit. Typical examples included statements such as these: “I will promote a culture of inquiry in my program area,” or “Read more, think more, pass along information to management and to staff.” A handful of participants indicated that they did not intend to do anything differently, sometimes because “We are doing it already! This reinforces our direction.”

A few responses indicated evidence of actual behavior change. For example, one person wrote this: “As a result of attending the Summit, we have already initiated a low-level evaluation project to primarily improve upon communication and customer service.” Another comment was promising, despite being framed as a future intention as opposed to a behavior already enacted: “I will include evaluation in all the planning processes happening in my park.” Through informal correspondence in the months following the Summit, it was discovered that the National Education Council is considering restructuring their own internal annual personnel evaluation to embody the idea of holding people accountable for learning as opposed to specific results. Further, the draft *Interpretation and Education Logic Model* that was introduced at the Summit is being included as a key reference in the Denali Education Plan that was being drafted as this report was being written.

¹² The item read: “What, if anything, do you plan to DO differently as a result of attending the Evaluation Summit?”

5. Develop action steps.

Intended short-term outcome for the Summit:

Develop clear, realistic, prioritized action steps for rolling out an evaluation strategy within the context of existing resources for Fiscal Year 2007.

In retrospect, it probably would have been more realistic to frame this as a medium-term outcome to follow the Summit rather than as a short-term outcome of the Summit itself. As far as strategic actions are concerned, the Summit functioned more as a starting gate than a finish line. Ultimate success in creating a culture of evaluation is probably far more dependent on what happens *after* the Summit than on what happened *at* the Summit. **Continuing to clearly frame, communicate, and accomplish tangible action steps is likely to be a critical factor in the future success of the implementation of the evaluation strategy.** Thus, despite the fact that 71 percent of respondents to the Summit Reaction Form indicated that they agreed with the statement, “I have a clear picture of the action steps for rolling out the evaluation strategy,” those charged with implementing the evaluation strategy should consider that some stakeholders still feel that plans and action steps remain less than clear. The following comments capture these sentiments:

“[The] logic model is confusing—[it] needs to be in plain language. As a partner, it reads as an internal NPS memo (shorthand). Concern: That this may turn into a top-down (where ‘top’ is the NPS I&E evaluation team) program.”

“My concern is that it will become very complex and GPRA-like with little evaluation empires established as the next new thing. I need reassurance from national and regional leaders that our tendency to grow like topsy will be held in check.”

This outcome highlights a core tension embedded in the process of trying to foster culture change within a large organization, especially one in which much of the service delivery is directed at the local level. On one hand, demonstration of results is essential to justify investment in major efforts such as reinvigorating interpretation and education through creating a culture of evaluation. On the other, investment of resources (such as those called for in the *Interpretation and Education Renaissance Action Plan*) is required to continually generate, refine, and achieve clear, realistic action steps. **Thus, this particular outcome may serve as a microcosm, or indicator of the overall progress of the evaluation strategy.**

Section C: Tracking a Culture of Evaluation Over Time

The Summit was framed as one initial action step in the roll-out of the multi-year evaluation strategy, which features this central goal:

Create a culture of evaluative thinking through interpretation and education staff of the NPS, characterized by continuous inquiry and learning, and decision making based on using various types of outcome data.

Long-term accomplishment of this central goal can be tracked by comparing the extent of a culture of evaluation now to similar measures at a future time. To enable such measurement, this section of the report provides a rough, concise snapshot of the current culture of evaluation within interpretation and education as a baseline for future comparison.

Three relatively replicable metrics or indicators make up this baseline measure. The first metric is the most discreet. It involves combining several items on a simple survey to create an average response number. The second measure is broader and involves documenting the behavior of a few key groups of stakeholders. The third measure is the broadest and entails making subjective inferences based upon the application of two different theoretical frameworks from the research literature. Taken as a group, these three measures can provide a reasonably clear overall picture of the level of the culture of evaluation at a given point in time.

The rest of this section presents evidence for the conclusion that, for the baseline snapshot as of the end of 2006, the interpretation and education function within the NPS is poised at a threshold of potential cultural change, but has not yet demonstrated systemic changes.

Measure 1. Readiness Survey Items

On a scale of 1 to 4, the “readiness for evaluation” index derived from 71 responses to the Summit reaction form yielded a 3.5 average (with a standard deviation [SD] of .4). **Thus, participants at the Summit reported a fairly high level of readiness for a culture of evaluation. Further details and discussion follow.**

Four of the numeric response items from the Summit reaction form¹³ were obtained directly from previous research that has been done on measuring how ready a non-formal education organization is to make effective use of program evaluation (Smith, 1992, as cited in Patton, 1997, p. 27). Averaging responses from these four items generated a “readiness for evaluation index” that can serve as an overall measure for any given sample.

For comparison, the readiness index value of 3.5 was higher than the 2.9 average (SD .7) for an item that asked about evaluation behavior. Responses to this item from the reaction form asked participants about the extent to which they agreed with this statement: “I habitually use outcome data to inform decisions in my everyday work.” This item about implementation is approximately a whole standard deviation lower than the readiness index described in the previous paragraph, suggesting that readiness in theory has not yet translated to regular evaluation practice. High readiness combined with room to improve on

¹³ The four items read: “A culture of evaluation within I&E would pave the way for better programs;” “Program evaluation would only increase our workload;” “‘Program evaluation’ and ‘accountability’ are just fads that hopefully will die down soon;” and “Now would be a good time to begin (or renew or intensify) work on program evaluation.” The response scale was a four-point agreement scale, and the items about “workload” and “fads” are reverse ordered.

Continuing to clearly frame, communicate, and accomplish tangible action steps is likely to be a critical factor in the future success of the implementation of the Systemwide Evaluation Strategy.

actual practice is a good place to be at the beginning stages of a major culture change effort.

As mentioned previously, this sample is probably not fully representative of the NPS community as a whole because this sample includes proportionately more leadership than field level staff. Members of this sample also had the benefit of participating in two days of dialogue about the potential of evaluation to serve the needs of interpretation and education within NPS. If a more rigorous and systematic measure of readiness for and practice of a culture of evaluation were deemed useful, these items from the Summit reaction form could be administered to a larger, more representative sample within the NPS.

Measure 2. Evaluation-related Behavior of Key Groups of Stakeholders

Broadening the data used to compile a composite snapshot of the current level of evaluative thinking within the NPS reveals three additional reference points in the observed behaviors of NPS leadership, researchers, and practitioners. Two of these reference points provide additional **evidence of beginning stages** of creating a culture of evaluation, and one indicates an **opportunity for continued growth** toward a more systemic culture of evaluation.

First, leadership (and other staff) attendance at evaluation-related events can be a quick indicator of the extent of a culture of evaluation. Summit facilitator Michael Patton made the following observation in his opening remarks:

“It is indeed an incredible baseline of the future of evaluative thinking in the National Park Service that the leadership has assembled here for this Summit, and people from across the Park Service at all levels [are] joining together in this activity.”

Second, the level of integration between the evaluation strategy and other measurement activities across the NPS can be seen as another piece of the puzzle. For instance, the Comprehensive Survey of the American Public is a project undertaken by the Social Science Program within the NPS every five years. The development of the evaluation strategy and accompanying Interpretation and Education Logic Models were too new to be integrated into the design process for the Comprehensive Survey that is due to be implemented in 2007. If the NPS continues to develop a systematic culture of evaluation, future administrations of the Comprehensive Survey might function more directly as a tool for measuring long-term outcomes as described in the Interpretation and Education Logic Model. Similarly, in future years and decades, the results of the wide range of existing evaluation and research activities currently occurring within the NPS¹⁴ could become more centrally organized and accessible, and perhaps even more coordinated.

A third additional data point for looking at the present level of a culture of evaluation is the findings from a recent informal survey of chiefs of interpretation. In June 2005, 81 park units responded to a questionnaire asking them to describe their evaluation questions, needs, and practices. Results from this survey were included in a targeted review of evaluation literature prepared by Dr. Nina Roberts and intern Jenni Mullins of the NPS Natural Resource Center, as well as several other interpretation practitioners, including members of the Education Evaluation Coordination Team. The executive summary is included in this report as an Appendix, and the complete results are available at www.nps.gov/interp/evaluation. Overall, the responses show a desire for more and better outcome data largely consistent with the evidence for enthusiasm for evaluation reflected in the evaluation data for the Summit.

It should also be noted that the results of this “Chiefs Survey” report exemplify one of the key pieces of advice that Summit panelists offered the NPS, and that Summit participants wholeheartedly embraced. Specifically, to create a genuine culture of evaluation, it is imperative to begin by asking local practitioners what they would like to know in order to do their jobs better.

In terms of establishing a current baseline for evaluative thinking within interpretation and education, the fact that this “Chiefs Survey” occurred more than a year before the Summit is further indication that at least some elements within the National Park Service are ready and willing (perhaps even hungry) to engage in useful evaluation. For the future, the extent of follow-up on the findings of this particular survey, as well as the number and type of similar efforts to collect and respond to the evaluation needs of the field staff, could be viewed as indicators of the level of NPS engagement in a culture of evaluative thinking. Further, this survey could perhaps even be replicated on a regular basis as a way to track changes in the extent of a culture of evaluation within interpretation and education.

Measure 3. Two Theoretical Frameworks for Understanding Cultural Change

For the broadest level view of charting the National Park Service’s progress toward creating a culture of evaluation, two conceptual frameworks are considered that emerge from research on change processes in other fields. Evidence reviewed for this evaluation is consistent with the types of **behaviors that are associated with**

¹⁴ For example, the work of Summit panelists Theresa Coble, Doug Knapp, Emma Norland, and Alison Druin, as well as other ongoing and episodic projects occurring within the NPS Social Science Program and/or conducted by other outside researchers.

the initial phases of the change processes described by each of these two theories.

The first framework is called “Diffusion of Innovations.” Research on this process (Moore, 1999; Gladwell, 2002; Rogers, 2003) suggests that people respond differently to new ideas and technologies based upon individual psychological and demographic characteristics. When faced with the uncertainty inherent in considering the adoption of a new technology or way of doing things, people tend to fall into one of the “adopter categories” described in Figures 9 and 10. The distribution of people in a given population tends to follow a normal, bell-shaped pattern with the early and late majority categories each comprising about a third of the population, and the innovators, early adopters, and laggards collectively making up the remaining third of the population.

National Park Service aspiration to create a culture of evaluation is a classic example of a diffusion of innovations process. Using the Evaluation Summit as one broad baseline indicator, it seems that all adopter categories were represented among Summit participants, and that the culture change process is probably hovering somewhere between having taken hold amongst the early adopters but not yet firmly established within the early majority.

The second conceptual framework that can help shed light on the current baseline status of a culture of evaluation within interpretation and education is called “Stages of Change.” This model for understanding intentional behavior change emerged from the field of behavioral psychology (Prochaska, Norcross, & DiClemente, 1992) and describes the ways that individuals and groups progress through a series of stages as described in Figure 11.

Figure 9. Adopter Categories (Moore, 1999, p. 17)

The Revised Technology Adoption Life Cycle

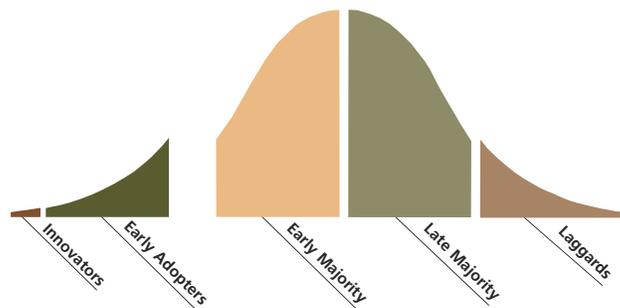


Figure 10. Adopter Category Summary Chart

Adopter Category	Descriptors (from Moore, 1999 from Rogers, 2003)	Core Wants	Strategies for Success
Innovator	Enthusiast Venturesome	<ul style="list-style-type: none"> • Straight facts, truth, no tricks • Be first 	<ul style="list-style-type: none"> • Don't expect immediate "profits" • Look for ones who can garner R&D support by virtue of being close to the "big boss"
Early Adopter	Visionary Respect	<ul style="list-style-type: none"> • Breakthrough technologies • Pursue a dream • Project orientation 	<ul style="list-style-type: none"> • Maintain frequent contact • Manage unrealistic expectations • Chunk innovations into discreet products or phases
Early Majority	Pragmatist Deliberate	<ul style="list-style-type: none"> • Incremental, predictable, measurable progress 	<ul style="list-style-type: none"> • D-Day analogy • Keystone species • Focus, focus, focus effort on strategic networkers and opinion leaders
Late Majority	Conservative Skeptical	<ul style="list-style-type: none"> • Smooth, easy change • Discount prices 	<ul style="list-style-type: none"> • Work the bugs out first • Plan for a customer service orientation
Laggard	Skeptic Traditional	<ul style="list-style-type: none"> • Keep status quo 	<ul style="list-style-type: none"> • Actively listen for "Emperor's New Clothes" phenomena (e.g. the Amish v. modern agribusiness) • Otherwise try to neutralize influence

According to the stages of change research, not all behavior patterns change at once. Rather, change happens in increments, and reversion to prior stages is normal and to be expected. Further, “the underlying structure of change is neither technique-oriented nor problem specific” (Prochaska, et al., p. 1110). Thus, success might be more accurately and usefully measured in terms of progression along a continuum of stages instead of solely in terms of having achieved the end goal in a linear, lockstep, theoretically predictable progression.

In reviewing the data collected from the Summit, it appears that there were participants at each stage of change regarding implementing evaluation. The Preparation stage was probably the most dominant at the Summit, and Contemplation or Pre-contemplation likely describes the current stage for most others at their regular duty stations.

Summary of Baseline Snapshot of Current Culture of Evaluation in the National Park Service

To be clear, this evaluation of the present status of evaluative thinking within the NPS is not intended to be a summative “standing in judgment of” the performance of NPS staff. Rather, the purpose is to provide some sense of a current baseline so that future measures will have something relatively tangible for comparison.

Another purpose of this discussion of baseline measures of a culture of evaluation is to promote critical thinking about the evaluation process. Evaluation projects should be held accountable to the standards they espouse.¹⁵ What evidence is there that investing resources in creating a culture of evaluation actually works? One important way to assess whether continued investment by the NPS in evaluation is warranted is to measure

progress regularly toward the central goal of the evaluation strategy, that is, to evaluate the evaluation.

The three indicators discussed in the preceding paragraphs are diverse. They range from narrowly focused survey items that afford precise (if somewhat thin) measurement, to broad conceptual frameworks that require a more holistic, subjective inference. Considered individually, none of these indicators is particularly compelling. However, regarded as a whole, they all point toward essentially the same conclusion. This triangulation among multiple sources and types of data warrants a relatively high degree of confidence in the overall claim.

In sum, as of the end of 2006, it appears that the interpretation and education function within the NPS is **poised at a threshold of potential cultural change, but has not yet demonstrated systemic changes.**

There are some hints of evidence in this evaluation suggesting that such a level of readiness may extend to other parts of the National Park Service as well. For instance, the Regional Director from the Alaska Region, Marcia Blaszak, made the following remarks:

“... while our focus in these two days was tied to the Interpretation and Education Program, from my perspective it is important that we take this culture of evaluation beyond just the Interpretation and Education Program. It has value to many of the programs we tend to stumble over.”

¹⁵ Evaluation should be useful, practical, ethical, and accurate.

Figure 11. Stages of Change Summary Chart

Stage	Description	Strategies for Success
Pre-Contemplation	<ul style="list-style-type: none"> Not intending to make a change in the next six months Not necessarily opposed, just not ready to start 	<ul style="list-style-type: none"> Focus on basic information of who, what, when
Contemplation	<ul style="list-style-type: none"> Thinking about making a change in next six months Ambivalent about costs v. benefits of the effort required 	<ul style="list-style-type: none"> Begin exploring why and how Present information, discussion in terms of “What’s in this for me?”
Preparation	<ul style="list-style-type: none"> Intending to make a change in the near future Convinced potential benefits outweigh the risks 	<ul style="list-style-type: none"> Provide coaching, skill and capacity building Pushing too quickly can lead to demoralization
Action	<ul style="list-style-type: none"> Overt behavior changes have been made 	<ul style="list-style-type: none"> Celebrate, encourage, support Be alert for overwhelm and slipping into previous stages
Maintenance	<ul style="list-style-type: none"> Maintained behavior change for at least six months Behavior has become more automatic The “old days” seem distant 	<ul style="list-style-type: none"> Regularly highlight progress made Continue some level of support Formalize self-support mechanisms



... change happens in increments, and reversion to prior stages is normal and to be expected.

Participants discuss how people respond differently to new ideas and technologies based on individual psychological and demographic characteristics.

However, claims of readiness for a culture of evaluation beyond interpretation and education within the NPS require further exploration before being asserted with confidence.

The following comment from a Summit participant captures the essence of this analysis. It also exemplifies the way that diffusion of innovations is a fundamentally social process (i.e., people in groups watch what their peers do):

“I’m the Executive Director of North Cascades Institute in Washington State, park partner, and I’m here for two reasons. One is because Jon Jarvis and Julia Washburn invited me, but the other was because of the language that was used in some of the materials. The one that really caught my eye was this talk about a culture of evaluation. What struck me today is the idea that that’s a culture of inquiry.”

So let me share a really brief poem by Robert Aiken, a Zen Roshi. It’s called *‘The Spirit of the Practice.’*

*One evening after meditation, Owl asked,
‘What is the spirit of the practice?’
Raven said, ‘Inquiry.’
Owl cocked his head and said,
‘What should I inquire about?’
Raven said, ‘Good start.’”*

Perhaps the highest purpose of presenting this baseline snapshot is to stimulate creative and effective thinking about the next steps in the long journey toward the ultimate vision of better serving the American public and our precious natural and cultural resources through embodying a culture of continuous learning within the NPS. The implications for next steps to follow up on this baseline snapshot are discussed in more detail in Part III of this report.



“ . . . it is important that we take this culture of evaluation beyond just the Interpretation and Education Program.”

—MARCIA BLASZAK, ALASKA REGIONAL DIRECTOR

Part III: Applying Lessons Learned

This part of the report is about moving beyond the findings of evaluating the Summit to inform next steps for implementing the long-term *Servicewide Evaluation Strategy*. In doing so, the goal is to *demonstrate* the kind of learning that can emerge from a culture of evaluation. Learning and evaluation are usually enhanced by collaboration, and this part of the report exemplifies that dynamic as well. A team consisting of NPS partners who provided evaluation expertise and NPS staff who provided in depth knowledge of NPS organizational needs and operations created this section of the report.

The one-page briefing statement prepared for the January 2007 meeting of the NPS National Leadership Council (NLC) is presented as Figure 12. It summarizes the way that the Education Evaluation Coordination Team translated findings from evaluation of the Summit into specific recommended action steps. Recommendations are divided into three categories (immediate, short-term, and long-term).

These recommendations reflect the next practical step in integrating the *Interpretation and Renaissance Action Plan* (that was previously endorsed by the NLC), the *Servicewide Evaluation Strategy* (which continues to evolve based on evaluation data and stakeholder input), and the lessons learned from evaluating the Summit. In general, findings from the Summit reinforce both the action plan and the strategy, while also adding some new insights and refinements to those preexisting frameworks. The remainder of this part of the report, following Figure 12, provides additional discussion, context, and rationale for each of the 13 recommendations. Rather than repeating the contents of the *Interpretation and Renaissance Action Plan* and *Servicewide Evaluation Strategy* documents, the discussion highlights the ways that evaluation of the Summit advanced thinking about relevant actions steps contained in those documents.

Opposite: Yellowstone National Park's Canyon Visitor Education Center features state-of-the-art, interactive exhibits that help visitors learn about and understand the geology of Yellowstone.

Figure 12. NLC Briefing Statement on Recommended Follow-up Actions

National Park Service
U.S. Department of the Interior



Briefing Statement

Bureau: National Park Service

Issue: Recommended follow-up actions to the Evaluation Summit

Date: 1/12/2007

Recommendations: All recommended actions integrate learning from the Summit and are consistent with the NLC endorsed National Education Council (NEC) action plan and the draft *Evaluation Strategy*.

Immediate actions (next 3 months): 1) Solicit feedback on recommended next steps from the National Leadership Council; 2) Complete, distribute to the field, and solicit feedback on Proceedings (including evaluation) and DVD of Summit; 3) Solicit feedback from partners on Summit Proceedings (including evaluation), the DVD of the Summit, and on the *Evaluation Strategy*.

Short term actions (up to 12 months from now): 1) Build a web portal/evaluation resources library with practical tools for parks and partners; 2) Enhance communication about evaluation within NPS and with partners; 3) Establish selection criteria and the process to identify pilot evaluation projects to be considered as funds become available; 4) Fill vacant GS-13 Evaluation and Visitor Studies Coordinator;

Long term, more comprehensive, actions (one to five years) include: 1) Require evaluation component for all funding sources in the Servicewide Consolidated Call (SCC) (selection criteria to include low cost options); 2) Systematically share lessons learned from existing evaluation projects; 3) Incorporate evaluation more tangibly into existing professional development opportunities, training programs, and reward systems; 4) Create a mini-grants program to promote small scale evaluation into questions of local interest; 5) Provide resources to systematically involve historically underserved audiences and communities in evaluation work; 6) Make available individualized technical assistance and support for evaluation "champions" in the field and at the national level.

Background: In early 2006 (in response to the draft Business Plan for Interpretation and Education) the NLC requested that the National Education Council (NEC) develop an *Action Plan for Interpretation and Education*. The developed Plan was endorsed by the NLC in August, 2006. A key tenet in the Plan calls for creating a culture of evaluation within Interpretation and Education in the NPS. As a first step in creating such a culture the Plan, and the draft *Servicewide Evaluation Strategy* for Interpretation and Education, recommended a Summit on Evaluation. In October, 2006 respected evaluation experts, NPS leaders at all levels of the organization, and partners participated in an Evaluation Summit at the University of Denver. The Evaluation Summit was an integral step in the *Servicewide Evaluation Strategy*. A collaborative team evaluated the Summit. Key findings include: strong enthusiasm and concern for creating culture of evaluation; attainment of all five intended short term Summit outcomes; the current status of a culture of evaluation within the NPS is that the agency is poised at a watershed of potential but has not yet demonstrated systemic change; and that future communication plans and action steps should address a wide range of readiness among NPS and stakeholders with respect to evaluation. The Summit and its evaluation reinforced and provided new insights into the action steps called for in section 4.0 of the *NEC Action Plan* and marks a transition to the Phase II objectives of the draft *Evaluation Strategy*.

Current Status: *NEC Action Plan* and *Business Plan* are in final draft and awaiting distribution to the field through a memo from the Director; draft *Evaluation Strategy* is out for review to the field; Evaluation Summit proceedings including evaluation findings, recommended next steps, and a DVD are being developed and will be distributed to leadership, the field, and key partners within the next three months.

Contacts: Sheri Forbes, Chief of Interpretation GLAC and Chair of the Evaluation workgroup; David Larsen, Co-Chair NEC; Julia Washburn, Co-Chair NEC.

Recommended Actions

Findings from the Interpretation and Education Evaluation Summit were very clear that the genuine needs of frontline field staff and partners must be served if the NPS is to succeed in creating a culture of evaluation . . .

RECOMMENDED IMMEDIATE ACTIONS

These three actions are recommended for implementation within the months immediately following publication of this report:

- A. Solicit feedback on recommended next steps from the National Leadership Council;
- B. Complete, distribute to the field, and solicit feedback on Proceedings (including evaluation) and DVD of Summit; and
- C. Solicit feedback from partners on summit proceedings (including evaluation), the DVD of the Summit, and on the evaluation strategy.

Action Steps A.–C. Solicit Feedback From Key Stakeholders

Considered as a group, these actions build directly upon the two major themes found in the Summit evaluation. By communicating to key stakeholders, the enthusiasm generated at the Summit (i.e., Major Theme 1) can continue to gather momentum. By gathering feedback *from* key stakeholders, concerns about implementation of a culture of evaluation (i.e., Major Theme 2) can be better understood and more directly addressed. Responding to feedback from these key stakeholders is a powerful way to practice reality testing, an important element of a strong culture of evaluation. The NLC was identified as a top-priority stakeholder during the creation of the plan for evaluating the Summit, so their continued input into next steps is crucial. Findings from the Summit were very clear that the genuine needs of frontline field staff and partners must be served if the NPS is to succeed in creating a culture of evaluation, indicating that NLC feedback and engagement is crucial as well.

The primary rationale for presenting these actions as most urgent is that they appear to be high-leverage, relatively low-cost, and strongly resonant with the values and strategies of stakeholder inclusion that many of the panelists and participants emphasized at the Summit.

RECOMMENDED SHORT-TERM ACTIONS

The following four actions are recommended for implementation within a year of the publication of this report.

- D. Build a Web portal and evaluation resources library with practical tools for parks and partners;
- E. Enhance communication about evaluation within the NPS and with partners;
- F. Establish selection criteria and the process to identify pilot evaluation projects to be considered as funds become available; and

- G. Fill vacant GS-13 Evaluation and Visitor Studies Coordinator position.

The rationale for naming these as short-term actions is that they follow evaluation findings from the Summit and are generally aligned with Phase II of the implementation matrix presented in the draft evaluation strategy (see Appendix F). Further, each of these actions requires some level of authorization from central NPS leadership, which is likely to require additional time to assure appropriate decision making thoroughness.

Action Step D. Build a Web Portal and Evaluation Resources Library

One important insight emerging from creating a baseline snapshot of the present status of a culture of evaluation within the NPS is the existence of a wide range of readiness levels. It seems that most NPS stakeholders are at “pre-contemplation” or “contemplation” stages of change with respect to creating a culture of evaluation, whereas a few are already in the “preparation” or “action” stages, ready to dive in and start building such a culture in practice.

A central, Web-based evaluation resources library could serve stakeholders at many different stages of readiness. This library could meet the needs of those who are contemplating getting started but need some concrete, tangible resources, and tools to begin. For those who are already conducting evaluation, such a portal or library could provide a mechanism for sharing their hard-earned insights and learning. It could be a cost-effective and time-efficient way for *all* parks to access at least *some* evaluation resources.

A convenient Web address, or URL, has been established for this resource: www.nps.gov/interp/evaluation. As content is added to this access point and mechanisms are put in place for NPS practitioners to include additional resources, the structure and functionality of the portal can evolve over time as more resources become available for coordination activities.

Creating an online evaluation library is presently described in the evaluation strategy as a Phase III activity to address objective No. 2 (establishing an evaluation information management system). Reflection on findings from the Summit suggests earlier (Phase II) initiation of the library, an activity that can serve as an active mechanism for peer-to-peer collaboration as well as sharing of practical and useful tools and insights for frontline staff.

Action Step E. Enhance Communication About Evaluation

The importance of communication and dialogue for maintaining enthusiasm and addressing implementation concerns was described in “Immediate Action Steps.” It warrants reiteration, however, that partners and frontline staff were repeatedly identified in the Summit evaluation data as stakeholders critical to the success of creating a culture of evaluation within the NPS.

A communication plan designed to support and inform partners and frontline staff about creating a culture of evaluation would likely be most effective if directed specifically at meeting the needs of “contemplators” or skeptics. People at that stage of readiness for change are best served by basic information (“Who?” “What?” “When?”) designed to answer proactively the question “What’s in this for me?” Another strategy for gaining credibility and promoting change with this audience is to acknowledge directly and publicly the potential barriers and challenges. This is a way of modeling risk taking, reality testing, and openness to learning from disconfirming information—all important elements of a strong culture of evaluation.

Action Step F. Begin Pilot Evaluation Process

Implementing a series of pilot evaluation projects is called for in the *Interpretation and Education Renaissance Action Plan* (Section 4.3) and in the *Servicewide Evaluation Strategy* (Phase II, objective No. 3—establish best practices in evaluation and interpretation and education). Discussions at the Summit about establishing selection criteria and processes for evaluation pilots were a rich source of data about enthusiasm for and engagement with creating a culture of evaluation. Envisioning potential pilot evaluation projects seemed to make the idea of evaluation more real and tangible for Summit participants. It was also a sensible way to show how the Interpretation and Education Logic Model can be used to support thinking about program design. The discussion about pilot project selection criteria and process was only begun at the Summit, and still needs to be finished.

The next step in this regard is to empower a person or group to facilitate completion of the pilot project criteria and selection process that was begun at the Summit. Regardless of when and from what source funds become available to begin implementing these pilots, the selection criteria and process should be securely in place and appropriately authorized as soon as possible in order to maintain momentum from the Summit. It should also be noted that involvement in developing pilot projects could be an excellent opportunity for engaging particularly enthusiastic partners, frontline staff, or senior leadership more deeply in the big picture, long-term goal of creating a culture of evaluation.

Action Step G. Fill Vacant GS-13 Evaluation, Visitor Studies Coordinator Position

Both the *Interpretation and Education Renaissance Action Plan* (Section 4.2) and the *Servicewide Evaluation Strategy* call for hiring a national evaluation coordinator.

The evaluation of the Summit presented evidence for accomplishment of all five intended short-term outcomes of the Summit. The Summit was essentially a success. However, the

evaluation also noted that the Summit was more of a starting gate than a finish line with respect to implementing the longer-term evaluation strategy. It is likely that the Summit raised more expectations than it satisfied. Thus, it is probably more important than ever that resources be invested to coordinate the growth and deepening of evaluation practice within the NPS. A national-level coordinator seems to be an advantageous way to centralize communication, growth, and efficiency in this ongoing effort.

An important first order of business for a national evaluation coordinator role could be to facilitate a thorough and inclusive process of evaluation pilot project selection and implementation.

RECOMMENDED LONG-TERM, MORE COMPREHENSIVE ACTIONS

The following six actions are recommended for implementation within one to five years of the publication of this report:

- H. *Systematically* share lessons learned from existing evaluation projects;
- I. Incorporate evaluation more tangibly into existing professional development opportunities, training programs, and reward systems;
- J. Require an evaluation component for all funding sources in the Servicewide Consolidated Call (SCC) (selection criteria to include low cost options);
- K. Create a mini-grants program to promote small-scale evaluation into questions of local interest;
- L. Provide resources to systematically involve historically underserved audiences and communities in evaluation work; and
- M. Make available individualized technical assistance and support for evaluation “champions” in the field and at the national level.

The first three of these action steps reflect ways that the Summit generally reinforced ideas that already existed in the evaluation strategy. The final three action steps, however, represent new refinements that emerged directly from evaluation of the Summit. All of these are placed in the long-term action steps category because they require more time and funding to implement effectively.

Action Step H. Systematically Share Lessons Learned From Evaluation

The core idea of this action step was already part of the Evaluation Strategy. The implications for practice that follow from the two main themes from the Summit evaluation (i.e., enthusiasm and concern) are largely consistent with this action. The Summit and its evaluation did, however, generate at least three related points that warrant mention.

First, multiple panelists explicitly highlighted the importance of moving from episodic to a more systemic approach to evaluation projects in order to maximize efficiency, learning, and utilization of evaluation results.

Second, there was much support at the Summit for the idea of making an organizational culture safe for experimentation,



... evaluation processes can promote and even lead the way in establishing culturally competent programs and practices.

Summit discussions about establishing selection criteria and processes for evaluation pilots provide a rich source of data about enthusiasm and engagement.

innovation, and “failure” as a way to maximize learning. A system for sharing evaluation results across the NPS that honored and celebrated the learning that resulted from things not going as planned could go a long way toward reducing anxieties and compromised performance associated with a compliance mentality and “gotcha” culture.

Third, preliminary analysis of the pilot evaluation project criteria that were generated at the Summit shows clearly that participants want to learn from each other’s evaluation efforts.

Action Step I. Build Evaluation Into Existing Training Programs

This action is woven throughout multiple objectives presented in the evaluation strategy. Evaluation of the Summit provided evidence in support of this approach in the form of many requests for evaluation tools and training. There were also a handful of specific suggestions to integrate evaluation into the IDP.¹⁶

Action Step J. Incorporate Evaluation Into Servicewide Consolidated Call

This action represents a creative synthesis of two key insights gleaned from the Summit. One insight followed from realizing that only a small amount of Summit evaluation data existed to support the (supposedly) “short-term” Summit outcome of finding evidence of park-level staff using outcome data to inform decision-making. The lesson learned was that this intended result of the Summit would have been more appropriately framed as a longer-term outcome. The other insight derived

from documenting the high degree of engagement that Summit participants demonstrated with respect to defining selection criteria for potential pilot evaluation projects. This reinforced the basic lesson that having clear criteria for a complicated task can help guide and coordinate behavior between different parts of an organization.

Integrating these two evaluation insights with the action step in the evaluation strategy that calls for linking reward systems to evaluation best practices yielded the idea of working toward an eventual requirement that all funded sources include some kind of evaluation component. The Servicewide Consolidated Call seemed like an efficient and advantageous place to begin to promote the use of outcome data concretely and systematically to inform decision-making.

As a way to stimulate implementation of this action step, a unit such as Harpers Ferry Center could develop a Request for Proposal (RFP) template for Indefinite Delivery, Indefinite Quantity (IDIQ) contractors¹⁷ for evaluation of interpretive services (both personal and media). If every funded project

¹⁶ The IDP (Interpretive Development Program) is a training and development system within NPS that provides mission-based curricula, tools, and peer review aimed at helping visitors have meaningful experiences. It has operated since 1995.

¹⁷ Essentially, a short list of pre-screened contractors that could be available for projects of “Indefinite Delivery, Indefinite Quantity.”

requires an evaluation component, enough work should exist for evaluation contractors to warrant development of such a list. This approach could achieve the following: (1) support parks in locating evaluation contractors; (2) increase the likelihood that parks get the most effective evaluators; and (3) increase efficiency and reduce the cost of evaluation services.

In any case, criteria for including evaluation in funded projects should include options for relatively low-cost, locally directed evaluation practices so that the evaluation does not become a burdensome, externally driven compliance activity. Tools to support such low-cost evaluation practices should be available through the Web portal and evaluation resources library described in Action Step D.

Action Step K. Create an Evaluation Mini-grants Program

This action step was not previously included in the evaluation strategy, but rather was a specific suggestion that emerged from small group discussions at the Summit about how to make pilot evaluation projects most successful. An advantage of such a program would be that it could reinforce the notion that evaluation does NOT have to be a big, externally mandated, high stakes process. Rather, a culture of evaluation can be characterized by continually applying an evaluation mindset to small tasks each day. Based on analysis of the Summit proceedings, building capacity for small-scale evaluations that are driven primarily, if not exclusively, by the questions that local, frontline staff have about how to improve their work is likely to be an essential component of a successful culture of evaluation within the NPS. An evaluation mini-grants program is one way to support such grassroots evaluation work.

Action Step L. Reach Out to Historically Underserved Populations

A powerful theme weaving through the Summit was the importance of engaging *all* Americans in the good work of the NPS. Several panelists suggested that evaluation processes can promote and even lead the way in establishing culturally competent programs and practices. Summit participants emphasized the importance of potential pilot projects embodying diversity, not just in park size, type, and budget, but also in terms of audiences reached. The Summit also included conversation about the particular challenges associated with reaching out to historically underserved populations. The process must require creativity, patience, collaborations with diverse groups, and a willingness to explore, think, and engage differently. These are all things that can be mutually reinforcing with a vibrant culture of evaluation. Systematically imbuing implementation of the evaluation strategy with cultural competence warrants attention as a distinct element of the work in the coming years.

Action Step M. Provide Technical Support to Evaluation “Champions”

This action step follows directly from the “Diffusion of Innovations” research that was used to characterize the current baseline culture of evaluation as evidenced by evaluation data from the Summit. The transition between early adopters and the early majority is a notoriously perilous time for many innovations.

In some business circles (Moore, 1999), this is referred to as “crossing the chasm.” It is precisely where most new business ventures fail, and it is also the threshold at which interpretation and education within the NPS appears to be poised. The reason this transition is so difficult is that people in the mainstream (i.e., early majority) tend not to trust the views of the enthusiastic and visionary innovators and early adopters. One group is more enthusiastic about potential benefits, and the other group is more concerned with potential risks and costs. Early adopters are not good salespeople for the early majority. Thus, the early majority tends not to cross a “tipping point” (Gladwell, 2002) in the culture until they see their peers in the early majority already doing it.

The trick, then, is to get those first few people in the early majority on board to start the snowball rolling. The Summit seems to have had some elements of this dynamic, but not enough to establish continued diffusion with confidence. More investment, energy, endorsement, and time are almost certainly needed. National Park Service field staff and partners who are particularly enthusiastic about implementing evaluation can promote contagious spread of a culture of evaluation, especially if they are supported with resources and recognition as evaluation “champions.”

CLOSING REFLECTION ON RECOMMENDED ACTION STEPS

An overall implication for practice emerging from evaluation of the Summit is that next steps for cultivating a culture of evaluation within the NPS should NOT assume a “one size fits all” approach. Activities and strategies should be segmented and designed specifically to meet the needs of stakeholders in different stages of change and innovation adopter categories (as described in Part II, Section C, Measure 3 of this report).

A key rationale for this segmented approach is that there was evidence from the Summit of a fairly wide range of readiness for evaluation at the level of individual stakeholders. In the terminology of “Stages of Change” research most Summit participants seemed to be at the Preparation stage. There was, however, also clear evidence that many individuals are still at the Contemplation or Pre-contemplation stage, and some have already moved past Preparation and well into the Action or even Maintenance stage. If the goal is to reach the National Park Service as a whole, the unique needs of stakeholders in *each* of these stages will need to be specifically addressed.

This interpretation is also consistent with the conceptual framework derived from research on “Diffusion of Innovations.” From this perspective, the innovation of “evaluative thinking” seems to have reached many of the early adopters, but has not yet become established in the mainstream group of the early majority. Again, the evaluation data for the Summit suggested that there were people spanning a wide range of attitudes with respect to a potential cultural shift toward evaluative thinking within interpretation and education. The prescription for supporting innovations at this early stage of diffusion is to be very targeted and intentional in action steps, investing heavily in a few key opinion leaders, networkers, or trusted champions within the mainstream.

References

*A convenient Web address has been established for evaluation resources:
www.nps.gov/interp/evaluation.*

- Gladwell, M. (2002).** *The tipping point: How little things can make a big difference*. Boston, New York, London: Little, Brown & Co.
- Miles, M. B., & Huberman, A. M. (1994).** *An expanded sourcebook: Qualitative data analysis (2nd ed.)*. California: Sage.
- Moore, G. (1999).** *Crossing the chasm: Marketing and selling high-tech products to mainstream customers*. New York: Harper Collins.
- Patton, M. Q. (1997).** *Utilization-focused evaluation: The new century text (3rd edition)*. Thousand Oaks: Sage.
- Prochaska, J., DiClemente, C., & Norcross, J. (1992).** In search of how people change: Applications to addictive behaviors. *American Psychologist*, 47(9), 1102-1114.
- QSR International. (2006).** *QSR NVivo version 7.0.247.0*. [Computer software]. Australia: Author.
- Rogers, E. (2003).** *Diffusion of innovations (5th ed.)*. New York: Free Press



Findings from the Summit were very clear that the genuine needs of frontline field staff and partners must be served if the NPS is to succeed in creating a culture of evaluation . . .

Appendices

An annotated table of contents is provided for the Appendices to facilitate streamlined hard-copy printing of this report without the full Appendices. The complete report (with full Appendices) is available at www.nps.gov/interp/evaluation.

APPENDIX A: Complete Summit Participant List A-2

An alphabetical listing of all Summit participants, including their title and affiliation.

APPENDIX B: Summit Reaction Form (including selected results) A-5

A copy of the Reaction Form instrument is included, as well as selected summary data. Descriptive statistics are provided for all Likert scale items, as well as complete responses to open-ended items as offered by panelists and guests experts only.

APPENDIX C: Summary of Pilot Project Criteria Discussions A-10

An analytical summary of the major, minor, and unique themes that emerged during small group discussions about selection criteria and process for potential pilot evaluation projects.

APPENDIX D: “Chiefs” Survey Executive Summary A-12

A summary of findings from a July 2005 informal survey sent to Chiefs of Interpretation, asking them about their evaluation needs.

APPENDIX E: Pre-Summit Statements From Panelists A-16

Prior to the Summit, each panelists crafted a three- to four-page summary of key ideas that they thought would be most useful to prepare participants for the Summit. They are provided in full as a provocative summary of important and relevant evaluation stories.

APPENDIX F: Evaluation Strategy Implementation Matrix A-48

This is a chart included in the draft *Servicewide Interpretation and Education Evaluation Strategy*. It shows the proposed phased implementation of various activities that support each of the four main objectives of the strategy. It is provided as a specific reference for the Applying Lessons Learned section of the report.

Opposite: Volunteers In Parks (VIP) and Student Conservation Association (SCA) volunteers play a vital role in park operations at Canyonlands.

Appendix A: Complete Summit Participant List

A

Craig Ackerman, Superintendent, Oregon Caves National Park
Deanne Adams, Chief of Interpretation & Education, Pacific West Region, NPS
Christie Anastasia, Learning Center Education Coordinator, Pt. Reyes National Seashore
Donna Asbury, Executive Director, Association of Partners for Public Lands

B

Karen Ballentine, Education Branch Chief, Great Smoky Mountains National Park
Les Baxter, Deputy Director of Evaluation, The Pew Charitable Trusts
Teresa Behrens, Director of Evaluation, W. K. Kellogg Foundation
Brad Bennett, Chief of Interpretation & Education, Alaska Region Office, NPS
Marcia Blaszak, Regional Director, Alaska Region, NPS
Beth Boland, Historian and Program Manager, TwHP Heritage Education Services NPS
Mary Bomar, Director, National Park Service
Tracy Bowen, Executive Director, Alice Ferguson Foundation
Jim Boyd, Technology Enhanced Learning Program Manager, National Park Service
Curt Buchholtz, Executive Director, Rocky Mountain Nature Association
Joanie Budzileni, Chief of Interpretation, San Juan National Historic Site

C

Linda Canzanelli, Acting Deputy Regional Director - Northeast, National Park Service
Lyn Carranza, Chief of Interpretation, Petrified Forest National Park
Diane Chalfant, Chief of Interpretation, Yellowstone National Park
Judy Chetwin, Interpretive Specialist, Intermountain Region, NPS
Vin Cipolla, President and Chief Executive Officer, National Park Foundation (did not attend)
Delia Clark, Director of Community Engagement, NPS Conservation Study Institute
Teresa Coble, Assistant Professor, Stephen F. Austin University
Cathleen Cook, Chief of Resource Education, Great Smoky Mountains National Park (did not attend)

D

Jack Davis, Member, Director's Council, National Park System Advisory Board
Neil DeJong, Chief of Interpretation & Education, Intermountain Region, NPS
Lynn Dierking, Associate Director, Institute for Learning Innovation
Allison Druin, Associate Professor, Director, Human-Computer Interaction Lab, Univ. of Maryland
Michael Duffin, Principal, Program Evaluation and Educational Research Associates

E

Lakita Edwards, Education Specialist, Harper's Ferry Center, NPS
Melissa English-Rias, Chief of Interpretation, Martin Luther King National Historic Site
Eric Epstein, Media Specialist, Harper's Ferry Center, NPS
Jennifer Epstein, Park Ranger, National Mall & Memorial Parks

F

Sheri Forbes, Chief of Interpretation & Education, Mt. Rainier National Park
Loran Fraser, Director, National Parks Centennial Project, National Geographic Society
Art Frederick, Deputy Director, Southeast Region, NPS
Marie Frias Sauter, Superintendent, Fort Union National Monument
Robert Fudge, Chief of Interpretation & Education, Northeast Region, NPS

G

James Gasser, Management Analyst, Office of Policy, National Park Service
David Given, Deputy Regional Director, Midwest Region, NPS
Jim Gramman, Visiting Social Scientist, NPS; Professor, Texas A & M University (did not attend)
Peter Grant, Board of Trustees, National Trust for Historic Preservation

H

Flip Hagood, Senior VP, Strategic Initiatives, Business Development, Student Conservation Assoc.
Sue Hansen, Chief of Interpretation & Education, National Capitol Region, NPS
Kathleen Harter, Chief of Interpretation & Education, Keweenaw National Historical Park

Gayle Hazelwood, Superintendent, National Capitol Parks East
Rita Hennessy, Outdoor Recreation Planner, NPS-AT Park Office
Elizabeth Hoermann, Education Specialist, Northeast Center for Education Services, NPS
Ginger Hollingsworth-Cox, Park Ranger, Cumberland Island National Seashore
Pat Hooks, Regional Director, Southeast Region, NPS

I

Catey Iacuzzi, Project Assistant, PEER Associates

J

Chris Jarvi, Associate Director, Partnerships, I&E, Volunteers, Outdoor Recreation, NPS
Jon Jarvis, Regional Director, Pacific West Region, NPS

K

Douglas Knapp, Associate Professor, Indiana University
Victor Knox, Deputy Regional Director, Alaska Region, NPS
Don Kodak, Director, Harper's Ferry Center, NPS
Richard Kohen, Interpretive Specialist, Intermountain Region, NPS
Amber Kraft, Park Ranger, Independence National Historical Park
Elisa Kunz, Education Program Specialist, Jefferson National Expansion Memorial

L

David Larsen, Co-Chair, National Education Council; Training Manager, Mather Training Center
Jennifer Lee, Special Assistant to the Director, National Park Service
Paul Lee, Interpretive Planner, National Park Service
Linda Lutz-Ryan, Interpretive Specialist, Intermountain Region, NPS

M

Cynthia MacLeod, Superintendent, Richmond National Battlefield Park
Fran Mainella, Former Director NPS; Member, Director's Council, NPS Advisory Board
Steve Martin, Deputy Director Operations, National Park Service
Jan Matthews, Associate Director Cultural Resources, National Park Service
Megan McBride, Senior Research Associate, Social Science Program, NPS
Lisa Mendelson-Ielmini, Deputy Regional Director, National Capitol Region, NPS
Tim Merriman, Executive Director, National Association for Interpretation
Alden Miller, Chief of Interpretation & Education, Sitka National Historical Park
Julie Miller, Interpretive Services Manager, Delaware North Company
Charles Money, President, Alaska Natural History Association
Martha Monroe, Associate Professor, Natural Resources Education & Extension, Univ. of Florida
Greg Moore, Executive Director, Golden Gate National Parks Conservancy
Chesley Moroz, President, Eastern National
Jason Morris, Director of Community Engagement, Yosemite National Institute
John Morris, Education Coordinator, Alaska Regional Office, NPS
Cicely Muldoon, Deputy Regional Director, Pacific West Region, NPS

N

Dom Nessi, Chief Information Officer, National Park Service (did not attend)
Polly Nordstrand, Curator of Native American Collection, Denver Art Museum
Emma Norland, Consultant, E-Norland Evaluation

O

Joseph O'Leary, Professor, Dept. of Recreation, Park, & Tourism Sciences; Texas A & M University
Anne O'Neill, Program Director for Education, National Park Foundation

P

Barbara Pahl, Regional Director, National Trust for Historic Preservation
Michael Patton, Founder & Director, Utilization-Focused Evaluation
Cherry Payne, Chief of Interpretation, Everglades National Park
Martin Perschler, Acting Manager, Park History, NPS Editor, CRM
Dwight Pettiford, Chief USPP

Q

John Quinley, Assistant Regional Director, Communications, Alaska Regional Office, NPS
Ernest Quintana, Regional Director, Midwest Region, NPS

R

Patti Reilly, Director, Northeast Center for Education Services, NPS
Jeff Reinbold, Acting Superintendent, Fort Necessity & Friendship Hill
John Reynolds, Executive Vice President, Grants & Strategic Alliances, National Park Foundation
Megan Richotte, Kennecott District Ranger, Wrangell-St. Elias National Park & Preserve
Tom Richter, Chief of Interpretation, Education, & Planning, Midwest Region, NPS
Dan Ritchie, Chancellor Emeritus, University of Denver

S

Diana Saathoff, Executive Director, Mount Rushmore National Memorial Society
Tony Schetzle, Deputy Regional Director, Intermountain Region, NPS
Shirley Sears Smith, Management Analyst, Office of Policy, National Park Service
Bruce Sheaffer, Comptroller, National Park Service, (did not attend)
Tessy Shirakawa, Chief of Visitor Service & Public Information Officer, Mesa Verde National Park
Kim Sikoryak, Interpretive Specialist, Intermountain Region, NPS
Jerry Simpson, Assistant Director, Human Capital, National Park Service
Carolyn Snowbarger, Director, Teacher to Teacher Program, U.S. Department of Education
Mike Snyder, Regional Director, Intermountain Region, NPS
David Sobel, Director, Teacher Certification Programs, Antioch University New England
Antonio Solorio, Park Ranger, Santa Monica Mountains National Recreation Area
Mike Soukup, Associate Director Natural Resource Stewardship and Science, NPS
Robert Stanton, Former Director NPS; Member, Director's Council, NPS Advisory Board
Carol Stapp, Director, Museum Education Program, George Washington University
Blanca Stransky, Superintendent, Agate Fossil Beds National Monument
Hazel Symonette, Professor, University of Wisconsin, Madison
David Szymanski, Detail, Pacific West Region, NPS

T

Karen Taylor-Goodrich, Associate Director Visitor and Resource Protection, NPS
Veronica Thomas, Professor, Human Development Program, Howard University
Naomi Torres, Outreach Specialist, Golden Gate National Recreation Area
Rich Turk, Value Analysis Program Coordinator, Washington Office, NPS

V

Sam Vaughn, Associate Manager, Interpretive Planning, Harper's Ferry Center, NPS
David Vela, Superintendent, George Washington Memorial Parkway
Gay Vietzke, Superintendent, Fort McHenry NM & Historical Shrine
Ken Voorhis, Executive Director, Great Smoky Mountains Institute at Tremont

W

Ronald Walker, Former Director NPS; Member, Director's Council, NPS Advisory Board
Sandy Walter, Acting Regional Director, Northeast Region, NPS
Bill Walters, Member, National Park System Advisory Board
Julia Washburn, Co-Chair NPS Education Council; Conservation Studies Institute, NPS
Sandy Weber, Cultural Resources Interpretive Specialist, Interpretation & Education, NPS
Saul Weisberg, Executive Director, North Cascades Institute
Dan Wenk, Acting Associate Director - Park Planning, Facilities, and Lands, NPS
Jon Wergin, Professor, PhD Program in Leadership and Change, Antioch University
Mike Whatley, Manager Information Services Branch, Natural Resources Program Center, NPS
Palma Wilson, Superintendent, Flagstaff Area National Monuments - WUPA
Don Wollenhaupt, Chief of Interpretation & Education, Southeast Region, NPS

Appendix B: Summit Reaction Form

National Interpretation & Education Evaluation Summit

Reaction Form

Your frank reactions to this Summit are essential for tracking progress toward our intended outcomes. By next week you will have electronic access to the compiled responses to the questions on this form. Your participation in this event is greatly appreciated, and **THANK YOU** for your thoughtful feedback.

How much do you disagree or agree?
For each item, please circle the number that best describes your opinion.

	Strongly DISAGREE	Tend to DISAGREE	Tend to AGREE	Strongly AGREE	Not Not Sure
I understand the Servicewide Interpretation and Education Evaluation Strategy	1	2	3	4	0
I have a clear picture of the action steps for rolling out the Evaluation Strategy	1	2	3	4	0
I enthusiastically endorse increased support for evaluation-related activities within I&E	1	2	3	4	0
A culture of evaluation within I&E would pave the way for better programs	1	2	3	4	0
Program evaluation would only increase our workload	1	2	3	4	0
I habitually use outcome data to inform decisions in my everyday work	1	2	3	4	0
"Program evaluation" and "accountability" are just fads that hopefully will die down soon	1	2	3	4	0
Overall, the format of the I&E Evaluation Summit was effective	1	2	3	4	0
Now would be a good time to begin (or renew or intensify) work on program evaluation	1	2	3	4	0
The I&E Evaluation Summit was a good use of NPS resources	1	2	3	4	0

What were the couple most important insights, ah-ha moments, or takeaway messages you got from this event?

My role in the NPS is best described as (please check one):

- Front line staff
- Park level Superintendent or manager
- Regional level manager
- Senior leadership
- Partner
- Summit panelist or external guest
- Other

Please fill out both sides

National Interpretation & Education Evaluation Summit

Reaction Form (continued)

What do you need to have or know in order to take the next step in building evaluation thinking into your everyday work? Or, what concerns do you have about the Evaluation Strategy?

What, if anything, do you plan to DO differently as a result of attending the Evaluation Summit?

What advice do you have about how to move the Evaluation Strategy forward most effectively?

Selected Summary Results From *Summit Reaction Form*
Administered October 26-November 19, 2006

(NOTE: The data presented in this table represent the complete sample, including participants in all different roles)	Strongly DISAGREE	Tend to DISAGREE	Tend to AGREE	Strongly AGREE	N/A or Not sure	Response Total
I understand the <i>Servicewide Interpretation and Education Evaluation Strategy</i> .	0% (0)	7% (5)	57% (40)	34% (24)	1% (1)	100% (70)
I have a clear picture of the action steps for rolling out the evaluation strategy.	1% (1)	23% (16)	61% (43)	10% (7)	4% (3)	100% (70)
I enthusiastically endorse increased support for evaluation-related activities within I&E.	1% (1)	0% (0)	12% (8)	87% (60)	0% (0)	100% (69)
A culture of evaluation within I&E would pave the way for better programs.	1% (1)	0% (0)	15% (11)	82% (58)	1% (1)	100% (71)
Program evaluation would only increase our workload.	19% (13)	59% (41)	17% (12)	3% (2)	1% (1)	100% (69)
I habitually use outcome data to inform decisions in my everyday work.	0% (0)	30% (21)	49% (34)	17% (12)	4% (3)	100% (70)
"Program evaluation" and "accountability" are just fads that hopefully will die down	66% (47)	27% (19)	4% (3)	0% (0)	3% (2)	100% (71)
Overall the format of the I&E Evaluation Summit was effective.	0% (0)	3% (2)	62% (43)	35% (24)	0% (0)	100% (69)
Now would be a good time to begin (or renew or intensify) work on program evaluation.	0% (0)	1% (1)	23% (16)	74% (52)	1% (1)	100% (70)
The I&E Evaluation Summit was a good use of NPS resources.	0% (0)	1% (1)	34% (24)	65% (46)	0% (0)	100% (71)

Anonymous responses from all four open-ended questions are available at <http://www.surveymonkey.com/Report.asp?U=276484043561>. Below are the responses from panelists and guest experts only.

(Panelists and Guest Experts) What advice do you have about how to move the evaluation strategy forward most effectively?
(David Sobel) It appeared to me that the Independence Park work is a possible gold mine of good evaluation. It appears they have a lot of pre/post content data that they're not quantifying, but is available. I also think this is possibly evaluation with a lot of political traction--i.e. it's data about students learning about our constitutional heritage. What could be more mom and apple pie than that. So, it's evaluation data that reaches across the aisle and would be persuasive to congressmen and senators of all political persuasions. Also, it appears that the program is serving mostly white suburban students while 90% of students within two miles of the school are students of color. This could translate into funding to support nearby student participation. It also appears that the programs are supported by local foundation funding that requires evaluation, but that none of that evaluation has been done. Therefore, the park staff needs to complete evaluation to fulfill grant requirements in order to make them eligible for future private funding. So conducting evaluation here is crucial.

Selected Summary Results From *Summit Reaction Form*
Administered October 26-November 19, 2006

<p>(Jon Wergin) Derive some key points and major themes from the Summit and use these as touchstones to come back to. Stay "on message." Follow up on the expressed desire of participants to inquire into questions of local interest. Determine what top management can do to encourage and facilitate the process. Tell stories about successful inquiries.</p>
<p>(Terri Behrens) Create opportunities (regional meetings, perhaps) for park personnel and their partners to share what questions (not answers) they have. The goal is to engage people's curiosity, rather than tapping their anxiety about evaluation.</p>
<p>(Polly Nordstrand) Look at all opportunities to conduct evaluation, especially those that involve the non-park visiting public</p>
<p>(Theresa Coble) Funding is critical; but equally critical is the effective engagement, empowerment and ownership of the initiative across all sectors (e.g., field personnel, interpretive supervisors, park management, partner organizations, and evaluators-- internal and external). I recommend that a mini-grant program be one component of the implementation of the SIEES plan. Parks should be able to apply directly for mini-grants so that they can do their own in-house evaluation. I also suggest that the dialogue about pilot projects that occurred on the second day of the Summit be continued post-summit and that a provision be made for the involvement of a variety of interested parties. Providing an opportunity for folks like myself, i.e., academics who do I & E evaluation in the parks, to participate in the discussion of project design would be beneficial.</p>
<p>(Veronica Thomas) Be consistent Start with small and realistic goals that can be realized within a reasonable period of time Provide adequate resources (e.g., financial, people power, etc) Get staff across the NPS to embrace the evaluation culture -- evaluation is everybody's job!</p>
<p>(Michael Patton) Set up some benchmarks for accountability on implementation of the evaluation portions of the action plan.</p>
<p>(Doug Knapp) For this answer, I defer to our break out group's recommendation to offer a variety of pilot projects that reflect a diverse set of sites / programs / and populations served. From small rural Parks to large urban Parks etc. Projects that offer a variety of assessment strategies from short term quantitative to long term qualitative.</p>
<p>(Allison Druin) I would suggest choosing some pilot projects to test out methods and outcome possibilities. I would try to choose projects that highlight different approaches so that NPS sees there are a variety of ways to approach evaluation. In actuality it shouldn't be considered one strategy, but many depending on goals and context.</p>
<p>(Martha Monroe) Move quickly toward hiring the Coordinator and identifying pilot projects. Aim for lots of diversity and people who can champion evaluation. Build partnerships with other agencies and organizations who can help and support NPS administrators and field folks as they move in this direction. Get the list of NPS staff who have attended evaluation courses and workshops recently, or who have been involved in evaluation, and get their input on pilot projects. It sounds like there is money in the new budget for doing this. I think you can assume you have approval -- now get started!</p>
<p>(Other) Careful thought needs to be put into how this strategy is actually rolled out - Who will conduct these pilots, how will front-line staff be involved, training is critical, etc</p>
<p>(Other) An external evaluation team should be hired to work with the pilot projects to do good design. Design an evaluation training program for staff of the pilot project. Think about political traction as a parameter for some of the pilot projects. Independence Park Institute needs an evaluation. They've got a goldmine of data and are required to do evaluation as part of external funding that hasn't been done. I nominate them as one of the pilot projects.</p>

Selected Summary Results From *Summit Reaction Form*
Administered October 26-November 19, 2006

(Other) Be patient. Try different approaches. Recognize people who make good faith efforts to apply evaluation. Share progress & challenges. Focus first on your willing partners. May need to do some training in evaluative thinking (e.g. cultivating a shared vocabulary).

(Other) Keep it front & center in communications & reports, budgets, discussions. Move quickly to hire coordinator & ID at least some of the pilot projects. Make sure some pilots are driven by staff and can be perceived as "doable"

(Other) Make sure resources and capacity building efforts are solidly in place

(Other) Highly publicized pilot projects real resources, new resources central point of contact - advocate, knowledgeable

(Panelists and Guest Experts) What were the couple most important insights, ah-ha moments, or takeaway messages you got from this event?

(David Sobel) That it would be really good to do further evaluation of the in depth Parks as Classroom projects. I'd be especially interested to see academic achievement data from the Pi Beta Phi Elementary School in Gatlinberg, TN and Great Smokies Nat'l Park. I most liked Jon Wergin's panel presentation and a couple of Michael Quinn Patten's comments. I thought the format for the afternoon panel presentation should have been changed to provide greater textual diversity. I think that ideas for possible evaluation pilot projects should have started to be discussed earlier. And I think we could have made further progress towards this goal by having less large group discussion and more focused small group discussion.

(Jon Wergin) The readiness expressed by most of those in attendance to nudge the NPS toward a "culture of inquiry," if not a "culture of evaluation."

(Terri Behrens) The complexity of the NPS was surprising -- and it can be an asset or a liability. It can be a tremendous asset if the strengths of all the partners are used. Many of the private sector partners already have a great deal of experience in evaluation that can be leveraged.

(Polly Nordstrand) This activity is an opportunity for NPS to embrace learning from their visitors

(Theresa Coble) I found John Wergin's comments about organizational evaluation and change insightful. As an academic, I could readily appreciate the focus of his research. In addition, I could see the applicability of his findings to agency contexts. On another note, I was encouraged, and remain encouraged to see the tremendous support that this initiative has within the agency, especially at the leadership level. With the implementation of the SIEES plan, and particularly if the requisite funding to fully implement the plan does in fact materialize in 2008, I am optimistic that the NPS can indeed create a culture of evaluative thinking. To see an increased emphasis on Interpretation and Education as a cornerstone of preparing for the NPS centennial makes sense to me. I hope I can work with the agency in the coming years to make it happen... Finally, a lot of people spent a lot of time gearing up for this event. I applaud and appreciate all their efforts.

(Veronica Thomas) The tremendous potential to move the evaluation agenda forward given all the "top level" NPS people in attendance and supportive of the Summit's goals

(Michael Patton) The Park Service leadership and staff commitment to evaluation is impressive and important.

Selected Summary Results From *Summit Reaction Form*
Administered October 26-November 19, 2006

(Doug Knapp) For me, there were two general impressions I left with the meeting. First, I was truly excited by the variety of high level people who were there and who were clearly interested in the notion of assessment / evaluation. As both a listener and a panel participant it was impressive! My second impression was the great analogies Dr. Patton used related to research and evaluation. They were very inspiring! Overall, one of the better meetings I have been a part of as both a panelist and participant!

(Allison Druin) I guess it was when we were taking questions and I was asked to address the technology-related questions. To me it was clear from the number of hands in the air to talk about these issues, that it is an area people are grappling with and want to discuss. I know people wandered off from thinking about evaluation per se, but it suggested to me that there may be some way to get people engaged in moving toward a culture of evaluation if we incorporate issues/uses/developments of technology.

(Martha Monroe) I greatly enjoyed Patton's version of Patton's stories that I've used in evaluation workshops! For my own workshops and evaluation teaching, I really appreciated the discussion about protecting creativity from accountability and the struggle with "negative evaluation results." I think that is really important to consider and was glad we spent some time on it. I think it will be difficult to implement, however, and more effort will be needed. I think starting with volunteer supervisors -- people who want to experiment with evaluation and all that it entails, will be helpful. I also perked up when Patton explained that evaluation tends to focus on the short term outcomes, NOT the long-term ones. It is research that provides the logical connections between the short term outcomes and allows us to project that there will be long term impacts. Unfortunately, I don't think we've done a good job of compiling and making accessible the research results that will help folks create meaningful logic models. Perhaps this is a project that could be funded to help the NPS and the field. Maybe all the panelists could contribute! I liked the idea of creating a culture of professionalism (incorporating the mushy/fuzzy concept of care and the punitive notion of evaluation). I hope they can run with that. I felt like twirling like Snoopy with my nose in the air when the NPS brass talked about how they were going to implement this new culture -- allowing them to give voice to evaluation for interp and education not only sets a good tone for the field, but created their verbal commitment to action. A Brilliant Idea!

(Other) Audience's as assets/participants in evaluation process - Linking evaluation with core NPS values - learning & integrity - Community of care rather than community of evaluation - Navajo/Hopi example of room arrangement and informal sharing as example of cultural competence

(Other) Relationship between evaluation accountability and innovation - Coercive evaluation vs. inquiry evaluation - Teresa Coble's presentation on research - Alison Druin's portrayal of technology as an engagement vehicle for children and nature -Michael's reframing of questions

(Other) Importance of culture history to NPS Challenge of how to approach I&E Distinction between evaluation questions & research questions

(Other) Evaluation reveals short-term outcome; research predicts long-term outcomes Innovation can be assessed & recorded, but not predicted with objectives assess = to sit beside

(Other) The tremendous potential of the NPS, through its programs and activities, to connect with and impact on the lives of children, families, and their communities and the potential partnerships

(Other) Defining end state - visioning - rather than strategic planning Self-to-self, self-to-others, self-to-system What question do we want to answer in 5 years? 36 Park study - no follow-up

Appendix C: Summary of Pilot Project Criteria Discussions

Following the Fishbowl Discussion during the morning of Day Two, participants were divided into 12 small groups. In each of these groups, members were asked to address four questions relative to selection criteria for the Pilot Projects. Each group had about 40 minutes to answer these questions and to bring back to the larger group one “ah-ha” moment that they wanted to share. Each group was asked to keep notes of their discussion and return these to the evaluators. The questions discussed by small groups were:

1. What did you hear about pilot project selection criteria and process that you want to reinforce?
2. What was missing or underrepresented in the fishbowl conversation in general?
3. What projects would you like to nominate for consideration for pilot evaluation?
4. Are there specific selection criteria or process items that you would propose to add?

What follows is a brief summary of the data collected from these small group discussions. This information is organized by question, and major, minor, and unique themes are highlighted within each question. Major themes have been defined as those points that were raised by most, if not all, small groups. Minor themes are those that were mentioned by more than two or three small groups, but not a majority. Unique themes are those points that were raised by only one or a small number of groups.

Question 1: What did you hear about pilot project selection criteria and process that you want to reinforce?

The major themes from question one responses are outlined in the table below.

In addition to these major themes, small groups indicated a range of minor themes and unique responses. Minor themes for selection criteria included projects that involved the public, that had buy-in at all levels of park hierarchy, and projects that had possible “champions” involved. Small groups also suggested that projects should be able to act as models for future evaluations, should model an “honor culture,” and should involve focused inquiry.

Regarding the pilot selection process, small groups highlighted the importance of clarity and transparency regarding the selection process as well as clarity regarding who will be responsible for selecting, completing, and supervising pilot projects.

Finally, several unique issues and ideas were raised by the small groups, ranging from selecting pilots that were user-friendly to pilots that would produce positive results, from concerns about funding issues and sustainability to concerns about removing obstacles to evaluations.

Question 2: What was missing or underrepresented in the fishbowl conversation in general?

Small group responses to Question 2 were far more varied, although several key topics were raised. First, small groups highlighted the important role that training will need to play in building and maintaining a culture of evaluation, and that pilot projects may have a role in this training. In addition, small groups indicated the importance of not duplicating work that has already been done, either within the NPS or within other organizations, which might be helpful (e.g., museums and higher education). Small groups also raised issues about the role of I&E in various parks—some parks have well developed programs while others

Major Themes

Diversity of Projects	This theme highlighted the importance of selecting a diverse range of projects, such as rural and urban parks, large and small parks, different types of programs, geographical diversity, budgets, etc.
Involve Partners	This theme universally expressed a desire to include park partners at the earliest opportunity.
Involve Field Staff	This theme communicated the importance of including field staff in the design, selection, and implementation of pilot projects.
Reasonable and Feasible	Small groups expressed the importance of pilot projects being reasonable and feasible to complete—concerns such as the size of the project, funding, and timeline for completion were the types of considerations raised.
Diverse Populations	Within this theme, small groups talked about pilots that reached diverse populations as well as pilots that addressed how to reach diverse populations.
Applicable Systemwide	Participants expressed a desire for the pilot projects to be applicable throughout NPS I&E—not just one park.

have virtually none. Small groups also asked for more information about the pilot projects, such as the goals of the pilots, how they fit into the logic model, and the process of selecting and implementing pilots. Concerns about how information collected from the pilots was also raised by small groups, as were questions about NPS commitment to I&E. Ultimately, many of the same concerns that had been voiced over the one and a half days of the Summit were reiterated in response to Question 2.

Question 3: What projects would you like to nominate for consideration for pilot evaluation?

Small group responses to Question 3 fell broadly into two types—specific projects and parks and specific questions to be answered. Specific projects that were mentioned include:

- Independence Park Institute
- Yellowstone and Technology
- Cuyahoga and Environmental Education
- WebRangers
- Lake Powell (diverse audience)
- Homestead’s Distance Learning
- Kenai Fjords/AK Sealife Center/Boat Tours
- IDP Effectiveness
- Civil War
- Freeman Tilden Programs
- Bridging Watershed Program
- Fort Laramie National Historic Site
- VIEWS
- Santa Monica Mountains
- Buffalo Soldiers – Yosemite
- Research Learning Centers

Small groups also suggested a range of project types and evaluation questions that could be addressed. Topics addressing who delivers I&E programs, such as examining the effects and value of the presence of a Park Ranger on visitor experience. This topic also included questions about partner- and volunteer-delivered programs. A range of questions addressing youth experiences in parks, including the effects of engaging students in the classroom, the effects of residential programs for youth, and the impact of involvement in curriculum-based programs on standardized test scores. Various questions arose regarding media, such as the effectiveness of types of media (film, print, exhibits, etc.) and the use and effectiveness of interactive media (e.g., websites and exhibits). Several questions regarding diversity were suggested, including diversity within the NPS and engaging diverse populations—why they are not visiting and what is effective when they do visit? There was a range of questions about large, episodic events (e.g., the NPS Centennial) including what has worked and what is the cost and benefit? Following up on “old” evaluation studies was also suggested (i.e., 36 parks). Other suggestions included evaluating I&E professional development, studying different stories (e.g., Oklahoma City and Flight 93), evaluating the relationships between parks and their various stakeholders (e.g., partners), and examining the effectiveness of informal contacts.

Question 4: Are there specific selection criteria or process items that you would propose to add?

Small groups responded to Question 4 with a range of suggestions, many of which echoed thoughts and concerns raised in response to previous questions. Others highlighted new themes. Topics that were raised again in response to this question included pilots that:

- Are reasonable to complete (time and money),
- Address underserved populations,
- Could act as models for future evaluations,
- Involve field staff, partners, and the public in design and implementation,
- Address diverse types of projects and parks,
- Build evaluation capacity within the NPS, I&E, and staff, and
- Are widely applicable.

Other criteria suggested include a range of issues such as the political traction of a pilot project, the ability to use results quickly and effectively, the importance of considering resource type in selecting pilots, whether to select pilots with known outcomes (positive) versus risky pilots, the importance of using training and evaluation tools, linking pilots to the NPS Centennial, the role of the pilot in moving I&E forward, the ability to use the pilot for advocacy purposes, and selecting pilots and parks with a demonstrated commitment to evaluation.

In addition to selection criteria, the small groups suggested a range of process questions and considerations. These included questions about how pilots will be advertised and how information from the pilots will be used. It was also suggested that the NPS Interpretation and Education Program develop some type of reward, recognition, and incentive program for parks implementing evaluation. Finally, small groups highlighted the importance of supporting (from start to finish) whatever pilots are selected. This included offering professional assistance, providing training and funding, encouraging collaboration between pilots, and ensuring that even if a pilot is “unsuccessful” that the momentum of the project is maintained.

Appendix D: "Chiefs" Survey Executive Summary

National Park Service, Education Council
Education Evaluation Coordination Team



A Review of the Literature and Brief Questionnaire Responses (from the field)

Evaluation



Prepared by:

Nina S. Roberts, Ph.D.
Education & Outreach Specialist
Natural Resource Program Center

Jenni Mullins, Intern
NPS Natural Resource Program Center

(July 2005)



Section II

Evaluation Informal Survey for Gathering “data” from the Field *Executive Summary*

In June 2005 a brief, informal questionnaire was sent with a memo to regional chiefs of interpretation/education. An email request was then distributed with copy of questionnaire to each of the site-specific chiefs of interpretation/education. Given the nature of this inquiry, responses from field staff were completely voluntary. As of July 5, *n*=81 responses were received and organized for a simple examination. A basic content analysis occurred seeking primary categories, major issues/concerns, and general program/education considerations. These highlights dissect education and interpretation as best as possible. (*“Park” is used to denote all NPS units*)

- **Regarding questions about impacts, effects, or outcomes of park programs, the following key themes emerged among the questions put forth:** Trends, Traditional programming vs. need for increased creativity, Connection to state/federal standards, Meeting school curriculum needs, Effectiveness with connecting visitors/students to significance/meaning of resources, Enhancement of stewardship ethic, program comparison, Outreach, Media effectiveness, Knowledge gained/retention, Attitude and/or behavior change and transference, Marketing and/or promotion, Impacts of Junior Rangers and Parks as Classrooms programs, Funding, Influence on career choices (students).
- Overall need for **“hard evidence”** of the general effects of park programs.
- There were many **typical responses** that would be expected from “career employees” who have been through the IDP or are trained park interpreters in a variety of ways: How do we “improve the visitor experience”? Are visitors “making a strong connection to the resource”? Do programs result in “better protection of park resources”? Are we effectively connecting visitors with the meaning/significance of the parks? Do programs enhance a stewardship ethic? Do visitors understand primary park themes?
- **Trends** questions and comments were very interesting: Examples relate to whether “tradition” educ/interp programs are still as effective today as they were 10-20 years ago; Addressing increasing use of technology by RV campers (e.g., promote attendance of programs or bring the programs to them); Is there movement away from personal service programs?; Employee/partner training and development—how is performance monitored? Continuing shift in visitation from overnight use to day-use (how to re-evaluate contacting and educating visitors); Diversity;
- **Limitations and/or decreasing number of staff** force difficult choices and decisions. Questions surfaced regarding how to prioritize programming given a variety of reductions. Fear of elimination of programs all together was an expressed concern.
- There is a majority **desire for longitudinal studies** that relate to visitors and/or students participating in park programs. Examples include: Retention of information 6-months or more (e.g., one-year) after visiting (general public or students); Long-term

- care/protection of park; Effects on choices and actions (transference: generally, at home, in neighborhoods, at school); stewardship/ethics;
- Many noted an interest in serving “**more visitors we are not presently reaching**” including diverse and underserved populations. Query about whether there are more effective methods of reaching the public.
 - **Curriculum needs:** 1) Meeting the goals and objectives set forth in State and/or Federal standards is very important to parks; 2) Meeting teacher needs—generally; 3) Influence of these programs on the affect local attitudes towards the NPS and conservation programs in general; 4) Desire to design or improve pre- and post-visit activities; 5) Effectiveness of school-based programs; 6) Constraints to participating; 7) Need for solid evaluation tools and/or techniques; 9) Measuring knowledge gained by students; 9) Have programs influenced career choices of the students? 10) Meeting the needs of home-schoolers; 11) Access to NPS information supporting curriculum/standards (sufficient?); 12) Do test scores improve/visits result in higher academic achievement? 13) Traveling Trunk use.
 - **Outreach** was put forward and commented on in a variety of ways. Evaluation of “off-site programming”; best way to serve local communities; expressed needs to develop a “plan”;
 - Several parks noted a need to have a “**formal plan**” for education to assist them with their future decision-making.
 - Youth programming: More questions related to the **Junior Rangers Program** and fewer asked about the **Parks as Classrooms**: Short-term and long-term learning needs to be documented and tracked. Frequent questions about true “impact” emerged. Query about what it takes for these kids to convey the message of preserving and protecting resources to their peer groups. Query about whether students actually develop a stewardship ethic or not. Do these programs “get kids excited to work in a park”?
 - Questions around the need for more “**creativity**” category relate to such aspects as: Are there more creative or effective methods of reaching the public than what is presently used? Effectiveness of web-based programs/electronic activities and information. Need for more creativity in terms of how resources are being interpreted. Strengthening campfire programs.
 - Annual **Visitor Survey Card** was challenged as to its true efficacy on numerous responses.
 - **Media effectiveness** category included questions primarily relating to: Impact of exhibits? Effectiveness of park literature? Electronic avenues (e.g., eField Trips)?
 - **Funding** queries primarily related to assisting schools in need of transportation, need for new/enhanced exhibits (e.g., “out of date”).

- **Marketing/promotion** revolved around examples such as generating more and diverse audiences, groups, visitors; encouraging more schools to visit; increasing participation in field seminars; getting teachers to attend pre-site visit teacher workshops;
- Suggested sites for pilot projects come from this list of park units responding to the questionnaire. These parks provide a variation of region, type of site, and were selected based on their responses to the questions posed:

Kenai Fjords National Park	Frederick Douglass NHS
Western Arctic National Parklands	Boston National Historical Park
Arches National Park	Cabrillo National Monument
Bent's Old Fort NHS	Great Basin National Park
Bandelier National park	Lava Beds National Monument
Bryce Canyon National Park	Congaree National Park
George Rogers Clark NHP	Little River Canyon National Preserve

#####

Appendix E: Pre-Summit Statements From Panelists

As part of the briefing process in preparation for the Summit, each panelist crafted a three-page statement of a few stories or main points that were relevant to the work of this event. Provided here are one-sentence summaries of the statements that follow.

* * * * *

Michael Quinn Patton – *An Example of Infusing Evaluation into an Organizational Culture*. By highlighting the use of evaluation within one organization, Dr. Patton discusses a real life example of a shift from episodic evaluations to infusing evaluative thinking throughout an organization.

Lester W. Baxter – *Evaluation at The Pew Charitable Trusts*. Mr. Baxter outlines the reciprocal relationship between program planning and evaluation. Further, he notes the importance of close collaboration with program staff in evaluation.

Teri Behrens – *NPS Meeting*. Dr. Behrens describes lessons learned at the W.K.Kellogg Foundation about the importance of evaluative thinking.

Theresa G. Coble – *Compelling Evidence, Competent Workforce: Research & Training for Effective NPS Programs*. Dr. Coble presents findings from evaluations conducted within NPS and suggests that these can serve as a foundation upon which to build.

Lynn Dierking – *If you don't know where you are going, any road will get you there*. Dr. Dierking presents a cogent argument for the value of evaluation in accomplishing goals in all aspects of program development, implementation, and assessment.

Allison Druin – *What Children Can Teach Us: Creating a Culture of Evaluation in Partnership*. Dr. Druin offers important insights regarding the use of technology in evaluation and the importance of involving a range of stakeholders, especially children, in evaluation.

Doug Knapp – *A Sample of What SIEES Can Do...* Dr. Knapp provides a sampling of findings from evaluations conducted within NPS and argues that evaluation is an important mechanism for meeting I&E goals.

Martha Monroe – *Creating a Culture of Evaluation: Building Capacity, Engaging Audiences, Practicing Curiosity, and Taking the Sting out of Failure*. Dr. Monroe offers humorous and poignant stories about the importance of learning from disconfirming evidence.

Richard Nichols – *Creating a Culture of Evaluation: An Indigenous Perspective*. Mr. Nichols provides a cultural perspective on evaluation and contends that evaluation needs to be seen as knowledge generation with input from all cultures served – and those not currently served.

Emma Norland – *Part or PARKS?* Dr. Norland offers some suggestions on making evaluations most useful, including focusing on utilization of evaluation, collaborating between evaluators and program staff, and conducting evaluations that fit into a bigger picture.

David Sobel – *Place-based Education and the Culture of Evaluation*. Dr. Sobel presents an example of evaluation in place-based education that led to a range of benefits, including better understanding of the needs, ability to replicate program components, and additional programming.

Carol B. Stapp – *Outcomes Matter: Evaluation Counts*. Dr. Stapp points out that while it is difficult to instill positive attitudes about evaluation, focusing on building competencies and capacity can be helpful.

Hazel Symonette – *Make Evaluation Work for the National Park Service Success Vision: Generative Pathways Towards a Culture of Evaluation*. Dr. Symonette reminds us of the importance of the “who” in evaluation, the need to make evaluation a learning opportunity, and the importance of evaluation being useful.

Veronica G. Thomas – *Blending Cultural Competence and Project Evaluation for Enhancing Place-Based Learning*. Dr. Thomas highlights the importance of attending to cultural competency when designing and conducting evaluations.

Jon F. Wergin – *When Evaluation Works, and When It Doesn't*. Dr. Wergin suggests the components of a successful evaluation versus one that simply fosters a “compliance mentality.”

An Example of Infusing Evaluation into an Organizational Culture

Michael Quinn Patton
Utilization-Focused Evaluation
Saint Paul, Minnesota
October 2006

When evaluation first emerged as a distinct profession in the 1960s and 1970s, the emphasis was on conducting specific evaluation studies for specific purposes. Correspondingly, research on utilization focused on how the findings from those studies were used. Now evaluation use has evolved to include infusing evaluative thinking into an organization's culture, the very focus of this upcoming Summit. This means looking beyond the results of isolated studies. It takes us into looking at how decision makers and staff incorporate *evaluative thinking* into everything they do as part of ongoing attention to mission fulfillment and continuous improvement. I'm often asked if I've ever consulted with an organization that has managed to do this. I have: the International Development Research Centre (IDRC) headquartered in Ottawa, Canada.

IDRC is a public corporation created by the Parliament of Canada in 1970 to help developing countries use science and technology to find practical, long-term solutions to the social, economic, and environmental problems they face. Support is directed toward developing an "indigenous research capacity" to sustain policies and technologies that developing countries need to build healthier, more equitable, and more prosperous societies. IDRC's mission is "Empowerment Through Knowledge" (IDRC 2006).

In 1993, IDRC established an Evaluation Unit to support the conduct of evaluations. The unit oversees various evaluation and monitoring functions, and produces an annual overview of each year's evaluations, synthesizing significant findings for senior management and the Board of Governors. During a retreat in 2001, IDRC's Senior Management Committee expanded the organization's evaluation commitment to include a framework for mission assessment at the overall corporate level. This involved the systematic collection of performance data regarding IDRC's strategic goals and operating principles. To do this, senior managers had to identify those principles, or fundamental ways of doing business, that were expected to permeate all of their work in accomplishing their two overall strategic goals: *Indigenous Capacity Building and Policy* and *Technology Influence*. They committed the organization to monitoring and evaluating not only results in these two strategic goal areas, but also the extent to which they were employing their fundamental operating principles. This is where it gets interesting from our point of view because one of those operating principles was *evaluative thinking*. This was the first organization I had encountered that made infusing evaluative thinking into the organizational culture an explicit dimension for performance measurement.

In essence, the senior management committed not only to supporting the conduct and use of specific high quality evaluation studies and management information system data, they made *evaluative thinking* a fundamental way of doing business, infused throughout the culture. What did they mean? Here's what they committed to monitoring as expressed in their working papers operationalizing evaluative thinking.

1. Evaluative thinking permeates our work so that we consciously and constantly reflect on project, program, regional, and corporate experience with a view to implementing improvements based upon what is learned. Evaluative thinking is evident in the way we clarify goals and design, conduct and interpret evaluations throughout the organization.
2. Within programs, evaluative thinking is demonstrated in the implementation of well-focused programs and in the use of high quality project and program evaluations which feed into program and project decision making. Time and resources are allocated for reflection on evaluation findings and for documenting use of the findings.
3. In program management, evaluative thinking is evident in our systematic use of evaluation to inform program design and implementation decisions.
4. Senior management demonstrates evaluative thinking in its support for the adoption of evaluation processes; in its routine demand for the generation of outcomes-based data; and in the use of this and other feedback to expand management viewpoints and inform decisions.
5. IDRC's partners in the South share our commitment to learning-based evaluation; they exhibit mastery over their own evaluation activities; they serve with competence as external and internal evaluators; and they support and promote advances in the science and art of performance assessment.
6. The Board of Governors will request and use outcomes-based data in their governance of the Centre.
7. Senior management will request and use outcomes based evaluation information in its decision making processes; engage in the Corporate Assessment Framework processes to reflect on past experiences, analyze empirical data related to the performance areas, and

- revise actions appropriately to improve performance; share learnings amongst themselves and with others in the organization; foster an organizational environment conducive to learning.
8. Program staff will implement, and use the results of high quality evaluations in order to improve project and program performance; establish regular processes of reflection (on both successes and failures) at the program level in order to share lessons and improve future performance; and support the building of capacity in evaluative thinking among partners.
 9. Resources Branch will develop reporting systems to facilitate reflective processes.
 10. Partners will develop effective evaluation systems directed to their own needs and purposes; collaborate with the Centre on implementing evaluations of relevance to both their and our needs; and have the capacity to operate as evaluators.

Evaluation Thinking in Practice

At the same time that IDRC was making evaluative thinking a priority area for overall organizational assessment, they were having to face a concrete reality at the most basic level: project managers were not completing required end-of-project reports. Indeed, they had accumulated a backlog of hundreds of unfinished project completion reports. A variety of carrot and stick efforts to get reports completed had failed. Evaluating these efforts, they found that the reports were viewed as an arduous paperwork requirement with no real utility. Project managers didn't get feedback when they did do reports and, given other workload priorities, there were no incentives to complete a report on work already done. All the energy was going into new initiatives rather than recording the details of yesterday's news.

As a part of rethinking the reporting function in the organization with an emphasis on creating opportunities for shared learning, they conducted an inquiry into how and when learning occurs in projects. Project staff said that the most learning takes place at the start of or during a project's life, while the least learning occurs at the end of a project, and different kinds of learning take place throughout a project's varying stages. Drawing on staff at different levels from throughout the organization, a working group was formed to redesign the project reporting process.

They developed a three-stage process dubbed the "Rolling Project Completion Report" (rPCR) and changed the format, timing, and information input approach. The new system emphasized learning rather than paperwork accountability. The new approach was pilot-tested on a sample of projects from different units throughout the organization. Early in the life of a project, a junior staff member interviews a project officer to gather data about project design, start-up lessons, and issues that will need attention going forward. In the middle of a project, team leaders interview project officers to capture lessons about implementation and interim outcomes, as well as update work on key issues. After the end of a project, senior managers interview project officers to complete the project reports, identify results, and capture any final learnings. Major learnings are highlighted at an Annual Learning Forum. This new rPCR process replaces the old paperwork requirement with an interview process that has people at different levels in the organization talking to each other, learning about each other's work, and sharing lessons. All those involved went through formal interview training, including senior managers, so people share language and understandings about quality interviewing and what kind of cross-organization learning is being sought. The process is designed so that interview responses are entered into the learning system in real time, as the interview takes place, with subsequent opportunities for project managers to make corrections and append supporting documentation and cross reference information sources.

The project report backlog was completely cleared and feedback about the process is highly positive. The organization-wide process of involving people in reflection and learning reinforces evaluative thinking as a core operating principle while also meeting accountability demands to get reports done in a timely and meaningful fashion. The capacity of staff to engage in evaluation thinking has been systematically enhanced, including deepening their interviewing skills, pattern recognition capabilities, and data interpretation skills. The attention garnered for projects featured at the Annual Learning Forum and the direct involvement of senior management provide additional incentives to take the process seriously and document both learning and results. The Project Completion Reports, long disdained, became a source of energy and enlightenment, and a manifestation of evaluative thinking infused into the organizational culture. This redesign of IDRC's reporting process illustrates nicely the insight of *Future Shock* author Alvin Toffler (1970) who observed:

The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn.

Evaluation at The Pew Charitable Trusts

Lester W. Baxter

Planning and Evaluation, The Pew Charitable Trusts

October 6, 2006

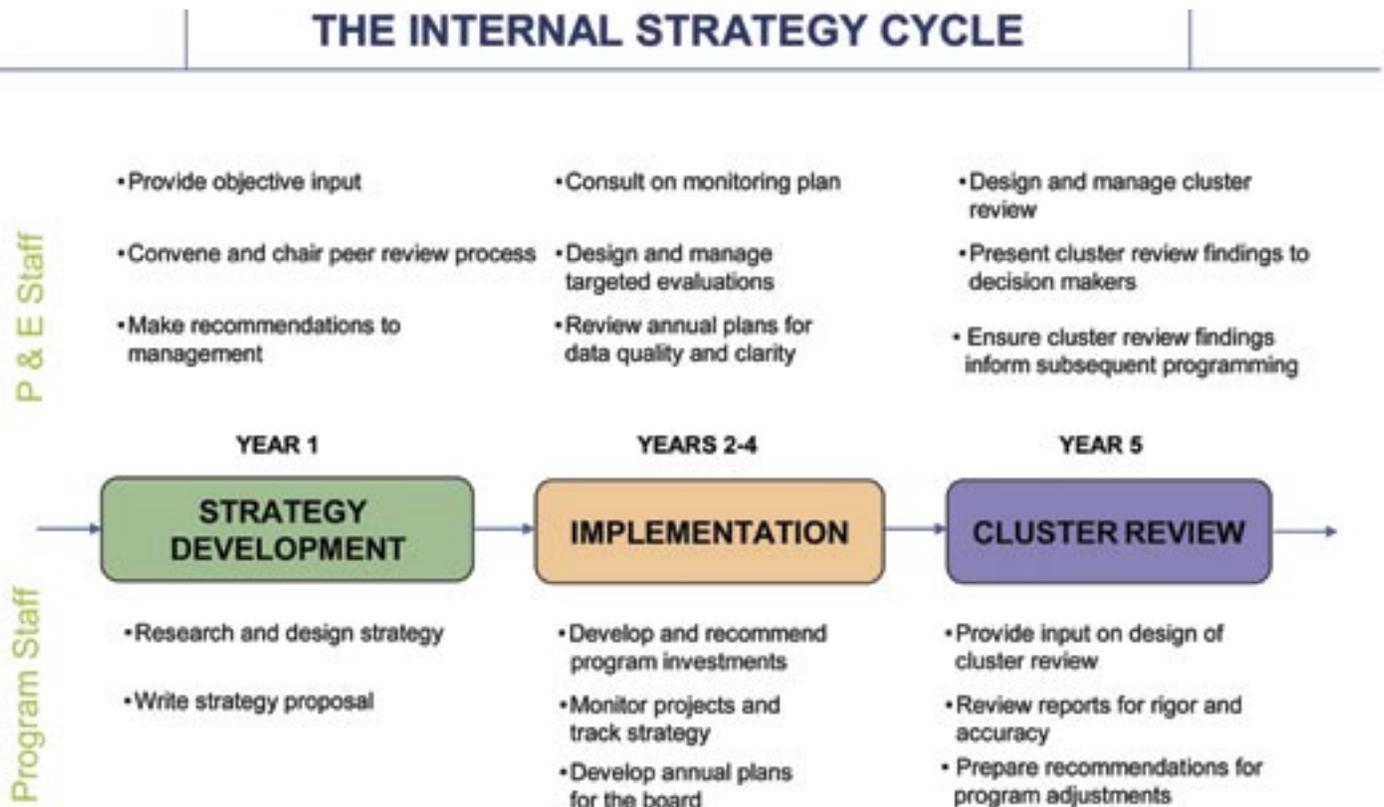
Achieving results is central to the mission of the Trusts. Evaluation contributes to this aim by enabling us to gauge the return on the Trusts' investments, test the effectiveness of specific strategies, and offer an informed perspective for adapting programs to upcoming opportunities.

In 1988, the Trusts established an internal department called Research and Evaluation. In the department's early years, evaluations served two main purposes: to ensure grantee accountability to the foundation and to facilitate grant renewal decisions. As the financial oversight of grants shifted to a growing Grants Administration department and grant monitoring duties shifted to the program areas, the Research and Evaluation department began to look at issues of effectiveness and organizational learning.

In 1992, the evaluation department commissioned its first review of an entire cluster, or program (a set of projects that are pursuing a collective goal), of grants. Evaluators hired to conduct those early cluster reviews suggested that if programs were more tightly focused their effect would be greater. The Trusts realized that such focus could not be accomplished simply through better evaluations, but rather by integrating the lessons learned from evaluations into program planning, sharpening programmatic focus from the start. Taking to heart such lessons, the evaluation unit evolved into what is today Planning and Evaluation.

Over time, this integration of evaluation with planning was explicitly linked to the Trusts' core activity: developing and implementing programs that lead to social benefits. This linkage occurs through a process we call the internal strategy cycle (depicted below), which has three major stages. The first, strategy development, involves creating a coherent and convincing plan, with feasible and measurable objectives, to address a specific problem. The second, implementation, entails turning the plan into action with our partners, carefully monitoring progress, and adjusting the plan as necessary. The final stage begins with a rigorous and independent evaluation of the overall strategy. Program staff then integrates the findings from this evaluation into a revised plan, triggering a new round of the internal strategy cycle. The entire cycle, from strategy development to cluster review, can take three to five years, or longer.

THE INTERNAL STRATEGY CYCLE



The internal strategy cycle starts with a desire to seize an opportunity or respond to a problem or issue (such as declining voter turnout among young people). Program staff then designs and presents to our board a strategy outlining a potential role for the Trusts. Two key steps in selecting an issue are understanding the root causes of a problem and then determining whether feasible approaches are available to address them. Moving too quickly to a solution is perhaps the most common mistake in program design; it usually means that we have not developed a complete sense of the causes or failed to fully review all of the options to tackle them. To help us avoid this error, program staff at the Trusts engages with colleagues, informally in brainstorming sessions and formally in internal peer-review meetings, to examine a problem from as many sides as possible. These teams benefit from the knowledge and counsel of outside experts as well as that of the Trusts' leadership.

If the proposed strategy is approved, program staff works with external partners to develop a coherent portfolio of projects to carry it out. As the strategy is implemented and projects are launched, program and evaluation staff develops a monitoring plan to provide staff the information needed to make good management decisions. When these observations raise issues for the program staff, a focused evaluation answers the question "why is this happening?" (Suppose that monitoring reveals an increase in voter turnout among college students. An evaluation could help explain whether turnout is increasing because voter registration campaigns are working, a particular issue is galvanizing the student community, or a close election has increased voter turnout generally.) Every year, program staff reports their progress and any necessary strategy adjustments to the board. After three to five years, a cluster review takes a look back at the effectiveness of the strategy and the lessons we learned from the experience, as well as what those lessons and changes in the field might mean for the Trusts' future investments in that area. Knowledge gained from the development, implementation, refinement and evaluation of the strategy then informs the decisions of program staff, the Trusts' management and the board going forward.

As with any approach, the process described above has both costs and benefits. On the cost side, it entails a close collaboration between program staff and evaluation staff at many points—and we do not pretend that negotiating those relationships is easy. The Trusts' approach is also undeniably resource-intensive, demanding both human and financial investment. But among the many benefits, this approach calls us to be accountable for results and for learning from our mistakes as well as our successes. It ensures that we do not ignore initiatives and strategies once they are launched, but continue to question our assumptions about the process of change, and gives us the room to make corrections when we find that we were wrong. It lays out a framework to help us target our resources to the places where they can have a tangible effect.

We do not pretend that this is the only approach to pursuing social benefits (or evaluation) or the best. We do believe, however, that this strategic approach has yielded stronger and more consistent results from the Trusts' programs. The above approach has helped us measure our success while improving the ability to develop and manage programs, acknowledge limitations and ultimately become more effective. These practices cannot guarantee success but we have abundant evidence that they improve the odds.

The Pew Charitable Trusts, an independent nonprofit, serves the public interest by providing information, advancing policy solutions and supporting civic life. Based in Philadelphia, with an office in Washington, D.C., the Trusts will invest \$248 million in fiscal year 2007 to provide organizations and citizens with fact-based research and practical solutions for challenging issues.

**Teri Behrens
Director of Evaluation
W. K. Kellogg Foundation
NPS Meeting**

The W. K. Kellogg Foundation funds domestically in the areas of Food Systems and Rural Development, Youth and Education, Health, and Philanthropy and Volunteerism. We also do “hometown” programming in Battle Creek, Michigan; in seven countries in southern Africa; and in Latin America and the Caribbean. Our grantmaking includes individual projects, clusters of related projects, and strategic initiatives, which are bodies of work intended to promote systems change.

WKKF was one of the first foundations to make a major commitment to evaluation. Beginning in the late 1980s, the Foundation began doing “cluster evaluation,” in which grantees working on a common issue were brought together to learn and share experiences. We both provide funding to the grantees to conduct their own evaluations, and contract directly with what we call “cluster evaluators” to conduct evaluations of the larger bodies of work.

To support individual grantees, we have produced a logic model development guide and an evaluation handbook that are both widely read and used. We are currently developing a handbook on cluster and initiative evaluation.

One of the keys to success for evaluation is the regular convening of grantees to learn from each other both about programmatic strategies and about how they are learning and evaluating their own work. Rather than make the evaluation a way to monitor our grantees performance, we introduce evaluation as a way to help them succeed. One example of this is in the work we are doing on community leadership – our Kellogg Leadership for Community Change (KLCC) program. KLCC is being funded as a series of sessions, each focused on a particular topic. For each session, a set of communities and grantees (“Host Agencies”) are funded to work with a group of 25 community fellows. In the first session, for example, six sites were funded to address “building public will for teaching and learning.” In session two, five sites were chosen to focus on “youth / adult partnerships for social justice.” In each community, there is a local evaluator, sometimes one of the fellows, who helps to track progress on the community change work.

The W. K. Kellogg Foundation has contracted with an ethnographic researcher who is both looking across sites to help us learn about how to promote collective community leadership, and working with the community level evaluators to help them develop and implement their own evaluation strategies. The evaluators meet in person 1 – 2 times per year, have a monthly telephone conference, and have a shared web site. The goal is to help each site learn to use evaluative thinking.

This evaluation process has helped to inform both how W.K.K.F. is operating the program and has helped individual communities to use data to change local policy and practices. At the W.K.K.F. level, some of the changes we have implemented as a result of our evaluation findings include providing funding over a longer time period; spending more up-front time in each community developing a management team before recruiting fellows; focusing more clearing on organizations and communities that are ready to tackle an issue; and (next session) being more specific about the topic.

At the community level, evaluation data have been used in a number of ways. In one community, young people surveyed the houses around a dump site to gather data on the incidence of health problems. When this information was shared with the city council, they applied for and received Superfund money to clean the site.



Compelling Evidence, Competent Workforce:

Research & Training for Effective NPS Programs

Theresa G. Coble

Stephen F. Austin State University (SFA)

This paper addresses a key question for the National Interpretation and Education Evaluation Summit: *What do we already know from evaluation within NPS I & E, and how might we build upon that foundation?* It also illustrates how SFA is partnering with the NPS to achieve the two goals outlined in the Servicewide Interpretation & Education Evaluation Strategy (SIEES). The paper highlights selected results from the first year of a 4-year, 16 park research project conducted in cooperation with the Intermountain Region of the NPS, SFA, and the Western National Parks Association (WNPA). The paper also introduces an NPS-SFA partnership graduate degree program (M.S. in Resource Interpretation) that is equipping the NPS workforce to integrate evaluation into their daily work.

SIEES Goal #1

The NPS will have a servicewide commitment to evaluation that facilitates coordination, fosters information exchange, and supports application of results.

Since 2004, the Intermountain Region has partnered with SFA and WNPA to jointly sponsor the Visitor Voices Project. The Visitor Voices Project is a 4-year, 16 park study to examine the interpretive outcomes visitors attain onsite at various NPS locations.

- The Visitor Voices Project explores a key question: *Do visitors benefit from park interpretive offerings in ways consistent with the stated goals and objectives of park interpretation?*
- The NPS Interpretive Development Program (IDP) establishes national standards for interpretive performance, linking interpretive effectiveness to cohesively developed, audience appropriate interpretive offerings that provide opportunities for visitors to form intellectual and emotional connections to the meanings and significance of park resources.
- To date, no study has measured visitor interpretive outcomes, defined as "connections," at a broad cross-section of NPS sites, or determined whether significantly different outcomes occur with respect to key socio-demographic variables.
- During the summer of 2005, year one of the proposed 4-year study, surveys were collected at four NPS sites, including Carlsbad Caverns National Park (CAVE), Lyndon B. Johnson National Historical Park (LYJO), Petrified Forest National Park (PEFO), and Timpanogos Cave National Monument (TICA).
- Researchers collected a total of 1,526 valid surveys, obtaining a 61.0% response rate overall.

- In this study, connections were the dependent variable. Connection data were obtained through participant self-report via two yes-no questions. Respondents reported the following interpretive outcomes:

	Neither	Intellectual Only	Emotional Only	Both	Total
CAVE	39 (28.3%)	109 (29.1%)	21(25.6%)	284 (30.6%)	453 (29.7%)
LYJO	16 (11.6%)	51 (13.6%)	12 (14.6%)	145 (15.6%)	224 (14.7%)
PEFO	31 (22.5%)	98 (26.1%)	18 (30.0%)	268 (28.9%)	415 (27.2%)
TICA	52 (37.7%)	117 (31.2%)	31 (37.8%)	231 (24.9%)	431 (28.3%)
Total	138 (9.1%)	375 (24.6%)	82 (5.4%)	928 (60.9%)	1523 (100%)

- Stepwise multinomial logistic regression identified a set of 15 factors which were significant at the $\alpha=0.05$ level. Approximately 70% of the time, the model correctly predicts interpretive outcomes using just the data associated with the 15 factors. A post-hoc analysis was conducted using two-way contingency tables in which each factor was independently analyzed with respect to the response variable (connections). In the post-hoc chi-square analysis, 13 of the 15 factors were significant. These 13 factors are examined in greater detail in a report submitted to the NPS, WNPA and the NPS Social Science Program [available: www.sfasu.edu/msri/teoble.htm].
- The 13 factors that explain observed interpretive outcomes are as follows:

Park Interpretation

1. Number of interpretive offerings experienced onsite ($\chi^2=100.897, df=39, p\text{-value} < .0001$).
2. Interpretive offering experienced immediately prior to completing survey ($\chi^2=69.381, df=21, p\text{-value} < .0001$).
3. Most meaningful interpretive offering experienced onsite ($\chi^2=81.501, df=21, p\text{-value} < .0001$).

Survey Logistics

4. Survey location ($\chi^2=28.591, df=18, p\text{-value}=0.0536$).

Respondent Group Size & Composition

5. Visiting park with myself and one or more children ($\chi^2=30.778, df=6, p\text{-value} < .0001$).
6. Visiting park with group other than family and friends ($\chi^2=27.242, df=6, p\text{-value}=0.0001$).

Why Respondent Visited the Park?

7. Interest in nature ($\chi^2=19.391, df=3, p\text{-value}=0.0002$).
8. Like visiting national parks ($\chi^2=18.655, df=3, p\text{-value}=0.0003$).

Respondent Life Experiences

9. Number of interpretive sites visited each year ($\chi^2=32.392, df=15, p\text{-value}=0.0078$).
10. Exposure during youth to natural and cultural areas ($\chi^2=39.932, df=12, p\text{-value} < .0001$).

Respondent Demographics

11. Gender ($\chi^2=19.719, df=6, p\text{-value}=0.0031$).
12. Race ($\chi^2=29.633, df=18, p\text{-value}=0.0412$).
13. Hispanic ethnicity ($\chi^2=19.702, df=6, p\text{-value}=0.0031$).

Table 1. Summary of Factor Effects (Visitor Voices Project, 2005 Data).

	Neither Type of Connection	Intellectual Only Connections	Emotional Only Connections	Both Types of Connection
More Likely to Form...	<ul style="list-style-type: none"> Attended 0-1 Programs Visit 1-2 Interp Sites/Year Control Group Exposed to TICA Exhibit Didn't Visit because of an Interest in Nature Didn't Answer Q's re: Most Meaningful Interp Offering, Group Composition (2), Youth Experiences, Race, Ethnicity 	<ul style="list-style-type: none"> Attended 0 Programs Visit 1-2 Interp Sites/Year Group Composition: Adult with ≥ 1 Child Didn't Visit because of an Interest in Nature Didn't Visit because He/She Likes Visiting National Parks Most Meaningful Interp Offering = Park Brochure 	<ul style="list-style-type: none"> Control Group 	<ul style="list-style-type: none"> Exposure to: Park Film @ PEFO Female Most Meaningful Interp Offering = Ranger-led Program
Less Likely to Form...	<ul style="list-style-type: none"> Attended 3-4 Programs Exposure to: Interp Talks, Geology 101 @ CAVE, Park Film @ PEFO, Cave Tour @ CAVE Most Meaningful Interp Offering = Ranger-led Program 	<ul style="list-style-type: none"> Attended 3-4 Programs Exposure to: Park Film @ PEFO 	<ul style="list-style-type: none"> Exposure to: Interp Talks 	<ul style="list-style-type: none"> Attended 0 Programs Control Group Exposure to: TICA Exhibit Didn't Visit because of an Interest in Nature Didn't Visit because He/She Likes Visiting National Parks Male Didn't Answer Q's re: Most Meaningful Interp Offering, Youth Experiences, Race, Ethnicity

SIEES Goal #2

The NPS workforce will have the motivation, knowledge, ability, and tools to thoroughly integrate evaluation practices into their daily work.

Since 2003, SFA has offered an online graduate degree program—*Master of Science in Resource Interpretation (MSRI)*—in partnership with the Stephen T. Mather Training Center [see website: www.sfasu.edu/msri]. Students take courses online and conduct thesis research. NPS student masters theses for 2005-2006 include (short titles provided):

Carla Beasley	Maximizing the Effectiveness of the Assessment Rubric for Informal Interpretation
Linda Chandler	The Effects of Crowding and Noise on Visitors' Interpretive Experiences at the Castillo de San Marcos NM
Amy Glowacki	Facilitating Exploration of Lowell's Acre as a Way to Integrate Interpretive and Educational Objectives
Vickie Mates	Visitor Meanings at Theodore Roosevelt National Park
Peggy Scherbaum	Skillful Delivery and Appropriateness: Elements that Influence the Effectiveness of Interpretive Techniques
Margaret Styles	Factors Influencing Literary Site and Local Community Relationships

**If you don't know where you're going, any road will get you there.
Flying Karamazov Brothers**

**Pre-Summit Statement by Lynn Dierking,
Vice President for Special Initiatives, Institute for Learning Innovation**

This is the quote I like to begin with when I work with an institution or organization that has decided to institutionalize the process of evaluation as the National Park Service is setting out to do as one component of its proposed Centennial Renaissance Plan. Although in this case the task of institutionalizing the evaluation process is extremely complex, given the size, breadth, scope and goals of the educational activities in which the Park Service engages, I still think that the point is apt. This idea resonates with folks in part because it is humorous and comes from a quarter that one does not expect (e.g. the Flying Karamazov Brothers are not what most folks conjure up when hearing the word evaluation!). However, I think the other fundamental reason that it resonates so much is that it makes such tremendous sense to any of us who care deeply about what we are doing and want to be effective.

The notion that evaluation is a tool for assisting us in successfully accomplishing our goals ensures that the focus of the evaluation is appropriate. After all, a major component of the evaluation process is changing one's thinking, to think more evaluatively. Further, it allows one to point out that effective practitioners practice aspects of evaluative thinking and evaluation all the time, for example, when we are reflective about our goals or modify and refine the development of a program or other "product" we are creating. Institutionalizing the evaluation process merely enables us to use evaluation more consistently and systematically as a decision-making tool in our day-to-day work.

This is a critical step in helping people debunk some often deeply held beliefs about evaluation. Evaluation Myth # 1: Evaluation is always about statistics, or grading someone or something. Rather, evaluation is the flip side of good planning. Evaluation Myth #2: There is a "right" method to use. Not so, because methods are selected based upon the questions that one has, the purpose of the evaluation, the intended use of the findings, and the resources available. Evaluation Myth #3: Evaluation design is a purely technical science. Actually, it is as much art as science, requiring creativity and an open mind.

Evaluation at its purest is applied research - the ultimate goal is the "use" of the information. It is key to effective development and implementation, be it a program, exhibition, web site, curriculum or. Evaluation is also a process -- a series of feedback loops within the development process which incorporates the user's/learner's perspective. Ideally it facilitates responsive, informed decision-making and enables "reflective practice" and institutional learning.

Evaluation should always be guided by the questions one has at a particular point in a project. Ideally it is integrated within all phases of an effort beginning with the very initial idea for the project/activity. During the early phases of projects, evaluation is often called front-end and is designed to provide input into the initial conceptualization and planning for the effort/activity. Front-end questions can include:

- Is there a need for the effort/activity?
- What do we know about the needs that this effort/activity will address?
- What is recognized as effective practice in this area?
- What does the relevant research or conventional wisdom tell us about this need?

Evaluation is also a critical tool during the development of any program, exhibition, web site, and so on. Often called formative or process evaluation, evaluation at this phase provides information during the course of development to improve and refine the effort. This is the type of evaluation that effective practitioners most often engage in informally. Formative questions can include:

- How is the effort/activity going?
- Is the approach working?
- Is the effort/activity consistent with the program plan?
- How can the effort/activity be changed to make it more effective?
- Are participants' needs being met?
- Will this effort/activity meet the needs of participants/stakeholders?

Evaluation is also a critical tool to use after an experience or "product" has been created and used and/or experienced. Often called summative or product evaluation at this phase, it is designed to document and understand the impact of a "settled" effort/activity. This is a critical form of evaluation in the field of free-choice (informal) learning, as suggested in the Draft Action Plan for the Centennial

Renaissance. The free-choice learning field as a whole, including the National Park Service, lacks strong scientifically valid information about the outcomes and impact of interpretation and education programs. Such data are essential when arguing for the importance of the kinds of experiences people have when they utilize the resources of places like national parks. Summative questions may include:

- Have the stated goals of the effort/activity been achieved?
- Have the needs of those served by the effort/activity been met?
- What are the long-term impacts of the effort/activity on participants?
- What are the unanticipated outcomes of the effort/activity?

Methods for answering the kinds of evaluation questions that arise at different points during a project vary greatly, depending upon the nature of the questions, the audience for the evaluation, the resources/expertise available and so on. However, some approaches include needs assessments, in which the needs of the particular desired group/audience and the ways in which the projected effort might respond to the perceived needs is determined; a review of relevant research, which synthesizes what is known about the problem or need based on available research, either published research or more “gray” literature (unpublished reports, email, planning documents, etc); review of effective practice and the creation of benchmarks, involving the selection and mining of exemplary practice which have relevance to the problem or need, observations of the effort in some kind of developmental form, usually a prototype, mock-up or some other in-process format; feedback forms; focused or in-depth interviews with participants/stakeholders; observations of the experience or “product” being used, embedded performance tasks, and on occasion when appropriate, measuring changes from pre-test to post-test on agreed-upon measures. These are just a few examples of potential methodologies. Methods should never drive the evaluation though, instead the questions should.

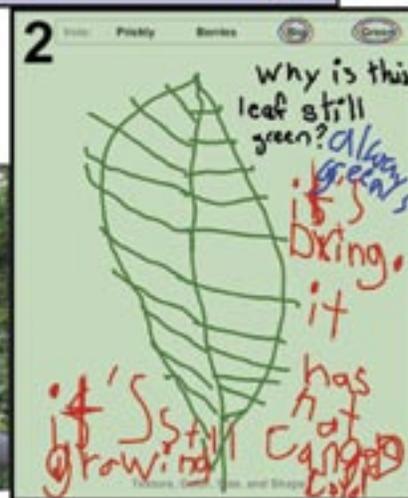
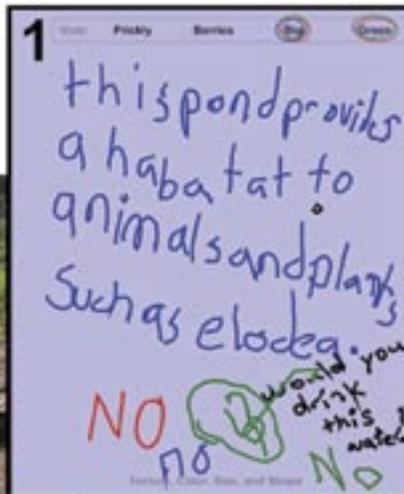
Clearly a major goal of the National Park Service is for evaluation to “become an integral part of program design and delivery to ensure ongoing program improvement, effectiveness, and efficiency.” However, despite all the previous rhetoric about evaluation merely being the flip side of good planning and stemming from questions that effective practitioners raise every day, it is one thing to state this goal and quite another to actually accomplish it in as large and diffuse an organization as the Park Service. There needs to be top-down and bottom-up support for this effort throughout the organization. The necessary resources, financial and intellectual, need to be invested in ensuring that those responsible for evaluating (or seeing that evaluation is conducted) understand the process and see it as a tool for improvement, rather than as a punitive measure. Most of all, there needs to be an appreciation for the difficulty in validly and meaningfully measuring many of the intangible outcomes that we know at a gut level result from people’s experiences in parks. Outcomes such as an appreciation for beauty, enhanced health, wonder, an understanding of democracy, and the struggles that have been inherent in the nation over time to name a few, are not easily documented and measured, certainly not with many of the scientific methods available to researchers in the social sciences. Many of us working to investigate the impacts, both short-term and long-term, of the kinds of experiences afforded by national parks are finding it necessary to create new metrics and methods for documenting the outcomes of these activities because tools primarily are available to measure more tangible outcomes. It is important to remember this statement from the draft Centennial Renaissance Plan: “To preserve only the tangible is to abrogate the power of the parks.”

What Children Can Teach Us:
Creating a Culture of Evaluation in Partnership

Dr. Allison Druin
Human-Computer Interaction Lab, College of Information Studies
University of Maryland

Imagine a child wanders through a National Park holding a tabletPC. He is writing a question and drawing a leaf. He notices that the Park Ranger is writing more about the leaf under his picture. He doesn't see her, but knows it's her handwriting. Another boy and his friend are also on the trail. They are wondering if they should drink the pond water. Their friend has already written 'no' to the Park Ranger's question, but they are debating. Later the Park Ranger brings together all the children on the trail to talk about what they have just seen. She tailors the interpretation experience to the places the children have pointed out in their writing and drawings. The children are captivated by their talk with the Park Ranger. With each stop on the trail, you can hear, "That's the leaf I drew!" "I told Justin about that Pond!"

This scenario is not waiting to happen. With today's technologies, 21st century interpretation and education are now evolving. This has come about because adults and children have partnered together to dream about what is possible.



Rock Creek Park, NPS Washington DC, October 2005

This partnership began before anything was built and before the good ideas were found. It began with local Maryland children and adults from the University of Maryland of Maryland and the National Park Service. Ideas and sketches were generated. Prototypes were built, tried out, refined, revised, and refined again. This iterative design and evaluation process led us to three afternoons in Rock Creek Park where the general public explored a trail using these technologies (Chipman, 2006). Their experiences were evaluated and are included in a Ph.D. dissertation that will be defended at the University of Maryland in December of this year.

A culture of evaluation permeated all parts of our work together. From idea generation to prototype use in a National Park, children and the general public were our partners in designing the future. To implement this vision, no single way of evaluation was used, but a variety of methods were integrated into the research experience. It could also be said that the roles for children in this research process was just as diverse. They served as expert consumers, test-drivers of technology, focus group informants, and partners in brainstorming and observing. In addition, the general public offered more generalizable feedback on their experiences.

By valuing the imagination, we empower ourselves to dream of the world becoming a better or more decent place, which provides an opening for us to act to transform it (Kohl, 2003, p.76).

Since 1999, children have been our partners at the University of Maryland's Human-Computer Interaction Lab. We have traveled to parks with mobile technologies (Chipman et al., 2006); explored making new storytelling worlds (e.g., Benford et al., 2000; Montemayor et al., 2004); taken new digital library journeys (Druin, 2005; Hutchinson et al., 2006; Hutchinson et al., 2005); and built bridges between children from different cultures (Komlodi et al., In Press). We continue today, twice a week to work in our lab with children, ages seven to eleven, and join researchers from computer science, education, psychology, and art. Over the summer, the lab team meets for two intensive weeks, eight hours a day to continue our research. Children have worked with the team as long as five years and as short as one year. Together we have become an *Intergenerational Design Team* pursuing projects together, writing papers, and creating new technologies.

Over the last two years, we have not only considered how to change the visitor experience in a National Park, but we have explored how to take those National Park experiences to children around the globe on the Internet. In a separate project initiative we partnered with the National Park Service in developing the WebRangers website, where our methods of evaluation with children have continued. From formative evaluation to feedback on evolving prototypes, our partnership has been rewarding in seeing the growth and change in what is possible for WebRangers. Below are some example activities from our sessions:

The lab that day was covered in big sheets of paper. Children were scattered all over the floor drawing with markers and talking to adults. Our goal for the day was to draw how we could collect stories about the people who work in the National Park Service (NPS). We began by wondering, who were the kinds of people that worked in NPS? We came up with everything from moose trackers to trail blazers, to planetarium guides. The question then became how we could we find out about these people? After a half an hour with markers and large paper, the group came back together with a wide range of ideas, "radio tagged rangers" "ranger cams" "stranger rangers"(guess the ranger) "video robot voting" (to change which cam was on, people would have to vote online). In summary, they wanted to see the here and now, and interact with the virtual rangers.

These were my notes from a spring session with our design team. With little more than traditional art supplies, energetic children, and knowledgeable adults, new ideas flourished. A few months later we continued our work together:

The WebRangers developers from the National Park Service came back with screen designs that suggested some new directions. We had the children click around on the screens to either hear stories or add stories. But what the children seemed to get stuck on was the drawing of the ranger's desk. They wanted to use every postcard, binocular, badge, radio, and piece of the bulletin board. So we decided on the spur-of-the-moment to blow up really big and print out the drawing of the desk area. We then really dug deep into those ideas surrounding the desk. Before the session began, we had no idea where the children would focus and what would come out of the session. The children added so many ideas (e.g., zooming to other places with the binoculars, calling other rangers with the radio, adding a first aid kit and a computer to the desk area). Then the children went off to lunch and I asked the adults (National Park Service staff and some of our grad students) to decide what 3 areas to focus on for brainstorming later that day. I expected them to pick the binoculars, radio or bulletin board. Instead, they saw the bigger picture—it wasn't about the little pieces of drawing—it was about "you and other places" (represented by the window and binoculars), "you and other people" (represented by b-board and postcards), and you and your activities (represented by the badges and desk area).

What started out as a critiquing experience, turned into a thinking experience. The children didn't do all the work, but the adults couldn't have reflected without seeing the children's focus. The process led us down paths we didn't expect in small and in big ways. This is the power of the evaluation process. We continue today to explore the future, one partnership at a time.

References

- Benford, S., Bederson, B., Akesson, K., Bayon, V., Druin, A., Hansson, P., Hourcade, J. P., Ingram, R., Neale, H., O'Malley, C., Simsarian, K., Stanton, D., Sundblad, Y., & Taxen, G. (2000). Designing storytelling technologies to encourage collaboration between young children. *CHI Letters*, 2(1), 556-563.
- Chipman, G., Druin, A., Fails, J., Guha, M. L., Simms, S., & Beers, D. (2006). A case study of Tangible Flags: A collaborative technology to enhance field trips. In *Proceedings of Interaction Design and Children (IDC'2006)*. Finland.
- Druin, A. (2005). What children can teach us: Developing digital libraries for children. *Library Quarterly*, 75(1), 20-41.
- Huchinson, H., Bederson, B.B., & Druin, A. (2006). The evolution of the International Children's Digital Library searching and browsing interface. In *Proceedings of Interaction Design and Children (IDC'2006)*. Finland.
- Huchinson, H., Rose, A., Bederson, B. B., Weeks, A., & Druin, A. (2005). The International Children's Digital Library: A case study in designing for a multilingual, multicultural, multigenerational audience. *Information Technology and Libraries*, 1, 4-13.
- Kohl, H. (2003). *Stupidity and tears: Teaching and learning in troubled times*. NY: The New Press.
- Komlodi, A., Albuero, J., Preece, J., Druin, A., Elkiss, A. & Resnik, P. (In Press). Evaluating a cross-cultural children's online book community: Lessons learned for sociability, usability, and cultural exchange. *Interacting with Computers*.
- Montemayor, J., Druin, A., Chipman, G., Farber, A., & Guha, M. L. (2004). Tools for children to create physical interactive StoryRooms. *Computers in Entertainment: Educating children through entertainment Part II*, 2(1). Retrieved July 7, 2004, from <http://www.acmcie.org>.

A Sample of What SIEES Can Do . . .

Doug Knapp Ph.D.
Associate Professor
Indiana University

For the past thirteen years we have evaluated interpretation and education programs. Our efforts have led to an array of findings that have been utilized by other researchers, university students and, most importantly, the people leading these programs. Much of the information that we have gained through our work (see bibliography) epitomizes the mission and spirit of the Service-wide Interpretation and Education Evaluation Strategy (SIEES). More importantly, our efforts offer a “sample” of the potential impact further types of research can bring to the National Park Service by developing extensive evaluation programs throughout the country.

Turning Research In To Practice

Within a culture of evaluation, compelling evidence is used to continually improve programming and demonstrate that programs are relevant, engaging, and effective. SIEES EXECUTIVE SUMMARY P. 3

Qualitative research conducted in several National Parks has yielded information, ideas, and concepts that have had direct impact on each of the Park’s programs. Below are three examples:

Research Findings

Six months following an interpretive canal boat tour through Lowell National Historic Park visitors were interviewed to learn what information / experiences could still be recalled. Findings were rich when tour topics had an association with the visitor (Knapp, accepted). Below is a representative quote from a participant:

My father married a French woman, so his parents disowned him. So I could relate to the strength of the immigrant mind at that particular point and what certainly the businessmen and administrators had to consider.

Park Application

The importance of the visitors’ own experiences related to the mills has helped the Park staff in adding more dialogue with the canal boat participants to find connections between Lowell’s history and the visitors.

Research Findings

A school field trip to George Washington Carver National Monument consists of a series of educational strategies including a Ranger led hike, a tour of Carver’s homestead, a visit to a hands-on science center and a movie of his impact on society. **Fourteen months following the school program** students were interviewed about their field trip. Several experiences and themes were strongly recalled by the students (Farmer, Knapp, & Benton, accepted). One, in particular, that was remembered was attempting to make peanut “milk”: *We got a partner and then we got these little white bowls with peanuts in them. And then we got these sticks and crushed them up and we put like milk in them and we shook them up but we didn’t have to drink it. . . thank God.*

Park Application

Findings from this study showed Ranger led activities had more impact than the static exhibits in the science center. This information was used to support more budgeting for staffing for a future science center at the Park.

Research Findings

A longitudinal analysis was conducted on the residential experience Expedition Yellowstone at Yellowstone National Park. Analysis yielded 20 general concepts that were associated with the program and recalled by the students one year following the program. Many of the students were able to retain specific information that was associated with park objectives (Knapp & Benton, 2006). Recollections included statements such as the following:

The Geyser in Steamboat only has a big eruption about every 50 years and one of the rangers got to see it a half hour after it erupted. . .we learned that you’re not supposed to put anything in the geyser cause it could harm the pattern of the geyser. . .and they made us test and see if some of the soil was acidic or not.

Park Application

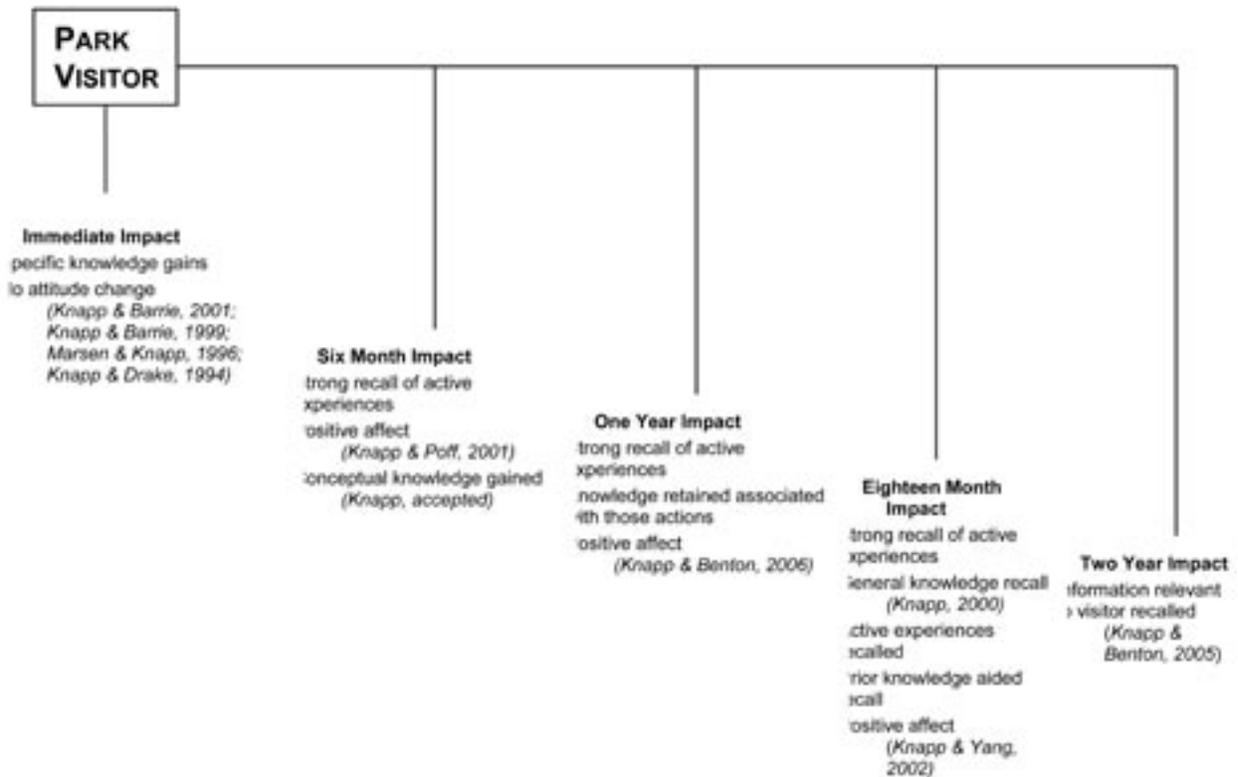
The rich findings from this study were used to support the case to lengthen the time of operation for *Expedition Yellowstone* from four months to nine months out of the year.

Using Research to Promote Accountability

Managers learn how best to invest funds and to apply rigorous accountability measures that support continual program improvement. SIEES EXECUTIVE SUMMARY P. 4

Some of the goals / objectives for National Park Service interpretive and educational programs are lofty ones. In particular, many at NPS feel that a crucial outcome of any program should be stewardship of the Park and the resources beyond its boundaries. Therefore, it is important to collect data that can potentially support this important goal. For the past fourteen years we have conducted a variety of quantitative and qualitative studies at federal, state and local sites (including 15 National Parks) to investigate the potential for such visitor outcomes. Below is an outline of the results of some of those studies and variables that were found to be important in impacting the visitor.

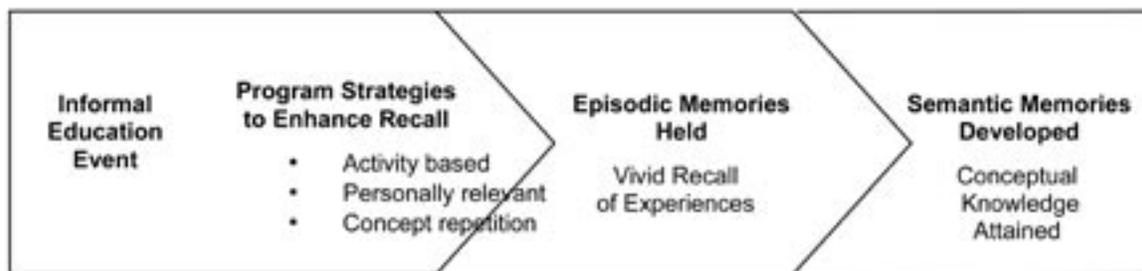
Immediate and Longitudinal Impacts of Interpretive and Education Programs



The outcomes from these studies are still far from proving NPS’s lofty intentions. But it is a start... and one that could be further developed through SIEES. More long term studies from a variety of institutions with a variety of research methods could only aid in the accountability and stature of both interpretation and education programs.

The results from several of our recent National Park Service studies have prompted the development of a learning model in interpretation and education. Its development was strongly influenced by the field of psychology and long term memory work. The model (see below) is based on the idea that an informal educational “event” would offer a set of experiences that would relate to one or more of the three variables that enhance episodic memory systems (active experiences, repetitive content, and information pertinent to the participants). The enhanced episodic event would then potentially aid in the development of these recollections into semantic memories. The development of semantic memories would then represent the potential for conceptual knowledge gained.

A Learning Model for Interpretation and Informal Education



This model has recently been published and/or accepted for publication in leading journals in science education, environmental education and interpretation. Its development and potential use by academicians and professionals beyond the Park Service can only help in promoting the accountability of NPS's interpretation and education programs. And, again, SIEES can promote the development of other models of learning which further increases the accountability and stature of interpretation and education.

In Conclusion

Evaluation at all levels of the NPS leads to sound decision making that ensures cost effectiveness, financial accountability, and interpretation and education that meet or exceed rigorous standards. SIEES EXECUTIVE SUMMARY P. 4

As noted at the outset, the information above is presented as a "sample" of what research has accomplished visiting a small number of Parks. It should, therefore, give a sense of the impact SIEES could have in creating more research and evaluation from others around the country. Pertinent findings from these projects could then be applied to interpretive and educational programs throughout the National Park Service. This, in turn, will lead to more effective and accountable programs.

References and Sample Bibliography

- Knapp, D.H. (accepted). The development of conceptual knowledge through interpretation. *Journal of Interpretation Research*.
- Knapp, D.H. (accepted). A longitudinal analysis of an out-of-school science experience. *School Science and Mathematics*.
- Farmer, J, Knapp, D.H., & Benton, G. (accepted). *The effects of primary sources and field trip experience on the knowledge retention of multicultural content*. Multicultural Education.
- Knapp, D.H., & Benton, G. (2006). Episodic and semantic memories of a residential environmental education program, *Environmental Education Research*, 12(2), 165-177.
- Knapp, D.H., & Benton, G. (2005). Long term recollections of an environmental interpretive program. *Journal of Interpretation Research*, 10(1), 51-55.
- Knapp, D.H. (2005). The case for a constructivist approach to interpretation. *Legacy: National Magazine for Interpretation*, 20-21.
- Knapp, D.H., & Benton, G. (2004). Analysis of interpretive programs at selected national park service units: Variables for successful interpretation. *Journal of Interpretation Research*, 9(2), 9-25.
- Knapp, D.H. (2003). Interpretation that works. *Legacy: National Magazine for Interpretation*, 14(6), 26-33.
- Knapp, D.H., & Yang, L. (2002). A phenomenological analysis of long term recollections of an interpretive program, *Journal of Interpretation Research*, 7(2), 7-17.
- Knapp, D.H., & Barrie, E. (2001). Content evaluation of an environmental science field trip. *Journal of Science Education and Technology*, 10(4), 351-357.
- Knapp, D.H. (2001). Content analysis of interpretive restoration programs. *Ecological Restoration*, 19(1), 58-60.
- Knapp, D.H., & Poff, R. (2001). A qualitative analysis of the immediate and short-term impact of an interpretive program. *Environmental Education Research*, 7(1), 55-65.
- Knapp, D.H. (2000). Memorable experiences of a science field trip. *School Science and Mathematics*, 100(2), 65-72.
- Knapp, D.H., & Barrie, E. (1999). Ecology versus issue interpretation: The analysis of two different messages. *Journal of Interpretation Research*, 3(1), 21-38.

**Creating a Culture of Evaluation:
Building Capacity, Engaging Audiences, Practicing Curiosity, and Taking the Sting out of Failure**

**Martha C. Monroe
School of Forest Resources and Conservation
University of Florida**

I applaud the leadership of the National Park Service as you explore how to support interpretation and education across the agency through evaluation. You are not alone in this adventure, as colleagues in other agencies and organizations are also moving toward greater accountability, enhanced professionalism, and more meaningful services. The fields of environmental education and interpretation have been working to train practitioners in the art and science of evaluation. Not many organizations, however, are exploring what it takes to create a *culture of evaluation*. Your discussions may catalyze improvements beyond the Service to the entire profession.

My experience in evaluation is in nonformal, environmental programs. I have worked with the U.S. Fish and Wildlife Service's four-day residential course on education program evaluation, taught graduate students to use evaluation in the development of environmental education programs, and worked with nature centers and organizations as they conduct evaluation. From my experience, I can provide these insights to the development of a culture of evaluation in the National Park Service.

Building Capacity

Perfectly fine program evaluations can be administered by outside experts, delivered to administrators, and used to make important decisions. They can help create a culture of using evaluation for decision-making. An NPS culture of evaluation, however, should engage field staff in defining, directing, implementing, and analyzing evaluations. The course participants with whom I've worked during the past ten years (from NPS and USFWS) have been eager to learn how to evaluate their work. They want to produce the very best programs possible, and they seek evaluation tools that will help provide evidence and insights for making improvements. In the process of learning evaluation skills, these individuals become more capable, competent professionals. I suggest that building a culture of evaluation will necessarily involve building capacity among interpretation and education staff across the system, from the field to headquarters.

An added benefit of building capacity among field interpreters is that good evaluation processes often require their participation to identify evaluation goals and audiences, create instruments, and interpret results. Like other participatory evaluation strategies, the success of your evaluations may reside with the ability of staff to assist.

Engaging Audiences

A culture of evaluation opens the door to asking questions such as "How well does this work?" and "How do we know it works?" As teams work on developing an evaluation, it is not unusual for someone to wonder, "Why do we think X is better than Y?" and "For whom is X appropriate?" or "Do all audiences agree that this ought to be our goal?" At that point, the development of the evaluation might move outside the agency to the stakeholders and audiences. Rather than merely completing forms or being interviewed, audiences can participate in the design of evaluations by helping to identify evaluation goals. Promoting involvement and engaging audiences in building civil society might begin, in a culture of evaluation, by working with audiences and stakeholders in the design, implementation, and interpretation of program evaluations.

Practicing Curiosity

Park Service interpreters and educators are naturally curious people. They ask questions about which insect made that gall, how long it takes that river to carve a channel, who pollinates that flower, and why didn't the Confederates destroy the Monocacy Aqueduct. They use their curiosity to create programs that spark the imagination of their audiences. They might need practice, however, to redirect their curiosity to ask questions about programs. A culture of evaluation will create opportunities for them to play with elements of their programs and determine how to best achieve their goals. Some interpreters, for example, stop short of giving their audiences the take-home message because they want to entice people to discover it on their own. A culture of evaluation would tickle the curiosity of those interpreters to ask "Did they get it?" If the answer is "Not really," a culture of evaluation would help those interpreters modify their program and ask again. Rather than being curious about what they are interpreting, a culture of evaluation enables interpreters to be curious about how they are interpreting.

Taking the Sting out of Failure

A culture of evaluation must make it safe for staff to ask hard questions and get disappointing results. Of course no one wants to have an evaluation reveal a fault or problem, but a culture of evaluation must communicate that such results are acceptable, and perhaps even more helpful at improving services than glowing reports. The trick will be in how these results are used. An evaluation of a countywide environmental education program in Florida, for example, has a prior commitment from the school district to use the results to improve the program through training or redesign, not in cancellation.

A small but well visited Flavor Graveyard at the Ben & Jerry's ice cream factory in Waterbury, Vermont, celebrates even the most short-lived combinations in clever limericks on tombstones. Sweet Potato Pie ice cream was not a hit, nor was Tennessee Mud (with a dash of whiskey), or Peanuts and Popcorn! By inviting us to laugh at their mistakes and mourn the retirement of tasty varieties that were no longer profitable (ever wonder about Rain Forest Crunch?), they help reinforce the notion that trying new ideas is good, even if they don't work.

R. I. P. Sugar Plum

*It swirled in our heads,
It danced in our dreams,
It proved not to be, though
The best of ice creams.
1989-1990*

Summary

A culture of evaluation builds the capacity of staff to conduct and participate in evaluations and opens the door to audiences and stakeholders to help in the design and implementation of evaluations. It fosters curiosity and supports staff who seek questions about the worth and value of their programs. A culture of evaluation enables staff to take risks as they learn more about the perceptions and attributes of their work, and rewards this type of learning with opportunities for improvement. A culture of evaluation empowers all staff with the tools that professionals use to reflect on their work and make changes.

Our Dearly Departed White Russian

*A concoction so to-die-for
We were forever in its debt
As the liqueur kicked the budget,
We finally had to just say "nyet."
1986-2002*

Ben and Jerry's Flavor Graveyard,
Waterbury Vermont

Creating A Culture of Evaluation: An Indigenous Perspective

Richard Nichols
President, Colyer Nichols, Inc.

As a co-Principal Investigator on a National Science Foundation-funded project to develop an Indigenous Evaluation Framework for STEM (science, technology, engineering, and mathematics) education, I'd like to proffer the notion of evaluation as knowledge creation rather than as judgment. In formulating our "framework," we conducted several regional and culturally diverse focus groups to get input from American Indian culturalists, traditional knowledge holders, educators and evaluators on what an evaluation model would look like from a truly indigenous perspective. Many of the participants emphasized the negative history of evaluation in Native communities, i.e., that evaluation is critical, judgmental and focuses on mostly finding what is wrong. From this concern about the judgmental nature of evaluation, we began to foster a conception of evaluation as knowledge creation — finding out what works and what doesn't in a program, how do we improve what doesn't work, how do we sustain and strengthen what does work.

My recommendations regarding creating a culture of evaluation within the National Park Service are drawn primarily from an evaluation my firm conducted of a 30+ million dollar American Indian higher education initiative of a well-known foundation targeting tribal colleges and universities, as well as mainstream universities. I use this example because the foundation did really make an effort to change it's organizational rules in designing and implementing the initiative and serves as a model for how other foundations might approach American Indian communities and tribes. The following tenets are taken from our examination of how that foundation changed its organizational culture.

Be open to real change in institutional rules and policies. In initially approaching tribal colleges and universities (TCUs), the foundation staff had proposed choosing certain TCUs for funding "centers of excellence" similar to what the foundation had done with historically black colleges and universities. The tribal colleges challenged the foundation and said that their values would not allow for this. The TCUs insisted that the foundation fund "all of us, or none of us" — i.e., that only a few could not benefit when the need was so great among all the colleges. The foundation was taken aback in that their formulation for funding was being refused. However, the CEO of the foundation — being privy to the "difficult conversations" that had taken place in these initial discussions — acceded and convinced the foundation's Board of Directors into working with the TCUs to design an initiative that they felt would work for the benefit of all. Similarly, the National Park Service should engender a willingness among its policy makers, as well as its internal workforce, to be open to the development of a culture of evaluative thinking.

Culture matters. Another lesson learned by the foundation was that culture matters, especially when dealing with many different tribal communities. Thus, the foundation sought representation from various constituencies involved in American Indian higher education to advise them in designing a comprehensive and systemic approach to addressing the higher education needs of Native students. The cluster evaluation team, thus, had to be culturally competent and at ease in "crossing borders." Similarly, the National Park Service in addressing the needs of its varied constituencies should be sensitive to the many cultural lenses through which its interpretation and educational programming will be viewed. Thus, its evaluation design should provide ways in which to engage these various cultures and integrate them into the evaluation process.

Context matters. The foundation, as it implemented its planning phase for the types of projects that might fall under the initiative, was responsive to the TCUs' and advisory committee's concerns that the special circumstances of each tribal college and/or mainstream university relationships be taken into consideration. Thus, early on it was decided that a "one-size-fits-all" approach to project design would not work. I think in developing its approach to engendering evaluative thinking among its many sites, the NPS must take into consideration their varying contexts. The evaluation design, approaches and instruments should be generated from the bottom up. Indeed, if evaluation is viewed as knowledge creation, the place to start is in engaging local park personnel in generating the types of evaluation questions that would be important to them and their visitors.

Evaluation must be meaningful to both NPS staff and their constituencies. Evaluation approaches that are responsive to a broad spectrum of stakeholders are necessary for capturing the views of the broad range of NPS concerns. Help NPS personnel learn more about those evaluation traditions that actively engage stakeholders — empowerment, responsive, participatory approaches. These varied stakeholders should be engaged early on in the evaluation process and actively engaged in helping design evaluation questions meaningful to them.

Allow for an indigenous perspective. Finally, as the program chair for the Indigenous People in Evaluation topical interest group of the American Evaluation Association, I would be remiss in not specifically addressing ways in which American Indian communities should be brought into the interpretation and education evaluation process. Many of the American Indian persons — former or current NPS personnel — with whom I have spoken noted that most national parks have an American Indian story that should be part of its interpretation and education programming, but often park service personnel limit their I & E programming to other topics, often limiting topics to those included in their enabling legislation. However, I would venture to say that just about every natural area and resource in this country has an indigenous story to tell, as well as the story of the interactions between indigenous peoples and non-Natives. Thus, NPS would do well to engage American Indian tribes in telling those stories and in evaluating current I & E programming to see what might be missing. Furthermore, when engaging tribes, do not only think local or nearby tribes. Remember that there has been a history of displacement among many tribes. Many tribes have stories about places that NPS may not think relate to them.

PART or PARKS?

Emma Norland
The Ohio State University

Except for the last six years, I have spent my entire life in Ohio being content in the knowledge that a buckeye can be both a tree and a football fan. When I moved to Washington DC to the US EPA in 2000, I had no idea how quickly a buckeye would become just a nut and PART, GPRA, OIG, and OMB would become the table talk of our nightly dinners. For my first three years in Washington I assisted the Inspector General in establishing an Office of Program Evaluation and the final three years were spent arm-wrestling OMB for Office of Research and Development in the Program Assessment Rating Tool process. I have quite a lot to share from those experiences...but not today.

Before my short public service stint in Washington, I was fortunate to hold a faculty appointment in both the Colleges of Agriculture and Education at the Ohio State University where I taught and conducted research of non-formal education including environmental education. During that time, I conducted various evaluations including NPF program evaluations. It is one such program evaluation, the PARKS (Parks As Resources for Knowledge in Science) Project Evaluation, I'd like to discuss today.

The NPS PARKS Evaluation

To date, the PARKS project evaluation is still one of the largest cluster evaluations conducted of National Parks education programs. Funded by the National Park Foundation through Exxon Mobil, the PARKS project not only supported park-school collaborative programs at 36 different National Parks but provided dedicated money for a cluster evaluation of the entire project. This evaluation included site visits to every park, multiple program observations, teacher and student interviews, park educator interviews, park-school administrator interviews, multiple written instruments, student questionnaires, and in addition, support and training for all PARKS participant programs to conduct their own mini-evaluations of their programs. Even though the parks were not randomly selected to receive a PARKS grant (they were competitively awarded) the recipient parks did vary by location, local culture/economy, cultural/historical/natural resources, budget, area, type of resource, education staff size and experience, education program history, and collaboration history with local schools. Thus, I believe there is much to be learned from both the process and the outcomes of the evaluation as the NPS moves forward in its planning for I&E Program Evaluation.

The actual PARKS project involved 36 park-school teams that worked over the three-year grant period to integrate the National Science Education Standards (NSES) into park-based education. Approximately 90,000 students in grades K-12 participated. Program formats included visits to the NPS sites, in-school activities, overnight trips, and web-based learning – many programs combined more than one activity.

Some Findings from the Cluster Evaluation:

- Park educators had success integrating the non-content NSES standards into park programming as evidenced by more student engagement, active participation, and use of inquiry-based activities.
- Students demonstrated higher levels of perceived science learning when NSES standards were better integrated. However, student attitudes, enjoyment of learning, and stewardship were not associated with the integration of NSES into learning activities.
- Strong park-school collaborations were related to higher levels of all student outcomes.
- Four components of student stewardship were discovered: long-term commitment to the resources and/or the places in which they are found; a desire to protect those resources and places for the future; socially-endorsed, situational environmental behaviors; and internal locus of control with respect to actions toward the environment.
- Students participating in longer programs demonstrated higher levels of stewardship.
- Students who had visited the park sites prior to participating in the PARKS programs showed higher levels of stewardship, more positive attitudes toward the National Parks, and the belief that science is in their everyday world.

- The above findings can be useful to the NPS and other non-formal education providers but what I'd like to focus on, from an evaluator's perspective, are the things I wonder about every time I conduct a *stand alone or one-time-only evaluation*, even one as large and as complicated as the PARKS project evaluation.
- Evaluations are hardly ever going to be as complicated and in depth as this one but they should all *fit into an organization's evaluation system* that is capable of containing one as complicated as the PARKS evaluation. The big picture, the template, the system, the puzzle frame, whatever we call it, should be comprehensive enough that any question that can be asked has the potential to be answered using one or more components of the evaluation system.
- The best working philosophy I've found is *the utilization-focused, stakeholder approach* (Patton, *dates too numerous to mention – ask him*). Every evaluation project begins with this thinking: Who wants the information? How are they going to use it? What decisions are they going to make? Who else wants the information? What are their questions specifically and are they really answerable? And if not, can we negotiate for different, additional, better information or questions?
- Example: Exxon Mobil wanted to know if the PARKS program increased student science knowledge. The problem was: that was a cause/effect question implying an experimental design with 36 different science contents (36 different science tests needing to be developed, tested for content validity and reliability) with ages from K-12, with no money for a longer testing time period, pilot tests, etc – we explained why the question could not be answered in this evaluation but how we could measure perceived student learning from several sources. . .and that was fine for their purposes. By involving the people and groups who really wanted information about the program, we could agree upon what they wanted to know and what we could find out, realistically and what the implications were, and then test all these assumptions throughout the process. We came back more than once with issues of budget and they came back with issues of needing more and different help. The flexibility of the 'active, reactive, adaptive' (a la Patton) flow of a stakeholder driven approach and evaluator worked well. Key – a seasoned evaluator and a project coordinator who trusted not only the evaluator but the process.
- Many organizations commissioning or conducting evaluations on a regular basis have a fairly consistent set of stakeholders asking fairly consistent questions. This would suggest that a *portfolio or an array of instruments* could be designed and tested for psychometric soundness (reliability, validity, usability) and then specific instruments could be selected and used depending upon the questions for a particular evaluation. The PARKS project evaluation staff developed and tested all the quantitative and qualitative instruments used in the study. Most of those instruments are still current and appropriate for use with NPS education programs and of particular interest are the stewardship instruments and the collaboration instruments.
- Along with the instruments is the *accompanying database* which, once developed will house and manage data long-term and allow analysis of data across subsets and across time. The PARKS project evaluation team created a data base and housed the quantitative data during the study in SPSS, a statistical package which allowed us to answer just about any question regarding the entire evaluation or the 8 park subset.

Some Final Thoughts

The National Park Foundation funded an author's retreat for the development of the *New Directions for Evaluation, No. 108, Winter, 2005*, titled "Evaluating Nonformal Education Programs and Settings". The editors, Cindy Somers and I are also joined at this summit by three other authors, Elizabeth F. Hoermann, Martha C. Monroe, and Julia Washburn. It is true that. . .

"Billions of dollars are spent annually on nonformal, informal, and nontraditional education programs and collaborative formal-nonformal efforts. Public and private dollars fund literally thousands of programs, and yet the field of program evaluation has provided little guidance for evaluating such efforts. There are precious few resources available to lead program administrators, staff, and evaluators through the maze of programs with the diversity of the constituencies that support them. The stakeholders and audiences of nonformal education programs are numerous. And these programs can range from a one-shot, hour-long lecture to an ongoing, one-day-a-week volunteer program, to a three-week study tour, to a four-weekends-across-one-year work camp, to a 'stop by when you can' museum collection."

As those who truly care about nonformal education and interpretation, we'd like to join with others in saying "congratulations and thank you, the leadership of NPS, for taking this monumental and historic step."

Place-based Education and the Culture of Evaluation

National Interpretation and Evaluation Summit
October 2006

David Sobel, Education Department
Antioch University New England

From 2000-2001, the Moore Foundation in San Francisco considered initiating a major new funding program focused on environmental education. The intent was to make a significant investment in environmental education in the Bay Area. They reviewed current literature, convened a symposium of prominent researchers, evaluators and educators in the field, and interviewed a wide range of environmental leaders from around the country. To everyone's consternation, the foundation directors decided not to focus on environmental education as a major funding initiative.

An article by Lew Coleman in the San Francisco Business Times entitled, "Moore's New Laws," articulated the foundation's decision.

"We screen out things that are not measurable, and there are lots of important things you can't measure. We end up saying 'no' to a lot of things that are important. For instance, after almost a year and a half of discussion, the foundation decided it would not make a big investment in environmental education for youth because it's not always clear if investments in environmental education result in changing a person's behavior."

In other words, they could find no evidence that environmental education really makes any significant, measurable difference—either in people's behavior or in the quality of the environment.

Practitioners in the fields of environmental education, and the emergent field of place-based education, intuitively knew they were making a difference in the quality of the environment and attitudes of people. But they didn't have good evaluation and research to convince the Moore Foundation of their convictions.

Simultaneous with the Moore Foundation's decision, the directors of a number of maturing place-based education initiatives in New England were facing overlapping evaluation needs for their programs. Recognizing the potential of strength in numbers, these programs and the Upper Valley Community Foundation joined together to create the Place-based Education Evaluation Collaborative (PEEC). Two of the programs were the CO-SEED project, a school improvement and community development initiative orchestrated by Antioch New England and the Forest for Every Classroom (FFEC) project which was run by a partnership of Marsh-Billings-Rockefeller National Historical Park, the Conservation Study Institute, Shelburne Farms, the National Wildlife Foundation and Green Mountain National Forest.

The goals of PEEC are to create a culture of evaluation among the partner organizations, disseminate evaluation techniques and tools to other place-based education initiatives and to contribute to the research base underlying the field of place-based education and school change. Of course, quietly, we also wanted to prove that the Moore Foundation had missed the boat.

The Forest for Every Classroom (FEEC) program is a professional development program for educators created and implemented by a unique partnership of public land management agencies and non-profit organizations. The partners work together to provide teachers with a yearlong workshop series in which they are trained in forest ecology, management and stewardship issues, service learning techniques and curriculum development. *"At the heart of the FEEC program is the belief that students who are immersed in the interdisciplinary study of place are more eager to learn and be involved in the stewardship of their communities and public lands."*

Since two national park units, Marsh-Billings-Rockefeller NHP and the Conservation Study Institute are partners, their experience with evaluation is illustrative. One of their first goals was to study the impact of the program on teacher behavior. Was the program contributing to changing the way teachers' used the natural environment? The evaluation provided both compelling evidence of teacher practice change, both quantitative and qualitative. Surveys showed that the more exposure to FFEC teachers had, the more likely they were to report using local resources for learning and increased engagement and professional growth. During interviews, comments like the following were illustrative and typical:

“The depth of training I received changed my understanding of forests and advanced my ability to develop curriculum. I was transformed as an educator and my commitment will not lessen with time. The opportunity to develop my unit under the guidance and encouragement of the community partners was crucial.”

“Our community is blessed to have a national park and we wanted our kids to feel a strong tie to it.”

Program evaluation also focused on signs of increased student engagement and stewardship behavior, case studies of effective practice and the value of having many partners contributing to the endeavor.

Numerous years of progressively more sophisticated evaluation has led to the following benefits:

A. Evaluation Facilitates an Understanding of Teachers’ Needs—As the result of comprehensive evaluation, MBR Education, Interpretation and Resource Management staff have a much clearer sense of what teachers want and the barriers they face. In the beginning, many were skeptical that teachers would want to be involved in long-term professional development. They were surprised to find that teachers thrived on deep involvement and wanted more. They also learned that working with teams of teachers was much more effective at creating change in the classroom than working with individual teachers.

B. Evaluation Becomes a Form of Civic Engagement—Rolf Diamant, Superintendent at MBR says that, “When teachers and community members are interviewed by formal evaluators, they feel respected and valued as a members of the community and as contributors to the park’s mission.” Knowing that the park is committed to rigorous, objective evaluation makes program participants more likely to continue using the park and to contribute to shaping other programs. Currently, local teachers are staffing a new Community Landscape Analysis initiative in the Prosper Valley north of Marsh-Billings Park. This project is related to the park’s mission, but would not be staffable by the park’s current staff. Thus, the potential of the park to expand its mission has been enlarged by educators whose commitment was secured during the evaluation process.

C. Evaluation Leads to a Fellowship Program—Evaluation findings led program and park staff to realize that alumni from the program wanted some way to stay involved with ongoing initiatives. As a result, FEEC has created a Master Teachers/Fellowship program to foster leadership skills in alumni who will continue to inspire other teachers in their schools and throughout the region.

D. Evaluation Shapes Replication—Evaluation reports have served to identify the salient elements of program success in the FEEC model. Based on these findings, FEEC directors asked evaluators to explore replication models in similar programs around the country. Now, two replication initiatives are underway. Project Learning Tree of New Hampshire is implementing a FEEC program in northern New Hampshire communities. And the Appalachian Trail Conservancy has used FEEC as a model to create A Trail to Every Classroom that will bring place-based education to schools along the trail from Georgia to Maine.

In sum, comprehensive evaluation has been an effective way for the staff at Marsh-Billings-Rockefeller and the Conservation Studies Institute to learn how best to engage the community in the park’s mission. If these NPS units had taken the same stance as the Moore Foundation, the Forest For Every Classroom likely would not have deepened and expanded as much as it has. Both the park and the teachers within its service area would have missed an opportunity to effect real change. As one teacher participant indicated to an evaluator,

“I never gave it much thought before, but now I see public lands as extremely valuable. Students get hands-on learning experiences, become motivated and learn about their environment when utilizing nearby lands—so important in our computer age.”

**OUTCOMES MATTER:
EVALUATION COUNTS**
National Interpretation and Education Evaluation Summit
October 2006
Denver, Colorado

Carol B. Stapp, Ph.D.
Museum Education Program
The George Washington University

Over the years, I have been in an extremely privileged position as the director of the Museum Education Program (MEP). The design of the curriculum allows me to work with the same cohort of Master's degree students for four sequential semesters—summer, fall, spring, and summer again—for about 13 months. While it has always been an extremely intensive curriculum that combines classroom instruction and supervised fieldwork, with extraordinary emphasis upon the students' sense of responsibility, we are nonetheless only in the second year (MEP dates back to 1974) of the "new paradigm"—which could be defined as "a culture of evaluative thinking... characterized by continuous inquiry and learning."

What has proven intriguing has been how difficult it is to institute and instill positive attitudes toward evaluation—new paradigm notwithstanding. Even if committed to effectiveness and curious about how to be more effective, students—like most practitioners, no doubt—are wary, having found in the past that evaluation has not been for improving effectiveness, has not been in their control, and/or has not had a useful return. It has required learning through experience—positive experience of practicing evaluation that lives up to its promise—that evaluation can be for improving effectiveness, evaluation can be in one's control, and evaluation can have a useful return.

The new paradigm posits evaluation as pre-eminent and integral, expressed in terms of keep and change, along with proposed revisions. The student frequently receives peer and/or participant reviews, then carries out a self review, before receiving a faculty review (which weighs the student's own weighing of his/her peer/participant reviews). Whatever the student has produced (gallery teaching presentation, self guide, grant proposal, evaluation report), however, does not represent the outcome for the project. Every project in all MEP core courses focuses on competencies—building capacity. The four semesters of inter-related projects were developed to foster reflective practitioners, as a model for graduates' efforts "to nurture an enlightened, humane citizenry" (*Excellence and Equity: Education and the Public Dimension of Museums, AAM, 1992*).

[Sample faculty reflection about ultimate outcomes, shared with students]

“ The Platonic* *Ideal* of the Museum Educator*** ”**

Key attributes of the museum educator:

1. Communicates effectively with all audiences using all media.
2. Understands museums in society fully.
3. Develops professionally continually.

***Platonic**

Plato: Greek philosopher, who lived 427(?)–347 BCE

Platonism: 3. the belief that physical objects are but impermanent representations of unchanging ideas, and that these ideas alone give true knowledge as they are known by the mind (American Heritage Dictionary, 3rd ed.)

*****Ideal***

1. a conception of something in its highest perfection [archetype, worthy of emulation, standard of excellence]

*****Museum Educator:** a competent professional who is able to

1. Articulate the vision inspiring and the rationale informing her/his decision-making and actions.
2. Reflect upon and take responsibility for the outcomes of the process and products of her/his decision-making and actions.

METAPRACTICE and METACOGNITIVE:

Museum educators as reflective practitioners: consciously and conscientiously processing their practice as a matter of professional self-development, as well as developing the profession—with the ultimate outcome of the museum’s nurturing “an enlightened, humane citizenry” (*Excellence and Equity*, 7).

[Sample project with evaluation—partial guidelines]

PROJECT GUIDELINES: Self-Guide

OUTCOME: After completing the Self-Guide Project, students will be better able to

- a. advocate for accessibility and accountability.
- b. cite the characteristics of effective self-guiding materials.
- c. develop, implement, and evaluate appropriate structured encounters between audiences and artifacts/sites that emphasize visual perception and enhance museum literacy.
- d. identify and implement the principles of working independently and productively.

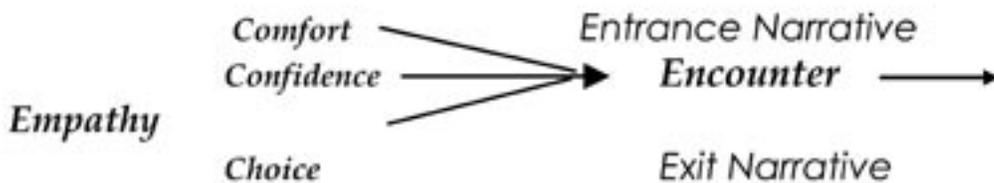
OUTPUT: A self-guide with interpretive plan

- ~~~~~
- a. Self-Guide framework:

OUTCOME:

OUTCOME:

OUTPUT:



- b. Self-Guide key elements (30/45 minutes):

- * Documentation of entrance narrative (benchmark)
- * *Structured encounter—audience + artifact(s)/site(s)*
- * Documentation of exit narrative (milestone)

- c. Reviews

- * Participant Review (Keep/Change)
- * Self-Review (Keep/Change) with Proposed Revisions
- * Faculty Review (Keep/Change) with Proposed Revisions

**Make Evaluation Work
National Park Success Vision:
Generative Pathways Towards a Culture of Evaluation
Hazel Symonette, Ph.D.**

There are voices from the future that are calling your names—in fact, all of OUR names! Those voices embody the hopes, dreams and wisdom of yesterday’s yesterday as well as the provocative possibilities of tomorrow’s tomorrow. Whose voices do you hear as we each stand and sit, as a detail of the earth’s landscape, among all our relations—the animate and non-animate, the 2-leggeds as well as the multitudes of others with whom we share this planet? Whose voices do you heed and how do you know—your evidential cues, clues and signposts? How do you know that you are hearing the full spectrum of voices in full voice? To what extent would which voices agree with your self-assessment? What leads you to believe that your “read” is accurate? These are foundational evaluative questions that we each need to ask ourselves.

Evaluation works best when we responsively work it for the greater good of the persons that our initiatives exist to serve. What claims are you making about the impact of the National Parks Service Interpretation and Education services for which you are responsible? How credible and compelling are those claims to your key stakeholders and how do you know? Savvy educators and service providers proactively embrace assessment and evaluation as rich self-diagnostic resources for critical and creative reflection, for empowered self-improvement and for strategic image management. Unleashing the potent powers of assessment and evaluation cultivates *inclusive excellence** when we resourcefully work these processes for the greater good of all the persons that we are expected to serve.

EXCELLENCE THROUGH KEEPING OUR EYES ON THE PRIZE. Assessment and evaluative judgments are inextricably bound up with culture and context so *engaging* diversity provides an essential resource. Excellence is the prize and diversity, a necessary prerequisite for its attainment. Excellence demands that we *know the prize* from multiple vantage points so that we can more fully focus our eyes upon it. The ultimate prize resides in persons who receive your Interpretation and Education services vis-à-vis your outcome promises. What does their and your success vision picture them experiencing, learning, being able to do as a result of exposure to and experience with your services? To what extent are those visions the same or at least congruent? In what ways and to what extent are your curricular, pedagogical and other intervention activities breathing life into success visions for all segments of your potential target population? How do you know what you have accomplished and to what extent do your evaluative judgments resonate with the lived realities of persons that you assess—experiential validity?

SELF AS INSTRUMENT AND INTERPERSONAL VALIDITY. Addressing these questions spotlights “interpersonal validity” issues which demand ongoing personal homework—expanding and polishing our sociocultural lenses and filters. Who are we as knowers, inquirers and engagers of others? Our lenses, filters and frames exert critical influences on assessment processes and evaluative judgment-making—whether intended or desired or not. Expanding our self-awareness is especially crucial when engaging and working across diversity divides—salient differences that make a substantive difference in access, process and success. Such interpersonal relations vary widely in their impacts on the quality, accuracy and trustworthiness of observations and interpretations. Data are neither self-evident nor do they speak for themselves. The same data can conjure up dramatically different meanings and interpretations depending upon where one stands and sits and, thus, one’s perceptual and interpretive prisms: i.e., what one looks for, actually sees and meaningfully discerns.

Without such deep self-development work, we cannot accurately judge the quality and resonance of our own perceptions, transactions and meaning-making interpretations—especially when data collection, analysis and interpretation processes involve boundary-spanning communications. Using self as responsive instrument summons understandings of self in dynamically diverse contexts within power and privilege hierarchies and also understandings of the contexts embodied in the self. Enhancing interpersonal validity—self as diversity-grounded knower, inquirer and responsive engager of others—calls for a lifelong learning and reflective practice journey because culture is itself organic, dynamic and ever-changing. This *self-as-responsive-instrument* learning and development journey is without end.

Evaluation for Inclusive Excellence commands us to deepen our awareness of “*interpersonal validity*” as a critical complement to the more conventional *methodological validity*—the soundness and trustworthiness of understandings warranted by one’s uses of the SELF vis-à-vis assessment/evaluation tools, techniques and strategies. To be sustainable, however, this needs to become a vibrantly responsive process that informs and improves as well as proves: notably, a valuable resource for relevant knowledge creation and continuous development towards excellence in addition to conventional accountability compliance verification.

CULTIVATING ENGAGEMENT AND COMMITMENT TO EXCELLENCE THROUGH EVALUATION. I suggest that the National Park Service embrace an approach to evaluation that magnifies the inform and improve drivers which works, first and foremost, for those closest to doing the work and delivering services. Such an approach is grounded in the belief that evaluation works best if you work it as engaged, active participants and collaborators rather than passive objects of others' evaluative judgments. My approach focuses on cultivating the *will* as well as the *skills* to engage in participant-centered program development and evaluation. I strive to mainstream assessment/evaluation processes through, for example, spotlighting systematic inquiry and judgment-making *in the service of* an envisioned intervention and its success vision. As capacity-enhancing resources, they enable the intervention as well as the interveners. Michael Patton's concept of developmental evaluation most closely reflects this approach. This form of assessment/evaluation involves mindful "R&D" that guides the design, implementation, and refinement of an intervention by "infusing evaluative questions, data, and logic" to support empirical evidence-based decision-making. (Patton, *Evaluation Encyclopedia*, 116). Developmental evaluation especially focuses on the *inform* and *improve* drivers vis-à-vis the *prove* drivers for evaluation.

PRIORITY EVALUATION DESIGN AND DEVELOPMENT PRINCIPLES

- Cultivate a vested interest in program data collection and education/interpretation outcomes evaluation because of the *internal* benefits for each program.
- Maximize the *natural* utility of program data collection, evaluation and reporting as a Parks staff resource for empowered self-improvement and strategic image management:
- Help program staff keep track of the *who, what, when, where, how* and *how much* aspects of their job duties.
- Encourage program staff to use program evaluations to discern and communicate their *data-grounded* understandings of how and why desired educational and interpretation outcomes *do or, do not*, occur for all program participants at the levels intended/expected.
- Promote program assessment and evaluation as iterative self-diagnostic processes for continuous improvement—notably, for double-loop learning.

- Cultivate program development and evaluation processes that facilitate the empowerment of staff and administrators through their uses as self-diagnostic resources to maximize the short- and long-term effectiveness of their programs. Empowerment grows through maximizing opportunities for *choices* among *meaningful alternatives* and providing opportunities to develop the *knowledge* and *skills* needed for effective performance. Cultivate this perspective: *Self-study evaluations allow each program (or unit) an opportunity to speak for itself—offering its vision and interpretation of the facts.*

- Cultivate trust and collaborative partnership through fostering assessment and evaluation as a "sit down beside" critical friend process with intrinsic benefits: more specifically, a *sit-in-deliberative-judgment* *with* rather than solely a *stand-in-judgment of* auditor-oriented process. To foster generative commitment and sustainability, I hope that the National Park Service will proactively and creatively address these considerations, even in the face of strong external accountability compliance mandates that pressure otherwise. Keep your eyes on the ultimate prize and let not the tail wag the dog!

- In general, help programs design participant-centered program information, developmental evaluation and reporting processes that are more useful, more user-friendly, more accessible and less onerous. Make compliance as natural a part of the service delivery process as possible.

Given the ennobling nature of the National Parks Service mission and much of your work (as I understand it), I think your approach to evaluation would probably benefit from the work of Barry Kibel in "Evaluating Activities that Ennoble" (<http://www.pire.org/resultsmapping/documents/EvalEnnobling.doc>) and his insightful distinction between evidential and evocative inquiry. Those distinctions have been elaborated and further developed through his work with an innovative leadership evaluation project and the resulting reference manual that is now available on the Kellogg Foundation website—*EvaluLEAD: A Guide for Shaping and Evaluating Leadership Development Programs*. http://www.wkkf.org/DesktopModules/WKF_DmaItem/ViewDoc.aspx?LanguageID=0&CID=281&ListID=28&ItemID=2813740&fld=PDFFile

***NOTE. INCLUSIVE EXCELLENCE**

The Association of American Colleges and Universities, with support from the Ford Foundation, spearheaded a research agenda that spotlights the integral interconnections among diversity and educational quality initiatives: **Making Excellence Inclusive: Diversity, Inclusion and Institutional Renewal**. It places these intersections at the center of campus planning and practice. The **Making Excellence Inclusive** project is designed to help colleges and universities fully integrate these efforts and embed them into the core of academic mission and institutional functioning: "*Through this initiative, AAC&U re-envision diversity and inclusion as a multilayered process through which we achieve excellence in learning; research and teaching; student development; institutional functioning; local and global community engagement; workforce development; and more.*"

Blending Cultural Competence and Project Evaluation for Enhancing Place-Based Learning

Veronica G. Thomas, Ph.D.
Howard University

In response to demands from the general public, employers, and policy makers that our nation elevate the learning outcomes of K-12 students, there have been serious attempts to improve public education. Particular attention has been given to reforming education for students who are most often placed at risk for academic failure -- low-income, minority students in urban schools. As a psychologist who teach in a School of Education and one who also worked in a University-based research center whose mission is to conduct research, development, and evaluation of activities needed to transform schooling for children who have historically fared poorly in our nation's schools, I am very excited about the potential of place-based education in raising students' academic achievement, increasing their social competence, and facilitating their sense of local place and civic engagement. With that in mind, I cannot overstate the reality that as increasing numbers of schools serving diverse populations partner with local communities and the National Park Service (NPS), it is essential that cultural competence is integrated into the design, implementation, and evaluation of place-based educational projects. Operationally this means infusing knowledge about groups' norms, values, customs, worldviews, and traditions into decisions about what projects are offered, how project components are delivered, who evaluates these projects, what evaluation approach(es) are adopted, and how evaluation findings are utilized.

Culture is central to all learning since it is the lens through which individuals' worlds are known, created, and experienced. It is a predominant force in shaping people's behaviors, values, and institutions. Because place-based projects draw from local culture, history, and geography, they can be powerful vehicles for delivery of culturally competent education that assist diverse students in crossing bridges between their home culture (the known) and the culture of the school (oftentimes, the unknown). In place-based learning projects, the local community becomes the context for, focus of, and significant element of the learning environment. Meaningful evaluation of place-based efforts must be sensitive to the cultural milieu of the project and the culture of the individuals (e.g., students, teachers, family and community members) being served by the project. As such, researchers who work in these setting should be culturally competent evaluators possessing the essential qualities of good technical skills, adequate cultural knowledge of the contexts and populations they are studying, strong interpersonal skills, and the ability to be self-reflective in an effort to be mindful about how their own biases, prejudices, and stereotypes might influence their evaluative work. Strong interpersonal skills, in particular, are useful as culturally competent evaluators prioritize and work toward relationship building in diverse communities (and not just viewing individuals/groups as data sources).

Hopefully, some lessons learned from our work at the Howard University Center for Research on the Education of Students Placed At Risk (HU/CRESPAR) (now referred to as the Capstone Institute) can provide the NPS with insights into a culturally competent approach to developing and evaluating place-based learning projects. HU/CRESPAR educational interventions were guided by the Talent Development (TD) model of school reform, which asserts that all students can learn to high standards when key stakeholders are committed to such a goal and hold themselves to high standards. The model has six signature themes: (1) building on students' assets, (2) providing students with transitional support across key developmental periods, (3) engaging students in constructivist and activist learning, (4) preparing students with skills for careers for the twenty-first century, (5) promoting the concept of school-as-community, and (6) focusing on meaning and connected learning. Our interventions sought to educate the "whole child." In other words, we did not simply focus on raising students' standardized test scores but also upon educating children for a vast array of intellectual, socio-emotional, and transformative competence including character building, personal fulfillment, and dual competence in the larger society as well as in their own local community. Our evaluations were not simply meant to be scientific endeavors in search of "truths (or more precisely "probably truths") and "solutions." They also represented social justice and critical enterprises as we argued for using culturally appropriate evaluation results to advocate change and restructuring for improving student outcomes. Our approach was rooted in several traditions of evaluation that intentionally seek engagements with contexts of practice (e.g., responsive, participatory, empowerment, and culturally competent approaches).

The Talent Development evaluation framework has five overlapping themes that guided our work in the diverse communities that we served. These themes include:

- *Engaging stakeholders.* Diverse stakeholders who are representative of the populations that the project served were engaged in authentic ways throughout the entire evaluation process. Obtaining genuine stakeholder engagement is a complex and labor-intensive task, especially when working in poor, diverse communities. However, it remained central to our work. We consistently entered the project under study gently, respectfully, and with a willingness to listen and learn in order to better plan and implement our interventions and evaluations.

- *Co-construction.* An extension of engaging stakeholders is co-construction. Specifically, co-construction is characterized by evaluators' collaborating and forming genuine partnerships with diverse stakeholder groups, project designers and implementers in order to conceptualize, implement, and evaluate the project in a manner that is customized to the local culture. Co-construction permits stakeholders to have meaningful input into framing evaluation questions, determining appropriate methodologies, and organizing results to assist in improving the project and building community capacity. Further, our co-construction efforts sought to democratize power dynamics by lessening the implicit, and sometimes explicit, hierarchical power relations between evaluators and project stakeholders.
- *Cultural competence.* Evaluating projects by making connections through culturally sensitive lens is a signature feature of the TD evaluation framework. If evaluators ignore how cultural context interacts with project implementation and impact, this could jeopardize the accuracy, validity, and believability of the evaluation findings. Cultural competence in TD evaluations was characterized by myriad features including, for example, evaluators' (a) ability and willingness to honor community-based values, traditions and customs and capitalize on opportunities to draw from cultural understandings, (b) infusion of multiple world views in evaluation planning and implementation, (c) sensitivity to methodological approaches and tools best suited for particular cultural groups, and (d) ability and willingness to share findings in culturally appropriate ways.
- *Responsiveness.* We actualized responsiveness by considering, as a point of departure, stakeholders' perspectives prior to planning, implementing, and evaluating any project. We would revisit these issues throughout the entire process to ensure that our work continued to be responsive to stakeholders' immediate needs in an environment of change and oftentimes competing demands.
- *Triangulation of perspectives.* A final theme of TD evaluation involved triangulation of perspectives. Here, we triangulated in multiple ways (e.g., investigator triangulation, multiple operationalism, methodological triangulation, analysis triangulation). For us, triangulation yielded not one, but many answers to a single evaluation question, which in turn, generated deeper and broader insights into the issue under study.

Clearly, blending cultural competence and evaluation is certainly not without challenges. Many urban settings that we worked lack a culture that valued and supported change, evidence-based practice, and ongoing inquiry. Timing was often a problem, since attending to cultural issues is both time-consuming and labor intensive, with much emphasis on relationship building, developing alliances, and co-constructing with stakeholder groups. Notwithstanding, we still found that crafting a shared vision through stakeholder engagement and co-construction and infusing cultural aspects into programs and evaluations have far-reaching benefits (e.g., empowering participants, building competence and capacity) beyond the intended project outcomes.

Blending cultural competence and evaluation of place-based learning initiatives is good practice consistent with the American Evaluation Association's *Guiding Principles for Evaluators*. In particular, the value-addedness of infusing cultural issues into these evaluations is that it can yield critical information for advancing knowledge on what makes a place-based learning project work (or not work) within diverse communities in particular settings and that information is more likely to be utilized in those settings. Based upon my experiences with TD projects and evaluations, I am concluding with the following recommendations to evaluators of place-based learning projects in diverse communities:

- Front load efforts and spend the necessary time to get to know the project setting
- Engage stakeholders in meaningful ways throughout the evaluation process
- Co-construct inquiry activities
- Give back to the community in tangible and intangible ways
- Ensure that evaluation staff are cultural competent
- Treat all stakeholders with respect and dignity
- Be patient and understanding, and have a high tolerance for ambiguity and change

When Evaluation Works, and When It Doesn't

Jon F. Wergin
Professor, PhD Program in Leadership and Change
Antioch University

Most of my experience with evaluation has been in colleges and universities. Program assessment in higher education inspires about as much enthusiasm as it does in other public sectors, which is not much. It makes little difference that the announced goal of evaluation is program improvement, not summative assessment: what predominates is what I've termed a "compliance mentality," a desire to do what is required and get the process over with as soon as possible. Needless to say, not much organizational learning results.

The research I've done on evaluation in higher education over the years suggests that, distressing as this scenario may appear, the compliance mentality can be overcome, and evaluation can work, even in cultures that are notoriously resistant to change. Here are the factors I've found to be most important – even critical – to the effectiveness of evaluation:

- *A leadership of engagement.* Leaders must be able to frame issues clearly, put clear choices before the staff, and be open to negotiation about what will inform decision-making, including the role evaluation will play.
- *Engaged organizational units.* Departments are encouraged to ask basic questions about themselves: "What are we trying to do? Why are we trying to do it? Why are we doing it that way? How do we know it works?" In essence, departments have created a climate for critical reflection.
- *A culture of evidence.* The organization has a spirit of reflection and continuous improvement based on data, an almost matter-of-fact acceptance for evidence as a tool for decision-making. A culture of evidence bears little relationship to the *amount* of evidence; the key lies rather in what the institution *does* with the evidence it collects.
- *A culture of peer collaboration and review.* Common criteria and standards for evaluation are negotiated based on a shared understanding by staff of one another's work, and differential contributions to the larger organizational mission.
- *Evaluation with consequence.* Evaluation has a tangible, visible impact on resource allocation decisions. "Consequence" has its limits, however: evaluation cannot be *so* consequential that it turns into a high-stakes political exercise. When that happens, incentives for improvement are lost in the rush to look good.

I should emphasize that these five factors emerge from the study of program evaluation in higher education, and of course significant differences exist between the academy and public service. I'd wager that transferability is high, however.

What these five factors add up to for me is that in order to create a "culture of evaluation" within NPS, evaluation must belong to the staff of the local unit. By this I mean that evaluation is not someone else's agenda, it's the staff's agenda – that it's inside-out, not outside-in. Evaluation should be a systematic process designed to help address questions that local staff care about, as long as these questions are negotiated with key stakeholders and are congruent with the NPS mission.

These factors lead to a second insight, one that may be more counter-intuitive than the first. Evaluation must be consequential, yes – but evaluation should not be conflated with accountability for results, which stifles creativity and experimentation. We engage in program initiatives because we think they will work, but we can't be sure. Such things are not predictable, and the paths to success are not linear. Holding program staff accountable for demonstrating the effectiveness of every new idea is a huge disincentive for trying *any* new idea. The risks of failure are too great. Instead, evaluation should encourage taking reasonable risks, and promote accountability for organizational learning. The NPS could then avoid much of the compliance mentality that has plagued evaluation in other settings, and concentrate instead on using evaluation data for the improvement of professional practice.

Appendix F: Evaluation Strategy Implementation Matrix

Objective	Phase I Foundation	Phase II Integration	Phase III Utilization
<p>Coordinate I&E Evaluation Processes</p>	<ul style="list-style-type: none"> Solicit input from a variety of NPS stakeholders Convene with the NPSAB an Evaluation Summit to guide and verify the SIEES 	<ul style="list-style-type: none"> Establish an I&E Evaluation Coordinator and Team at the national level Identify and tap into existing evaluation mechanisms and offices 	<ul style="list-style-type: none"> Engage the I&E networks at the park, regional, and national levels to implement the SIEES Link on-going evaluation efforts with other assessment tools and requirements Review the SIEES annually
<p>Establish Information Management System</p>	<ul style="list-style-type: none"> Compile NPS evaluation case studies Review related literature 	<ul style="list-style-type: none"> Inventory and catalogue existing evaluation efforts 	<ul style="list-style-type: none"> Create an online evaluation library
<p>Establish Best Practices in Evaluation and in I&E</p>	<ul style="list-style-type: none"> Create a comprehensive program model for Interpretation and Education 	<ul style="list-style-type: none"> Identify pilot evaluation projects Review existing research to identify lessons learned, best practices, gaps, and evaluation needs 	<ul style="list-style-type: none"> Support pilot evaluation projects to establish best practices; conduct evaluation Design mechanisms to track utilization of results, and develop guidelines for generalizable results Establish criteria to assess condition of media
<p>Train Employees and Develop Tools</p>	<ul style="list-style-type: none"> Develop Vision, Goals, Objectives, and Actions for the SIEES Seek employee feedback on SIEES 	<ul style="list-style-type: none"> Roll out SIEES in the context of other NEC efforts 	<ul style="list-style-type: none"> Link professional development opportunities and reward systems to evaluation best practices Create an interpreters' and managers' toolkit of evaluation materials

Service-wide Education Evaluation Strategy document completed December 2006

A convenient Web address has been established for evaluation resources: www.nps.gov/interp/evaluation.

Printed on recycled paper with a minimum of 30% post-consumer waste using vegetable-based inks. Please reduce, reuse, and pass this publication along to other readers.

