2013

18(2)

journal of *interpretation* RESEARCH

journal of *interpretation* RESEARCH

Volume 18, Number 2
2013

Editor

Carolyn Widner Ward
Chief Executive Officer
Blue Ridge Parkway Foundation
Asheville, North Carolina

Editorial Assistant

Jason Urroz
Project Director
Healthy Kids/Healthy Parks
Blue Ridge Parkway Foundation
Asheville, North Carolina

Associate Editors

James Absher, Ph.D.
Pacific Southwest Research Station
USDA Forest Service
Riverside, California

Robert Bixler, Ph.D.
Department of Parks, Recreation, and Tourism Management
College of Health, Education, and Human Development
Clemson University
Clemson, South Carolina

Rosemary Black, Ph.D.
Senior Lecturer,
School of Environmental Sciences
Charles Sturt University
Albury, New South Wales
Australia

Alan D. Bright, Ph.D.
Department of Parks, Recreation and Tourism
Colorado State University
Fort Collins, Colorado

John H. Burde, Ph.D.
Department of Forestry
Southern Illinois University
Carbondale, Illinois

Theresa Coble, Ph.D.
Arthur Temple College of Forestry
Stephen F. Austin State University
Nacogdoches, Texas

Lesley Curthoys, Ph.D.
School of Outdoor Recreation, Parks and Tourism
Lakehead University, Canada

William E. Hammitt, Ph.D.
Department of Parks, Recreation and Tourism Management
College of Health, Education, and Human Development
Clemson University
Clemson, South Carolina

Doug Knapp, Ph.D.
Department of Recreation and Park Administration
Indiana University
Bloomington, Indiana

Brenda Lackey, Ph.D.
College of Natural Resources
University of Wisconsin–Stevens Point
Stevens Point, Wisconsin

Ross J. Loomis, Ph.D.
Department of Psychology
Colorado State University
Fort Collins, Colorado

Steven Martin, Ph.D.
Environmental and Natural Resource Sciences Department
Humboldt State University
Arcata, California

Mark Morgan, Ph.D.
Department of Parks, Recreation and Tourism
University of Missouri
Columbia, Missouri

James Pease, Ph.D.
Deptartment of Natural Resource Ecology and Management
Iowa State University
Ames, Iowa

Erin Seekamp, Ph.D.
Department of Parks, Recreation & Tourism Management
North Carolina State University
Raleigh, North Carolina

Dave Smaldone, Ph.D.
Division of Forestry
West Virginia University

Gail A. Vander Stoep, Ph.D.
Department of Community, Agriculture, Recreation, and Resource Studies
Michigan State University
East Lansing, Michigan

John A. Veverka
John Veverka & Associates
Okemos, Michigan

NAI Staff

Margo Carlock
Executive Director

Paul Caputo
Art & Publications Director

Jamie King
Membership Manager

Deb Tewell
Events Manager

Beth Bartholomew
Business Manager

Emily Jacobs
Certification & Training Program Manager

Carrie Miller
Certification Program Specialist

Francine Maestas
Administrative Assistant

Advisory Board

Cem Basman
Assistant Professor
Department of Recreation and Park Administration
Indiana University
Bloomington, Indiana

Larry Beck
Professor
Department of Recreation, Parks and Tourism
San Diego State University
San Diego, California

Ted Cable
Assistant Department Head and Professor
Dept of Horticulture, Forestry, and Recreation Resources
Kansas State University
Manhattan, Kansas

Sam Ham
Professor and Director
Center for International Training & Outreach (CITO)
Department of Resource Recreation and Tourism
College of Natural Resources
University of Idaho
Moscow, Idaho

Bernard Lane
Director
Rural Tourism Unit and Architectural Conservation Programme
Co-editor, *Journal of Sustainable Tourism*
Center for the Historic Environment
Department of Archaeology
University of Bristol, England

Gary Machlis
Professor of Forest Resources and Sociology, National Coordinator for the CESU
College of Natural Resources
Department of Forest Resources
University of Idaho
Moscow, Idaho

Michael Manfredo
Professor, Department Chair
Colorado State University
Natural Resource Recreation and Tourism
Fort Collins, Colorado

Joseph Roggenbuck
Professor
Natural Resource Recreation
Virginia Polytechnic Institute and State University
Department of Forestry
Blacksburg, Virginia

Sam Vaughn
Associate Manager, Interpretive Planning
Harpers Ferry Center
Denver Service Center
Denver, Colorado

Betty Weiler
Professor of Tourism
Department of Management
Berwick Campus
Monash University
Narre Warren, Australia

NATIONAL ASSOCIATION FOR INTERPRETATION
P.O. Box 2246
Fort Collins, CO 80522
888-900-8283 toll free
970-484-8283
970-484-8179 fax
www.interpnet.com

interpPress

Copyright ©2013 ISSN 1092-5872

EBSCO PUBLISHING
A Part of the EBSCO Information Services Group. EBSCO Publishing/EBSCOhost is the registered trademark of EBSCO Publishing.

Contents

Special Issue: *Interpretation: Making a Difference on Purpose*

3 Introduction to Special Issue
Sam H. Ham

9 What Leads to Better Visitor Outcomes in Live Interpretation?
Marc J. Stern, Robert B. Powell

45 Is It the Program or the Interpreter? Modeling the Influence of Program Characteristics and Interpreter Attributes on Visitor Outcomes
Robert B. Powell, Marc J. Stern

61 Speculating on the Role of Context in the Outcomes of Interpretive Programs
Robert B. Powell, Marc J. Stern

79 The Difference Between Good Enough and Great: Bringing Interpretive Best Practices to Life
Marc J. Stern, Robert B. Powell, Kevin D. McLean, Emily Martin, Jennifer M. Thomsen, Bethany A. Mutchler

101 Book Review: *Interpretation: Making a Difference on Purpose*
Carolyn Ward

Appendix

105 Submission Guidelines for Authors

Introduction to Special Issue

by Sam H. Ham

I read a blog a few weeks ago that lamented a "lack of evidence" that interpretation can produce certain outcomes. Unfortunately—as many such discussions go—what these outcomes might be was left vague. This is a problem sometimes in discourse about interpretation, especially when it comes to conversations about what "interpretation" is trying to achieve—what I've called its "endgame." If you can't describe in words what "success" looks like, then discussions about interpretation achieving "something" seem sort of pointless.

However, when the outcomes of interest are clear, the discussion is focused on concrete assertions that can be examined in light of the actual research record. In the past 25 years, compelling evidence has come from studies on cognitive and social psychology that interpretation (when it's done well) does indeed stand a good chance of enhancing the experiences of visitors to free-choice learning settings; it can shape attitudes and foster appreciation of the places being interpreted and of the features or concepts that make those places important; and in certain instances, interpretation can successfully influence how audiences decide to behave, at least in the immediate time frame. If this is the kind of evidence the blog demands, then it is readily available to anyone who is willing to search for it.

This special issue of the *Journal of Interpretation Research* features four articles by two of interpretation's most prominent scholars, Dr. Robert Powell of Clemson University and Dr. Marc Stern of Virginia Tech University. One of the articles is co-authored with their graduate students (Emily Martin and Jennifer Thomsen at Clemson, and Kevin McLean and Bethany Mutchler at Virginia Tech). The articles resulted from perhaps the most comprehensive examination of interpretation ever attempted—an empirical analysis of hundreds of face-to-face interpretive programs conducted by the US National Park Service at 24 sites. In my opinion, this research is unprecedented both in scope and in terms of the insights it has produced and the questions it raises.

The four articles examine the methods and approaches considered colloquially as "best practices," and they explore the correlates of these methods to program characteristics and ultimately to the outcomes of interpretation in the parks studied. The research reveals how achieving such outcomes might vary depending on an array of different contexts and program characteristics. And it elucidates not only how different approaches to interpretation perform in terms of promising mission-relevant

outcomes for the U.S. NPS, but also how interpreter characteristics such as charisma, enthusiasm, self-confidence, and sincerity figure into things. Stern, Powell, and their student co-authors also address the question of which factors distinguish interpretation that is merely adequate from interpretation that is outstanding. Their findings lead them to conclude that interpretation at its best can serve a purpose far beyond simple infotainment or entertaining fact giving for pleasure-seeking audiences. As I have long held (Ham, 2013), they call for interpretation that stimulates audiences to a "eudaimonic" state—one in which they are provoked to deep personal thought and to the making of personal connections with the place, and to the features and stories interpreted there. This is an important conclusion and it is consistent with substantiated theory from many areas of cognitive science.

For the reader in search of evidence related to interpretation's potential outcomes, there is plenty to consider in these articles. But there is also evidence telling us we still have a long way to go in terms of understanding the totality of interpretation with all its myriad influences—from contextual and setting factors to communication approach to audience factors, interpreter factors, and other factors research almost certainly hasn't yet identified. Using a robust quantitative procedure (structural equation modeling), the authors tested the ability of three different models (each involving a suite of 18 interpreter and program characteristics) to explain the variation in interpretive program success related to three oft-cited outcomes of interest (audience satisfaction, visitor experience and appreciation, and behavioral intentions of the audience). They find (and I must add here, *not surprisingly*) that even though some program and interpreter characteristics did emerge that were most predictive of program success, the models themselves were capable of explaining very little of the variation in the three outcomes examined. In fact, the analyses left somewhere between 73% and 90% of the actual variation in visitor-reported outcomes unexplained by the models.

Why would I interject *not surprisingly* in the sentence above? I did so simply because interpretation is complicated stuff. Here, I'm again reminded of the blogger demanding "hard facts" and implicitly chastising social science for not yet having produced them. However, when you objectively consider the sheer number of possible combinations between and among dozens of source factors, message factors, delivery-system factors, and audience factors thought to be involved in any act of interpretation, an inescapable conclusion is that a communication process like interpretation is mind-boggling in its complexity. Expectations that such a young area of applied social science should have, by now, captured all of this complexity seem quite a tall order. Indeed, as the earliest communication researchers in the 1950s unwittingly discovered (see Ajzen, 1992), comprehending the sheer number of possible combinations among all these factors probably isn't going to happen anytime soon.

This seems to me to be the same sort of scenario Powell and Stern have elucidated through their modeling study. And I believe it's an important (albeit humbling) premise for anyone who chooses to study the interpretation process. John Falk (2004: 85) put it well for museum researchers: "Theoretically, the total number of factors that directly and indirectly influence learning from museums number in the hundreds, if not thousands."

Communication between human beings is arguably one of the most complex phenomena in the universe—ultimately more complicated than atoms or quarks or astrophysics. And we are, of course, hampered by the inherent limitations of using science to study ourselves as a species. Add to this the fact that no current theory

captures the totality of human communication, even when the context is defined (e.g., interpretation in US national parks).

Stern and Powell have done a masterful job of systematically digging well below the surface of this complexity. Recognizing the potential shortcomings of relying solely upon visitors' own assessments, they also examine their own third party assessment of the overall quality of each program. Here, similar factors emerge that differentiate between what they distinguish as "good enough" and "great" interpretation, the latter associated with that eudaimonic state I mentioned above. But in this third-party analysis, the explanatory power is much stronger, differentiating good from great in 88% of the observed cases.

Although Stern and Powell rightly acknowledge the limitations of their study and that their list of best practices is incomplete, each of their analyses point toward a similar set of interpretive practices and delivery styles that appear to move visitors toward more meaningful experiences. By any measure, theirs is both a groundbreaking effort and an unprecedented foundation for future researchers to build on. How fortunate tomorrow's Ph.D. students are that these studies will be available to inform their thinking and research directions.

Powell and Stern conclude that "accounting for all factors seems a near impossibility." Yet, this is precisely why we do research. We know for every answer there will be more than one new question, and we know that in their quest for excellence there will always be the impatient professional who wants answers now. This is laudable. When the day arrives that interpreters feel they have all the answers, professional growth and maturation will no longer be possible, and sadly, even the blogging will stop. In the meantime, let the research and theory building continue.

Sam H. Ham
Emeritus Professor of Conservation Social Sciences
University of Idaho, USA

August T. Larsson Visiting Research Fellow
Environmental Communication Unit
Swedish Centre for Nature Interpretation
Swedish University of Agricultural Sciences

References

Ajzen, I. (1992). Persuasive communication theory in social psychology: A historical perspective. In Manfredo, M, (Ed.), *Influencing human behavior*. Champaign: Sagamore Publishing Co., Inc., 1–27.

Falk, J. (2004). The director's cut: Toward an improved understanding of learning from museums. Wiley InterScience, DOI 10.1002/sce.20014, retrieved September 12, 2013 from www.interscience.wiley.com.

Ham, S. (2013). *Interpretation: Making a difference on purpose*. Golden: Fulcrum Publishing.

Acknowledgment

I am writing this introduction from the Swedish Centre for Nature Interpretation, which is based in the Environmental Education Unit of the Swedish University of Agricultural Sciences in Uppsala. I want to thank my Swedish colleagues not only for allowing me the time to work on this important issue of *JIR*, but also for giving me the opportunity to work with some of the most brilliant interpretation practitioners and researchers in the world.

SPECIAL ISSUE

What Leads to Better Visitor Outcomes in Live Interpretation?

Marc J. Stern
Department of Forest Resources and Environmental Conservation, Virginia Tech

Robert B. Powell
Department of Parks, Recreation and Tourism Management and School of Agricultural and Forest Environmental Sciences, Clemson University

Abstract
We conducted a study to empirically isolate the factors that are most consistently linked with positive outcomes for the attendees of live interpretive programs. We examined the relationships between interpreter and program characteristics and three visitor outcomes—visitor satisfaction, visitor experience and appreciation, and intentions to change behaviors—across 376 programs in 24 units of the U.S. National Park Service. The analyses revealed a list of 15 characteristics associated with these outcomes across a wide range of program types and contexts. Some of these characteristics constituted commonly promoted practices in the interpretation literature (e.g., thematic communication, Tilden's principles, and appropriate organization). However, certain characteristics of the interpreter, in particular their confidence, passion, sincerity, and charisma, were also strongly correlated with positive visitor outcomes. We discuss the study's implications for both interpretive practice and future research.

Keywords
appreciation, behavior, evaluation, interpretation, National Park Service, research, satisfaction, visitor experience

Introduction
Live interpretive programs can serve multiple purposes (Ham, 2013). These include enhancing the experiences and the enjoyment of visitors to special places (Moscardo, 1999; Stern et al., 2011), increasing visitors' knowledge and understanding of natural and cultural resources and places (Ham, 1992; Tilden, 1957), fostering a sense of appreciation or other attitudes toward those resources (Powell et al., 2009), and promoting stewardship behaviors, both on-site and after visitors leave the site of the interpretation (Ham, 2009).

While volumes have been published outlining what might be considered best practices for producing such outcomes, a recent review of the empirical literature suggests that the linkage between these best practices and visitor outcomes have only circumstantial support, despite strong theoretical grounding (Skibins et al., 2012). This is largely due to a lack of comparative studies, which can empirically isolate which practices are the ones most likely causing desired outcomes. Most research studies have evaluated the outcomes of single programs rather than mixtures of programs with varying characteristics. While findings of positive outcomes across multiple studies suggest the broad efficacy of interpretation in general, no study has yet isolated the influence of different interpretive practices and approaches upon visitor outcomes.

This study aims to close this gap in the literature through a comparative study of live interpretive programs across the National Park Service (NPS), by identifying which practices and approaches most consistently lead to more positive outcomes, including visitor satisfaction, enhancement of visitor experience and appreciation of the park unit and its resources, and intentions to change behaviors resulting from program attendance.

Hypothesized best practices for interpretation

Skibins et al. (2012) identified consensus-based best practices of the field in a recent review article. Many of these practices stem from Freeman Tilden's (1957) original six principles first identified in 1957. The principles generally highlight the importance of making communication relevant to the audience; of telling holistic stories; of practicing the art of revelation based on information rather than information dissemination; of provoking the audience to want to do something, whether it be to reflect more deeply, learn more, or act upon new information; and of tailoring interpretation to different audiences. Many others have expanded upon those original best practices to provide insights into how to best craft stories; how to organize content; how to make interpretation relevant, engaging, and entertaining; and how to achieve particular outcomes (see Skibins et al., 2012, for a summary of this work). We drew upon this broad body of literature to develop many of the key program characteristics of interest in this study (see Table 3 for full list).

The role of the interpreter

In addition to characteristics of programs, the characteristics of the interpreters and their delivery styles also likely influence program outcomes. Passion on behalf of the interpreter, for example, has long been recognized as an important element of successful interpretive programs (e.g., Beck and Cable, 2002; Ham & Weiler 2002; Ward & Wilkinson, 2006). We supplement this concept with additional theories from education and communication to further explore the impact of the interpreter on visitor outcomes in addition to the content and format of the program.

The concepts of immediacy, credibility, and clarity have been studied extensively in the communications and education fields (Finn et al., 2009). Immediacy behaviors are those that tend to enhance the familiarity and reduce psychological distance between the communicator and his or her audience (Mehrabian, 1969). Such behaviors might include friendly physical gestures, small talk, calling people by name, or the sharing of personal information (Myers et al., 1998). These behaviors may also be related to "affinity-seeking," or the process through which communicators attempt to get listeners to like

them (McCroskey et al., 1986). Studies suggest that such behaviors can enhance the openness of audiences (most studies involve students and their teachers) to the content of lessons (Finn et al., 2009). Others have also assumed that general likeability may be an important factor in audience response (Ward & Wilkinson, 2006).

Credibility refers to audience members' perceptions of the believability or legitimacy of the communicator. Credibility has been found to be important in predicting the responses of message recipients in multiple fields (e.g., Ajzen 1992; Rogers 1995; Stern 2008). Within the education and communications fields, Finn and others (2009) suggest that this credibility is composed of three dimensions: competence, trustworthiness, and caring. Competence can be related to the apparent knowledge, confidence, and eloquence of the communicator. Trustworthiness can be based on multiple factors, including the interpreter's appearance, performance, degree of comfort and/or authority, title or position, and/or personal interactions with the audience. Caring is primarily related to the sincerity with which the interpreter communicates as well as his or her interactions with the audience.

Clarity is not only related to eloquence, but also to the consistency, or "fidelity" of the communicative experience (Chesebro & Wanzer, 2006). Finn and others' review (2009) found that lessons taught with any combination of these characteristics (clarity, credibility, and immediacy) tend to be more effective for learners than those exhibiting only one of them.

Interpreters also have the ability to assume particular roles as communicators. These range from friend to authority figure to the "walking encyclopedia" that Enos Mills warned future nature guides against becoming nearly 100 years ago (Mills, 1920). Each of these identities may be differentially appropriate in different situations and with different audiences (Wallace & Gaudry, 2005). Other items of interest include any apparent bias, misinformation, or false assumptions about the audience made by the interpreter, which could detrimentally influence audience responses.

Interpreters' planning processes and psychological states might also influence the quality of their programs (see Stern at al., this issue). As noted above, interpretation can be used for many purposes, ranging from teaching to entertainment to persuasion. Interpreters' intentions may drive, at least to some extent, audience responses to their programs (Ham, 2013).

Methods

Selection of sites

We aimed to select park units that reflected the diversity of locations, types, and resources of the U.S. NPS system. Criteria for selecting park units for the study included annual visitation numbers, park location (region of the country and distance from population centers), programming focus, number of programs offered to the public, and willingness to participate in the study. In order to ensure adequate visitor attendance at interpretive programs, we only considered parks that received at least 35,000 annual recreation visits. Parks were categorized as urban, urban-proximate, or remote based on their proximity to metropolitan centers. Metropolitan areas were defined as having an urban core of at least 50,000 residents. Urban parks were located within the limits of these metropolitan areas. Urban-proximate parks were located outside these cores, but within a 60-mile radius of these areas. As such, they were typically in rural or

Table 1. Park units included in the study.

Park Unit	Resource Focus	Park Location	Annual Recreation Visits[a]
Aztec Ruins National Monument	Cultural	Remote	37,437
Badlands National Park	Natural	Remote	977,778
Bryce Canyon National Park	Natural	Remote	1,285,492
Chaco Culture National Historical Park	Cultural	Remote	34,226
Ford's Theater National Historic Site	Cultural	Urban	662,298
Fort McHenry National Monument and Historic Shrine	Cultural	Urban	611,582
Gettysburg National Military Park	Cultural	Urban-Proximate	1,031,554
Grand Canyon National Park	Natural	Remote	4,388,386
Great Smoky Mountains National Park	Mix	Urban-Proximate	9,463,538
Harpers Ferry National Historical Park	Cultural	Urban-Proximate	268,822
Independence National Historical Park	Cultural	Urban	3,751,007
Jefferson National Expansion Memorial	Cultural	Urban	2,436,110
Jewel Cave National Monument	Natural	Remote	103,462
Lincoln Home National Historic Site	Cultural	Urban	354,125
Manassas National Battlefield Park	Cultural	Urban-Proximate	612,490
Mesa Verde National Park	Mix	Remote	559,712
Mount Rushmore National Memorial	Cultural	Remote	2,331,237
National Mall	Cultural	Urban	1,363,389
Navajo National Monument	Mix	Remote	90,696
Point Reyes National Seashore	Natural	Urban-Proximate	2,067,271
San Francisco Maritime National Historical Park	Cultural	Urban	4,130,970
Ulysses S. Grant National Historic Site	Cultural	Urban	39,967
Wind Cave National Park	Natural	Remote	577,141
Yosemite National Park	Natural	Remote	3,901,408

[a] Annual visitation from 2010 (http://www.nature.nps.gov/stats/)

suburban areas. Remote parks were located at least 60 miles from any metropolitan area. Parks were placed into one of three categories based on their primary resource base: predominantly cultural, predominantly natural, or a mix of the two. We aimed to have our selection of units mirror the makeup of the NPS system and also allow us to observe at least 10 programs in each park (or within nearby clusters of parks in cases such as Aztec Ruins and Navajo National Monuments) in five days or less. Twenty-four park units were selected for inclusion in the study (Table 1).

We observed programs in 14 predominantly culturally focused park units, seven predominantly nature-focused park units, and three park units with a mixed focus. This roughly mirrors the distribution of these different types of park units throughout the NPS, where roughly 30% of park units are predominantly nature-focused and roughly 60% are predominantly culturally focused.[1] We visited 11 remote park units, five urban-proximate parks, and eight urban park units. This variability provides a

Table 2. Programs observed and total number of surveys collected.

Park unit	Programs attempted	Programs observed	Surveys collected	Used in analyses	
				Programs	Surveys
Aztec Ruins National Monument	4	2	4	2	4
Badlands National Park	22	19	157	14	118
Bryce Canyon National Park	12	12	133	12	127
Chaco Culture National Historical Park	9	8	85	7	70
Ford's Theater National Historic Site	20	20	519	18	448
Fort McHenry National Monument and Historic Shrine	23	14	133	11	113
Gettysburg National Military Park	26	21	206	18	186
Grand Canyon National Park	30	30	384	28	363
Great Smoky Mountains National Park	19	14	96	12	86
Harpers Ferry National Historical Park	21	15	100	12	79
Independence National Historical Park	36	22	156	17	122
Jefferson National Expansion Memorial	22	16	146	14	135
Jewel Cave National Monument	20	20	190	18	177
Lincoln Home National Historic Site	18	14	89	10	72
Manassas National Battlefield Park	20	17	88	15	80
Mesa Verde National Park	14	14	301	14	290
Mount Rushmore National Memorial	23	19	171	9	101
National Mall	47	22	65	16	49
Navajo National Monument	8	3	23	3	23
Point Reyes National Seashore	12	9	34	8	32
San Francisco Maritime National Historical Park	20	16	69	14	64
Ulysses S. Grant National Historic Site	15	9	40	8	36
Wind Cave National Park	18	18	215	13	175
Yosemite National Park	29	22	199	19	172
Totals	**488**	**376**	**3,603**	**312**	**3,122**

reasonable sample from which to make generalizations to the broader population of live interpretive programs across the NPS. Park units were organized for logistical purposes by geographic region into six clusters. Teams of two researchers collected data from each park unit. One team of researchers sampled Great Smoky Mountains National Park and the mid-Atlantic, Washington D.C., and California locations. The other team sampled the Southwest, Midwest, and South Dakota locations.

Sampling and data collection

Individual live interpretive programs served as the unit of analysis for this study. Programs were selected within each park based on variability (with regard to subject matter—natural vs. cultural—and types of delivery—guided walks vs. campfire programs vs. hands-on activities, etc.) and their time and location to maximize the number of programs observed at each park unit. Regular programs were selected over children's programs whenever possible, as adult respondents were the targets of visitor surveys. We attempted to attend 488 scheduled programs, of which only 376 occurred. From these 376 programs, we collected 3,603 surveys from visitors (Table 2). Data from 312 programs were used in the analyses contained within this paper (see "*Interpretive program sample development and data cleaning*" below for more detail).

Throughout the research, the same procedure was followed for observing all

programs. Upon arrival at the program site, a brief interview was conducted with the interpreter. Interview questions included interpreters' intended programmatic outcomes, questions about program development, and others about the preparation and the level of enthusiasm of the interpreter. The interviews also collected basic background information about the interpreter, which included age, gender, and interpretation experience. These interviews were conducted on all but 15 programs. In those cases, time did not allow for the interviews to take place. Basic information about the program itself was recorded by the observer, including time, location, type, topic focus, and size and age breakdown of the audience.

At the end of the program we asked visitors over the age of 15 to complete a short survey regarding their opinions of the program and its influence on them. For programs with fewer than 50 participants, we attempted a census of all eligible attendees. In programs that were particularly large (more than 50 attendees), the researchers employed systematic sampling whenever possible—for example, selecting every nth row to complete surveys at Ford's Theatre. In these cases, the researchers chose the sample interval in attempt to target at least 20 respondents.

During each program, researchers maintained an unobtrusive presence within the group, acting simply as another member of the audience. The researchers completed observation sheets during and immediately following each program.

Throughout the duration of all field work, researchers would periodically attend programs together to ensure reliability and consistency in scoring each variable. Occasional check-ins were also completed between team members to ensure that observation techniques were consistent, to clarify questions about scoring certain variables, and to add variables that were deemed relevant to the research. No new variables were added after the first week of fieldwork.

Measurement

Dependent variables: outcomes

The dependent variables in the study were composed of retrospective assessments provided by program attendees on surveys administered immediately following their programs.[2] While interpretation may produce multiple outcomes, we focused primarily on visitor satisfaction and shifts in knowledge, attitudes, and behavioral intentions relevant to the park experience.

Overall satisfaction with the program was measured on a scale from 0 to 10, with 0=Terrible and 10=Excellent. An additional battery of survey items provided response prompts for the following question: "To what degree did the program you just attended influence any of the following for you?" Response categories were composed of a five-point Likert-type scale, with answer choices: Not at all (1), A little (2), Somewhat (3), A moderate amount (4), and A great deal (5). The survey items included:

- Made me think deeply
- Made me reflect on my own life
- Enhanced my appreciation for this park
- Enhanced my appreciation for the National Park Service
- Made me more likely to avoid harming park resources

- Increased my knowledge about the program's topic
- Made my visit to this park more enjoyable
- Made my visit to this park more meaningful
- Changed the way I will behave while I'm in this park
- Changed the way I will behave after I leave this park
- Made me want to tell others about what I learned
- Made me care more about this park's resources
- Made me care more about protecting places like this

These items were developed based on key literature (e.g., Ham, 1992; Moscardo, 1957; Tilden, 1957; Ward & Wilkinson, 2006) and extensive input from NPS staff. This input included interviews and focus groups with the NPS National Education Council; a focus group and associated surveys conducted with NPS interpreters at the National Association for Interpretation (NAI) National Workshop in Las Vegas, November 2010; and two surveys conducted in 2010 and 2011 with NPS superintendents and supervisors of interpretation, respectively (see Stern & Powell, 2011). The resulting responses were analyzed to reduce the items into fewer latent factors reflecting the key outcomes of programs for visitors (see *Results* section).

Independent variables: predictors

Our primary independent, or *predictor*, variables of interest included both interpreter characteristics and the interpretive practices employed during a program. These practices were primarily drawn from an extensive literature review aimed at identifying best practices in the field (Skibins et al., 2012) as well as characteristics identified by interpretive experts within the NPS and ranked highly by interpretive staff in surveys (Stern and Powell, 2011). Additional items emerged as potentially important in pilot tests (e.g., consistency of tone and quality throughout a program) and were also measured.

Program characteristics were based in theory found in key texts within the interpretation literature (Table 3). A subset of these characteristics, however, were based primarily within the field of social psychology and relate to programs that explicitly aim to influence the behavior of participants. In short, the Theory of Planned Behavior (Ajzen, 1991) suggests that people base their behaviors upon three types of evaluations they make about the likely outcomes of performing that behavior: the benefits vs. the costs of the expected outcomes of the behavior (behavioral beliefs), what they perceive their peers might think about the behavior (normative beliefs), and the degree of control and/or ability they feel with regard to carrying out the behavior (control beliefs). We translated the theory into observable characteristics that would theoretically address these evaluations (see "Behavioral theory elements," Table 3).

Interpreter characteristics, meanwhile, focused upon the appearance, identity, and overall styles of the interpreters themselves, drawn largely from the communications and education literature, though many of these factors are also referenced in the interpretation literature (Table 4). Citations are provided where characteristics were drawn from the literature. Additional insights and examples can be found in a companion article in this same issue (Stern et al., this issue).

Table 3. Program characteristics observed in the study, their definitions, and operationalization.

Program characteristic	Definition	Scoring
Introduction quality (Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999)	Degree to which the introduction captured the audience's attention and oriented (or pre-disposed) the audience to the program's content and/or message.	3= Oriented audience and captured attention 2= Minimally oriented audience; did not necessarily capture attention 1= Poorly executed
Appropriate logistics (Jacobson, 1999; Knudson et al., 2003)	Degree to which basic audience and program needs were met (i.e., restrooms, weather, technology, accessibility, shade, etc.).	4= Well planned and appropriate 3= Audience/program needs mostly addressed 2= Needs marginally addressed 1= Needs not met
Appropriate for audience (Beck and Cable, 2002; Jacobson, 1999; Knudson et al., 2003)	Degree to which the program aligned with audience's ages, cultures, and level of knowledge, interest, and experience.	5= Very appropriate 4= Appropriate 3= Moderately appropriate 2= Only slightly appropriate 1= Not appropriate
Appropriate sequence (Beck and Cable, 2002; Ham, 1992; Jacobson, 1999; Larsen, 2003)	Degree to which the program followed a logical sequence.	4= Enhanced messaging 3= Appropriate 2= Choppy 1= Detracted from messaging
Transitions (Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999; Larsen, 2003)	Degree to which program used appropriate transitions that kept the audience engaged and did not detract from the program's sequence.	4= Enhanced messaging and were smooth 3= Appropriate 2= Forced or irrelevant 1= Detracted from messaging or not present
Links to intangible meanings and universal concepts (NPS Module 101; Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Knudson, et al., 2003; Larsen, 2003; Lewis, 2005; Moscardo, 1999; Tilden, 1957; Ward and Wilkinson, 2006)	Communication connected tangible resources to intangible meanings and universal concepts. Intangibles: stories, ideas, meanings, or significance that tangible resources represent Universals: concepts that most audience members may identify with	5= Extensively developed; powerful concepts 4= Well developed 3= Present but weak 2= Difficult to detect or slightly used 1= Clearly not present
Multisensory (Beck and Cable, 2002; Knudson et al., 2003; Lewis, 2005; Moscardo, 1999; Tilden, 1957; Veverka, 1998; Ward and Wilkinson, 2006)	Degree to which the program intentionally and actively engaged more than just basic sight and sound.	3= Explicit/purposeful inclusion of two sense beyond sight and sound 2= Actively incorporated a sense beyond passive use of sight and sound, or actively focused upon either of these senses as a vehicle for conveying the message (e.g., "close your eyes and listen") 1= Primarily a talk in which the ranger did not explicitly use multiple sense beyond passive use of sight (scenery/objects) and sound (words)
Physical engagement (Beck and Cable, 2002; Knudson, et al., 2003; Lewis, 2005; Moscardo, 1999; NPS Module 101; Sharpe, 1976; Tilden, 1957)	Degree to which the program physically engaged audience members in a participatory experience; i.e., through touching or interacting with resource.	4= Central programming element 3= Occurred multiple times 2= Minimal effort to engage 1= No efforts
Verbal engagement (Knudson, et al., 2003; Moscardo, 1999; Sharpe, 1976; Tilden, 1957; Veverka, 1998)	Degree to which the program verbally engaged audience members in a participatory experience; i.e., dialogue (a two-way discussion).	5= Central programming element 4= Occurred multiple times 3= Modestly engaged 2= Minimal effort to engage 1= No efforts
Cognitive engagement (Knudson, et al., 2003; Moscardo, 1999; Sharpe, 1976; Tilden, 1957; Veverka, 1998)	Degree to which the program cognitively engaged audience members in a participatory experience beyond simply listening; i.e. calls to imagine something, reflect, etc.	5= Central programming element 4= Occurred multiple times 3= Modestly engaged 2= Minimal effort to engage 1= No efforts
Multiple activities (Knapp and Benton, 2004; Moscardo, 1999; Ward and Wilkinson, 2006)	Degree to which the program consisted of a variety of activities and opportunities for direct audience involvement (not including dialogue).	4= 2+ primary activities included 3= 2+ secondary activities included 2= One secondary activity included 1= One activity only
Props (Jacobson, 1999; Knapp and Benton, 2005; Ham, 1992; Ward and Wilkinson, 2006)	A visual aide beyond a screen-based slideshow.	1 = Prop(s) used 0 = Not used

Relevance to audience (Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999; Knapp and Benton, 2004; Lewis, 2005; Moscardo, 1999; NPS Module 101; Sharpe, 1976; Tilden, 1957; Veverka, 1998)	Degree to which the program explicitly communicated the relevance of the subject to the lives of the audience.	5= Major focus of messaging 4= Well developed efforts 3= Moderate efforts 2= Minimal efforts 1= No efforts
Affective messaging (Jacobson, 1999; Lewis, 2005; Madin and Fenton, 2004; Tilden, 1957; Ward and Wilkinson, 2006)	Degree to which the program communicated emotion (in terms of quantity, not quality).	5= Central programming element 4= Frequent and repeated messages 3= Occasional messages 2= Minimal effort to include messages 1= Messages absent
Fact-based messaging (Frauman and Norman, 2003; Jacobson, 1999; Lewis, 2005; Tilden, 1957; Ward and Wilkinson, 2006)	Degree to which the program communicated factual information.	1 = Messaging was solely fact-based 0 = Messaging was not solely fact-based (incorporated affective messaging)
Surprise (Beck and Cable, 2002; Moscardo, 1999)	Degree to which the program used the element of surprise in communication. This could include "aha" moments or unexpected or contrasting messages.	3= Major element 2= Minor element 1= Not used
Novelty (Beck and Cable, 2002; Frauman and Norman, 2003; Knapp and Benton, 2004; Moscardo, 1999)	Degree to which the program presented novel ideas, techniques, or viewpoints as an element of communication; i.e., using a device not usually associated with or related to resource.	3= Major element 2= Minor element 1= Not used
Provocation (Beck and Cable, 2002; Brochu and Merriman, 2002; Knudson, et al., 2003; Tilden, 1957)	Degree to which the program explicitly provoked participants to personally reflect on content and its deeper meanings.	4= Powerful and explicit inclusion 3= Occasional inclusion 2= Isolated or vague inclusion 1= No attempt made
Multiple viewpoints (Beck and Cable, 2002; Brochu and Merriman, 2002; Tilden, 1957)	Degree to which the program explicitly acknowledged multiple perspectives or uncertainty within a theme or message. (Primarily for controversial messaging; when an argument is made, was a relevant counter-argument provided?)	3= Multiple viewpoints developed; none given clear priority 2= Primarily one viewpoint, with some focus on others 1= No effort NA = not applicable
Holistic storytelling (Beck and Cable, 2002; Larsen, 2003; Tilden, 1957)	Degree to which the program aimed to present a holistic story (with characters and a plot) as opposed to disconnected pieces of information.	5= Holistic story used throughout; all messaging tied to story 4= Holistic story present; some info did not relate to story 3= Equal mix of storytelling and factual information, no single, holistic story 2= Factual information primarily used; some stories used to create relevance. 1= Facts and information primarily; no attempt at storytelling.
Place-based messaging (Beck & Cable, 2002; Knudson, et al., 2003; Lewis, 2005; Moscardo, 1999; NPS Module 101; Sharpe, 1976)	Degree to which the program emphasized the connection between the visitor and the site/resource.	5= Central focus of messaging 4= Well-developed connection through repetition and engagement 3= Moderately emphasized through repetition or engagement 2= Slightly developed verbally 1= Not developed
Introduction and conclusion linkage (Beck and Cable, 2002; Brochu and Merriman, 2002; Larsen, 2003)	Degree to which program connected conclusion back to the introduction in an organized or cohesive way (i.e., program "came full circle.")	4= Intro and conclusion were linked in a cohesive way that enhanced messaging 3= Intro and conclusion were linked, but didn't necessarily enhance messaging 2= Intro and conclusion were weakly linked 1= Intro and conclusion were disconnected from each other

Clear theme (Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999; Knudson, Cable, and Beck, 2003; Larsen, 2003; Lewis, 2005; Moscardo, 1999; Sharpe, 1976; Veverka, 1998; Ward and Wilkinson, 2006)	Degree to which the program had a clearly communicated theme(s). A theme is defined as a single sentence (not necessarily explicitly stated) that links tangibles, intangibles, and universals to organize and develop ideas.	4= Theme is clearly developed and communicated 3= Easy to detect, but not well developed 2= Difficult to detect, present but at least somewhat ambiguous 1= Unclear/not present
Central message (Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999)	Degree to which program's message(s) was clearly communicated; i.e., the "so what?" element of the program.	4= Clearly communicated and well developed 3= Easy to detect, but not well developed 2= Difficult to detect, ambiguous 1= Unclear/not present
Consistency (Beck and Cable, 2002; Ham, 1992)	Degree to which the program's tone and quality were consistent throughout the program	3=Consistent 2=Some shift in either tone or quality during the program 1= Shift in both tone and quality
Pace (Jacobson, 1999)	Degree to which the pace of the program allowed for clarity and did not detract from the program.	Categorical: Too fast Too slow Just fine
Quality of the resource	Degree to which the resource where program took place is awe-inspiring or particularly iconic.	3= Contextually iconic or grandiose 2= Pleasant but not iconic 1= Unimpressive/generic
Unexpected negative circumstance	Were there any unexpected interruptions or emergencies during the program, such as a sudden change in weather, medical emergency, technical difficulties, or hazardous conditions that detracted from the quality of the program?	1 = Yes 0 = No
Unexpected positive circumstance	Was there an unexpected experience that occurred during the program, such as seeing charismatic wildlife or other unique phenomena that added significantly to the quality of the experience?	1 = Yes 0 = No
Behavioral theory elements The following were only measured for programs in which a behavioral change was expressed by the interpreter as a desired program outcome.		
Benefits of action (Ajzen, 1991; Ham et. al., 2007; Jacobson, 1999; Knudson, et al., 2003; Moscardo, 1999; Peake et. al, 2009)	Degree to which the program emphasized the potential benefits resulting from performing a particular action(s).	4= Explicitly/purposefully emphasized 3= Mentioned a moderate amount 2= Explained a little 1= No mention NA = not applicable
Costs of action (Ajzen, 1991; Ham et. al., 2007; Jacobson, 1999; Knudson, et al., 2003; Moscardo, 1999; Peake et. al, 2009)	Degree to which the program emphasized the potential costs resulting from performing a particular action(s).	4= Explicitly/purposefully emphasized 3= Mentioned a moderate amount 2= Explained a little 1= No mention NA
Norms of action (Ajzen, 1991; Ham et. al., 2007; Jacobson, 1999; Knudson, et al., 2003; Moscardo, 1999)	Degree to which the program emphasized the social acceptability of performing a particular behavior or desired action.	4= Explicitly/purposefully emphasized 3= Mentioned a moderate amount 2= Explained a little 1= No mention NA
Ease of action (Ajzen, 1991; Ham et. al., 2007; Jacobson, 1999; Knudson, et al., 2003; Moscardo, 1999; Tilden, 1957)	Degree to which the program communicated the ease (or difficulty) of performing a particular behavior or desired action.	4= Explicitly/purposefully emphasized 3= Mentioned a moderate amount 2= Explained a little 1= No mention NA
Demonstrates action (Ajzen, 1991; Beck and Cable, 2002; Knudson, et al., 2003; Moscardo, 1999; Sharpe, 1976; Widner Ward and Wilkinson, 2006)	Degree to which the program provided examples of, or opportunities for, performing a desired action.	4= Majority of audience engaged 3= Demonstration by ranger or small proportion of audience 2= Verbal description 1= No mention/demonstration NA

Table 4. Interpreter characteristics observed in the study, their definitions, and operationalization.

Interpreter characteristic	Definition	Scoring
Professional appearance	The extent to which the interpreter appears properly dressed and groomed.	0 = Interpreter appears disheveled or unkempt and is not professionally dressed 1 = Interpreter appears well-groomed and is professionally dressed
Comfort of the interpreter (Lewis 2008; Moscardo, 1999; Ward and Wilkinson, 2006)	Degree to which the interpreter presenting the program seems comfortable with the audience and capable of successfully presenting the program without apparent signs of nervousness or self-doubt.	1 = Interpreter seems scared, nervous, or unable to lead the program 2 = Interpreter seems nervous and struggles with much of the program 3 = Interpreter seems comfortable, but might become uncomfortable at times 4 = Interpreter is not nervous and handles the program with ease
Responsiveness (Jacobson, 1999; Knudson et al., 2003; Lewis, 2008)	The extent to which the interpreter interacts with the audience, collects information about their interests and backgrounds, and responds to their specific questions and requests or non-verbal cues.	NA = Not able to observe (e.g., large programs in dark theatres) 1 = Interpreter is aloof or averse to the visitors' presence 2 = Interpreter is somewhat responsive to visitors' questions/body language 3 = Interpreter was very responsive to the audience
Inequity (Ham and Weiler, 2002)	The presence of unequal attention devoted to certain attendees and not others through greater interaction or attentiveness.	1 = Interpreter did not pay equal attention to all audience members. 0 = No inequity issues.
Humor quality (Ham and Weiler, 2002; Knapp and Yang, 2002; Regnier et al., 1992)	How funny is the interpreter overall? Does the audience react positively to the interpreter's use of humor and seem to enjoy it?	1 = Not funny at all 2 = A little funny 3 = Moderately funny 4 = Hilarious
Humor quantity	The extent to which the interpreter attempts to use humor, sarcasm, or jokes to share the topic with the visitor, regardless of their success.	1 = Interpreter attempts no humor throughout the presentation 2 = Interpreter rarely uses humor 3 = Interpreter uses an equal mix of humor and non-humor to convey the message 4 = Interpreter is mostly trying to be humorous 5 = Interpreter uses humor as the primary vehicle to convey their message
Sarcasm	The degree to which the interpreter used sarcasm (the use of mocking, contemptuous, or ironic language or tone) or self-deprecation that was not meant to be serious, as a part of presenting their program.	1 = Not at all 2 = Done to some extent 3 = A central feature of the delivery style
Charisma (Ward and Wilkinson, 2006)	A general sense of the overall likeability/charisma of the interpreter, commonly recognized by seemingly genuine interaction with the visitors, including smiling, looking people in the eye, and having an overall appealing presence.	1 = Not likeable/found interpreter irritating 2 = Somewhat off-putting 3 = Neither liked or disliked interpreter 4 = More or less liked interpreter 5= Found interpreter very likeable/charismatic
Sincerity (Ham, 2009)	The degree to which the interpreter seems genuinely invested in the messages he or she is communicating, as opposed to reciting information, and seems sincere in the emotional connection they may exude to the message and/or the resource. In other words, the extent to which the interpretation was delivered through authentic emotive communication.	1 = Interpreter seemed to only be going through the motions, with no real emotional connection or sincerity 2 = Interpreter seemed somewhat connected through the words they used, though their mannerisms or intonation didn't corroborate their words. 3 = Interpreter seemed mostly sincere with authentic emotive communication for most of the program 4 = Communication was clearly sincere and authentic throughout the program, as evidenced by words, gestures, intonation, or other mannerisms
Passion (Beck and Cable, 2002; Ham and Weiler, 2002; Moscardo, 1999)	The interpreter's apparent level of enthusiasm for the material, as opposed to a bored or apathetic attitude toward it. The overall vigor with which the material is presented.	1 = Interpreter seems completely detached/disinterested from the program 2 = Low levels of passions 3 = Interpreter shows moderate levels or sporadic instances of passion 4 = Pretty high levels of passion overall 5 = Interpreter seems extremely passionate about the program

Personal sharing (Jacobson, 1999; Myers et al., 1998)	The degree to which the interpreter shared personal insights or experiences, answered questions about themselves for the audience, or provided their own opinion on topics or events relevant to the program.	1 = Interpreter did not share any personal information about themselves with the audience 2 = Interpreter shared minimal personal information or viewpoints 3 = Interpreter shared a large amount of personal information and perspective 4 = Interpreter's personal life/point of view is explicitly the central focus of the experience (used themselves as the primary framework for the program)
Apparent knowledge (Ham and Weiler, 2002; Lewis, 2008; Ward and Wilkinson, 2006)	The degree to which the interpreter appears to know the information involved in the program, the answers to visitors questions, and has local knowledge of the area and its resources.	1 = Interpreter seems not at all knowledgeable (unsure of facts or has a hard time recalling the information intended for the program) 2 = Interpreter seems somewhat knowledgeable, but appears to forget a few things or leave out important details 3 = Interpreter appears more or less knowledgeable without any major hiccups or uncertainty throughout the program. 4 = Interpreter's presentation of facts and information during the program is flawless
Audibility	The extent to which the interpreter can clearly be heard and understood by the audience.	1 = Interpreter could not be heard by the audience during the majority of the program 2 = Interpreter could be clearly heard for the majority of the program, but wasn't audible during some parts 3 = Interpreter could be clearly heard throughout the entire program
Eloquence (Lewis, 2008)	The extent to which the interpreter spoke clearly and articulately, and did not mumble or frequently use filler words such as "um" or "like."	1 = Interpreter stumbled on their speech throughout their entire program and was hard to understand 2 = Interpreter had some minor issues with mumbling or unclear speech 3 = Interpreter had no such issues during the program 4 = Interpreter was exceptionally eloquent
Impatience	Did the interpreter show any explicit impatience toward audience members?	1 = Interpreter was explicitly impatient with the audience 0 = No issues noted
Formality	The degree to which the interpreter was very formal and official vs. casual and laid back about the presentation.	1 = Interpreter was extremely casual 2 = More casual than formal 3 = Interpreter was neither explicitly casual nor formal 4 = More formal than casual 5 = Interpreter was entirely formal
False assumption of the audience	At any point during the program, did the interpreter make assumptions of the audience's attitudes or knowledge that could have easily been false?	1 = No problem with false assumptions 2 = Some minor false assumptions that likely did not detract from the quality of the program 3 = Obvious false assumptions that made the experience less enjoyable or meaningful
Character acting	The degree to which role playing or character acting is incorporated into the program, either to add authenticity or to help tell a story.	0 = Interpreter does no character role playing during the program, he/she is simply leading the program 1 = Interpreter acts like one or more characters during parts of the program 2 = Interpreter is in full costume or does not break character at any point during the program
Primary identity (Ham and Weiler, 2002; Ham, 2002; Knapp and Yang, 2002; Larsen, 2003; Mills, 1920; Wallace and Gaudry, 2005)	**Friend:** outwardly friendly, casual, approachable, mingles informally	1 = primary identity; 0 = not
	Authority figure: emphasizes own role as a park ranger and focuses on rules, regulations, and/or authority to communicate	1 = primary identity; 0 = not
	Walking encyclopedia: Focused on conveying a large volume of facts	1 = primary identity; 0 = not
Questionable information	Obvious factual inaccuracy (incorrect or inaccurate information) or false attribution (unfounded claims about others, e.g., "the native people were happy to hand over their land so a National Park could be formed.")	1 = present 0 = not present
Bias	Did the interpreter share any apparent bias or strong opinion with potential effects on relationships with audience members?	1 = yes 0 = no

We also collected details pertaining to the experience level and demographics of the interpreter, their intended outcomes for their programs, and their level of excitement about the particular program they were about to deliver. In addition, we tracked information on the context for the program including location (e.g., indoors vs. outdoors), type of program, its focus (natural vs. cultural/historical vs. both), and other unexpected circumstances that could impact program outcomes (e.g., weather). In addition, we estimated the number of attendees at each program and the ratio of youth (ages 15 and under) to adults. Each of these contextual variables is examined in another article within this issue (Powell and Stern, this issue).

Pilot testing

Extensive pilot testing aided instrument development and refinement and enhanced the reliability of measurement across the research team. Prior to the field research, we observed video-recorded interpretive programs from an undergraduate interpretation class. These programs were used to develop consistent measurement of each relevant characteristic. Programs were viewed repeatedly and scores were compared among team members on each characteristic. These exercises were also used to refine the scoring of several variables.

From this testing, a preliminary assessment sheet was developed. These assessment sheets were further pilot tested at Great Smoky Mountains National Park in May of 2011, where the research team observed three live interpretive programs. Extensive discussion allowed us to further refine definitions and observation techniques for each of the characteristics under study. For each measure, we aimed to maximize the number of points in each scale to differentiate practices/attributes and enhance variability in the findings. However, existing definitions from the literature and results of pilot-testing limited most scales to four or fewer points. Pilot testing revealed that the middle-points on larger scales for many variables were not easily differentiated in a consistent manner by the research team. As a result, the scoring for each item varies to maximize the potential range of scores while maintaining inter-rater reliability. Binary scores were used in cases where the most appropriate measure was to indicate presence or absence.

Reliability and calibration

We built a calibration phase into the research design to ensure that each researcher's scores of each observed characteristic were consistent and reliable and therefore could be interpreted similarly. This involved three steps. First, immediately upon the completion of the field research and data entry, we carefully examined differences in the average scores of each variable between each member of the research team using a one-way ANOVA with posthoc tests. We identified all statistically significant differences between the mean scores for observations by different members of the research team. Second, through detailed examination of field notes and group discussions, we determined whether any of these differences might be attributed to systematic differences in observation techniques as opposed to differences in the unique sets of programs observed by each researcher. Two types of systematic differences emerged. In the first case, one researcher was systematically higher or lower than the other three on a particular measurement scale. In these cases, scoring procedures were reviewed, consensus definitions were refined, and that one researcher re-coded the variable based on these definitions and their qualitative program notes. Variables that were re-coded in

this manner included comfort of the interpreter, passion, apparent knowledge, sincerity, provocation, holistic story, and appropriateness for the audience. In the second case, a researcher had misinterpreted the response scale (scoring values) of the variable being coded. Again, a consensus definition was clarified and re-coding of that variable took place. These variables included cognitive engagement, clear theme, and central message. In one case, a variable was removed due to inconsistent interpretation of its definition in the field: place-based messaging.

Data entry and cleaning

Post-program surveys and program audits were coded and entered into Microsoft Access Database and Microsoft Excel to facilitate data entry. Data were then transferred to SPSS for screening and analysis. The visitor survey data were first screened for missing values and any surveys missing more than 50% of the items per factor were removed. A total of 118 respondents were removed as a result. Data were then screened for univariate and multivariate outliers on outcome variables following Tabachnick and Fidell (2007) using Mahalanobis Distance (MAH) and studentized deleted residuals (SDRESID). A total of 58 cases were removed for exceeding +/- 3 standard deviations, or the criterion Mahalanobis Distance value. This reduced our sample to 3,427 individual surveys from 376 interpretive programs.

Interpretive program sample development and data cleaning

Because the interpretive program is the unit of analysis in this study, we aggregated individual data at the program level by calculating the mean score of each visitor outcome for each program. To do so, we first needed to determine how many completed surveys within a particular program would serve as a viable reflection of the quality of that program and its impacts on visitors. Prior research suggests that programs with particularly small numbers of attendees may be inherently different than programs with larger numbers of attendees (Forist, 2003; McManus, 1987, 1988; Moscardo, 1999). In particular, programs with fewer than five attendees may have a high likelihood of serving only a single cohesive group (e.g., a single family). Meanwhile, programs with five or more have a higher likelihood of being composed of multiple groups. Moreover, a greater number of survey responses enhances the reliability of the research findings. Based on this rationale, we separated programs with fewer than five attendees from those with five or more attendees, and analyzed them separately.

For groups with five or more attendees, we included in the analysis all programs with 10 or more respondents to the surveys. We only included those programs with fewer than 10 respondents if the number of respondents represented at least half of the eligible respondents at the program (those over the age of 15). This yielded a total of 272 programs with five or more attendees for analysis.

For programs with fewer than five attendees (n=45), we only included those in which all eligible respondents (those over the age of 15) completed a survey. If a census was not achieved, the program was dropped from further analysis. This resulted in the removal of five of these smaller programs, leaving 40 in the sample for further analysis.

Table 5. Outcome indexes developed through confirmatory factor analyses.

OUTCOME INDEXES
Program outcome: Visitor Experience and Appreciation (Cronbach's α = 0.89) To what extent did the program you just attended influence any of the following for you? • Made my visit to this park more enjoyable • Made my visit to this park more meaningful • Enhanced my appreciation for this park • Increased my knowledge about the program's topic • Enhanced my appreciation for the National Park Service
Program outcome: Behavioral intentions (Cronbach's α = 0.94) To what extent did the program you just attended influence any of the following for you? • Changed the way I will behave while I'm in this park • Changed the way I will behave after I leave this park

Results

Index development: Dependent variables

Before conducting further analyses, we conducted exploratory and confirmatory factor analyses to explore the relationships between items and form factors made up of multiple items that represent a concept. The items that vary together as part of a factor can be combined to create scales or composite indexes that represent coherent concepts for use in subsequent analyses (DeVellis, 2003). Following procedures outlined by DeVellis (2003) we conducted exploratory and confirmatory factor analysis on dependent variables using the individual respondent data. Exploratory factor analyses and reliability analyses revealed the presence of two latent factors. Confirmatory factor analysis (CFA), which is a form of structural equation modeling, further refined the structure of these two factors. The resulting CFA model confirmed two factors while also providing a more parsimonious solution. Model fit statistics were all within the acceptable range (S-B χ^2=338.41; CFI=.96; RMSEA=.08). We labeled the resulting factors *Visitor Experience and Appreciation* and *Behavioral Intentions* (Table 5).

These factors form two of the three outcomes employed in this study. The first factor reflects an overall assessment of the impact of the program on the individual's experience, attitudes, and knowledge. Taken as a whole, it may be the best reflection of the first two elements of the classic statement from an old NPS manual quoted by Tilden (1957), "Through interpretation, understanding; through understanding, appreciation; through appreciation, protection." The *Behavioral intentions* factor relates to the third part of the classic quote, actually influencing the behavior of visitors in some way. The third outcome, satisfaction, was measured through a single survey item: "On a scale of 0 to 10, 10 being the best, please rate your overall level of satisfaction with the program you just attended."

Composite indexes were created for each of the factors by equally weighting each item and taking the average of all items within the index. Table 6 shows the individual items that comprise each resulting index, as well as Cronbach's alpha scores for each. Cronbach's alpha is a measure of internal consistency of each index and can range from 0 to 1. Cronbach's alpha scores above 0.7 are considered acceptable for developing indexes (DeVellis, 2003). Higher Cronbach's alpha scores indicate greater internal consistency of the index. Both indexes were found to be highly reliable.

Table 6. Independent variable indexes developed through exploratory factor analyses.

INDEPENDENT VARIABLE INDEXES
Interpreter characteristic: Confidence (Cronbach's α = 0 .70) • Comfort of the Interpreter • Apparent knowledge • Eloquence
Interpreter characteristic: Authentic emotion and charisma (Cronbach's α = 0.85) • Passion • Charisma • Sincerity
Program characteristic: Organization (Cronbach's α = 0.82) • Quality of the introduction • Appropriate sequence • Effective transitions • Holistic story • Clarity of theme • Link between introduction and conclusion
Program characteristic: Connection (Cronbach's α = 0.88) • Links to intangible meanings and universal concepts • Cognitive engagement • Relevance to audience • Affective messaging • Provocation

Index development: Independent variables

To explore the relationships between the individual program characteristics, we conducted exploratory factor analyses and reliability analyses on program observations. We did not conduct confirmatory factor in this case because program characteristics are formative variables that are observed and represent a specific practice or attribute that is thought to directly influence a dependent variable. This is opposed to reflective indicators, which are thought to represent a broader concept and are not directly observed (see Kline, 2005; Diamantopoulis & Siguaw, 2006; Jarvis et al., 2003; Padsokoff et al., 2003, for further explanation). Exploratory factor analyses and reliability analyses on program level data revealed the presence of four latent factors: two interpreter characteristics and two program characteristics. We have named the two resulting interpreter characteristics factors *confidence* and *authentic emotion and charisma*. We labeled the two resulting program characteristics factors *organization* and *connection*. The items making up each factor are included in Table 6.

The *confidence* factor generally reflects the notion that the interpreter appears in control of the program and is comfortable with what they are presenting. We use the term *authentic emotion and charisma* to denote a special sort of identity that the interpreter exudes to his or her audience. Interpreters scoring high on this factor showed apparent and obvious passion and care for what they were interpreting and were generally likeable. Organization reflects many of the best practices taught by the National Park Service's Interpretive Development Program in addition to the writings of Sam Ham (e.g., Ham, 1992). Meanwhile, Connection strongly reflects the core elements of Tilden's classic core principles (Tilden, 1957).

While the factor analyses revealed that *confidence, authentic emotion and charisma*,

Table 7. Means and standard deviations of outcome variables measured in visitor surveys.

Variable (Scale)	Means (with standard deviations)	
	Five or more attendees	Fewer than five attendees
Satisfaction (0 to 10)	8.96 (0.68)	9.02 (0.89)
Visitor experience and appreciation (1 to 5)	4.41 (0.32)	4.57 (.042)
• Made my visit to this park more enjoyable (1 to 5)	4.55 (0.30)	4.70 (0.43)
• Made my visit to this park more meaningful (1 to 5)	4.49(0.32)	4.69 (0.45)
• Enhanced my appreciation for this park (1 to 5)	4.36(0.37)	4.51 (0.51)
• Increased my knowledge about the program's topic (1 to 5)	4.45(0.34)	4.62 (0.47)
• Enhanced my appreciation for the National Park Service (1 to 5)	4.27(0.36)	4.38 (0.58)
Behavioral intentions (1 to 5)	2.92 (0.64)	3.02 (0.98)
• Changed the way I will behave while I'm in this park (1 to 5)	2.92(0.67)	3.08 (0.97)
• Changed the way I will behave after I leave this park (1 to 5)	2.92(0.61)	2.97 (1.04)

organization, and *connection* are separate constructs, they are also moderately correlated with each other (r ranges from .357 to .623). This suggests that when an interpreter scores highly on any one of these indexes, he or she is likely to score highly on the others as well.

Visitor characteristics

All descriptive statistics reported below are calculated *only* from the 312 programs that met our sampling criteria. More than half of the respondents to the surveys were female (56.4%). The ages of respondents ranged from 16 to 88, with a mean of 45 and a median of 46. Eighty-seven percent of respondents described themselves as White and not of Hispanic descent. Roughly 7% described themselves as Hispanic (3.6%) or Asian (3.6%). Only 34 respondents (1.1%) described themselves as Black and not of Hispanic descent; 15 respondents identified themselves as Native American and 25 respondents identified themselves as "other." Twenty-five respondents marked more than one category. Roughly 5% were from a country other than the United States. For comparison, a 2009 survey of U.S. residents conducted by the National Park Service estimated that roughly 78% of all visitors to National Park units were White; roughly 9% were Hispanic; roughly 7% were African American; roughly 3% were Asian; and roughly 1% were Native American (Taylor et al. 2010). Less than 5% of survey respondents attended the program alone. More than half (50.8%) were visiting with children. Most (59.1%) had been in the park less than one full day when they attended the program, and 37.4% had attended a ranger-led program in the same park prior to the one they were attending on the day they were surveyed.

Descriptive statistics: Outcomes

Table 7 displays the means and standard deviations of each outcome variable for programs with five or more attendees and for smaller programs. While *satisfaction* and *visitor experience and appreciation* consistently scored highly, items associated with *behavioral intentions* were more evenly distributed. Visitor satisfaction scores ranged

from 5 to 10 on the 0 to 10 scale and 95% of respondents scored above the midpoint on the *visitor experience and appreciation* index. Meanwhile, 43% percent of respondents scored above the midpoint on the *behavioral change* index. There were no statistically significant differences in visitor outcome scores between larger programs and programs with fewer than five attendees.

Descriptive statistics: Program types and attendees

We attempted to investigate 488 programs. Only 376 programs actually occurred. Programs were cancelled for a range of reasons including weather, no visitor attendees, or failure of the interpreter to appear. Data from 312 programs were used for analyses in this paper. Advertised program lengths for these programs ranged from 15 minutes to four hours. Actual program lengths ranged from 10 minutes to three and a half hours. The average program length was just over 48 minutes. One-hundred and ninety-eight (64%) of the programs focused primarily on cultural heritage; 74 (24%) had a primary focus on the natural environment. Thirty-three (11%) had a dual focus and others had neither central focus (for example, general orientation talks). Programs included guided tours, talks, demonstrations, hands-on activities, and multi-media presentations. Guided tours and stationary talks made up over 80% of the programs we observed. Seventy-two percent of programs took place outdoors; 20% took place indoors; and others used both indoor and outdoor settings. The breakdowns of program lengths and types were roughly similar for programs in the two different size classes described above.

The number of attendees at each program ranged from one person to approximately 600 people. The median number of attendees was 17. Only 17% of the programs had no children in their audiences. Forty programs (13%) ended with fewer attendees than they had begun with. Forty-eight programs (15%) were at least 20% shorter than advertised; 53 programs (17%) were at least 20% longer than advertised. Thirteen (4%) of the programs experienced notably bad weather. No significant differences were noted in program length or weather-related variables when comparing small (fewer than five attendees) with larger programs.

Descriptive statistics: Interpreter characteristics

Two-hundred and seventy-one (87%) of the observed interpreters were park rangers; 37 were volunteers, and five were concessionaires. Sixty-four percent were male. Nineteen percent were under the age of 25; 23% were between the ages of 25 and 34; 24% were between the ages of 35 and 50; and 34% were over 50 years old. The interpreters averaged 9.6 years of experience in the NPS and 7.1 years in interpretation at their current park unit. Nearly one quarter of the interpreters (24.7%) had presented the program we observed at least 100 times before. More than one-third (36.0%) had presented the program at least 50 times before. Nearly one-third (32.6%) had presented the program 10 or fewer times. For seven interpreters, this was their first time presenting the program we observed.

We asked interpreters prior to their programs to indicate their intended visitor outcomes for that program (Table 8). The most commonly noted intended outcome was providing the audience with new knowledge. Most (90%) noted more than one intended outcome. We also asked interpreters how their programs were developed (Table 9). Most reported developing their own programs with little guidance beyond a suggested topic.

We asked a subset of interpreters (n=188) about their level of excitement about the

Table 8. Intended outcomes expressed by interpreters immediately prior to their programs.

I want my audience to . . .	Proportion expressing each outcome
Have an increased knowledge of the program topic	79.5%
Have an increased appreciation for this park	56.4%
Have an increased understanding of the park's resources	39.1%
Want to learn more about the program topic	24.8%
Be entertained	15.6%
Have an increased appreciation of the NPS	14.1%
Have an increased concern for a specific topic	11.5%
Change their attitudes toward something	10.6%
Change a certain behavior in the future	7.0%
Develop and practice a new skill	3.5%

Table 9. How interpretive programs were developed.

Program development	Proportion expressing each
Program provided for ranger with full script planned out	< 1%
Program provided for ranger with some freedom to inject own style	14%
Program topic provided, little restrictions on information or style to be presented	20%
General topic suggested, but wrote own script and selected information	53%
Interpreter selected and developed entire program free of restrictions	13%

Table 10. Means and standard deviations of ordinal interpreter delivery styles.

	Means (with standard deviations)	
Variable (Scale)	**Five or more attendees**	**Fewer than five attendees**
***Confidence* index (1 to 4)**	3.28 (0.49)	3.12 (0.41)
• Comfort of the interpreter (1 to 4)	3.49 (0.60)	3.25 (0.63)
• Apparent knowledge (1 to 4)	3.45 (0.63)	3.40 (0.59)
• Eloquence (1 to 4)	2.99 (0.65)	2.83 (0.50)
***Authentic emotion and charisma* index (1 to 5)**	3.57 (0.85)	3.46 (0.70)
• Passion (1 to 5)	3.23 (1.02)	3.08 (1.04)
• Charisma (1 to 5)	3.82 (0.86)	3.68 (0.69)
• Sincerity (1 to 4)	2.93 (0.77)	2.88 (0.65)
Responsiveness (1 to 3)[a]	2.81 (0.41)	2.82 (0.45)
Humor quality (1 to 4)	2.08 (0.73)	1.92 (0.58)
Humor quantity (1 to 5)	2.08 (0.72)	1.85 (0.53)
Personal sharing (1 to 4)	1.68 (0.72)	1.79 (0.73)
Audibility (1 to 3)	2.86 (0.36)	2.85 (0.36)
Formality (1 to 5)	3.21 (0.86)	3.00 (0.68)
Sarcasm (1 to 3)	1.23 (0.46)	1.15 (0.36)
False assumptions of audience (1 to 3)	1.17 (0.40)	1.08 (0.27)

[a] Responsiveness was not observable in every case. For larger programs, n = 245.

Table 11. Descriptive statistics of interpreter delivery styles (categorical variables).

Interpreter delivery style	% of programs in which delivery style occurred	
	Five or more attendees	Fewer than five attendees
Professional appearance of the interpreter	98.2	100.0
Inequitable treatment of audience	2.9	2.5
Impatience	1.8	2.5
Primary identity: Friend	18.0	37.5
Primary identity: Authority	4.4	2.5
Primary identity: Walking encyclopedia	76.8	67.5
Character acting: partial	2.6	2.5
Character acting: complete	2.9	0.0
Interpreter bias	3.3	7.5
Questionable information	9.9	2.5

Table 12. Means and standard deviations of ordinal program characteristics.

Variable (Scale)	Means (with standard deviations)	
	Five or more attendees	Fewer than five attendees
***Organization* index (1 to 5)**	3.34 (0.71)	3.14 (0.65)
• Quality of introduction (1 to 3)	2.13 (0.45)	1.93 (0.42)
• Appropriate sequence (1 to 4)	2.79 (0.69)	2.70 (0.69)
• Transitions (1 to 4)	2.72 (0.76)	2.55 (0.71)
• Holistic story (1 to 5)	2.78 (1.01)	2.78 (0.77)
• Conclusion linked to intro (1 to 4)	2.63 (0.86)	2.48 (0.75)
• Clear theme (1 to 4)	2.82 (0.86)	2.58 (0.90)
***Connection* index (1 to 5)**	2.77 (0.78)	2.74 (0.55)
• Links to intangible meanings and universal concepts (1 to 5)	2.88 (0.94)	3.00 (0.80)
• Cognitive engagement (1 to 5)	2.85 (0.94)	2.78 (0.83)
• Relevance to audience (1 to 5)	2.86 (0.86)	2.70 (0.69)
• Affective messaging (1 to 5)	2.43 (0.95)	2.38 (0.71)
• Provocation (1 to 4)	2.24 (0.72)	2.25 (0.67)
Clear message (1 to 4)	2.20 (0.94)	2.00 (0.85)
Appropriate logistics (1 to 4)	3.11 (0.93)	3.15 (0.89)
Appropriate for the audience (1 to 5)	3.93 (0.70)	4.15 (0.83)
Multisensory (1 to 3)	2.39 (0.51)	2.35 (0.48)
Physical engagement (1 to 4)	1.42 (0.69)	1.50 (0.75)
Verbal engagement (1 to 5)	2.51 (1.02)	2.68 (0.80)
Surprise (1 to 3)	1.10 (0.31)	1.03 (0.16)
Novelty (1 to 3)	1.18 (0.43)	1.10 (0.30)
Consistency (1 to 3)	2.88 (0.37)	2.88 (0.34)
Resource quality (1 to 3)	2.37 (0.70)	2.13 (0.69)
Multiple viewpoints (1 to 3)[a]	2.63 (0.51)	2.61 (0.50)
Behavioral theory elements [b]		
Benefits of action (1 to 4)	2.52 (0.63)	2.80 (0.45)
Costs of action (1 to 3)	1.97 (0.75)	2.40 (0.89)
Norms of action (1 to 3)	1.48 (0.57)	1.40 (0.55)
Ease of action (1 to 3)	1.81 (0.65)	1.20 (0.45)
Demonstrates action (1 to 4)	2.13 (0.96)	2.20 (1.30)

[a] Multiple viewpoints were not appropriate or relevant in every case (e.g., a talk on butterfly life cycles). We only observed this variable where it seemed potentially relevant (n = 94 for larger programs; n = 22 for smaller programs).

[b] These variables are explicitly associated with behavioral change theory. As such, they were only observed on a small subset of cases within the sample where specific behaviors were discussed by the interpreter (n = 31 for larger programs; n = 5 for smaller programs).

Table 13. Descriptive statistics of program characteristics (categorical variables).

Program characteristics	% of programs w program characteristic was observed	
	Five or more attendees	Fewer than five attendees
Fact-based messaging	26.8%	25.0%
Use of props	30.5%	27.5%
Pace too fast	6.2%	5.0%
Pace too slow	9.2%	5.0%
Pace just right	84.6%	90.0%
Unexpected positive circumstance	1.8%	2.5%
Unexpected negative circumstance	15.8%	10.0%

program they were about to present. The level of excitement averaged 7.81 on a 10-point scale, with responses ranging from 2 to 10 on the scale. Seven percent ranked their level of excitement below the midpoint (5) on the scale; 4% selected the midpoint; and 89% rated their level of excitement above the midpoint.

Descriptive statistics: Interpreter delivery styles
Tables 10 and 11 display descriptive statistics of each of the interpreter delivery styles observed in the study. Table 10 contains ordinal variables (variables that are measured on an increasing scale). Table 11 contains binary and categorical variables, or those in which the presence or absence of the characteristics is the essential feature being measured. Means comparisons, chi-square tests, and effect size calculations revealed few meaningful differences between the two size classes of programs. Interpreters typically scored slightly lower on the *confidence* index in smaller groups ($t=2.0$; $p=0.042$; Cohen's $d=0.38$). We also more commonly observed the "friend" identity in smaller groups ($\chi^2=8.0$; $p=0.005$).

Descriptive statistics: Program characteristics
Tables 12 and 13 display descriptive statistics for each of the program characteristics observed in the study. Table 12 displays ordinal variables, while Table 13 displays categorical variables. No statistically significant differences were observed between the two size classes of programs.

Which practices and approaches most consistently lead to more positive outcomes for visitors?

Interpreter and program characteristics
Table 14 displays (in rank order) correlations between all ordinal independent variables (program and interpreter characteristics) and visitor outcomes for programs with five or more attendees. Statistical significance is displayed in two ways within the table. A single asterisk indicates that the correlation is statistically significant at $p < 0.05$. A double asterisk indicates that the correlation is statistically significant at $p < 0.01$. As such, the stronger relationships are those with two asterisks. These are bolded and italicized for ease of interpretation. Cells with no asterisks represent no statistically significant relationships between the variables.

Behavioral theory elements were observed in 42 programs overall, including 31

Table 14. Pearson correlations between ordinal independent variables and visitor outcomes for programs with five or more attendees.

Variable	Satisfaction	Visitor experience and appreciation	Behavioral intentions
Interpreter style: Confidence index	***.479***	***.277***	***.174***
Interpreter style: Authentic emotion and charisma index	***.423***	***.303***	***.182***
Program characteristic: Approp. for audience	***.381***	***.378***	.153*
Program characteristic: Organization index	***.362***	***.219***	.132*
Program characteristic: Connection index	***.342***	***.259***	.124*
Interpreter style: Humor quality	***.288***	***.233***	.155*
Program characteristic: Consistency	***.271***	***.281***	.034
Program characteristic: Clear message	***.255***	***.281***	***.187***
Interpreter style: Responsiveness	***.241***	***.245***	.061
Program characteristic: Verbal engagement	***.234***	***.240***	***.162***
Program characteristic: Multisensory engagement	***.216***	.115	.141*
Interpreter style: Audibility	***.197***	.134*	.104
Interpreter style: False assumption of audience	***-.172***	***-.197***	-.088
Program characteristic: Appropriate logistics	***.170***	***.245***	***.165***
Program characteristic: Surprise	.150*	.151*	.127*
Program characteristic: Novelty	.145*	.024	.014
Interpreter style: Humor quantity	.144*	.097	.062
Program characteristic: Physical engagement	.074	.120*	.061
Interpreter style: Formality	-.069	-.155*	-.023
Interpreter style: Sarcasm	.105	.053	-.114
Program characteristic: Quality of the resource	.077	.068	.065
Interpreter style: Personal sharing	.035	.048	.112
Program characteristic: Multiple points of view	.031	.157	.128

with five or more attendees. Only one behavioral theory element showed a statistically significant correlation with the *behavior change* index, "costs of action" ($r=.597$, $p < .001$). This suggests that programs that explicitly addressed the costs of undertaking a potential behavior were generally more successful at influencing behavior change intentions than others.

T-tests and ANOVAs were performed to examine the relationships of categorical variables upon visitor outcomes. These variables included *fact-based messaging, unexpected positive* and *negative circumstances, pace, bias, impatience, inequitable treatment of the audience, questionable information, use of props*, and *interpreter identities*. Tables 15 and 16 summarize only the statistically significant relationships observed in the data. To facilitate interpretation of the t-tests, we calculated Cohen's d for each of the statistically significant associations. *Cohen's d* is an effect size measure that provides an assessment of the meaningfulness of the difference between groups. Cohen (1988) suggested that even statistically significant differences may not be meaningful in a practical sense. They may rather be an artifact of large sample sizes. Cohen posited that meaningful differences begin at $d=0.2$. Differences near 0.2 may be considered small, while those approaching 0.5 are considered medium and 0.8 large.

Programs in which the interpreter outwardly expressed impatience with the audience received lower *satisfaction* and *visitor experience and appreciation* scores than others, as did programs with an *unexpected negative occurrence*. Programs in which

Table 15. Statistically significant t-tests results, comparing the means of visitor outcome scores for selected categorical variables for programs with five or more attendees.

Observed category	Satisfaction				Visitor experience and appreciation				Behavioral intentions			
	Mean diff.	t	p	Cohen's d	Mean diff.	t	p	Cohen's d	Mean diff.	t	p	Cohen's d
Impatience	-0.36	-2.2	0.031	0.68	-0.47	-3.3	0.001	1.28				
"Friend"	0.23	2.3	0.023	0.36								
"Walking encyclopedia"									-0.20	-2.2	0.031	0.32
Fact-based messaging	-0.34	-3.9	< 0.001	0.50	-0.12	-2.6	0.011	0.36				
Unexpected neg. circumstance	-0.29	-2.8	0.006	0.45	-0.19	-3.6	< 0.001	0.60				
The following categorical variables yielded no statistically significant differences in visitor outcomes: Inequitable treatment of the audience, questionable information, "Authority" identity, unexpected positive circumstances, use of props.												

Table 16. One-way ANOVA comparing outcome variables for programs of different pace with five or more attendees. *Items not sharing the same superscript are statistically different from one another.*

Pace	Means		
	Satisfaction	Visitor experience and appreciation	Behavioral intentions
Too fast	8.62^{A}	4.27^{AB}	2.56^{A}
Too slow	8.43^{A}	4.23^{A}	2.84^{AB}
Appropriate	9.03^{B}	4.44^{B}	2.96^{B}
Statistics	F = 12.9; p < 0.001 *Cohen's d (appropriate pace vs. others): 0.78*	F = 6.9, p = 0.001 *Cohen's d (appropriate pace vs. others): 0.57*	F = 3.2, p = 0.042 *Cohen's d (appropriate pace vs. others): 0.34*

the interpreter employed the "friend" identity manifested higher satisfaction scores than others. Meanwhile, programs in which the interpreter employed the "walking encyclopedia" identity yielded lower behavioral intention scores than others. Paces that felt too fast or too slow resulted in lower satisfaction scores. A too-slow pace was related to lower *visitor experience and appreciation* scores, and a too-fast pace was associated with weaker *behavioral intentions*. No statistically significant differences were observed for smaller programs (fewer than five attendees).

Program attrition and outcomes

Program attrition (people leaving a program before it was completed) was related to both *satisfaction* and *visitor experience and appreciation* for programs with five or more attendees (see Table 17), suggesting that program attrition may serve as another reasonable indicator of program quality. Thirty-six of programs with five or more attendees experienced attrition. The best predictors of program attrition for programs with five or more attendees included interpreters' lack of *responsiveness* to the audience, inaudibility, *false assumptions about the audience*, the identity of the *walking encyclopedia*, inappropriate logistics, the use of props, slow pace, lack of interpreter *confidence*, a lack of organization of the program, and an *unexpected negative circumstance* (see Tables 17 and 18).[3] No other interpreter or program characteristics exhibited any statistically significant relationship with program attrition at $p < 0.05$.

Table 17. Independent samples t-tests comparing means of characteristics for programs that experienced attrition (people left the program early) vs. those that did not.

Characteristic	Program attrition?	Means	t	p	Cohen's d
Responsiveness of the interpreter	Yes	2.62	-2.4	0.020	0.46
	No	2.83			
Audibility	Yes	2.72	-2.3	0.025	0.49
	No	2.91			
False assumption of the audience	Yes	1.31	2.4	0.020	0.50
	No	1.11			
Appropriate logistics	Yes	2.44	-5.0	<0.001	0.86
	No	3.23			
Confidence	Yes	3.08	-2.8	0.006	0.46
	No	3.32			
Organization	Yes	3.09	-2.2	0.031	0.32
	No	3.36			
Outcomes	**Program attrition?**	**Means**	**T**	**p**	
Satisfaction	Yes	8.49	-3.9	<.001	0.79
	No	9.04			
Visitor experience and appreciation	Yes	4.26	-2.6	.014	0.51
	No	4.44			
Behavioral intentions	Yes	2.73	-1.8	.070	0.34
	No	2.95			

Table 18. Chi-square tests comparing programs that experience attrition vs. those that did not.

Characteristic	Pearson χ^2 statistic	p	Relation to attrition
Interpreter identity: walking encyclopedia	3.6	.058	More attrition
Use of props	12.4	.001	More attrition
Slow pace	5.8	.026	More attrition
Unexpected negative occurrence	8.9	.006	More attrition

Table 19. Statistically significant t-tests results, comparing the means of visitor outcome scores for interpreters who expressed different intended outcomes for their interpretive programs.

	Satisfaction				Visitor experience and appreciation				Behavioral intentions			
Intended outcome	Mean diff.	t	p	Cohen's d	Mean diff.	t	p	Cohen's d	Mean diff.	t	p	Cohen's d
Increased knowledge					-0.12	2.4	0.019	0.37				
Increase desire to learn	0.20	2.2	0.029	0.30	0.14	3.2	0.002	0.46				
Change attitude	0.18	2.0	0.048	0.31	0.16	4.3	< 0.001	0.45				
Increase appreciation for Park	0.22	2.7	0.007	0.34	0.09	2.2	0.028	0.28				
Increase understanding of resource					0.08	2.1	0.040	0.26				
Increase level of concern									0.27	2.2	0.032	0.41
Change visitor behavior									0.41	2.7	0.008	0.66

Relationship between interpreter and program characteristics and outcomes in programs with fewer than five attendees
Fewer statistically significant correlations ($p < 0.05$) were observed in programs with fewer than five attendees. In rank order, they included:

Correlated with *Satisfaction:*

- Connection index: r=.492, p=.001
- Organization index: r=.420, p=.007
- Appropriate for the audience: .337, p=.033
- Humor quality: r=.323, p=.045

Correlated with *Visitor experience and appreciation:*

- Connection index: r=.438, p=.005
- Organization index: r=.368, p=.020
- Appropriate for the audience: .348, p=.028

Correlated with *Behavioral intentions:*

- Novelty: r=.408, p=.009

Thus, a subset of the variables that predicted positive outcomes in larger programs predicted similar outcomes in smaller programs. Because only four programs within this sample experienced attrition, no additional analyses were conducted pertaining to attrition.

Interpreters' background, excitement, and intentions
For the smaller program sample (those with fewer than five attendees), no statistically significant relationships were observed between interpreter backgrounds, level of excitement, program origin, or intended outcomes and visitor outcomes. Some differences were noted, however, in the larger sample.

For larger group sizes (five or more attendees), program outcomes were not related to the age, gender, or experience of interpreters, nor their degree of autonomy in program development. The interpreters' degree of excitement about the program was positively correlated with visitor *satisfaction* (r=.186; p=0.013) and *visitor experience and appreciation* (r=.153; p=0.041). Interpreters expressing higher degrees of excitement also exhibited higher levels of *confidence* (r=.324, $p < .001$) and *authentic emotion and charisma* (r=.475; $p < .001$). Volunteers tended to achieve lower degrees of visitor satisfaction than did park rangers (means: 8.70 vs. 8.98; t=-2.4; p=.019; Cohen's d=0.42).

We examined the relationships between interpreters' intended outcomes and visitor-reported outcomes by conducting independent samples t-tests, which compare the means of two groups. In these cases, groups were defined by the presence of an intended outcome or not. Table 19 summarizes only the statistically significant relationships between interpreters' intended outcomes and visitor survey responses. Cohen's d statistics are also provided as effect size estimates. *Visitor experience and appreciation* was the most sensitive to interpreters' intended outcomes, with five different desired

outcomes related to more positive visitor responses. *Satisfaction* was related to a subset of these items. Only one intention was negatively related to visitor outcomes. Interpreters who were aiming to increase visitors' knowledge as a primary outcome of their program generally achieved lower *visitor experience and appreciation* scores. Two intended outcomes were positively related to reported *behavioral intentions* by visitors: increasing the audience's level of concern and changing visitors' behaviors.

Discussion

The study sought to determine which practices and approaches most consistently lead to more positive outcomes for live interpretive programs' attendees. In this manuscript, we have limited our analyses to bivariate relationships between practices and outcomes rather than employing multivariate statistics. We did this for two reasons. First, we wished to examine the individual relationship of each observed practice and interpreter characteristic with visitor outcomes. Second, multivariate analyses are used to provide the most parsimonious statistical model of observed phenomenon. In multivariate processes, certain observed characteristics may be removed from the best explanatory model if they explain a similar portion of the variance as another variable, despite being an important part of influencing a particular outcome (Byrne, 2006). As a result, the multivariate approach may lead to misinterpretation of the importance (or lack thereof) of particular practices and program characteristics. If one were to focus only on the variables contained in the multivariate statistical model, at the expense of others that covaried with those same variables, there would be a danger of inappropriately assuming that practices not in the model are unimportant. In a companion piece, we use structural equation modeling to develop more parsimonious causal models (see Powell and Stern, this issue). These multivariate analyses help to illuminate the inter-relationships of different interpreter and program characteristics and their roles in influencing outcomes. However, they do not negate the bivariate relationships shared in this article.

Understanding outcomes

Live interpretive programs across the NPS generally seem to produce consistently high levels of satisfaction in their attendees. Eighty-five percent of the analyzed sample rated the program as an 8 or better on the 0 to 10 satisfaction scale. Such satisfaction skewness is common in customer satisfaction surveys, and the modal response is typically the most positive response allowed by the scale (Peterson & Wilson, 1992). The mode in our case was a 9 out of 10. Prior research suggests that satisfaction assessments may be influenced by social desirability bias or acquiescence (Peterson & Wilson, 1992). In our case, such social factors might include some degree of gratitude or sympathy toward the interpreter regardless of the program quality, leading respondents to check a positive response. High satisfaction scores might also be attributed in part to what is known as assimilation effects (Sherif & Hovland, 1961). In the context of tourism, this means that expectations are often a stronger driver of satisfaction ratings than the quality of the actual experience (del Bosque & San Martín, 2008). In other words, if visitors strongly expect an experience to be positive, they have a high tendency to rate it as such regardless of its specific qualities. This may of course be the case with visitors to national parks. Still, the particularly high satisfaction values observed in this study suggest that few visitors were dissatisfied with their interpretive experiences. *Visitor experience and appreciation* also showed similar trends.

Despite the skewness of the data, we observed significant statistical relationships between certain program characteristics and visitor outcomes. The positively skewed dependent variables, however, suggest that our findings do not necessarily identify the practices that separate good programs from bad programs. Rather, the findings illuminate which characteristics most commonly move programs along a scale from good to better from a visitor's standpoint (see Stern et al., this issue).

The *behavioral intentions* outcome was centered closer to the midpoint of the five-point scale. This is likely due to widely varying baselines in terms of visitors' behaviors prior to programs (some visitors wrote on the survey cards things like "I already respect the parks"). For example, if a visitor is a major park supporter and an environmentally sensitive visitor, we might expect them to report no change, despite experiencing what may have been an outstanding program. Meanwhile, an inexperienced visitor to the same program might have reported a great deal of change. As such, we might expect muted results regarding program and interpreter characteristics' associations with the behavioral intentions outcome. This may in part explain the smaller number of independent variables associated with intentions to change behaviors. Other authors have also expressed concern when measuring intentions and behavior change, especially in nature-based settings (see Beaumont, 2001; Powell et al, 2008).

What leads to better outcomes?

Interpreters who expressed that a primary goal of their program was to increase the knowledge of the audience about their program's topic achieved lower visitor experience and appreciation scores than others. Those aiming to change their audience's attitudes, appreciation, understanding, and/or desire to learn achieved more positive attitudinal outcomes. Interpreters who explicitly aimed to increase their audience members' levels of concern or change their behavior were more likely to achieve more positive post-program behavioral intentions than others.

The best predictors of positive outcomes varied somewhat for different outcomes. In programs with at least five attendees, the outcomes *Satisfaction* and *visitor experience and appreciation* were correlated with a similar list of program and interpreter characteristics, including: *confidence, authentic emotion and charisma, appropriateness for the audience, organization, connection, humor quality, consistency, a clear message, responsiveness, verbal engagement, audibility,* and *appropriate logistics and pace*. Multisensory engagement and fact-based messaging (negative relationship) were additionally related to *satisfaction*.

Behavioral theory suggests that interpretation (and other communication/educational experiences) should not be expected to change behavior unless a specific behavior is explicitly targeted and communication is designed to address attitudes relevant to that behavior (e.g., Ajzen, 1991; Ham et al., 2007). Programs in which the interpreter explicitly targeted behavior change as an intended outcome (7%) were more successful at doing so. Programs of this nature that explicitly addressed the costs of taking that action were the most successful, supporting Ajzen's (1991) emphasis on both ability and trade-offs in predicting behavior. Moreover, *confidence, authentic emotion and charisma, a clear message, verbal engagement,* and *appropriate logistics* showed the strongest statistically significant correlations with the *behavioral intentions* outcome. These items mirror theoretical constructs from multiple disciplines known to be predictive of behavior change, including credibility and trust in the communicator

Figure 1. Best practices for live interpretive programs observed in the study.

1. Confidence
    - Comfort, eloquence, apparent knowledge
2. Authentic emotion and charisma
    - Passion, sincerity, charisma
3. Appropriateness for audience
4. Organization
    - Quality of introduction, appropriate sequence, effective transitions, holistic story, clear theme, link between introduction and conclusion
5. Connection
    - Links to intangibles and universal concepts, cognitive engagement, relevance to audience, affective messaging, provocation
6. Consistency
7. Clear message
8. Responsiveness
9. Audibility
10. Appropriate logistics
11. Verbal engagement
12. Multisensory engagement
13. Appropriate pace
14. Avoid focusing on knowledge gain as the program's central goal and communicating solely factual information
15. Avoid making uncertain assumptions about the audience

(Rogers, 1995; Stern, 2008), empowerment of the message recipient and verbal engagement (Ajzen, 1991; Stern, 2008), and the elimination of distraction and clear orientation to place (Moscardo, 1999). For a broader discussion of behavior change and interpretation see Ham et al. (2007) and Ham (2009).

A smaller subset of interpreter and program characteristics were correlated with outcomes for smaller programs (those with fewer than five attendees). *Connection, organization,* and *appropriateness for the audience* were each correlated with *satisfaction* and *visitor experience and appreciation*. *Humor quality* was additionally correlated with *satisfaction*. Only *novelty* was correlated with post-program behavioral intentions for these smaller programs.

Implications for live interpretation

The study carries implications for both the practice of live interpretation as well as future research pertaining to best practices. Figure 1 provides a list of the program characteristics most strongly associated with the outcomes measured in this study. These "best practices" cut across multiple contexts (see Powell & Stern, this issue) and constitute elements of interpretation that could inform interpretive training both within the National Park Service and beyond. While *humor quality* also was positively related to outcomes, we don't list it as a best practice, as not all programs *should* necessarily be funny.

Although each of the practices listed in Figure 1 was statistically correlated with better outcomes, variability within the sample suggests that the entire suite of best practices is not a necessary precursor to a high-quality program. Rather, each of these practices in various combinations was found to enhance outcomes across a majority of programs in which they were practiced. A wide range of diverse approaches led

to positive visitor outcomes. As such, we recommend maintaining the freedom for interpreters to be creative and innovative in their presentations. This is further supported by correlations between interpreters' own excitement about a program and positive visitor outcomes.

While many of the "best practices" in Figure 1 speak to specific interpretive techniques, some, at first glance, appear to exist outside of the famous "interpretive equation" used in NPS trainings (Lacome, 2013). The interpretive equation is presented as a "foundation" for NPS interpretive training and as a tool for identifying "the elements of successful interpretation" and the relationships between them. In its simplest form, the equation states that an interpretive opportunity (IO: "one that provides a favorable set of circumstances for a meaningful moment of connection between audience and resource," p. 5) is brought about by knowledge of the resource (KR), knowledge of the audience (KA), and appropriate techniques (AT).

The Interpretive Equation: KR + KA x AT=IO

Many of the "best practices," in particular *confidence, authentic emotion and charisma*, and avoiding a focus on knowledge gain, do not clearly constitute "knowledge of the resource," "knowledge of the audience," or "appropriate techniques" directly. They are rather the observable manifestations of internal states specific to individual interpreters during their programs. Their significance speaks to the importance of the appropriate translation of the interpretive equation into action. While knowledge of the resource is critical, it should not necessarily be the focus of communications within an interpretive setting. Rather, knowledge of the resource may play a more important role in enhancing the confidence of the interpreter and allowing his or her own positive emotions and connections to the resource to show through. Presenters who are more familiar with their topics generally experience less anxiety (Daly et al. 1989). When coupled with knowledge of the audience and appropriate techniques, feelings of self-confidence and freedom to express oneself might be instrumental in moving from good, or adequate, visitor outcomes toward more powerful ones. This also suggests that the general organizational culture in which the interpreter finds herself is likely important as well. More supportive and empowering cultures may lead to better performance (Pearce & Sims, 2002; Rafferty & Griffin, 2006). The particular roles of interpreter characteristics vs. program characteristics are examined in greater detail in a companion article within this issue (Powell & Stern, this issue).

Implications for future interpretive research

This research suggests that certain interpretive practices are statistically linked to desired outcomes across a range of contexts. Without the ability to compare a large sample of programs, this identification would not have been possible. We thus urge others to undertake similar forms of research and to learn from our shortcomings. Even comparative research of just a few programs can shed additional light on what practices and approaches are linked to more positive visitor outcomes (see Ballantyne & Packer 2002, for example).

Our limitations and shortcomings were many in this effort, including both controllable and uncontrollable factors. Those most relevant to future research involve the selection and measurement of the key independent and dependent variables of the

study. The treatment (an interpretive program in a national park setting) is a complex phenomenon that is influenced by an interaction between the resource and its qualities, the social environment, including the makeup of social groups, the characteristics of the interpreter and the individual attendees, and the topic and characteristics of the program (Powell et al, 2009). This research focuses on the relationships between visitor outcomes and selected interpreter and program characteristics. As such, other potential influences are not accounted for.

Our experience revealed that it required considerable and iterative training, feedback, and adjustment for our team to produce consistent and reliable monitoring results. This is a well-known challenge in any research using a team of human observers, who have a tendency to cling to their own personal biases or sometimes idiosyncratic interpretations of similar events (Jacobs et al., 2012). In an ideal situation, additional pilot testing and assimilation of the team toward consistent definitions could take place and programs would be consistently observed in pairs, rather than by individuals.

Our selection of dependent variables was quite challenging due to the wide diversity of program content and formats included in this study. Visitor survey items were designed to be rather general in their content so as to be appropriate and relevant to all programs. The general nature of outcome measures may have also contributed to a "ceiling effect," which describes the phenomenon when individuals (in this case, NPS visitors) come into an experience with already high scores on the outcomes considered (in this case the specific attitudes and intentions measured in the study). As such, some respondents would report little to no change for an outcome measure because their attitudes or intentions may already be at the high end of the spectrum for the outcome in question. In these cases, the survey items may not be sensitive enough to detect the influence of a program. We urge future researchers to develop more sensitive dependent variables, and, if possible, include a control group. In particular, other researchers have found that multiple measures of satisfaction with both positive and negative wording can produce more variability (Peterson & Wilson, 1992). A rigorous approach to control group sampling might involve a similar design as our own (see endnotes) with a larger sample of non-participants. Alternatively, researchers might consider comparison groups exposed to similar interpretation with the exception of only a few variables (or ideally one experimental variable) at a time.

Conclusions

Overall, our analysis suggests that Tilden (1957), writing over 50 years ago, was right about a lot of things. Programs that are relevant to the audience, tell holistic stories, provoke the audience to reflect, and move beyond facts into the realm of revelation tend to produce better visitor outcomes than programs that are fact-based and detached from the audiences' lives. It also suggests that more recent interpretive texts and training programs include numerous ideas that can enhance the interpretive experience, including the passion of the interpreter (e.g., Beck & Cable, 2002; Ward & Wilkinson, 2006), the organization of the material (e.g., Ham, 1992; Larsen, 2003), the importance of a central message (e.g., Ham, 1992; Jacobson, 1999), the connection of tangible objects to intangible meanings and universal concepts (NPS, 2003), and multiple forms of engagement and responsiveness (Beck & Cable, 2002; Knudson et al., 2003; Lewis, 2005; Moscardo, 1999). The study also revealed some factors that appear less regularly in existing training programs, but are certainly not surprising. In essence, the study

revealed the importance of the sincerity, passion, confidence, and delivery style of individual interpreters, as much as the planning and content of the program itself. We echo Tilden (1957) in believing that "interpretation is an art … and that any art is in some degree teachable." We hope that the results of this study can contribute to the learning process of the committed individuals around the world who care deeply enough about our world to call themselves "interpreters."

Literature Cited

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Decision-Making Processes,* 50(2), 179–211.

Ajzen, I. (1992). Persuasive communication theory in social psychology: a historical perspective. In: Manfredo, M. (Ed.). *Influencing Human Behavior. Theory and Applications in Recreation, Tourism, and Natural Resources Management*. Champaign, IL: Sagamore Publishing, Inc.

Babbie, E. (2007). *The Practice of Social Research*. Belmont, CA: Wadsworth.

Ballantyne, R. and J. Packer. (2002). Nature-based excursions: School students' perceptions of learning in natural environments. *International Research in Geographical and Environmental Education 11,* no. 3: 218–236.

Beck, L., & Cable, T. T. (2002). *Interpretation for the 21st century: Fifteen guiding principles for interpreting nature and culture* (2nd ed.). Champaign: Sagamore.

Beaumont, N. (2001). Ecotourism and the conservation ethic: Recruiting the uninitiated or preaching to the converted? *Journal of Sustainable Tourism,* 9(4):317–341.

Brochu, L., & Merriman, T. (2002). *Personal interpretation: Connecting your audience to heritage resources.* Fort Collins, CO: InterpPress.

Chesebro, J. L., & Wanzer, M. B. (2006). Instructional message variables. *Handbook of instructional communication: Rhetorical and relational perspectives* (pp. 89–116). Boston: Allyn & Bacon.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Daly, J.A., Vangelisti, A.L., Neel, H.L., & Cavanough, P.D. 1989. Pre-performance concerns associated with public speaking anxiety. *Communication Quarterly,* 37(1); 39–53.

del Bosque, I. R., & San Martín, H. (2008). Tourist satisfaction: A cognitive-affective model. *Annals of Tourism Research,* 35(2): 551–73.

DeVellis, R.F. 2003. *Scale development: Theory and applications applied social research methods.* 2nd ed. Thousand Oaks, CA: Sage Publishing.

Diamantopoulos, A., & Winklhofer, H.M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research,* 38(2); 269–277.

Eppley Institute for Parks and Public Lands. (2012). "About interpretation." http://idp.eppley.org/about-interpretation. Last accessed July 16, 2012.

Finn, A. N., Schrodt, P., Witt, P. L., Elledge, N., Jernberg, K. A., & Larson, L. M. (2009). A meta-analytical review of teacher credibility and its associations with teacher behaviors and student outcomes. *Communication Education,* 58(4), 516–537.

Forist, B. (2003). *Visitor Use and Evaluation of Interpretive Media. A Report on Visitors to the National Park System*. National Park Service Visitor Services Project. http://nature.nps.gov/socialscience/ docs/Visitor_Use_and_Evaluation.pdf. Accessed May 24, 2012.

Frauman, E., & Norman, W.C. (2003). Managing visitors via "mindful" information services: One approach in addressing sustainability. *Journal of Park and Recreation Administration*, 21(4): 87–104.

Ham, S. H. (1992). *Environmental interpretation: A practical guide for people with big ideas and small budgets*. Golden: Fulcrum.

Ham, S. (2009). From interpretation to protection: is there a theoretical basis? *Journal of Interpretation Research* 14(2): 49–57.

Ham S. H. (2013). *Interpretation: Making a difference on purpose*. Golden, CO: Fulcrum.

Ham, S. H., Brown, T. J., Curtis, J., Weiler, B., Hughes, M., & Poll, M. (2007). *Promoting persuasion in protected areas: A guide for managers. Developing strategic communication to influence visitor behavior.* Southport, Queensland, Australia: Sustainable Tourism Cooperative Research Centre.

Ham, S H. & Weiler, B.M. (2002). Tour guide training: a model for sustainable capacity building in developing countries. *Journal of Sustainable Tourism* 10(1): 52–69.

Interpretive Development Program. (2008). *National Park Service: Professional Standards for Learning and Performance*. U.S. Department of the Interior.

Jacobs, W.J., Sisco, M., Hill, D., Malter, F., and Figueredo, A.J. 2012. On the practice of theory-based evaluation: information, norms, and adherence. *Evaluation and Program Planning 35*, no. 3: 354–369.

Jacobson, S. K. (1999). *Communication skills for conservation professionals*. Washington, DC: Island Press.

Jarvis, C.B., MacKenzie, S.B., & Podsakoff, P.M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(2); 199–218.

Kline, R.B. (2005). *Principles and practice of structural equation modeling.* New York: The Guilford Press.

Knapp, D., & Benton, G. M. (2004). Elements to successful interpretation: A multiple case study of five national parks. *Journal of Interpretation Research*, 9(2), 9–25.

Knapp, D. & Yang, L. (2002). A phenomenological analysis of long-term recollections of an interpretive program. *Journal of Interpretation Research*, 7(2), 7–17.

Knudson, D. M., Cable, T. T., & Beck, L. (2003). *Interpretation of cultural and natural resources* (2nd ed.). State College: Venture Publishing.

Lacome, B. (2013). *The Intepretive Equation*. http://idp.eppley.org/IDP/sites/default/files/EquationEssay.pdf. Accessed April 6, 2013.

Larsen, D. L. (2003). *Meaningful Interpretation: How to Connect Hearts and Minds to Places, Objects, and Other Resources*. Eastern National.

Lewis, W. J. (2005). *Interpreting for park visitors* (9th ed.). Fort Washington: Eastern National.

Madin, E. M. P., & Fenton, D. M. (2004). Environmental interpretation in the Great Barrier Reef Marine Park: An assessment of programme effectiveness. *Journal of Sustainable Tourism*, 12(2), 121–137.

McCroskey, J.C., Richmond, V.P., & Stewart, R.A. (1986). *One on one: The foundations of interpersonal communication*. Englewood Cliffs, NJ: Prentice-Hall.

McManus, P.M. (1987). It's the company you keep … The social determination of learning-related behaviour in a science museum. *The International Journal of Museum management and Curatorship,* 6: 263–270.

McManus, P.M. (1988). Good companions. More on the social determination of learning-related behaviour in a science museum. *The International Journal of Museum Management and Curatorship* 7: 37–44.

Mehrabian, A. (1969). Attitudes inferred from non-immediacy of verbal communications. *Journal of Verbal Learning & Verbal Behavior*, 6, 294–295.

Mills, E. 1920. *Adventures of a Nature Guide*. New York: Doubleday, Page & Company.

Moscardo, G. (1999). *Making visitors mindful: Principles for creating quality sustainable visitor experiences through effective communication*. Champaign: Sagamore.

National Association for Interpretation (NAI). 2012. *Mission, Vision and Core Values*. http://www.interpnet.com/about_nai/mission.shtml. Accessed May 24, 2012.

National Park Service 2003. *Module 101: Fulfilling the NPS mission: The process of interpretation*. Washington, DC: National Park Service.

Pearce, C.L. & Sims, H.P. (2002). Vertical versus shared leadership as predictors of the effectiveness of change management teams: an examination of aversive, directive, transactional, transformational, and empowering leader behaviors. *Group dynamics: theory, research, and practice,* 6(2), 172–197.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2007). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology,* 88(5) 879–903.

Powell, R.B., Skibins, J.C. & Stern, M.J. (2010) *Linking interpretation best practices with outcomes: A review of literature*. Clemson University and U.S. National Park Service, National Education Council.

Powell, R.B., Kellert, S.R., & Ham, S.H. (2008). Antarctic tourists: Ambassadors or consumers? *Polar Record,* 44(230): 233–241.

Powell, R. B., Kellert, S. R., & Ham, S. H. (2009). Interactional theory and the sustainable nature-based tourism experience. *Society & Natural Resources,* 22(8), 761–776.

Rafferty, A.E. & Griffin, M.A., (2006). Refining individualized consideration: distinguishing developmental leadership and supportive leadership. *Journal of Occupational and Organizational Psychology,* 79(1), 37–61.

Regnier, K., Gross, M., & Zimmerman, R. (1992). *The Interpreter's Guidebook: Techniques for Programs and Presentations*. Interpreter's Handbook Series. Stevens Point: UW-SP Foundation Press.

Rogers, E. M. (1995). *Diffusion of Innovations* (Fourth Edition ed.). New York: Free Press.

Skibins, J.C., Powell, R.B., & Stern, M.J. (2012). Exploring empirical support for interpretation's best practices. *Journal of Interpretation Research*.

Sharpe, G.W. (1976). *Interpreting the environment*. New York: John Wiley & Sons.

Sherif, M., and C. Hovland (1961). *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. New Haven: Yale University Express.

Stern, M.J. (2008). Coercion, voluntary compliance, and protest: the role of trust and legitimacy in combating local opposition to protected areas. *Environmental Conservation,* 35(3): 200–210.

Stern, M.J. & Powell, R.B. (2011). *The Views of National Park Service Superintendents and Interpretation and Education Supervisors on Interpretation and Education in the National Park Service. Submitted to the National Education Council of the U.S. National Park Service*. October, 2011.

Stern, M.J., R.B. Powell, and K.S. Hockett (2011). Why do they come? Understanding interpretive program attendance at Great Smoky Mountains National Park. *Journal of Interpretation Research* 16(2): 35–52.

Tabachnick, B.G. and L.S. Fidell. (2007). *Using multivariate statistics* (5th ed.) Needham Heights, MA: Allyn & Bacon.

Taylor, P.A., Grandjean, B.D., Gramann, J.H. (2011). *National Park Service Comprehensive Survey of the American Public 2008–2009: Racial and Ethnic Diversity of National Park System Visitors and Non-Visitors*. Natural Resource Report NPS/NRSS/SSD/NRR-2011/432. Wyoming Survey & Analysis Center, University of Wyoming, Laramie, Wyoming. Online: http://nature.nps.gov/socialscience/docs/CompSurvey2008_2009RaceEthnicity.pdf.

Tilden, F. (1957). *Interpreting our heritage* (3rd ed.). Chapel Hill: The University of North Carolina Press.

Veverka, J. A. (1998). *Interpretive master planning: the essential planning guide for interpretive centers, parks, self guided trails, historic sites, zoos, exhibits and programs* (2nd ed.). Tustin: Acorn Naturalists.

Wallace, G. N., & Gaudry, C. J. (2002). An evaluation of the "authority of the resource" interpretive technique by rangers in eight wilderness/backcountry areas. *Journal of Interpretation Research*, 7, 43–68.

Wearing, S., & Wearing, B. (2001). Conceptualizing the selves of tourism. *Leisure Studies*, 7, 11–23.

Ward, C.W., & Wilkinson, A. E. (2006). *Conducting meaningful interpretation: A field guide for success*. Golden: Fulcrum.

Notes

1. Based on a review of web pages of all park units at the time of the research (www.nps.gov).
2. Our original research design also included administering shorter pre-experience surveys at different, but similar programs across the parks in our sample. These surveys contained two batteries of survey items that could be compared to the post-experience surveys to create a control group against which to compare outcomes. Unfortunately, an insufficient number of these surveys were administered at most parks to create a reliable control group. As a result, we did not include these data in further analyses.
3. Our field observations suggest that the association between the use of props and increased attrition may be influenced by cases in which not all visitors were able to engage with the prop(s). This may have motivated their departure.

Is It the Program or the Interpreter? Modeling the Influence of Program Characteristics and Interpreter Attributes on Visitor Outcomes

Robert B. Powell
Department of Parks, Recreation and Tourism Management and School of Agricultural and Forest Environmental Sciences, Clemson University

Marc J. Stern
Department of Forest Resources and Environmental Conservation, Virginia Tech

Abstract
This study modeled the relative influence of program characteristics and interpreter attributes on three visitor outcomes (satisfaction, visitor experience and appreciation, and behavioral intentions) using Structural Equation Modeling (SEM). The three resulting models accounted for between 10% and 27% (R^2) of the variance in the outcomes. The models suggest that both program and interpreter characteristics, as well as other unaccounted for factors, influence these outcomes. We discuss the implications of the findings for researchers and practitioners, calling for greater attention to both interpreter attributes and context.

Keywords
interpretation, communications, structural equation modeling, communication theory

Introduction
Much has been written regarding the interpretive techniques that should be employed to enhance visitor outcomes (e.g., Ham, 1992, 2013; Moscardo, 1999; Knudson, Cable, & Beck, 2003; Brochu & Merriman, 2002; Lewis, 2005). These interpretive techniques, which we refer to as program characteristics, are the focus of training efforts offered by organizations such as the National Park Service (see NPS, 2003a,b,c,d,e,f) and the National Association for Interpretation (NAI) as well as college courses offered around the world. These characteristics are believed to improve the quality of interpretive communications and to contribute to reaching desired outcomes, such as inspiring audiences to form intellectual and emotional connections with interpreted resources, influencing attitudes, and in some cases motivating behaviors. Researchers and field interpreters also recognize that there are

other factors, such as the attributes of the interpreter (confidence, charisma, enthusiasm, passion, apparent knowledge, etc.) that may influence the effectiveness of an interpretive program (e.g., Ham & Weiler, 2002a). However, these attributes are often overlooked in research and interpretive training. Stern and Powell (article 1, this issue) and Powell and Stern (article 3, this issue) investigated live interpretation programs provided by the U.S. National Park Service (NPS) to examine the relationship between 56 different interpretive practices and interpreter attributes and visitor satisfaction, enjoyment and appreciation for resources, and behavioral intentions. The results suggest that only certain program characteristics and interpreter attributes were significantly related with these outcomes. This study seeks to extend these findings by modeling the relative influence of these program characteristics and interpreter attributes on visitor outcomes using structural equation modeling (SEM).

The interpretive techniques promoted by professional associations and organizations have evolved over many decades and are based on experience, expert consensus, theory, and peer-reviewed research (Skibins et al., 2012). However, the empirical support for many of these best practices is largely anecdotal, because few studies to-date have attempted to isolate the influence of particular practices on outcomes through the use of experimental (or quasi-experimental) designs or comparative approaches (Skibins et al., 2012). The isolation of the influence of particular practices is challenging even with these designs, as program outcomes inevitably emerge from a dynamic interaction between the interpreter, the audience, the content, the setting/context, and the delivery (Powell et al., 2009, 2012; Archer & Wearing, 2003; Wearing & Wearing, 2001). Accounting for all factors seems a near impossibility. In this paper, we explore the relative influence of two of these elements, interpreter characteristics and program characteristics, on visitor outcomes. We define interpreter characteristics as those that may be entirely unique to the individual interpreter in any given context. These elements might include their mood, personality, or particular style of presentation. While program characteristics may also be highly dependent upon the interpreter, they could also be incorporated by design into a pre-packaged program, such as the sequence, content, theme, or logistics of the program.

We use structural equation modeling (SEM) for two reasons. First, the models give us a sense of the relative strength of influence of interpreter and program characteristics on visitor outcomes. The models can reveal the percentage of the observed variance in each outcome that can be explained by the predictors (Byrne, 2006). Second, the models allow for an examination of the interactions between interpreter characteristics and program characteristics. SEM also reveals the most parsimonious causal models for each outcome. As such, only the most predictive combination of variables remains in the final models. Examining which variables are present in the final models and their inter-relationships allows for consideration of the relative influence of program design vs. interpreter attributes. For example, if only interpreter characteristics are present in the final models, we would consider them dominant drivers of visitor outcomes. If both interpreter and program characteristics are present, it would support a view that outcomes are produced more by the interactions between interpreter and program design rather than by one or the other.

Methods

Selection of sites
We observed 376 diverse interpretive programs provided by 24 NPS park units across six regions of the NPS that generally reflected the current makeup of the NPS system (see Stern and Powell, this issue). The criteria for selecting NPS units included:

- Annual visitation greater than 35,000
- Geographic distribution across the county
- Variable distances from urban centers (urban, urban proximate, remote)
- Resource-base (cultural, natural, mixed)
- The ability to observe multiple programs in a short period of time
- Willingness to participate

The 24 selected units varied widely in terms of visitation, resource base, and locations, providing a reasonable sample from which to make generalizations regarding interpretation provided across the NPS system.

Sampling and data collection
Four researchers collected field data. Prior to each program one researcher conducted a short interview with the interpreter to collect demographic and background information regarding the program. During the program, this same researcher monitored 56 different program and interpreter characteristics and recorded these details on standardized observation sheets. After the program, we surveyed attendees that were age 15 or older using a standardized questionnaire. For programs with fewer than 50 participants, we attempted a census of all eligible attendees. In programs with more than 50 attendees, we systematically sampled attendees. From the 376 programs, we collected 3,603 surveys from visitors (for more detail, see Stern & Powell, this issue).

Data cleaning
Post-program surveys and program audits were coded and entered into Microsoft Access Database and Microsoft Excel to facilitate data entry. Data were then transferred to SPSS and EQS v6.1 software (Bentler, 2005) for screening and analyses. The visitor survey data were first screened for cases missing more than 50% of the items per factor (Tabachnick & Fidell, 2007). A total of 118 respondents were removed as a result. Data were then screened for univariate and multivariate outliers on outcome variables following Tabachnick and Fidell (2007) using Mahalanobis Distance (MAH) and studentized deleted residuals (SDRESID). A total of 58 cases were removed for exceeding +/- 3 standard deviations, or the criterion Mahalanobis Distance value (Fox 1991). This reduced our sample to 3,427 individual surveys from 376 interpretive programs.

Next we reviewed the number of valid respondents per individual interpretive program. Prior theory and research suggest that programs with a low number of attendees may be inherently different than programs servicing a larger number of attendees (Forist, 2003; McManus, 1987, 1988; Moscardo, 1999; Stern & Powell, this issue). We observed 272 programs with five or more attendees (see Stern & Powell,

Table 1. Outcome (dependent) variables with descriptive statistics.

Outcomes	N	Mean	St. Dev.
Satisfaction: 0 to 10 scale	272	8.94	0.64
Visitor experience and appreciation **(α=.89): 1 to 5 scale**	272	4.41	0.32
• Made my visit to this park more enjoyable		4.55	0.30
• Made my visit to this park more meaningful		4.49	0.32
• Enhanced my appreciation for this park		4.36	0.37
• Increased my knowledge about the program's topic		4.45	0.34
• Enhanced my appreciation for the National Park Service		4.27	0.36
Behavioral intentions **(α=.94): 1 to 5 scale**	272	2.92	0.64
• Changed the way I will behave while I'm in this park		2.92	0.67
• Changed the way I will behave after I leave this park		2.92	0.61

article 1, this issue for more extensive description). We chose this sample for the analyses conducted herein because it is most representative of programs in general and it provides a sample large enough to conduct structural equation modeling (Byrne, 2006). Because the program was our unit of analysis, our final step in data preparation included aggregating individual data at the program level by calculating the mean score of each visitor outcome for each program. For SEM purposes, all data was then grand mean centered (Tabachnick & Fidell, 2007)

Dependent variables: outcomes

Based on extensive input from the NPS and a review of literature, we developed three dependent variables (Table 1). The first dependent variable served as a measure of visitor satisfaction with the program on a scale from 0 to 10, with 0=Terrible and 10=Excellent. Two indexes were developed from other survey items following procedures outlined by DeVellis (2003) to represent visitor experience and appreciation and behavioral intentions. The items comprising each index were measured using a five-point Likert-type scale, with answer choices: Not at all (1), A little (2), Somewhat (3), A moderate amount (4), and A great deal (5). Composite scores were created for each of the scales by taking the mean of all items (for more detail, see Stern & Powell, this issue).

Program and interpreter characteristics

The independent variables used in this SEM analyses included both interpreter and program characteristics that met two criteria. We included ordinal variables that were correlated ($p < 0.01$) to the particular outcome in question in any context (See Stern & Powell, article 1, and Powell and Stern, article 3, this issue). We also included categorical variables with at least "moderate" effect size in association with the particular outcome in question in any context (Cohen's $d > 0.5$). The program characteristics (Table 2) were originally drawn from an extensive literature review aimed at identifying best practices in the field (see Skibins et al., 2012). The interpreter characteristics were developed from the communications and education literature, though many of these factors are also referenced in the interpretation literature (Table 3). The tables also contain descriptive statistics. For more detail, see Stern and Powell (this issue).

Table 2. Program characteristics, their definitions, and descriptive statistics.

Program characteristic	Definition
Organization (α = 0.82) Scale: 1 to 5 Mean: 3.34 S.D.: 0.71	**Equally weighted composite mean score of 6 program characteristics:** • **Quality of the introduction** (Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999): Degree to which the introduction captured the audience's attention and oriented (or pre-disposed) the audience to the program's content and/or message. • **Appropriate sequence** (Beck and Cable, 2002; Ham, 1992; Jacobson, 1999; Larsen, 2003): Degree to which the program followed a logical sequence. • **Effective transitions** (Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999; Larsen, 2003): Degree to which program used appropriate transitions that kept the audience engaged and did not detract from the program's sequence. • **Holistic story** (Beck and Cable, 2002; Larsen, 2003; Tilden, 1957): Degree to which the program aimed to present a holistic story (with characters and a plot) as opposed to disconnected pieces of information. • **Clarity of theme** (Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999; Knudson, Cable, and Beck, 2003; Larsen, 2003; Lewis, 2005; Moscardo, 1999; Sharpe, 1976; Veverka, 1998; Ward and Wilkinson, 2006): Degree to which the program had a clearly communicated theme(s). A theme is defined as a single sentence (not necessarily explicitly stated) that links tangibles, intangibles, and universals to organize and develop ideas. • **Link between introduction and conclusion** (Beck and Cable, 2002; Brochu and Merriman, 2002; Larsen, 2003): Degree to which program connected conclusion back to the introduction in an organized or cohesive way (i.e., program "came full circle.")
Connection (α = 0.88) Scale: 1 to 5 Mean: 2.77 S.D.: 0.78	**Equally weighted composite mean score of 5 program characteristics** • **Link tangibles to intangible meanings and universal concepts** (NPS Module 101; Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Knudson, et al., 2003; Larsen, 2003; Lewis, 2005; Moscardo, 1999; Tilden, 1957; Ward and Wilkinson, 2006): Communication connected tangible resources to intangibles and universal concepts. • Cognitive engagement (Knudson, et al., 2003; Moscardo, 1999; Sharpe, 1976; Tilden, 1957; Veverka, 1998): Degree to which the program cognitively engaged audience members in a participatory experience beyond simply listening; i.e. calls to imagine something, reflect, etc. • **Relevance to audience** (Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999; Knapp and Benton, 2004; Lewis, 2005; Moscardo, 1999; NPS Module 101; Sharpe, 1976; Tilden, 1957; Veverka, 1998): Degree to which the program explicitly communicated the relevance of the subject to the lives of the audience. • **Affective messaging** (Jacobson, 1999; Lewis, 2005; Madin and Fenton, 2004; Tilden, 1957; Ward and Wilkinson, 2006): Degree to which the program communicated emotion (in terms of quantity, not quality). • **Provocation** (Beck and Cable, 2002; Brochu and Merriman, 2002; Knudson, et al., 2003; Tilden, 1957): Degree to which the program explicitly provoked participants to personally reflect on content and its deeper meanings.
Appropriate logistics Scale: 1 to 4 Mean: 3.11 S.D.: 0.93	Degree to which basic audience and program needs were met (i.e., restrooms, weather, technology, accessibility, shade, etc). (Jacobson, 1999; Knudson et al., 2003)
Appropriate for audience Scale: 1 to 5 Mean: 3.93 S.D.: 0.70	Degree to which the program aligned with audience's ages, cultures, and level of knowledge, interest, and experience. (Beck and Cable, 2002; Jacobson, 1999; Knudson et al., 2003)
Multisensory Scale: 1 to 3 Mean: 2.39 S.D.: 0.51	Degree to which the program intentionally and actively engaged more than just basic sight and sound. (Beck and Cable, 2002; Knudson et al., 2003; Lewis, 2005; Moscardo, 1999; Tilden, 1957; Veverka, 1998; Ward and Wilkinson, 2006)
Physical engagement Scale: 1 to 4 Mean: 1.42 S.D.: 0.69	Degree to which the program physically engaged audience members in a participatory experience; i.e., through touching or interacting with resource. (Beck and Cable, 2002; Knudson, et al., 2003; Lewis, 2005; Moscardo, 1999; NPS Module 101; Sharpe, 1976; Tilden, 1957)
Verbal engagement Scale: 1 to 5 Mean: 2.51 S.D.: 1.02	Degree to which the program verbally engaged audience members in a participatory experience; i.e., dialogue (a two-way discussion). (Knudson, et al., 2003; Moscardo, 1999; Sharpe, 1976; Tilden, 1957; Veverka, 1998)
Fact-based messaging Binary: 27%	Program communicated only fact-based information. (Frauman and Norman, 2003; Jacobson, 1999; Lewis, 2005; Tilden, 1957; Ward and Wilkinson, 2006)
Clear message Scale: 1 to 4 Mean: 2.20 S.D.: 0.94	Degree to which program's message(s) was clearly communicated; i.e., the "so what?" element of the program. (Beck and Cable, 2002; Brochu and Merriman, 2002; Ham, 1992; Jacobson, 1999)
Consistency Scale: 1 to 3 Mean: 2.88 S.D.: 0.37	Degree to which the program's tone and quality were consistent throughout the program. (Beck and Cable, 2002; Ham, 1992)

Table 3. Interpreter characteristics observed in the study, their definitions, and descriptive statistics for cases analyzed in this paper.

Interpreter characteristic	Definition
Confidence (a = 0 .70) Scale: 1 to 4 Mean: 3.28 S.D.: 0.49	**Equally weighted composite mean score of 3 interpreter characteristics**: • **Comfort of the Interpreter** (Lewis 2005; Moscardo, 1999; Ward and Wilkinson, 2006): Degree to which the interpreter presenting the program seems comfortable with the audience and capable of successfully presenting the program without apparent signs of nervousness or self-doubt. • **Apparent knowledge** (Ham and Weiler, 2002a; Lewis, 2005; Ward and Wilkinson, 2006): The degree to which the interpreter appears to know the information involved in the program, the answers to visitors questions, and has local knowledge of the area and its resources. • **Eloquence** (Lewis, 2005): The extent to which the interpreter spoke clearly and articulately, and did not mumble or frequently use filler words such as "um" or "like."
Authentic emotion and charisma (a = 0.85) Scale: 1 to 5 Mean: 3.57 S.D.: 0.85	**Equally weighted composite mean score of 3 interpreter characteristics**: • **Passion** (Beck and Cable, 2002; Ham and Weiler, 2002b; Moscardo, 1999): The interpreter's apparent level of enthusiasm for the material, as opposed to a bored or apathetic attitude toward it. The overall vigor with which the material is presented. • **Charisma** (Ward and Wilkinson, 2006): A general sense of the overall likeability/charisma of the interpreter, commonly recognized by seemingly genuine interaction with the visitors, including smiling, looking people in the eye, and having an overall appealing presence. • **Sincerity** (Ham, 2009): The degree to which the interpreter seems genuinely invested in the messages he or she is communicating, as opposed to reciting information, and seems sincere in the emotional connection they may exude to the message and/or the resource. In other words, the extent to which the interpretation was delivered through authentic emotive communication.
Responsiveness Scale: 1 to 3 Mean: 2.81 S.D.: 0.41	The extent to which the interpreter interacts with the audience, collects information about their interests and backgrounds, and responds to their specific questions and requests or non-verbal cues. (Jacobson, 1999; Knudson et al., 2003; Lewis, 2005)
Humor quality Scale: 1 to 4 Mean: 2.08 S.D.: 0.73	How funny is the interpreter overall? Does the audience react positively to the interpreter's use of humor and seem to enjoy it? (Ham and Weiler, 2002b; Knapp and Yang, 2002; Regnier et al., 1992)
Sarcasm Scale: 1 to 3 Mean: 1.23 S.D.: 0.46	The degree to which the interpreter used sarcasm (the use of mocking, contemptuous, or ironic language or tone) or self-deprecation that was not meant to be serious, as a part of presenting their program.
Audibility Scale: 1 to 3 Mean: 2.86 S.D.: 0.36	The extent to which the interpreter could be clearly heard and understood by the audience.
Impatience Binary: 1.8%	Exhibition of explicit impatience toward audience members.
Goal: Behavior Change Binary: 7%	Intention of the interpreter for the program to influence audience's behavior. (Ham, 2013)

Table 4: Variables included in hypothesized models for each outcome.

Variable	Satisfaction	Visitor Experience and Appreciation	Behavioral Intentions
Interpreter characteristics			
Audibility	X	X	
Authentic emotion and charisma	X	X	X
Confidence	X	X	X
False assumption about audience	X	X	
Goal: Behavior change			X
Humor quality	X	X	X
Impatience	X		
Responsiveness	X	X	
Sarcasm			X
Program characteristics			
Appropriate for audience	X	X	X
Appropriate logistics	X	X	X
Clear message	X	X	X
Consistency	X	X	
Connection	X	X	
Multisensory engagement	X		
Organization	X	X	
Verbal engagement	X	X	X
Fact-based messaging	X		

Structural equation modeling

We used structural regression modeling (a.k.a. path analysis), a form of SEM, to examine the influence of different program and interpreter characteristics on three outcomes. We used SEM for this analysis because it is confirmatory (as opposed to exploratory) in nature and requires the researcher to have an explicit hypothesized model; it can model measurement error, which reduces inaccuracies; it allows for the analysis of a complete multivariate model including direct and indirect effects and in this case it can assess causal relationships between independent variables and a dependent variable (Byrne, 2006; Kline, 2005). In this study, all independent variables are formative (as opposed to reflective). That is, they were observed and represent a specific practice or attribute that is thought to directly influence the dependent variables (see Kline, 2005; Diamantopoulis et al., 2008; Diamantopoulis & Winklhofer, 2001; Jarvis et al., 2003, Padsakoff et al., 2007 for further explanation).

We used the EQS v6.1 software (Bentler 2005) to perform the statistical analyses, which progressed in several stages. First, the data were screened for univariate and multivariate deviations from normality. Next, we used structural regression modeling to assess the causal relationships between independent variables and each dependent variable (three separate models). For each outcome, we began with a model that contained all interpreter and program characteristics that met the criteria described

above for that outcome. The starting list of program practices and interpreter attributes used in the hypothesized models are in Table 4. To develop the final structural regression models we used an iterative process in which diagnostics (modification indices: Lagrange Multiplier Test (LM), Wald Test) indicated potential modifications, including removal of independent variables from the model, to improve fit and parsimony.

Structural regression analysis provides multiple statistics that can be used to evaluate the "fit" of a specified model (Byrne, 2006). In this paper we report the Satorra-Bentler Scaled Chi-Square (S-B χ^2), Robust Comparative Fit Index (CFI), Standardized Root Mean Square Residual (SRMR), the Robust Root Mean Square Error of Approximation (RMSEA) and its associated 90% confidence interval (Bentler & Yuan, 1999; Byrne, 2006). The S-B χ^2, which should be interpreted like a χ^2, is reported because it corrects for the degree of kurtosis in the data (Satorra & Bentler, 1994). The Robust CFI accounts for non-normality in the data and is an "incremental or comparative fit index" that evaluates the change in fit between the hypothesized model and the "independence model" (Byrne, 2006, 97; Bentler, 1990; Kline, 2005, 140). The independence model assumes that all the variables in the model are unrelated. The CFI represents the total covariation in the data and is measured on a scale of 0 to 1 with values greater than .9 indicating an acceptable fit and values greater than .95 indicating an excellent fit (Byrne, 2006; Hu & Bentler, 1999). The SRMR statistic provides the average difference between the sample and the predicted correlation matrices and thus is not susceptible to non-normality (Byrne, 2006). The SRMR uses standardized values with the range of scores between 0 and 1; values less than .1 are considered acceptable and less than .05 are considered a good fit (Hu & Bentler, 1995; Kline, 2005). The Robust RMSEA also accounts for non-normality in the data and is based on the average lack of fit per degree of freedom; therefore, as the fit improves, the RMSEA decreases. As such, this measure is sensitive to the degrees of freedom and the complexity of the model (Byrne, 2006). Like the SRMR, the scores range between 0 and 1, with values of .05 to .08 deemed acceptable and values less than .05 considered excellent (Browne & Cudeck, 1993; Hu & Bentler, 1999).

Beta weights in structural regression models reflect the effect size of an independent variable on the dependent variable. R^2 values gauge the predictive validity of the structural model, explaining the proportion of the total observed variance in the dependent variable explained by the model. It is recommended to assess R^2 values independently of fit indices, as the latter do not pertain to predictive validity (Kline, 2005).

Results

Three models were created based on the list of variables in Table 4. All independent variables (interpreter and program characteristics) were first entered as direct predictors of each outcome. In each case, the initial fit of each model was deemed unacceptable (Byrne, 2006). Through an iterative process, we adjusted the models using diagnostics that indicate potential model changes that would improve fit and parsimony. This generally involves removing variables one at a time based on statistical indicators produced at each stage of the modeling process. As the iterative modeling continues, it also can include adding or changing the nature of relationships between variables. In the end, a single "best fit" model is produced that represents the most parsimonious and predictive model for each outcome. The resulting models are displayed in Figures 1, 2, and 3.

Figure 1 represents the final model pertaining to how the interpreter and program

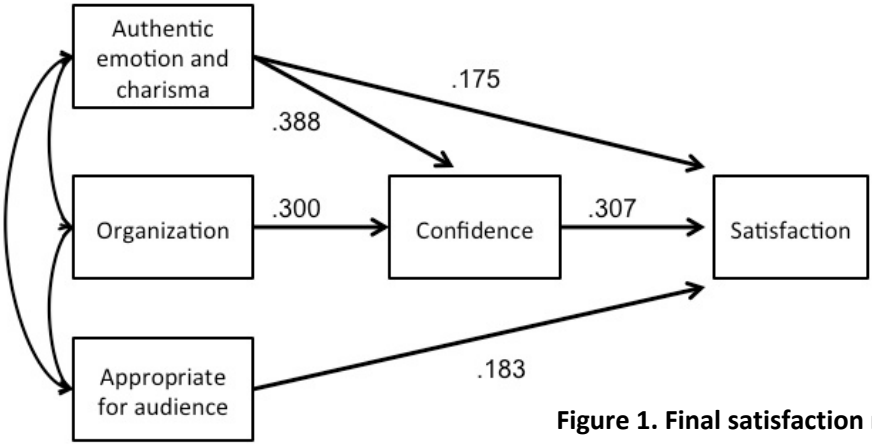


Figure 1. Final satisfaction model.

characteristics influenced *visitor satisfaction*. Fit indices for the final "satisfaction" model ($SB\chi^2$=5.39, p < .07; CFI=.99; SRMR=.029; RMSEA=.08) indicated the model was an acceptable representation of the relationships present in the data (Byrne, 2006; Kline, 2005). *Authentic emotion* was a strong predictor of interpreter's confidence (β=.388, p < .05) and a weaker predictor of visitor *satisfaction* (β=.171, p < .05). Organization was also a strong predictor of interpreter's *confidence* (β=.300, p < .05), but not a direct predictor of visitor *satisfaction*. Confidence was a strong predictor of visitor *satisfaction* (β=.307, p < .05). *Appropriate for the audience* was also a significant predictor of visitor *satisfaction* (β=.183, p < .05). The model accounted for 35% (R^2) of the variance in confidence and 27% (R2) of the variance in visitor *satisfaction*.

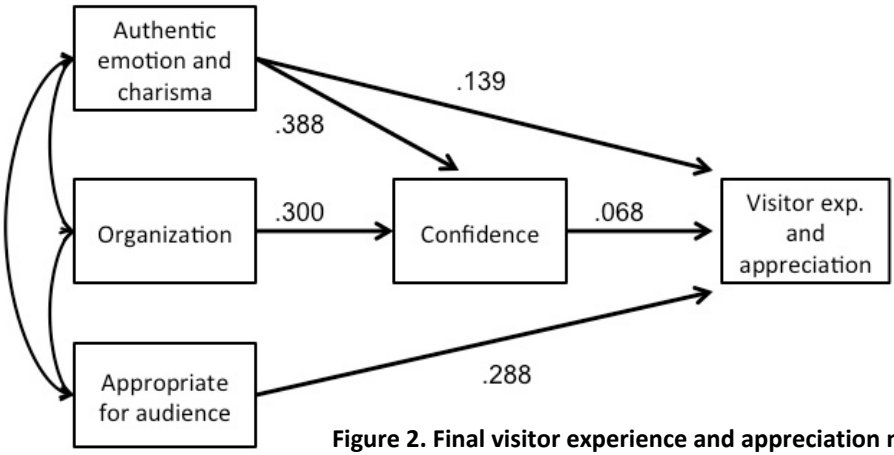


Figure 2. Final visitor experience and appreciation model.

The final structural regression model for visitor *experience and appreciation* had the same structure as the final visitor satisfaction model (Figure 2). Fit indices for the model ($SB\chi^2$=4.45, p < .1; CFI=.99; SRMR=.027; RMSEA=.069) indicated the model was an acceptable fit of the data (Byrne, 2006; Kline, 2005). The only structural differences between this model and the satisfaction model involved the relative strength

of *confidence* and *appropriateness for the audience*. *Appropriate for audience* (β=.288, p < .05) was the strongest predictor of visitor experience and appreciation, followed by *authentic emotion* (β=.139, p < .05) and *confidence* (β=.068, p < .05). The model accounted for 35% (R^2) of the variance in *confidence* and 16% (R^2) of the variance in *visitor experience and appreciation*.

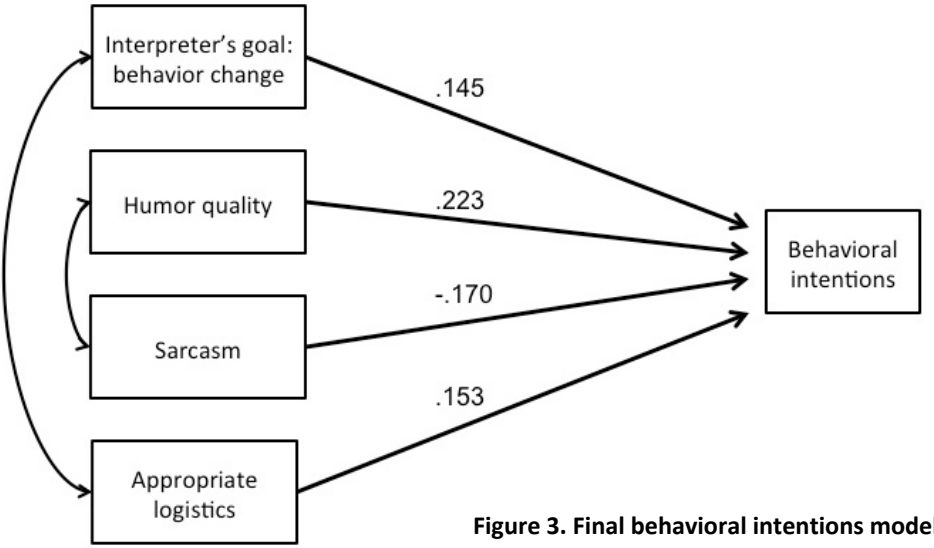


Figure 3. Final behavioral intentions model.

The model in Figure 3 represents how interpreter and program characteristics predicted *intentions to change behaviors*. Fit indices for the model in Figure 3 (SBχ^2=7.38, p < .05; CFI=.96; SRMR=.040; RMSEA=.03) indicated the model was an acceptable representation of the relationships present in the data. Having a *goal to influence behavior* (β=.145, p < .05), *appropriate logistics* (β=.153, p < .05), and *humor quality* (β=.223, p < .05) were significant positive predictors of intentions to *change behaviors*. Use of *sarcasm* (β=-.170, p < .05) was a significant but negative predictor of *intentions to change behaviors*. The model accounted for 10% (R^2) of the variance in intentions to change behaviors.

Discussion: Is it the interpreter or the program?

We used structural equation modeling to examine the relative influence of interpreter and program characteristics upon visitor outcomes at live interpretation programs across the U.S. National Park Service. The resulting models reveal three main lessons. First, it appears in each case that *both* interpreter and program characteristics influenced visitor outcomes. Second, depending on outcome, certain program practices and interpreter attributes provided the best model fit and predictive power. Third, the final models accounted for a relatively low percentage of the overall variance in visitor outcomes. We explain each finding and some important limitations in the interpretation of the analyses below.

In each model, both interpreter and program characteristics influenced outcomes. The *satisfaction* and the *visitor experience and appreciation* models each contained *authentic emotion and charisma, organization, confidence,* and *appropriate for the*

audience. In each model, *authentic emotion and charisma* and *organization* were mediated by *confidence*. In other words, the model suggests that authentic emotion and charisma and organization each help to create interpreter confidence, which in turn enhances visitor outcomes. *Authentic emotion and charisma* also served as a direct causal predictor of each outcome, as did the *appropriate for the audience variable*.

The final structural regression model of intentions to change behaviors suggests that *humor quality, appropriate logistics,* and *intending to influence* behaviors through a program positively influenced intentions to change stewardship behaviors. The use of *sarcasm* was associated with weaker intentions to change stewardship behaviors. In other words, interpreters that successfully employed *humor*, ensured that their audience's needs were met, and explicitly intended to influence their audience's behaviors were more successful at doing so. Meanwhile, overly sarcastic interpreters were less likely to influence changes in behavioral intentions. Interestingly, only 7% of all interpreters interviewed in the study explicitly intended to influence audience behaviors (Table 3). Ham (2013) reminds interpreters that outcomes, such as behavior change, do not happen magically; instead a program should be planned and developed with an outcome in mind. When focusing on behavior change, numerous techniques may increase the likelihood of influencing specific behaviors (Ham et al., 2007; Powell & Ham, 2008; Stern & Powell, this issue).

Certain limitations in the data and analyses are important to consider when interpreting these findings. First, structural equation modeling explicitly aims to produce the most parsimonious predictive or in this case causal model for selected outcomes. As such, independent variables that may be strongly related to outcomes are commonly removed during the modeling process due to their relationships with other variables. For example, the connection variable is highly correlated with organization, authentic emotion and charisma, and confidence (see Stern et al., this issue). As a result, it may be removed from a model because it explains a redundant proportion of the variance in the outcome as the other independent variables. This is the case with many of the program practices and interpreter characteristics tested in this analysis. It would be inappropriate to assume that their absence in the final models reduces their significance in influencing more positive outcomes.

Second, the models accounted for 10% to 27% of the variance in the outcomes. The strongest model accounted for 27% of the variance in *satisfaction*. The weakest model accounted for 10% of the variance in *behavioral intentions*. This suggests that much more is at play than simply the interpreter and the program elements. Interpretive programs are complex phenomena, and audience outcomes can be influenced by characteristics of the individual audience members, the makeup of the group, and the location and context of the program, in addition to characteristics of the program and the interpreter (Powell et al., 2009). Past research into communications (see Ajzen, 1992, for more) suggests that few consistent trends emerge when attempting to examine the range of source (interpreter) factors, receiver (audience) factors, channel (program) factors, and message (content) factors that influence outcomes resulting from communications. These factors vary with each program and produce an almost unlimited number of interactions and potential combinations (Falk, 2004). We examine a small portion of these additional factors in a separate article in this issue (Powell & Stern, this issue).

Relatively low R^2 values may also be a product of the lack of variance observed in *satisfaction* and *visitor experience and appreciation* scores. We discuss this issue in

greater depth in a separate article in this issue as well (Stern et al., this issue). Predictive ability may be particularly low for *behavioral change* for a number of reasons. As noted earlier, few programs actually targeted behavioral change as an outcome. As such, changing behavioral intentions may have been more of a side effect than an intended outcome of a program. Moreover, many interpretive program goers may already perform many of the behaviors discussed in interpretive programs, leaving little room for change (see Stern & Powell, this issue, for a more detailed discussion).

Despite the limitations, the results suggest that outcomes are influenced by both program *and* interpreter characteristics and that these characteristics interact and influence each other. For example, confidence may ultimately emerge from an interpreter's passion for the resource and careful planning, which leads to good organization. Because most prior research and formal training have focused on what we have categorized as "program characteristics" (Skibins et al., 2012), we urge future researchers, trainers, and practitioners to give some meaningful attention to interpreter attributes and delivery styles. Training programs might add elements that could improve interpreters' abilities to project confidence and authentic emotion. Some lessons for doing so might be found in the formal education field, where "affinity-seeking" and immediacy behaviors have garnered some attention (e.g., Finn et al., 2009). These practices involve efforts to ingratiate teachers with their students by reducing the social distance between them (see also Stern & Powell, this issue; Stern et al., in press). Interpretive organizations might also consider these findings in light of the role of the individual interpreter in program development. If organizations can provide opportunities for creating and sustaining authentic connections between interpreters and the resources they interpret, they might enhance interpreters' abilities to convey their own passions to their audiences. Finally, we urge researchers to consider how different program and interpreter characteristics may function differently in varying contexts.

References

Ajzen, I. (1992). Persuasive communication theory in social psychology: A historical perspective. In M. J. Manfredo (Ed.), *Influencing human behavior* (pp. 1–27). Champaign, IL: Sagamore Publishing.

Archer, D., & Wearing, S. 2003. Self, space, and interpretive experience: The interactionism of environmental interpretation. *Journal of Interpretation Research,* 8 (1): 7–23.

Beck, L., & Cable, T. T. (2002). *Interpretation for the 21st century: Fifteen guiding principles for interpreting nature and culture* (2nd ed.). Champaign: Sagamore.

Bentler, P.M. 1990. Comparative fit indexes in structural models. *Psychological Bulletin,* 107: 238–46.

Bentler, P.M. 2005. *Eqs 6 structural equations program manual*. Encino, CA: Multivariate Software (www.mvsoft.com).

Bentler, P.M. and K.H. Yuan. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research,* 34: 181–97.

Brochu, L., & Merriman, T. (2002). *Personal interpretation: Connecting your audience to heritage resources*: Fort Collins, CO: InterpPress.

Browne, M.W., & Cudeck, R. (1993). *Alternative ways of assessing model fit. In Testing structural equation models*, ed. K.A. Bollen and J.S. Long, 445–55. Newbury Park, CA: Sage.

Byrne, B.M. (2006). *Structural equation modeling with eqs: Basic concepts, applications and programming*. Second ed. Mahwah, NJ: Erlbaum.

DeVellis, R.F. (2003). *Scale development: Theory and applications Applied social research methods*. 2nd ed. Thousand Oaks, CA: Sage Publishing.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, 61(12), 1203–1218.

Diamantopoulos, A., & Winklhofer, H.M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2); 269–277

Falk, J. (2004). The director's cut: Toward an improved understanding of learning from museums. *Science Education*, 88(S1), S83–S96. doi: 10.1002/sce.20014

Finn, A. N., Schrodt, P., Witt, P. L., Elledge, N., Jernberg, K. A., & Larson, L. M. (2009). A meta-analytical review of teacher credibility and its associations with teacher behaviors and student outcomes. *Communication Education*, 58(4), 516–537.

Forist, B. (2003). *Visitor Use and Evaluation of Interpretive Media*. A Report on Visitors to the National Park System. National Park Service Visitor Services Project. http://nature.nps.gov/socialscience/ docs/Visitor_Use_and_Evaluation.pdf. Accessed May 24, 2012.

Fox, J. 1991. *Regression diagnostics*. Newbury Park, CA: Sage.

Frauman, E., & Norman, W.C. (2003). Managing visitors via "mindful" information services: One approach in addressing sustainability. *Journal of Park and Recreation Administration*, 21 (4): 87–104.

Ham, S. (2009). From interpretation to protection: Is there a theoretical basis? *Journal of Interpretation Research* 14(2): 49–57.

Ham, S. H. (1992). *Environmental interpretation: A practical guide for people with big ideas and small budgets*. Golden, CO: Fulcrum Publishing.

Ham, S. H. (2013). *Interpretation: Making a difference on purpose*. Golden, CO: Fulcrum Publishing.

Ham, S. H., Brown, T. J., Curtis, J., Weiler, B., Hughes, M., & Poll, M. (2007). *Promoting persuasion in protected areas: A guide for managers*. Developing strategic communication to influence visitor behavior. Southport, Queensland, Australia: Sustainable Tourism Cooperative Research Centre.

Ham, S.H., & Weiler, B.M. (2002a). Toward a theory of quality in cruise-based nature guiding. *Journal of Interpretation Research,* 7 (2): 29–49.

Ham, S H., & Weiler, B.M. (2002b). Tour guide training: a model for sustainable capacity building in developing countries. *Journal of Sustainable Tourism,* 10(1): 52–69.

Hu, L.-T., & P.M. Bentler. (1995). Evaluating model fit. In *Structural equation modeling: Concepts, issues, and applications*, ed. R. H. Hoyle, 76–99. Thousand Oaks, CA: Sage.

Hu, L.-T., & P.M. Bentler. (1999). Cutoff criteria for fit indices in covariance structure analysis: Guidelines, issues, and alternatives. *Structural Equation Modeling,* 6: 1–55.

Jacobson, S. K. (1999). *Communication skills for conservation professionals*. Washington, D.C.: Island Press.

Jarvis, C.B., MacKenzie, S.B., & Podsakoff, P.M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(2); 199–218.

Kline, R.B. (2005). *Principles and practice of structural equation modeling.* 2nd ed. New York: The Guilford Press.

Knapp, D., & Benton, G. M. (2004). Elements to successful interpretation: A multiple case study of five national parks. *Journal of Interpretation Research*, 9(2), 9–25.

Knapp, D. & Yang, L. (2002). A phenomenological analysis of long-term recollections of an interpretive program. *Journal of Interpretation Research*, 7(2), 7–17.

Knudson, D. M., Cable, T. T., & Beck, L. (2003). *Interpretation of cultural and natural resources,* (2nd ed.). State College: Venture Publishing.

Larsen, D. L. (2003). *Meaningful Interpretation: How to Connect Hearts and Minds to Places, Objects, and Other Resources*. Fort Washington: Eastern National.

Lewis, W. J. (2005). *Interpreting for park visitors* (9th ed.). Fort Washington: Eastern National.

Madin, E. M. P., & Fenton, D. M. (2004). Environmental interpretation in the Great Barrier Reef Marine Park: An assessment of programme effectiveness. *Journal of Sustainable Tourism*, 12(2), 121–137.

McCroskey, J.C., Richmond, V.P., & Stewart, R.A. (1986). *One on one: The foundations of interpersonal communication*. Englewood Cliffs, NJ: Prentice-Hall.

McManus, P.M. (1987). It's the company you keep . . . The social determination of learning-related behaviour in a science museum. *The International Journal of Museum management and Curatorship,* 6: 263–270.

McManus, P. M. (1988). Good companions: More on the social determination of learning-related behaviour in a science museum. *International Journal of Museum Management and Curatorship*, 7(1), 37–44.

Moscardo, G. (1999). *Making visitors mindful: Principles for creating quality sustainable visitor experiences through effective communication*. Champaign: Sagamore.

National Park Service. (2003a). *Interpretive Development Program. Module 101: Fulfilling the NPS mission: The process of interpretation*.

National Park Service. (2003b). *Interpretive Development Program. Module 103: Preparing and presenting an effective interpretive talk*.

National Park Service. (2003c). *Interpretive Development Program. Module 210: Prepare and present an effective conducted activity.*

National Park Service. (2003d). *Interpretive Development Program. Module 220: Prepare and present an interpretive demonstration or other illustrated program*.

National Park Service. (2003e). *Interpretive Development Program. Module 230: Interpretive writing.*

National Park Service. (2003f). *Interpretive Development Program. Module 311: Interpretive media development*.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., & Podsakoff, N.P. (2007). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5) 879–903.

Powell, R.B., Brownlee, M.T.J., Kellert, S. R. & Ham, S.H. (2012) From awe to satisfaction: Immediate affective responses to the Antarctic tourism experience. *Polar Record*. 48(2): 145–156.

Powell, R.B., Kellert, S. R., & Ham, S.H. (2009). Interactional theory and the sustainable nature-based tourism experience. *Society and Natural Resources*, 22(8): 761–776.

Powell, R. B., & Ham, S. H. (2008). Can ecotourism interpretation really lead to pro-conservation knowledge, attitudes and behavior? Evidence from the Galapagos Islands. *Journal of Sustainable Tourism*, 16(4): 467–489.

Regnier, K., Gross, M., & Zimmerman, R. (1992). *The Interpreter's Guidebook: Techniques for Programs and Presentations*. Interpreter's Handbook Series. Stevens Point: UW-SP Foundation Press.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 399–419). Thousand Oaks, CA: Sage.

Skibins, J.C., Powell, R.B., & Stern, M.J. (2012). Exploring empirical support for interpretation's best practices. *Journal of Interpretation Research*. 17(1): 25–44.

Sharpe, G. W. (1976). *Interpreting the environment*. New York: John Wiley & Sons.

Stokols, D., & Altman,I. eds. (1987). *Handbook of environmental psychology*. New York: John Wiley & Sons.

Tabachnick, B.G. and L.S. Fidell. (2007). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon.

Tilden, F. (1957). *Interpreting our heritage* (3rd ed.). Chapel Hill: The University of North Carolina Press.

Veverka, J. A. (1998). *Interpretive master planning: the essential planning guide for interpretive centers, parks, self guided trails, historic sites, zoos, exhibits and programs* (2nd ed.). Tustin: Acorn Naturalists.

Wearing, S., & Wearing, B. (2001). Conceptualizing the selves of tourism. *Leisure Studies*, 7: 11–23.

Ward, C.W., & Wilkinson, A. E. (2006). *Conducting meaningful interpretation: A field guide for success*. Golden: Fulcrum.

Speculating on the Role of Context in the Outcomes of Interpretive Programs

Robert B. Powell
Department of Parks, Recreation and Tourism Management and School of Agricultural and Forest Environmental Sciences, Clemson University

Marc J. Stern
Department of Forest Resources and Environmental Conservation, Virginia Tech

Abstract
Based on data from 272 live interpretive programs conducted across 24 units of the U.S. National Park Service, we investigate the influence of context upon interpretive programs and visitor outcomes. We first examined whether outcomes vary based upon the size of the audience and its age makeup; program characteristics such as duration, topic, and type; and characteristics of the setting including proximity to urban centers, program location (indoor vs. outdoor), and resource quality. We then examine whether different program or interpreter characteristics operate differently in different contexts by examining their relationships to visitor outcomes in four context pairings: programs with mostly children vs. mostly adults in the audience; culturally focused vs. environmentally focused programs; programs conducted in remote vs. urban parks; and indoor vs. outdoor programs. The findings suggest that a small number of program and interpreter characteristics may operate differently within different contexts. Based on these results, we propose hypotheses regarding which program characteristics appear to be more or less beneficial (or harmful) to generating desired visitor outcomes in different contexts.

Keywords
interpretation, communications, evaluation

Introduction
The interpretive equation suggests that successful interpretation requires that an interpreter must have knowledge of not only the resource, but also of their audience (Lacome, 2003). With this knowledge, interpreters can select and use appropriate

techniques to make meaningful connections for visitors. In other words, interpretation is not a "one size fits all" prospect; selection and use of appropriate techniques depends upon the characteristics of the audience including their age, background, expectations, and motivations for attendance. Although not explicitly accounted for in the interpretation equation, setting and other context elements may also meaningfully influence interpretive programs and their outcomes (Larsen, 2003; Merriman & Brochu, 2005; Moscardo, 1999). Some suggest that characteristics of the setting, attributes of the resource, and the collective characteristics of the audience form integral parts of the interpretive experience and should be accounted for in the planning and implementation phases (Larsen, 2003; Merriman & Brochu, 2005; Moscardo, 1999).

The other articles in this special issue explore which interpretive techniques are most strongly associated with visitor outcomes across a wide range of programs. But do certain techniques or approaches work better or worse in particular contexts and with certain audiences? To what extent does "context" influence visitor outcomes? This paper explores interactions between the duration, topic, type, and setting of programs, the nature of the interpreted resources, the size and age makeup of the audience, and visitor outcomes. The results of this study support the idea that context matters. We explore data collected from 272 programs across 24 diverse units of the U.S. National Park Service to build speculative hypotheses about which interpreter and program characteristics may be more or less important in producing positive visitor outcomes in different contexts.

Interactional theory

Interpretive programs and resulting visitor outcomes can be thought of as an interaction between the characteristics of the audience, the site/setting, the interpreter, and the interpretive program (Archer & Wearing, 2003; Mayer & Wallace, 2008; Merriman & Brochu, 2005; Powell, Kellert, & Ham, 2009; Wearing & Wearing, 2001). This notion of interactions between humans and their social and physical environments influencing cognition and behavior is the main premise of interactional theory (Altman & Rogoff, 1987; Stokols & Altman, 1987). Through the lens of interactional theory, visitor outcomes associated with attending interpretive programs result from the interaction of the characteristics of the program, the interpreter, other audience members, and the setting in which the program occurs (Archer & Wearing, 2003; Arnould & Price, 1993; Falk & Deirking, 2000; Wearing & Wearing, 2001). This theoretical approach acknowledges that interpretive programs are complex and promotes a holistic view of the relationships between multiple factors that together produce experiential outcomes (Altman & Rogoff, 1987; Archer & Wearing, 2003; Brochu & Merriman, 2002; Wearing & Wearing, 2001).

Potential influences of context: Audience, program, and setting characteristics

Research and theory suggest that the makeup of the audience should influence the techniques that are used as well as the outcomes of a program (Ham, 2013; Larsen, 2003). Although it is assumed that audience size and the age ranges of an audience will influence the selection of interpretive techniques, few have examined which techniques work best for different audience makeups (from all children to all adults) or how audience makeup may influence outcomes. Coble and others (2013) provide one exception, finding that the presence of children in an audience reduced the formation

of intellectual and emotional connections made by audience members in U.S. National Park Service interpretive programs.

The bulk of the research on the effects of group size comes from the formal education literature and suggests that smaller class sizes in formal settings tend to produce improved student outcomes (Boozer & Rouse, 2001; Finn & Achilles, 1999; Glass, 1982). In informal settings, such as in the case of interpretation and environmental education, there is less conclusive evidence. Powell and others (2009) examined visitors who received interpretation while rafting down the Colorado River through Grand Canyon National Park and found that group size was negatively associated with knowledge gain. Coble and others (2013) also found that as group size increased, intellectual connections decreased in attendees to NPS interpretation. However, Stern and others (2008) investigated the influence of group size at a residential environmental education center for elementary school children and found that larger groups were associated with improved awareness and interest in discovery and learning.

It is often assumed that the longer someone engages with an interpretive opportunity, whether an exhibit or a live interpretive program, the better the outcomes. While some empirical research supports this assertion, most have studied the influence of the number of interpretive programs attended or the number of days of a residential program and not the influence of duration of a single live interpretive program (Stern et al., in press). For example, Powell and others (2009), Stern and others (2008), Ballantyne and Packer (2005), and Coble and others (2013) have all found that greater exposure led to more positive outcomes. Museum and exhibit visitor studies also support the notion that the longer one engages an exhibit or collection of exhibits, the better (Falk, 2004).

We found few studies that examined whether particular types of interpretive programs were more or less effective in producing positive audience outcomes. Coble and others (2013) found that interpretive films were not as successful at producing intellectual connections as other interpretive program types such as live interpretation, illustrated programs, exhibits, and other conducted activities; no other trends were found. Van Winkle (2012) also examined the differences between electronic audio vs. live interpretation and found no differences in learning outcomes. We also examined whether particular interpretive techniques were more effective in programs interpreting natural resources vs. cultural resources and were unable to find prior research.

Other factors that may influence cognitive, affective, and behavioral outcomes include park setting, program location, and quality of the resource (Archer & Wearing, 2003; Mayer & Wallace, 2008; Merriman & Brochu, 2005; Powell et al., 2009; Wearing & Wearing, 2001). We refer to "park setting" in this study as a description of where the park unit that provided the interpretation program falls on the urban to remote spectrum. Different park units in different settings have different resources and may attract different visitors, each arriving with different motivations. However, it is still unclear if certain program practices work better in particular locations.

Natural environments, as opposed to built or indoor environments, are thought to enhance affective outcomes such as interests, emotions, and attitudes; cognitive outcomes such as learning; and psychological restoration (Crompton & Sellar, 1981; Kahn & Kellert, 2002; R. Kaplan & Kaplan, 1989; R. Kaplan, Kaplan, & Ryan, 1998; Kellert, 2005; Stern, Powell, & Hill, in press). However, several reviews of the literature suggest that indoor settings can be more effective than outdoor settings and other non-traditional settings for producing certain student outcomes (Zelezny, 1999; Zink &

Table 1. Description and mean score of outcomes.

Outcomes	N	Mean	S.D.
Satisfaction	272	8.94	0.64
Visitor experience and appreciation **(Cronbach's α =.89)**	272	4.41	0.32
• Made my visit to this park more enjoyable		4.55	0.30
• Made my visit to this park more meaningful		4.49	0.32
• Enhanced my appreciation for this park		4.36	0.37
• Increased my knowledge about the program's topic		4.45	0.34
• Enhanced my appreciation for the National Park Service		4.27	0.36
Behavioral intentions **(Cronbach's α =.94)**	272	2.92	0.64
• Changed the way I will behave while I'm in this park		2.92	0.67
• Changed the way I will behave after I leave this park		2.92	0.61

Table 2. Description of context variables.

Context Variable, Definition, and Measurement	Mean or Frequency
Audience: Group size* Number of total participants	Mean= 48 Median=17
Audience: Ratio of children to adults Categorized the ratio of children to adults in the audience using 4 point scale: 1=Mostly Children; 2=Even Distribution; 3=Mostly Adults; 4=All Adults.	Mostly Children=25 (9%) Even Distribution=82 (31%) Mostly Adults=132 (49%) All Adults=29 (11%)
Program: Duration Duration of interpretation program defined by time in minutes.	Mean= 49 minutes
Program Topic Nature-focused, culturally-focused, or dual focus.	Natural=170 (63%) Cultural=70 (26%) Dual Focus=29 (11%)
Program Type Guided Walk/Tour, Activity, Demonstration, or talk/slideshow/presentation	Guided Walk/Tour=161 (59%) Activity=8 (3%) Demonstration=5 (2%) Talk/slideshow/presentation=98 (36%)
Setting: Urban-Remote Parks were categorized as urban (within the limits of metropolitan areas with < 50,000 residents), urban proximate (outside urban area, but within a 60 mile radius), or remote (60 miles or more from any metropolitan area).	Urban= 91 programs (33%) Urban-proximate= 50 programs (18%) Remote=131 programs (48%)
Setting: Location Indoors, outdoors, or both.	Indoors=55 (20%) Outdoors=195 (72%) Both Inside and Outside=22 (8%)
Resource quality Degree to which the resource where the program took place is awe inspiring or particularly iconic: 1= Unimpressive/generic; 2= Pleasant but not iconic; or 3= Contextually iconic or grandiose.	Mean=2.37 Iconic or grandiose=134 (49%) Pleasant but not iconic=104 (38%) Unimpressive/generic=34 (13%)
Intervening Variable: Unexpected negative event Any unexpected interruptions or emergencies during the program, such as a sudden change in weather, medical emergency, technical difficulties, or hazardous conditions that detracted from the quality of the program: 1=Occurred; 0=No Issues.	Bad Weather=9 (3%) Negative events=34 (13%)
Intervening Variable: Unexpected positive event An unexpected experience that occurred during the program, such as seeing charismatic wildlife or other unique phenomena that added significantly to the quality of the experience: 1=Occurred; 0=Did not occur.	Positive events=5 (2%)

* Analyses pertaining to group size used all 312 valid programs. Because we deemed programs with 5 or more attendees (n=272) to be different phenomena from programs with 5 or less attendees (n=40), all analyses pertaining to the other context variables used the sample of programs with 5 or more attendees.

Burrows, 2008). Therefore the influence of conducting live interpretation in indoor vs. outdoor locations and ascertaining which program practices work best in each may be more nuanced than previously thought.

Another aspect of the setting with potential to influence the outcomes of interpretation includes the quality of the resource and setting. Larsen (2003) suggests that the basis of most interpretation is a tangible resource, which has some iconic value that anchors the program. In fact, research suggests that some resources and settings with unique iconic or symbolic qualities may have powerful impacts on visitors affective, cognitive, and behavioral domains. For example, extreme aesthetic natural and built environments have been associated with peak, spiritual, extraordinary, and transformative experiences (S. Kaplan, 1993; Laski, 1961; Otto, 1958; Powell, Brownlee, Kellert, & Ham, 2012), increased feelings of satisfaction and enjoyment (Arnould & Price, 1993; Powell et al., 2012), enhanced ethical concern for nature and commitment to stewardship (Kellert, 1996; Powell et al., 2012), enhanced emotional and cognitive connections (Kellert, 2005; Powell et al., 2012), and feelings of awe and wonder (Kellert & Farnham, 2002; Powell et al., 2012). Expansive, grand, and austere landscapes also may promote feelings of humility, spirituality, and even fear (Brown & Raymond, 2007; Galagher, 1993; Heintzman, 2009; Heintzman & Mannell, 2003; Koecni, 2005; Powell et al., 2012; Williams & Harvey, 2001). Therefore it seems appropriate to examine whether the quality of a program's resource influences the participant's outcomes.

Finally, three intervening variables—the occurrence of accidents or other negative events; the occurrence of positive events, such as the sighting of a charismatic animal; and extreme weather—are also consider in this study because of their potential to influence the interpretive experience, and because they are considered largely outside the control of the interpreter and the audience (Powell et.al., 2009).

This study sought to better understand 1) the extent to which the context variables discussed above influence visitor outcomes and 2) whether certain forms of program delivery appear to work better or worse in particular contexts. These forms of program delivery are divided into interpreter characteristics and program characteristics and are described in detail in Stern and Powell (article 1, this issue).

Methods

We observed 376 live interpretation programs conducted by the NPS across 24 different park units. During these programs we recorded the occurrence and extent of a wide-range of characteristics pertaining to program practices, interpreter attributes, and context (audience, program, and setting). Program practices were drawn from an extensive literature review that identified recommended practices (Skibins, Powell, & Stern, 2012). Interpreter attributes were largely identified from a review of the communications and education literature, although many are also referenced in the interpretation literature (see Stern & Powell, article 1, this issue). For a complete list, see Stern and Powell (article 1, this issue).

Immediately after each interpretive program, we administered short questionnaires to attendees who were over the age of 15 to gauge the influence of these programs on three dependent variables (Table 1). The first dependent variable measured program attendees' level of satisfaction, using a single survey item that asked visitors to rate their overall level of satisfaction with the program they had just attended on a scale ranging from 0 ("terrible") to 10 ("excellent"). The second dependent variable, "visitor experience

and appreciation," was composed of five survey items. The third dependent variable, "behavioral intentions," was composed of two survey items that gauged the program's influence on attendees' intentions to change future behaviors in the park and at home. The items comprising the two scales were measured using a five-point Likert-type scale, with answer choices: Not at all (1), A little (2), Somewhat (3), A moderate amount (4), and A great deal (5). Composite scores were created for each of the scales by taking the mean of all items.

From the 376 live interpretation programs, 64 were eliminated from analyses because of missing data or low response rates. We then divided the remaining 312 programs into those that served fewer than five people (n=40) and those that served five or more (n=272) because literature suggests that small programs are inherently different phenomenon than larger programs (Forist, 2003; McManus, 1987, 1988). We use the five-and-over sample in this paper because of the larger sample size, except for when examining the influence of group size. In this study, the interpretive program served as our unit of analysis. Therefore, all dependent variables were aggregated to the program level by calculating the mean score for each program (Table 1). For further information regarding sampling, data collection, data cleaning, dependent variable development procedures, program practices, and interpreter characteristics see Stern and Powell (article 1, this issue).

The audience, program, and setting characteristics under investigation included two continuous variables, four categorical variables, and two ordinal variables (Table 2). The two continuous variables included group size and program duration. The four categorical variables included the program topic, the program type, the park setting, and the location of the program. The two ordinal descriptors—the ratio of children to adults in the audience and quality of the resource—were recorded by the researchers in the field. Finally three intervening variables—the occurrence of extreme weather, the occurrence of accidents or other negative events, and the occurrence of positive events—were recorded because of their potential for influencing the interpretive experience. Table 2 provides a definition for each variable, an explanation of its measurement, and the mean or frequency depending upon the type of variable.

Results

How did context influence outcomes?

We first examine whether particular context variables are directly related to different outcomes. In other words, do certain contexts tend to produce different results? We also examine whether certain program characteristics or interpreter delivery styles are more prevalent in different contexts.

Group Size: The number of attendees to the 312 interpretive programs included in this analysis ranged from one person to approximately 600 people. The mean audience size was 48 and the median number of attendees was 17. When examining the correlation between the size of the audience and outcomes, we found no consistent relationships with *satisfaction* or the *visitor experience and appreciation* program outcomes. However, as audience numbers increased, programs tended to record greater audience *intentions to change behaviors* ($r=.127$; $p=.031$). As audience sizes increased, interpreters also tended to score higher in *confidence* ($r=.237$; $p < .001$), *organization* ($r=.167$; $p=.002$) of their programs, and *humor quality* ($r=.213$; $p < .001$). However, they also tended to be

more *formal* (r=.346; p < .001) and provide less *physical* (r=-.140; p=.009) and *verbal engagement* (-.308; p < .001).

Ratio of children to adults: In programs with five or more attendees, 9% of the programs (n=25) had mostly children present; 31% (n=82) had roughly an equal mix of adults and children; 49% (n=132) had mostly adults; and 11% (n=29) had all adults. The higher the ratio of children to adults, the higher the *behavioral intentions* score (r=.182; p=0.003). In other words, the more children present, the more likely adult participants were to report that the program had changed their behavioral intentions. Programs with higher ratios of children to adults were more commonly *multisensory* (r=.143; p=.019) and contained elements of *novelty* (r=-.133; p=.029). Interpreters were more likely to share their own personal stories (r=.151; p=0.014) when more adults were present relative to children. Programs with all adults were more commonly solely *fact-based* than those where children were present (Pearson χ^2=7.6; p=.006).

Program duration: Advertised program lengths ranged from 15 minutes to four hours. Actual program lengths ranged from 10 minutes to three hours. The average program length was just under 49 minutes. No statistically significant relationships were observed between program duration and visitor outcomes.

Program focus: One-hundred and seventy (63%) of the programs focused primarily on cultural heritage; 70 (26%) had a primary focus on the natural environment. Twenty-nine (11%) had a dual focus. *Behavioral intentions* scores were statistically higher for nature-based programs (means: 3.05 vs. 2.84, t=2.2, p=0.026; Cohen's d=0.33). No other statistically significant differences were noted in overall outcomes. In interviews prior to the programs, interpreters were more likely to express *behavioral change* as an intended outcome for nature-focused programs as opposed to culturally focused programs (χ^2=7.4; p=.007).

Program type: Programs included guided walks and tours (n=161); talks, slide shows, and multi-media presentations (n=98); demonstrations (n=5); and activities (n=8). Guided walks/tours and stationary talks made up 95% of the programs we observed. No statistically significant differences in outcomes between program types were observed.

Urban vs. remote: Within our sample of programs with five or more attendees, 91 (33%) programs took place in urban parks, 50 (18%) took place in urban-proximate parks, and 131 (48%) took place in remote parks. There were no significant differences in outcomes based upon proximity to urban centers.

Indoors vs. outdoors: Seventy-two percent (n=195) of programs took place outdoors; 20% (n=55) took place indoors; and 8% (n=22) used both indoor and outdoor settings. *Visitor experience and appreciation* scores tended to be greater following programs that took place entirely outdoors when compared to programs that took place entirely indoors (means: 4.45 vs. 4.33; t=2.6; p=0.011; Cohen's d=0.36) or programs that had both indoor and outdoor components (means: 4.45 vs. 4.25; t=2.1; p=0.039; Cohen's d=0.55). Indoor programs also tended to have larger audiences than programs conducted outdoors (means: 171.79 vs. 24.87; t=8.8; p<.001; Cohen's d=0.95).

Resource quality: We rated the quality of the resource where the program occurred. Forty-nine percent of program resources were rated as iconic or grandiose; 38% were rated as pleasant but not iconic; and 13% were rated as unimpressive or generic. The mean on the scale was 2.37 (s.d.=0.69). The quality of the resource did not exhibit any consistent relationships with program outcomes.

Exceptional events: Thirty-four programs (13%) experienced negative events such as interruptions, technical difficulties, and accidents. Nine (3%) of the programs experienced notably bad weather. Only five programs (2%) experienced unexpected positive events, such as a rare animal sighting. We combined bad weather and negative events and conducted a means comparison between these programs and those without negative circumstances. Programs with negative circumstances (n=43) exhibited significantly lower satisfaction (means: 8.70 vs. 8.99; t=2.8; p=0.006; Cohen's d=0.33) and visitor experience and appreciation scores (means: 4.25 vs. 4.44; t=3.6; p <.001; Cohen's d=0.43) than programs without these distractions. The small number of programs that experienced positive unexpected events precluded further analysis.

Which programmatic practices and interpreter attributes appear to work better in different contexts?

To examine whether different programmatic practices and interpreter attributes influence outcomes better in particular contexts and settings, we split the sample in the following ways: programs with larger and smaller proportions of children in the audience, culturally focused vs. environmentally focused programs, programs conducted in remote vs. urban parks, and indoor vs. outdoor programs. To ensure adequate sample sizes, we used the sample of programs with more than five attendees for each analysis. We examined the relationships between interpreter and program characteristics and visitor outcomes within each context. We report only characteristics that show at least one statistically significant relationship with an outcome.

When a correlation coefficient for a particular program practice was significant in one context and not in another, we used *Fisher r to z transformation* to assess the significance of these differences. *Fisher r to z transformation* compares correlation coefficients of different groups, taking into account their respective sample sizes. The test yields a z-score and associated p-value. These statistics provide a more stringent criteria for distinguishing differences in correlation coefficients across the subsamples and helped us avoid Type I errors (cases in which a real relationship is assumed, but sufficient evidence is lacking to support it). We have bolded and shaded these significant differences (z-score at $p < 0.05$) in the subsequent correlation tables. To further evaluate differences in binary variables' relationships to outcomes, we only highlight instances where the mean score in one subsample is significant at $p < 0.01$ and the other is not statistically significant ($p > 0.05$). Our goal in these analyses is to take a conservative approach to identifying practices that appear to operate differently in different contexts. Because the sample sizes shrink rapidly as we split the data into subsamples, we acknowledge that the emergent patterns are speculative rather than definitive trends.

Adult audiences vs. audiences with children: Tables 3 and 4 summarize relationships between program and interpreter characteristics and visitor outcomes in programs with different ratios of children to adults in their audiences. The column labeled "adults"

Table 3. Correlation coefficients for programs with mostly adult audiences (n=161) vs. those containing an equal or larger proportion of children (n=107).

	Satisfaction		Visitor experience and appreciation		Behavioral intentions	
Characteristic	**Children**	**Adult**	**Children**	**Adult**	**Children**	**Adult**
Interpreter characteristics						
Audibility	***.317*** **	***.104***	***.290*** **	***.005***	.215 *	.034
Authentic emotion and charisma	.450 **	.403 **	.410 **	.199 *	.203 *	.192 *
Confidence	.523 **	.455 **	.386 **	.186 *	***.336*** **	***.096***
False assumption about audience	-.167	-.184 *	-.258 **	-.179 *	-.139	-.036
Humor quality	.313 **	.263 **	***.382*** **	***.099***	.199 *	.135
Humor quantity	.184	.100	***.236*** *	***-.043***	.099	.044
Personal sharing	.097	-.001	.174	-.068	.235 *	.101
Responsiveness	.302 **	.195 *	.267 *	.208 **	.000	.087
Program characteristics						
Appropriate for audience	.404 **	.267 **	.397 **	.313 **	***.365*** **	***.039***
Appropriate logistics	***.317*** **	***.038***	***.396*** **	***.055***	.279 **	.104
Clear message	.312 **	.229 **	.274 **	.101	.302 **	.167 *
Connection	.403 **	.308 **	.350 **	.180 *	.153	.141
Consistency	.374 **	.178 *	.316 **	.223 **	.028	.064
Multisensory engagement	.182	.240 **	.072	.169 *	.107	.134
Novelty	.213 *	.080	.090	-.042	-.066	.085
Organization	.380 **	.359 **	.278 **	.177 *	.122	.167 *
Physical engagement	.075	.078	.214 *	.029	.187	-.001
Surprise	.201 *	.101	.193 *	.116	.104	.142
Verbal engagement	.230 *	.227 **	.265 **	.192 *	.162	.170 *

** Significant at $p \leq 0.01$
* Significant at $p \leq 0.05$

Table 4. T-tests for programs with mostly children vs. mostly adult audiences.

	Satisfaction				Visitor experience and appreciation				Behavioral intentions			
	Children		Adult		Children		Adult		Children		Adult	
Program characteristics	**Mean diff.**	**t**	**Mean diff.**	**t**	**Mean diff.**	**t**	**Mean diff.**	**t**	**Mean diff.**	**t**	**Mean diff.**	**t**
Fact-based messaging	-0.52	-2.8**	-0.25	-2.3*	-0.24	-2.5*	-0.06	-1.1	-0.21	-1.6	-0.07	-0.6
Appropriate pace	0.73	4.2**	0.41	3.1**	0.25	2.6*	0.18	2.9**	0.35	2.2*	0.19	1.3

** Significant at $p \leq 0.01$
* Significant at $p \leq 0.05$

represents programs in which adults made up a clear majority of the audience (60% of programs). The column labeled "children" represents programs with an equal or greater number of children compared to adults (40% of programs). Only characteristics showing at least one statistically significant relationship with an outcome are presented. While several program practices and interpreter attributes were consistently important irrespective of audience, there were several that appeared to be only significant for audiences with a large number of children and were significantly different from the mostly and all adult subsample. To determine which of these differences might be the most meaningful, we conducted Fisher r to z transformations to compare the correlation coefficients of different groups. We have bolded and shaded these differences in Table 3 (and subsequent correlation tables) that yielded a statistically significant z-score at $p < 0.05$.

These analyses reveal that four characteristics had stronger relationships to outcomes in programs with more children than they did in programs with little or no children. *Confidence* of the interpreter was more strongly linked with positive changes in behavior intentions in programs with more children (z=2.01; p=0.01). Appropriate for the audience was more strongly linked with behavioral intentions as well (z=2.72;

Table 5. Correlation coefficients for natural (n=170) vs. cultural programs (n=70).

Characteristic	Satisfaction		Visitor experience and appreciation		Behavioral intentions	
	Natural	Cultural	Natural	Cultural	Natural	Cultural
Interpreter characteristics						
Audibility	.029	.221**	.014	.190*	.056	.120
Authentic emotion and charisma	.440**	.394**	.294*	.316**	.291*	.070
Confidence	.503**	.437**	.297*	.270**	.330**	.112
False assumption about audience	**-.368****	**-.040**	-.273*	-.133	-.206	-.041
Humor quality	.202	.277**	.150	.248**	.204	.131
Humor quantity	**-.024**	**.217****	**-.093**	**.198****	-.033	.039
Responsiveness	.207	.208*	.319**	.213*	.035	.015
Program characteristics						
Appropriate for the audience	.458**	.355**	.492**	.351**	.269*	.122
Appropriate logistics	.286*	.115	.222	.247**	.252*	.156*
Clear message	.310**	.243**	.212	.201**	.186	.128
Connection	.335**	.360**	.311**	.288**	.215	.090
Consistency	.302*	.271**	.319**	.253**	.131	.045
Multisensory engagement	.282*	.244**	.245*	.109	.183	.031
Novelty	.261*	.111	.147	-.069	-.029	-.009
Organization	.266*	.431**	.276*	.247**	.190	.128
Sarcasm	-.068	.128	-.083	.074	**-.322****	**-.049**
Surprise	.174	.130	.161	.134	.261*	.041
Verbal engagement	.290*	.212**	.457**	.177*	.247*	.089

** Significant at $p \leq 0.01$
* Significant at $p \leq 0.05$

Table 6. T-tests for cultural vs. natural programs.

Program characteristics	Satisfaction				Visitor experience and appreciation				Behavioral intentions			
	Cultural		Natural		Cultural		Natural		Cultural		Natural	
	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t
Fact-based messaging	-0.34	-2.6*	-0.31	-2.1*	-0.11	-1.9	-0.11	-1.3	0.01	0.1	-0.30	-1.9
Appropriate pace	0.52	3.8**	0.46	2.4*	0.17	2.5*	0.11	2.2*	0.29	2.1*	0.11	0.5
Use of props	0.07	0.5	0.13	1.0	0.01	0.1	0.17	2.2*	0.02	0.2	-0.04	-0.2

** Significant at $p \leq 0.01$
* Significant at $p \leq 0.05$

$p < 0.01$). Appropriate logistics and audibility were more strongly linked with *satisfaction* (z=2.30; p=0.01 and z=1.71; p=0.04, respectively) and visitor experience and appreciation (z=2.88; $p < 0.01$ and z=2.32; p=0.01, respectively) in programs with more children. *Humor quality* (z=2.40; $p < 0.01$) and humor quantity (z=2.25; p=0.01) were also more predictive of visitor experience and appreciation in programs with more children. Differences noted in t-tests did not meet our threshold.

In short, the results suggest that most of the key best practices identified in Stern and Powell (article 1, this issue) cut across contexts. However, certain program characteristics may be particularly beneficial with audiences dominated by children. These include exhibiting confidence, using humor, ensuring audibility, gearing program content and delivery style to the specific audience, and paying careful attention to appropriate logistics.

Natural vs. cultural focused programs: We ran a similar set of analyses for nature-focused vs. culture/history-focused programs (Tables 5 and 6). For this analysis, we removed programs with equally balanced nature-based and cultural-based content because of their small sample size (n=29). There were 70 nature-focused programs and

Table 7. Correlation coefficients for programs that took place in urban (n=91) vs. remote parks (n=131).

Characteristic	Satisfaction		Visitor experience and appreciation		Behavioral intentions	
	Urban	Remote	Urban	Remote	Urban	Remote
Interpreter characteristics						
Audibility	.238*	.159	**.267***	**.043**	.163	.000
Authentic emotion and charisma	.415**	.432**	.352**	.280**	.069	.262**
Confidence	.453**	.519**	.264*	.294**	.191	.265**
False assumption about audience	-.096	-.308**	-.189	-.259**	-.039	-.176*
Formality	-.046	-.132	-.259*	-.086	.100	-.039
Humor quality	.373**	.275**	.355**	.207*	.198	.141
Humor quantity	**.355****	**-.019**	**.372****	**-.061**	.163	-.027
Personal sharing	-.027	.060	.073	.044	-.024	.107
Responsiveness	.230	.235**	.213	.304**	.123	.120
Program characteristics						
Appropriate for the audience	.371**	.366**	.391**	.344**	.165	.233**
Appropriate logistics	.186	.162	.307**	.240**	.233*	.167
Clear message	.285**	.250**	.267*	.201*	.107	.202*
Connection	.394**	.364**	.270*	.285**	.080	.154
Consistency	.385**	.300**	.347**	.353**	.095	.022
Multisensory engagement	**.316****	**.076**	.076	.066	.047	.194*
Novelty	.276**	.084	.127	-.082	-.025	-.077
Organization	.466**	.307**	.239*	.245**	.178	.148
Sarcasm	**.290****	**.007**	**.259***	**-.070**	**.051**	**-.214***
Surprise	.109	.197*	.068	.190*	**-.150**	**.278****
Verbal engagement	.285**	.190*	.279**	.199*	.047	.147

** Significant at $p \leq 0.01$
* Significant at $p \leq 0.05$

Table 8. T-tests for programs that took place in urban vs. remote parks.

Program characteristics	Satisfaction				Visitor experience and appreciation				Behavioral intentions			
	Urban		Remote		Urban		Remote		Urban		Remote	
	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t
Fact-based messaging	-0.57	-3.5**	-0.35	-3.0**	-0.23	-2.5*	-0.10	-1.5	-0.06	-0.4	-0.21	-1.8
Appropriate pace	0.46	2.2*	0.43	3.4**	**0.19**	**1.8**	**0.23**	**3.2****	0.39	1.9	0.14	1.1

** Significant at $p \leq 0.01$
* Significant at $p \leq 0.05$

170 cultural/history-focused programs. The results suggest a consistent list of program elements that are significant in both natural and cultural programs. However, three interpreter characteristics appeared to have different influences on outcomes according to our criteria. *Humor quantity* was positively linked with *satisfaction* (z=1.69; p=.04) and *visitor enjoyment and appreciation* (z=2.03; p=.02) in cultural programs but not in nature-based programs. Making a *false assumption about the audience* was negatively related to visitor *enjoyment and appreciation* (z=-2.39; $p < 0.01$) in nature-based programs but not in cultural programs. Sarcasm (z=-1.97; p=.02) was negatively related to *behavioral intentions* in the nature-based programs but not cultural programs. Differences noted in t-tests did not meet our threshold.

In summary, it appears that making false assumptions about the audience and sarcasm may be more damaging to visitor outcomes in nature-focused programs than in cultural programs. Meanwhile, additional attempts at humor may have more positive influences on visitor outcomes in cultural programs as opposed to nature-based programs.

Table 9. Correlation coefficients for indoor (n=55) vs. outdoor (n = 195) programs.

Characteristic	Satisfaction		Visitor experience and appreciation		Behavioral intentions	
	Indoor	Outdoor	Indoor	Outdoor	Indoor	Outdoor
Interpreter characteristics						
Audibility	.052	.236**	.254	.152*	.134	.097
Authentic emotion and charisma	.284*	.442**	.221	.266**	.119	.180*
Confidence	.273*	.551**	**.093**	**.337****	.017	.199**
False assumption about audience	-.278*	-.163*	-.302*	-.189*	-.049	-.103
Humor quality	.145	.330**	.092	.222**	.115	.132
Responsiveness	.284	.194**	.183	.195**	.049	.037
Program characteristics						
Appropriate for the audience	.330*	.375**	.214	.368**	.149	.112
Appropriate logistics	.284*	.118	.427**	.148*	.190	.126
Clear message	.345*	.217**	.124	.116	.279*	.131
Consistency	.125	.290**	**-.080**	**.338****	-.099	.041
Connection	.286*	.332**	.117	.242**	.248	.055
Multisensory engagement	.145	.196*	-.188	.113	**-.107**	**.178***
Novelty	.045	.192**	-.164	.068	-.054	.024
Organization	.273*	.385**	**-.098**	**.297****	.001	.142*
Physical engagement	**-.266***	**.120**	**-.296***	**.141***	-.125	.080
Sarcasm	.068	.098	-.078	.043	-.003	-.210**
Surprise	.063	.174*	-.013	.179*	.047	.141*
Verbal engagement	.025	.228**	-.008	.182*	.023	.139

** Significant at $p \leq 0.01$
* Significant at $p \leq 0.05$

Table 10. T-tests for indoor (n=55) vs. outdoor (n = 195) programs.

Program characteristics	Satisfaction				Visitor experience and appreciation				Behavioral intentions			
	Indoor		Outdoor		Indoor		Outdoor		Indoor		Outdoor	
	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t	Mean diff.	t
Fact-based messaging	-0.58	-2.5*	-0.18	-1.7	-0.20	-1.7	-0.01	-0.3	-0.32	-1.6	-0.03	-0.3
Appropriate pace	**0.36**	**1.3**	**0.61**	**5.2****	**0.14**	**0.9**	**0.22**	**3.9****	-0.1	-0.3	0.25	2.1*

** Significant at $p \leq 0.01$
* Significant at $p \leq 0.05$

Urban vs. remote parks: Within our sample of programs with five or more attendees, 91 programs took place in urban parks, 50 took place in urban-proximate parks, and 131 took place in remote parks. Because of the small number of programs within the urban-proximate park subsample, we dropped this group from the analysis. We thus explored only differences between programs occurring in urban and remote park units. When examining the relationship between location, outcomes and program and interpreter characteristics, certain variables appeared more predictive of outcomes in certain areas.

Tables 7 and 8 summarize relationships between program and interpreter characteristics and outcomes in both urban and remote parks. Again, most previously identified "best practices" (Stern & Powell, this issue) cut across park types. However, four interpreter delivery styles and two program characteristics displayed potentially meaningful differences in their relationships to outcomes. Sarcasm showed more positive relationships with *satisfaction* (z=2.11; p=0.02) and *visitor experience and appreciation* (z=2.44; $p < 0.01$) in urban parks and a negative relationship with changes in *behavioral intentions* in remote parks (z=-1.94; p=0.03). *Surprise* exhibited more positive relationships with changes in behavioral intentions in remote park units (z=3.15; $p < 0.01$). *Humor quantity* was more positively linked with *satisfaction* (z=2.82; $p < 0.01$)

and *visitor experience and appreciation* ($z=3.26$; $p < 0.01$) in urban settings. *Multisensory engagement* was positively linked to *satisfaction* in urban settings ($z=1.01$; $p=0.04$), and audibility was more positively linked to *visitor experience and appreciation* in urban settings ($z=3.15$; $p=0.05$). Moreover, t-tests revealed that appropriate pace was more positively related to *visitor experience and appreciation* in remote settings than in urban settings.

In summary, sarcasm appears to be significantly more effective with audiences who visit urban parks than those who visit remote parks. In fact, it actually exhibited positive relationships with attitudinal outcomes (*satisfaction* and *visitor experience and appreciation*) in urban settings and a negative relationship with *behavioral intentions* in remote settings. Meanwhile, the element of surprise may be more effective for audiences who visit remote parks. Maintaining an *appropriate pace* may also be a more relevant concern for programs in remote parks than in urban parks. Focusing more heavily on *humor* and *multisensory engagement* may be more effective in urban settings. Moreover, *audibility* may be more of a meaningful issue in urban settings than in remote settings.

Indoor vs. outdoor programs: We also compared programs that took place indoors vs. programs that took place outdoors (Tables 9 and 10). For this analysis, we removed programs that took place both indoors and outdoors because of the small sample size ($n=22$). There were 55 programs that took place completely indoors and 195 programs that occurred solely outdoors. Six program and interpreter characteristics showed significantly different relationships with observed outcomes across the two contexts. *Confidence* ($z=1.65$; $p=0.05$), *consistency* ($z=2.76$; $p < 0.01$), and *organization* ($z=2.59$; $p < 0.01$) were each more strongly related to more positive *visitor experience and appreciation* in outdoor programs. *Physical engagement* exhibited a significant positive relationship with *visitor experience and appreciation* in outdoor programs and a significant negative relationship in indoor programs ($z=2.86$; $p < 0.01$). *Multisensory engagement* showed a more positive relationship with behavioral intentions in outdoor settings than in indoor settings ($z=1.84$; $p=0.03$). T-tests revealed that appropriate pace was more positively related to both *satisfaction* and *visitor experience and appreciation* in outdoor settings.

In summary, *confidence, consistency, organization,* and *pace* may be more important drivers of outcomes in outdoor settings than in indoor settings, though *confidence* and *organization* appear to be clearly important in both. Indoor audiences may less commonly feel comfortable with higher degrees of *physical engagement* when compared to outdoor audiences. *Multisensory engagement* was also more positively linked with changes in behavioral intentions for outdoor audiences than for indoor audiences. Finally, maintaining an appropriate pace was a better predictor of attitudinal outcomes (*satisfaction* and *visitor experience and appreciation*) in outdoor programs than it was in indoor programs.

Discussion and Conclusion

This study sought to better understand 1) the extent to which context influences outcomes for interpretive program attendees and 2) which program practices and interpreter attributes may work best in particular contexts. We first explored the potential influence of context. We examined the size of the audience and its age makeup, program characteristics such as duration, topic, and type, and characteristics of the setting including proximity to urban centers, program location (indoor vs. outdoor),

Table 11. Program and interpreter characteristics with different relationships to outcomes in different contexts.

Context	Satisfaction	Visitor Experience and Appreciation	Behavioral Intentions
More children in the audience	Appropriate logistics (+) Audibility (+)	Appropriate logistics (+) Audibility (+) Humor quality (+) Humor quantity (+)	Confidence (+) Appropriate for audience (+)
Nature-focused programs	False assumption about the audience (-)	False assumption about the audience(-)	Sarcasm (-)
Culturally-focused programs	Humor quantity (+)	Humor quantity (+)	
Urban parks	Audibility (+) Sarcasm (+) Humor quantity (+) Multisensory (+)	Audibility (+) Sarcasm (+) Humor quantity (+)	
Remote parks		Appropriate pace (+)	Surprise (+) Sarcasm (-)
Indoor programs	Physical engagement(-)	Physical engagement(-)	
Outdoor programs	Physical engagement (+) Appropriate pace (+)	Confidence (+) Consistency (+) Organization (+) Appropriate pace (+) Physical engagement (+)	Multisensory (+)

and resource quality by testing their relationship to three outcomes, *satisfaction, visitor experience and appreciation,* and *behavioral intentions*. In these analyses, there were several trends. First, we found that as group size increased, intentions to perform stewardship behaviors also increased. One explanation for this trend could be the exertion of normative pressure from peers or other audience members to change behaviors (see Ajzen, 1992; Ham et al., 2007). However, we did not test this hypothesis. Second, we found that as the number of children in an audience increased, intentions to change behaviors increased. One explanation for this trend may be that an audience with more children may foster intergenerational learning (Ballantyne, Fien, & Packer, 2001; Duvall & Zint, 2007). Also, programs that served audiences with more children tended to be less fact-based and were more commonly multisensory and novel. Theory and research on behavior change supports the notion that presenting facts, or attempting to increase knowledge, has little to do with whether someone will change their behavior (e.g., Ham, 2013; Stern & Powell, this issue). We also found that programs that occurred outdoors produced greater *visitor experience and appreciation* in their audiences. This finding supports the notion that outdoor settings may enhance more emotive and affective outcomes, such as enjoyment and appreciation in participants (e.g., Kahn & Kellert, 2002; R. Kaplan et al., 1998; Kellert, 2005). These outdoor programs also tended to have smaller audiences. This combination of a more intimate social environment coupled with an outdoor setting may further enhance outcomes.

To investigate and then develop hypotheses about whether certain practices might work better or worse in particular contexts, we split our sample of interpretive programs based on four contextual variables: programs with greater vs. lesser proportions of children in the audience; culturally focused vs. nature-focused programs; programs conducted in remote vs. urban parks; and indoor vs. outdoors programs. We compared

relationships between program practices and interpreter attributes and outcomes within each subsample. We then examined these differences using more stringent thresholds to determine which might be indicative of a potentially meaningful trend warranting the development of a hypothesis. Several trends emerged across these four comparisons. First, a consistent list of programmatic practices and interpreter attributes appear important for achieving better visitor outcomes across most contexts. These include *confidence, authentic emotion and charisma, organization, connection, verbal engagement, appropriate for audience, clear message, responsiveness,* and *fact-based messaging (negative*). These findings largely corroborate the results of our analyses in Stern and Powell, articles 1 and 4 this issue, and Powell and Stern, article 2 this issue. Despite the consistent performance of some program practices across context, we did identify program characteristics that appeared to perform differently in particular contexts (Table 11).

While most program and interpreter characteristics performed similarly in programs containing different adult-to-child ratios, certain characteristics appeared to be more beneficial with younger audiences. These included confidence, using humor, ensuring audibility, gearing program content and delivery style to the specific audience, and paying careful attention to appropriate logistics. Similarly, few potentially meaningful differences surfaced between nature-focused and culturally focused programs in terms of the characteristics most strongly associated with outcomes. Making false assumptions about the audience met with less positive attitudinal visitor outcomes (satisfaction and visitor experience and appreciation) and using sarcasm exhibited a negative relationship with changes in behavioral intentions in nature-focused programs. Meanwhile, humor met with more positive attitudinal visitor outcomes in cultural programs.

We found similar trends with the relative influence of sarcasm and humor when comparing urban vs. remote parks. Each exhibited stronger positive links with attitudinal outcomes in urban parks and sarcasm was negatively related to behavioral outcomes in remote parks. Focusing more heavily on humor and multisensory engagement may be more effective in urban settings. Moreover, audibility may be more of a meaningful issue in urban settings than in remote settings. Our analyses suggest that maintaining an appropriate pace may not only be more important in remote settings as opposed to urban settings, but also in outdoor settings as opposed to indoor settings.

Confidence, consistency, organization, and pace may also be more important drivers of outcomes in outdoor settings than in indoor settings, though confidence and organization appeared to be clearly important in both. Physical engagement was positively linked to attitudinal outcomes in outdoor programs and negatively associated with the same outcomes in indoor programs. This suggests that audiences of indoor programs may have different expectations than audiences of outdoor programs and may not be as comfortable with physical engagement.

Overall, our analyses suggest that most of the "best practices" identified in the broader sample (Stern & Powell, this issue) are important regardless of context. However, some program and interpreter characteristics may operate differently in different settings and across contexts. However, we submit that *all of the contextual differences explained herein are speculative and would require additional targeted investigation to validate*. While we are confident that our overall sample represents a reasonable approximation of the diversity of interpretive programs across the NPS, we are less confident in the representativeness of each subsample. As our sample size is reduced, generalizability is weakened. As such, we suggest that *the results of these contextual analyses should*

be thought of as hypotheses that could be further investigated to test their validity. The results, however, suggest that we can be confident in saying: context matters! Thus we urge researchers to design studies that can refine our understanding of how context influences outcomes, and which program practices and interpreter attributes work best in particular contexts.

References

Ajzen, I. (1992). Persuasive Communication Theory in Social Psychology: A Historical Perspective. In M. J. Manfredo (Ed.), *Influencing Human Behavior* (pp. 1–28). Champaign, IL: Sagamore Publishing, Inc.

Altman, I., & Rogoff, B. (1987). World view in psychology: Trait, interactionist, organismic, and transactionalist approaches. In D. Stokols & I. Altman (Eds.), *Handbook of Environmental Psychology* (pp. 1–40). New York: John Wiley.

Archer, D., & Wearing, S. (2003). Self, space, and interpretive experience: The interactionism of environmental interpretation. *Journal of Interpretation Research,* 8(1), 7-23.

Arnould, E. J., & Price, L. (1993). River Magic: Extraordinary Experience and the Extended Service Encounter. *Journal of Consumer Research,* 20(1), 24–45.

Ballantyne, R., Fien, J., & Packer, J. (2001). Program Effectiveness in Facilitating Intergenerational Influence in Environmental Education: Lessons From the Field. *The Journal of Environmental Education,* 32(4), 8–15. doi: 10.1080/00958960109598657

Boozer, M., & Rouse, C. (2001). Intraschool variation in class size: Patterns and implications. *Journal of Urban Economics,* 50(1), 163–189.

Brochu, L., & Merriman, T. (2002). *Personal interpretation: Connecting your audience to heritage resources*. Fort Collins, CO: InterpPress.

Brown, G., & Raymond, C. (2007). The relationship between place attachment and landscape values: Toward mapping place attachment. *Applied Geography,* 27(2), 89–111. doi: http://dx.doi.org/10.1016/j.apgeog.2006.11.002

Crompton, J. L., & Sellar, C. (1981). Do Outdoor Education Experiences Contribute to Positive Development in the Affective Domain? *The Journal of Environmental Education,* 12(4), 21–29. doi: 10.1080/00958964.1981.9942638

Duvall, J., & Zint, M. (2007). A Review of Research on the Effectiveness of Environmental Education in Promoting Intergenerational Learning. *The Journal of Environmental Education,* 38(4), 14–24. doi: 10.3200/joee.38.4.14-24

Falk, J. (2004). The director's cut: Toward an improved understanding of learning from museums. *Science Education,* 88(S1), S83–S96. doi: 10.1002/sce.20014

Falk, J., & Deirking, L. D. (2000). *Learning from Museums*. Walnut Creek, CA: AltaMira Press.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis,* 21(2), 97–109.

Forist, B. (2003). *Visitor use and evaluation of interpretive media. A report on visitors to the National Park System*. National Park Service Visitor Services Project.

Galagher, W. (1993). The power of place. New York: Poseidon.

Glass, G. V. (1982). *School class size: Research and policy.* Beverly Hills, CA: Sage Publication.

Ham, S. H., Brown, T. J., Curtis, J., Weiler, B., Hughes, M., & Poll, M. (2007). *Promoting persuasion in protected areas: A guide for managers*. Developing strategic communication to influence visitor behavior. Southport, Queensland, Australia: Sustainable Tourism Cooperative Research Centre.

Heintzman, P. (2009). Nature-Based Recreation and Spirituality: A Complex Relationship. *Leisure Sciences,* 32(1), 72–89. doi: 10.1080/01490400903430897

Heintzman, P., & Mannell, R. C. (2003). Spiritual Functions of Leisure and Spiritual Well-Being: Coping with Time Pressure. *Leisure Sciences,* 25(2–3), 207–230. doi: 10.1080/01490400306563

Kahn, P. H., & Kellert, S. R. (Eds.). (2002). *Children and nature: Psychological, sociocultural, and evolutionary investigations*. Cambridge, MA: MIT Press.

Kaplan, R., & Kaplan, S. (1989). *The experience of nature: A psychological perspective*. Cambridge: Cambridge University Press.

Kaplan, R., Kaplan, S., & Ryan, R. L. (1998). *With people in mind: Design and management of everyday nature*. Washington DC: Island Press.

Kaplan, S. (1993). *The role of natural environment aesthetics in the restorative experience*. St. Paul, MN: U.S. Department of Agriculture, U.S. Forest Service.

Kellert, S. R. (1996). *The Value of Life*. Washington, D.C.: Island Press.

Kellert, S. R. (2005). *Building for life: Designing and understanding the human-nature connection*. Washington DC: Island Press.

Kellert, S. R., & Farnham, T. (Eds.). (2002). *The Good in Nature and Humanity: Connecting Science, Religion, and Spirituality with the Natural World*. Washington, DC: Island Press.

Koecni, V. (2005). The aesthetic trinity: Awe, being moved, thrills. *Bulletin of Psychology and the Arts,* 5, 27–44.

Lacome, B. (2003). The interpretation equation. In D. Larsen (Ed.), *Meaningful interpretation: How to connect hearts and minds to places, objects, and other resources.* Eastern National.

Larsen, D. (Ed.). (2003). *Meaningful interpretation: How to connect hearts and minds to places, objects, and other resources*. Eastern National.

Laski, M. (1961). *Ecstasy: A study of some secular and religious experiences*. London: Cressett Press.

Mayer, C. C., & Wallace, G. (2008). The interpretive power of setting: Identifying and protecting the interpretive postential of the internal and external setting at Copan Archaeological Park, Honduras. *Journal of Interpretation Research,* 13(2), 7–29.

McManus, P. M. (1987). It's the company you keep: The social determination of learning,Äêrelated behaviour in a science museum. *International Journal of Museum Management and Curatorship,* 6(3), 263–270. doi: 10.1080/09647778709515076

McManus, P. M. (1988). Good companions: More on the social determination of learning, related behaviour in a science museum. *International Journal of Museum Management and Curatorship,* 7(1), 37–44. doi: 10.1080/09647778809515102

Merriman, T., & Brochu, L. (2005). *Management of interpretive sites: Developing sustainable operations through effective management*. Fort Collins, CO: InterpPress.

Moscardo, G. (1999). *Making Visitors Mindful: Principles for Creating Quality Sustainable Visitor Experiences through Effective Communication*. Champaign, IL: Sagamore Publishing.

Otto, R. (1958). *The Idea of the Holy* (J. W. Harvey, Trans. second ed.). New York: Oxford University Press.

Powell, R. B., Brownlee, M. T. J., Kellert, S. R., & Ham, S. H. (2012). From awe to satisfaction: immediate affective responses to the Antarctic tourism experience. *Polar Record,* 48(02), 145–156. doi: doi:10.1017/S0032247410000720

Powell, R. B., Kellert, S. R., & Ham, S. H. (2009). Interactional theory and the sustainable nature-based tourism experience. *Society and Natural Resources,* 28(8), 761–776.

Skibins, J. C., Powell, R. B., & Stern, M. J. (2012). Linking interpretation best practices with outcomes: A review of literature. *Journal of Interpretation Research,* 17(1), 25–44.

Stern, M. J., Powell, R. B., & Hill, D. Environmental education program evaluation in the new millennium: What do we measure and what have we learned? *Environmental Education Research*. doi: 10.1080/13504622.2013.838749

Stokols, D., & Altman, I. (Eds.). (1987). *Handbook of Environmental Psychology.* New York: John Wiley & Sons.

Van Winkle, C. M. (2012). The effect of tour type on visitors' perceived cognitive load and learning. *Journal of Interpretation Research,* 17(1), 45–58.

Wearing, S., & Wearing, B. (2001). Conceptualizng the selves of tourism. *Leisure Studies,* 20, 143–159.

Williams, K., & Harvey, D. (2001). Transcendent experience in forest environments *Journal of Environmental Psychology,* 21(3), 249–260. doi: http://dx.doi.org/10.1006/jevp.2001.0204

Zelezny, L. C. (1999). Educational interventions that improve environmental behaviors: a meta-analysis. *The Journal of Environmental Education,* 31(1), 5–14. doi: 10.1080/00958969909598627

Zink, R., & Burrows, L. (2008). "Is what you see what you get?" The production of knowledge in-between the indoors and the outdoors in outdoor education. *Physical Education and Sport Pedagogy,* 13(3), 251–265. doi: 10.1080/17408980701345733

The Difference Between Good Enough and Great: Bringing Interpretive Best Practices to Life

Marc J. Stern
Department of Forest Resources and Environmental Conservation, Virginia Tech

Robert B. Powell
Department of Parks, Recreation and Tourism Management and School of Agricultural and Forest Environmental Sciences, Clemson University

Kevin D. McLean
Department of Forest Resources and Environmental Conservation, Virginia Tech

Emily Martin
Department of Parks, Recreation and Tourism Management, Clemson University

Jennifer M. Thomsen
Department of Parks, Recreation and Tourism Management, Clemson University

Bethany A. Mutchler
Department of Forest Resources and Environmental Conservation, Virginia Tech

Abstract
The purpose of this paper is to illuminate in both a quantitative and qualitative sense the practices that distinguish great interpretive programs from those that may merely be adequate to satisfy the visitor's basic desires to learn, be entertained, or spend time with a ranger. Great programs, like great works of art, have the potential to impact audiences in a deeper sense by providing memorable experiences that may influence multiple aspects of visitors' lives. This paper draws on experiences from three months of fieldwork, observing 376 interpretive programs across 24 units of the U.S. National Park Service, to illustrate examples of program elements that distinguished what we considered to be the best programs we observed.

Keywords
evaluation, eudaimonic satisfaction, hedonic satisfaction, interpretation, National Park Service, research

Introduction
Effective interpretation may produce multiple positive outcomes for program attendees. These include enhancing their knowledge and/or appreciation for the resource, site, or agency, influencing their behavior both on-site and off-site, and providing inspiration, both in a general sense and a more specific sense to enhance desires to explore further, learn more, or otherwise take self-directed action (Ham, 2009, 2013; Ward & Wilkinson, 2006). These outcomes may result from high-quality orientation, skills-building, persuasive communication, and/or effective storytelling that creates meaningful cognitive and emotional connections (Ham, 1992, 2009, 2013; Tilden, 1957; Ward & Wilkinson, 2006). But what makes the difference between good, or adequate, interpretation and great interpretation? This article serves as the final article in this special issue and focuses on this distinction, both in terms of the outcomes that might differentiate the two and the characteristics that appear to influence those outcomes.

Our research team observed 376 interpretive programs across 24 units of the U.S. National Park Service (NPS), tracking 56 independent variables that we later tested for relationships with outcomes measured in surveys administered to program attendees (Stern & Powell, this issue). The results indicated that certain practices and interpreter characteristics were statistically linked with more positive visitor outcomes. Perhaps the most striking finding of the study, however, was that over 85% of the people we surveyed rated the program they had attended an 8 or above on a 0-to-10 scale depicting their level of satisfaction. This led us to conclude that our results based on visitor surveys could not clearly distinguish good programs from bad programs. Rather, they could only identify characteristics that appear to move the scale from good to better.

Despite these consistently high ratings, our team witnessed dramatic variability in what we perceived to be the quality of these programs. In this paper, we draw upon our qualitative observations and an additional subjective quantitative measure made in the field by the research team about the overall quality of each program in an attempt to draw a clearer distinction between "good enough" and "great" interpretive programs.

We first explore theory relevant to understanding visitors' generally high levels of satisfaction in the study, elucidating the role of visitors' expectations on their evaluations of the programs they attended. Second, we contrast visitor expectations with their experiences, drawing a distinction between what it means to meet expectations and what it means to provide a more meaningful experience. We then present analyses of the factors that drove our own judgments of each program. Finally, we provide examples from our field notes of the factors that appear to distinguish between programs sufficient to satisfy visitors' basic expectations and those that might do something more.

Meeting expectations vs. making meaningful connections
Visitors' expectations may play two primary roles in their assessments of interpretive experiences (del Bosque & San Martin, 2008). First, they provide a basis for assessing performance. That is, a visitor's satisfaction can, in part, be based on the comparison of their experience with their pre-conceived notions. If the experience meets or exceeds

their valuation of that pre-conceived notion, we would expect a positive evaluation. Others suggest that expectations may also serve as a direct antecedent to satisfaction evaluations (Szymanski & Henard, 2001). This can be explained by Assimilation Theory (Sherif & Hovland, 1961) as well as the Theory of Cognitive Dissonance (Festinger, 1957). Individuals suffer cognitive dissonance (psychological conflict) when their experiences do not match their pre-conceived notions. In these cases, individuals make efforts to resolve the dissonance they feel. One common response is to adjust (or assimilate) their perception of the experience to match their pre-conceived notions. An example would be a family that saved up for an annual vacation that didn't meet their expectations, yet convinced themselves that the vacation was still well worth the effort expended. Del Bosque and San Martin (2008) investigated these two roles of expectations in tourism satisfaction and found that expectations in this latter sense were the dominant drivers of satisfaction. Expectations in the comparative sense were linked to positive and negative emotions. However, positive emotions were more strongly based on expectations alone, rather than how well the program met those expectations.

Understanding motivations for program attendance can help shed light on the likely expectations of attendees. Stern, Powell, and Hockett (2011) explored the primary motivations of visitors at Great Smoky Mountains National Park for attending interpretive programs. The most common motivations included entertainment, a chance to see something the visitor might otherwise miss, accommodating others in the visitor's group, and learning more about a specific topic or place. Other researchers have uncovered similar motivations for program attendance (Veverka, 1978; Srisomyoung, 2000; Galloway, 2002; Irving, 1986; Packer, 2004). These motivations provide insights into the probable expectations of the program attendees in our recent study—that programs should draw visitors' attention to unique resources in an entertaining and educational way.

These basic expectations may be relatively easy to meet. As such, programs may not need to inspire, provoke, or have deep meaning for the visitor to achieve a basic level of satisfaction. Still, we witnessed during our time in the parks what we felt to be some dreadfully boring talks and others where the interpreter struggled to recall facts about the resources they were attempting to interpret. Del Bosque and San Martin's (2008) expectancy theory helps to explain why visitors might still rate a mediocre program with moderately high scores.

We also witnessed programs that brought visitors to tears or clear expressions of elation and/or epiphany. Other programs elicited obvious displays of satisfaction and clear expressions of what one might call "pleasant surprise" or basic "wow" moments indicative of interpreters' clearly exceeding visitors' expectations. Despite the clear differences we observed in visitor expressions, actions, and emotions on-site, only minor differences were apparent in quantitatively measured satisfaction and visitor experience and appreciation scores. However, our qualitative observations and the quantitative assessments shared in this paper indicate to us a meaningful difference between programs that produce basic short-term satisfaction versus those that might approach what positive psychologists call eudaimonic well-being (Ryan & Deci, 2001).

Hedonic vs. eudaimonic satisfaction

Ryan and Deci (2001) define two perspectives on assessing human well-being. The hedonic perspective is based on the short-term satisfaction of basic needs and desires (e.g., pleasure attainment and pain avoidance). The eudaimonic perspective is more akin to Maslow's (1943) concept of self-actualization and Tilden's (1957) concept of provocation. In the context of interpretive programming or other similar experiences, eudaimonic satisfaction goes beyond short-term pleasure and enjoyment toward touching the personal values and/or provoking the deeper thoughts of the audience member (Oliver & Bartsch, 2010; Wirth et al., 2012). Oliver & Bartsch (2010, p. 76) use the term "appreciation" to describe a eudaimonic audience response to a powerful movie as "an experiential state that is characterized by the perception of deeper meaning, the feeling of being moved, and the motivation to elaborate on thoughts and feelings inspired by the experience." This is similar to the psychological concept of elaboration, which is generally seen as a precursor to cognitive changes in a message recipient that can lead to long-term attitude or behavior change (Ham, 2009; Petty & Cacciopo, 1986). We posit that, like a great movie, excellent interpretation can lead to this eudaimonic state, and that this process delineates the space between adequate interpretation, which primarily satisfies short-term hedonic interests, and great interpretation.

In short, adequate, or even mediocre, interpretation may achieve substantial hedonic satisfaction, but great interpretation is also capable of achieving eudaimonic satisfaction. Like a great movie or work of art that stays with an audience in some form for days, months, or even years, great interpretation also has the potential to have meaningful influence on how audience members perceive the world after it (Ham, 2013). This distinction may be particularly relevant in interpretive programs in national parks, where visitor expectations may be quite basic for most interpretive program goers (Stern et al., 2011), particularly for those who have never been exposed to a "great" program.

While satisfying basic expectations (e.g., providing some degree of entertainment or satisfying a general curiosity) appears to be common (Stern & Powell, this issue), achieving more meaningful, eudaimonic impacts for the visitor may be more challenging. Yet, NPS training documents and various other textbooks, trainings, and guidance documents regularly reference the importance of meaningfully connecting audiences to resources in ways that go beyond mere knowledge provision (Brochu & Merriman, 2002; Ham, 1992, 2009, 2013; Knudson et al., 2003; Larsen, 2003; NPS Module 101; Lewis, 2005; Skibins et al., 2012; Ward & Wilkinson, 2006). Each program presents an opportunity to do so.

We focus the rest of this article on illustrating the characteristics that appear to differentiate programs that merely satisfy basic visitor expectations from those that seize the opportunity to move visitors toward eudaimonic satisfaction.

What makes a great program?

The research effort with which this paper is associated uncovered a number of specific practices that were statistically linked with enhanced visitor experience and appreciation, greater satisfaction, and even reported changes in behavioral intentions in some cases (Stern & Powell, this issue). These included both characteristics of the interpreter and of the program itself:

Interpreter characteristics

- Confidence (comfort, eloquence, and apparent knowledge)
- Authentic emotion and charisma (passion, sincerity, and charisma)
- Responsiveness of the interpreter to the audience's interests, questions, needs, etc.
- Audibility
- Avoiding a focus on knowledge gain as the program's central goal and communicating solely factual information
- Avoiding making uncertain assumptions about the audience

Program characteristics

- Appropriateness for the audience
- Organization (quality of introduction, appropriate sequence, effective transitions, holistic story, clear theme, link between introduction and conclusion)
- Connection (links to intangibles and universal concepts, cognitive engagement, relevance to audience, affective messaging, provocation)
- Consistency of tone and quality
- Clear message
- Appropriate logistics
- Verbal engagement
- Multisensory engagement
- Appropriate pace

To further explore the notion of separating good, or adequate, programs from excellent programs, we make use of an additional measure made by our research team in the field. Immediately following each program, the researcher observing the program scored its overall quality on a scale from 1 to 10. This score was based on two factors. The first was the researcher's personal opinion of the quality of the program. The second was based on the researchers' observations of audience responses. Did the interpretation achieve an appropriate response from the audience? Conversely, was the audience visibly disinterested? Each researcher witnessed more than 90 live interpretive programs over the course of the study. In an effort to ensure reliability, researchers were instructed to revisit their overall quality scores periodically throughout the field season to ensure that the scale provided adequate comparisons from program to program.

To keep consistent with our analyses of visitor responses (Powell & Stern, this issue; Stern & Powell, this issue), we limited this analysis to programs with five or more attendees. Scores ranged from 2 to 10, with a mean of 5.9 and a standard deviation of 1.9. Only three programs were rated a 10 out of 10. The research team collectively agreed that a score of eight represented a clear threshold for what we would consider to be excellent programs, as described in the eudaimonic sense above. Twenty-three percent of the programs we observed were placed into this category (scoring 8 or higher on the overall quality measure).

Table 1. Relationships between visitor-reported outcomes and researchers' overall assessments of program quality.

Visitor-reported outcomes	Pearson correlation with researchers' assessments	Comparisons of visitor-reported outcome scores with programs rated "excellent" (≥ 8) or less than excellent (< 8) by research team				
		Overall score	Means	t	p	Cohen's d
Satisfaction (0 to 10)	.543**	≥ 8	9.36	7.6	< .001	0.97
		< 8	8.83			
Visitor experience and appreciation (1 to 5)	.412**	≥ 8	4.54	3.7	< .001	0.56
		< 8	4.37			
Behavioral intentions (1 to 5)	.218**	≥ 8	3.08	2.3	.024	0.34
		< 8	2.87			

** p < .001

Table 2. Independent samples t-tests comparing means of characteristics for programs that were rated by the research team as "excellent" (≥ 8) or "less than excellent" (< 8).

Characteristic	Overall score	Means	t	p	Cohen's d
Authentic emotion and charisma (1 to 5)	≥ 8	4.38	12.1	< .001	***1.57***
	< 8	3.34			
Connection (1 to 5)	≥ 8	3.42	8.7	< .001	***1.29***
	< 8	2.56			
Organization (1 to 5)	≥ 8	3.94	8.2	< .001	***1.24***
	< 8	3.17			
Confidence (1 to 4)	≥ 8	3.66	9.2	< .001	***1.21***
	< 8	3.17			
Appropriate for the audience (1 to 5)	≥ 8	4.47	7.2	< .001	***1.12***
	< 8	3.78			
Humor quality (1 to 4)	≥ 8	2.59	6.5	< .001	***0.94***
	< 8	1.94			
Clear central message (1 to 4)	≥ 8	2.82	6.3	< .001	***0.90***
	< 8	2.02			
Verbal engagement (1 to 5)	≥ 8	3.15	6.1	< .001	***0.87***
	< 8	2.34			
Multisensory engagement (1 to 3)	≥ 8	2.70	5.8	< .001	***0.84***
	< 8	2.30			
Self-reported level of excitement of the interpreter prior to the program (0 to 10)	≥ 8	8.55	4.7	< .001	0.75
	< 8	7.08			
Humor quantity (1 to 5)	≥ 8	2.44	4.5	< .001	0.65
	< 8	1.99			
Surprise (1 to 3)	≥ 8	1.26	3.5	.001	0.60
	< 8	1.04			
Responsiveness (1 to 3)	≥ 8	2.96	4.8	< .001	0.58
	< 8	2.76			
Novelty (1 to 3)	≥ 8	1.39	3.6	.001	0.57
	< 8	1.12			
Multiple activities (1 to 4)	≥ 8	1.37	2.9	.005	0.50
	< 8	1.13			
Personal sharing (1 to 4)	≥ 8	1.95	3.5	.001	0.49
	< 8	1.60			
Appropriate logistics (1 to 4)	≥ 8	3.41	2.9	.004	0.45
	< 8	3.02			
Consistency (1 to 3)	≥ 8	2.97	3.3	.001	0.38
	< 8	2.85			
False assumption of the audience (1 to 3)	≥ 8	1.08	-2.5	.013	-0.34
	< 8	1.20			
Formality (1 to 5)	≥ 8	2.98	-2.4	.018	-0.34
	< 8	3.26			
Physical engagement (1 to 4)	≥ 8	1.61	2.4	.019	0.34
	< 8	1.37			
Not statistically related to achieving an excellent outcome rating (≥ 8): Prior experience of the interpreter, audibility, sarcasm, multiple viewpoints, quality of the resource.					

Table 3. Differences in binary characteristics of programs that the research team scored as "excellent" (≥ 8) or "less than excellent" (< 8).

Characteristic	Pearson χ^2	p	Direction of relationship
Interpreter identity: friend	***35.7***	***< .001***	***Positive***
Interpreter identity: encyclopedia	***13.6***	***< .001***	***Negative***
Fact-based messaging	***13.5***	***< .001***	***Negative***
Appropriate pace	***11.3***	***.001***	***Positive***
Interpreter's intended outcome: get audience to want to learn more	9.8	.002	Positive
Program 20% shorter than advertised	8.0	.005	Negative
Props	6.6	.010	Positive
Pace too slow	5.2	.023	Negative
Interpreter's intended outcome: increase knowledge of audience	5.0	.026	Negative
Not statistically related to achieving an excellent outcome rating (≥ 8):Location of park (urban vs. urban-proximate vs. remote), indoor vs. outdoor program, program 20% longer than advertised, pace too fast, questionable information, other intended outcomes (see Stern and Powell, this issue), whether interpreter was a volunteer, park ranger, or paid concessionaire, professional appearance, inequitable treatment of audience, impatience, interpreter identity: authority, bias, false attribution, unexpected negative or positive circumstances.			

Table 4. Binary logistic regression model predicting an "excellent" overall score (≥ 8) by the research team (Nagelkerke R^2 = 0.57).

		Predicted score		Percentage Correct
		< 8	≥ 8	
Observed score	< 8	191	12	94.1%
	≥ 8	19	40	67.8%
			Overall Percentage	88.2%
Predictors:			**p**	**Exp (β)**
Authentic emotion and charisma			< .001	4.2
Confidence			.034	3.9
Organization			.005	2.9
Appropriate for the audience			.010	2.6
Verbal engagement			.006	1.8

Our subjective assessments of overall quality were significantly correlated with each of the outcomes measured in the visitor surveys (Table 1). Moreover, scores above eight also showed strong statistical relationships with more positive visitor-reported outcomes, particularly for satisfaction and visitor experience and appreciation. As such, our subjective assessments were validated to some extent by the visitor surveys, yet they provide a far more sensitive measure of program quality, accounting for the enhanced expectations of more experienced interpretive program audience members.

Tables 2 and 3 show the results of t-tests and chi-square tests that examine the statistical differences in interpreter and program characteristics between programs that scored an 8 or above on our overall quality measure and programs scoring lower. Characteristics with statistical relationships are further explained in Stern and Powell (this issue) and in Tables 5 and 6. In Table 2, bold and italicized items are those with a "large" statistical effect on membership in the "excellent" category (Cohen's d > 0.8). In Table 2, bold and italicized items represent those with the smallest probability of occurring by chance (p < .001). These characteristics in each table generally mirror those that predicted better visitor-reported outcomes (Stern & Powell, this issue). In this case, however, they explicitly distinguish what we considered to be great programs from all others.

Table 5. Qualitative field notes describing interpreter characteristics observed during programs with statistically significant relationships with measured outcomes.

Characteristic	Examples
Characteristics comprising "confidence"	
Comfort of the Interpreter Degree to which the interpreter presenting the program seems comfortable with the audience and capable of successfully presenting the program without apparent signs of nervousness or self-doubt (Lewis 2005; Moscardo, 1999; Ward & Wilkinson, 2006).	HIGH: The interpreter used a very conversational tone when interacting with the audience. At each stop he would sit down on a fence post or lean against a sign while continuing his story. He asked visitors to stop him with questions and to suggest answers to various questions he posed. Following engagement with the audience (or any type of interruptions), he would continue his story seamlessly with effective transitions. LOW: The interpreter was clearly unnerved by a large crowd consisting of a mix of adults and very distracted children who were bored by the historical topic of the talk. He mentioned that Civil War history was not his area of expertise and struggled to remember certain numbers and facts. He was unable to answer most visitors' questions and did not maintain the large group very well when moving from location to location. He tried several times to stop visitors from leaving the program and looked clearly saddened each time more people left. LOW: The interpreter seemed very nervous and was visibly shaking and had to pause several times to collect thoughts and recall what came next. The interpreter apologized frequently for forgetting what she had scripted and relied on "um, yeah, and like" to fill in the gaps.
Apparent Knowledge The degree to which the interpreter appears to know the information involved in the program, the answers to visitors questions, and has local knowledge of the area and its resources (Ham & Weiler, 2002; Lewis, 2005; Ward & Wilkinson, 2006).	HIGH: Not only did the interpreter know facts and scientific details about every plant, but also stories about their connection to humans and how people have used them in the past. She answered every question posed by visitors, including scientific names, habitat ranges, and various vascular functions. She never paused before answering and appeared entirely confident in every response she gave. LOW: The interpreter attempted to tell us the name of the man who designed a certain memorial, the date it was commissioned, and who funded its construction, but could not remember any of these things. He referred to his notes continually throughout the program and sometimes spent an extended period of time looking through them, searching for a particular fact to share. When visitors asked questions, he would again refer to his notes and even then could rarely provide an answer. LOW: The interpreter mentioned halfway through the program that it was her first time giving it, which was evidenced by her difficulty recalling facts/figures, her regular use of notes, and long walks between stops without talking to visitors at all while she reviewed her notes.
Eloquence The extent to which the interpreter spoke clearly and articulately, and did not mumble or frequently use filler words such as "um" or "like" (Lewis, 2005).	HIGH: Each story told by the interpreter was clearly illustrated through a strong vocabulary and a purposeful use of words. Pauses were only used when necessary for effect and the interpreter never seemed unsure of what to say next. The manner of speaking was concise and to the point but conversational enough to not feel explicitly scripted. LOW: The interpreter said "like" often and used "um" as filler every time he paused or tried to think of an answer. He commonly used the phrase "y'know," followed by long pauses. He mumbled at times when he didn't seem confident in what he was saying. Visitors were visibly confused.
Characteristics comprising "authentic emotion and charisma"	
Passion The interpreter's apparent level of enthusiasm for the material, as opposed to a bored or apathetic attitude toward it. The overall vigor with which the material is presented (Beck & Cable, 2002; Ham & Weiler, 2002; Moscardo, 1999).	HIGH: The interpreter explicitly told us that he was excited to share information with us about the natural resources found within the park. He said things like "let me tell you why I love this plant so much" and "I bet you can see why this is such a cool place." He had the audience look at things and feel them, tell the group what they liked best about it, and share their own reasons why the park was so special to them. HIGH: The interpreter told us why the park makes him feel inspired, what he loves most about it, and makes him come alive. He had us reflect on our own feelings about the place by sharing stories. He jumped from rock to rock with an obvious excitement in his step and clearly couldn't wait to share his next story. When the topic called for a more somber and reflective tone he slowed down subtly, removed his hat, and reminded us why we should care about this place. LOW: This interpreter shared facts about the battles that unfolded in the park with a flat tone of voice, very quietly. At one point she apologetically said, "the Civil War isn't really my area of expertise, but it's worth knowing something about." She would point out things along the way and say "I think this is where ____ happened" or "some people find this interesting."

Charisma A general sense of the overall likeability/charisma of the interpreter, commonly recognized by seemingly genuine interaction with the visitors, including smiling, looking people in the eye, and having an overall appealing presence (Ward & Wilkinson, 2006).	HIGH: The interpreter was kind and smiling throughout the program, like a sweet grandmother figure telling stories about her childhood. The audience leaned in to hear what she had to say and observe what she was doing. Both the interpreter and audience had smiles on their faces throughout the program. HIGH: The interpreter had a deep laugh that put smiles on the faces of visitors. He used friendly, casual banter throughout the program to keep visitors engaged and to inquire about their specific interests and hobbies. Visitors were clearly engaged throughout the program because of his interactions. LOW: The interpreter had a very abrupt manner of speaking to visitors and sounded annoyed to have them on the program. He ignored questions entirely and clearly hurried through the program. He made no effort to engage the audience or carry on a conversation; rather, he seemed focused on presenting what he had prepared and getting away from visitors as soon as he was finished.
Sincerity The degree to which the interpreter seems genuinely invested in the messages he or she is communicating, as opposed to reciting information, and seems sincere in the emotional connection they may exude to the message and/or the resource. In other words, the extent to which the interpretation was delivered through authentic emotive communication (Ham, 2009).	HIGH: While leading a tour of a war memorial, this interpreter maintained a very solemn and respectful demeanor throughout. He told us about the hard work, sacrifice, and heartache of people at home and abroad that made the war effort possible. Upon entering the memorial, he removed his hat and stood silently for a moment to take it all in. As he talked about each feature of the memorial he would touch it gently and slowly shake his head. His emotional connection to the resource was clearly demonstrated. LOW: This interpreter spoke in a very monotone, droning manner. At each stop, she listed several facts and then moved on to the next stop. She didn't wait for visitors to observe or enjoy the various resources and seemed to have no interest in looking at them herself. She seemed bored. Her cold and scripted delivery of facts and numbers about the battle that took place there made her seem almost callous to the topic.
Individual interpreter characteristics	
Humor Quality How funny is the interpreter overall? Does the audience react positively to the interpreter's use of humor and seem to enjoy it? (Ham & Weiler, 2002; Knapp & Yang, 2002; Regnier et al., 1992).	HIGH: The interpreter poked fun at the notorious love life of a Civil War general. He told us about pranks that soldiers would play on one another and had us laughing. This helped the program not only avoid being far too sad/somber, but also connected us with the fact that these were regular people just like us. LOW: The interpreter tried to use corny jokes and silly metaphors throughout the program to get laughs out of the audience. The audience clearly did not find these funny. He relied so heavily on these jokes that the rest of his program was largely devoid of worthwhile information. The audience seemed tired and uninterested by the end of the program, but he kept cracking bad jokes anyway.
Responsiveness The extent to which the interpreter interacts with the audience, collects information about their interests and backgrounds, and responds to their specific questions and requests or non-verbal cues (Jacobson, 1999; Knudson et al., 2003; Lewis, 2005).	HIGH: The ranger talked to people ahead of the program to ask them about their specific interests in the tour. He addressed these particular interests on the tour and actually addressed the people by name who were interested in the topic to engage them directly. When asked a question, the ranger gave both the factual answer and another question, which caused the visitor to think. LOW: When a member of the audience raised their hand, the ranger simply said "Please hold all questions until the end of the program."
False Assumption of Audience *(negative impact)* At any point during the program, did the interpreter make assumptions of the audience's attitudes or knowledge that could have easily been false?	PRESENCE: The interpreter regularly referred to names and dates very specific to events during the Civil War. These were used without any further explanation. The interpreter rather assumed that the audience already had a fairly thorough knowledge of the Civil War. There was a small group of war "buffs" who seemed to follow and enjoy the program, but most of the rest of the audience seemed somewhat lost and disconnected without this extra knowledge.

Table 6. Qualitative field notes describing program characteristics observed during programs with statistically significant relationships with measured outcomes.

Characteristic	Examples
Characteristics comprising "organization"	
Intro Quality Degree to which the introduction captured the audience's attention and oriented (or pre-disposed) the audience to the program's content and/or message (Brochu & Merriman, 2002; Ham, 1992; Jacobson, 1999).	HIGH: Interpreter began the program by saying "It is the morning of the first battle of -_____. It's hot and muggy. You've just finished breakfast and you're preparing for a long march over these fields you see before you. But before the day is done, half of your company will be brought down by confederate cannon and musket fire..." This captured our attention, set the tone for the program, and led directly into the theme of the program. HIGH: As the program began, the ranger asked the visitors to close their eyes and imagine themselves transported back in time. She painted a picture with words, describing a battle at sea and the sound of munitions exploding all around. She caused visitors to jump when she yelled "Man overboard!" LOW: The interpreter arrived just in time to start the program and did not interact with the audience at all or provide any information about the program before it started. The first thing he said to the audience was "OK, let's get started," at which point he walked off to our first stop. When we arrived at the first stop, while much of the group was still walking, he started talking about trees and listing facts about them. There was no introduction to the talk, nothing to capture our attention, and nothing to let us know that we were even on the right program.
Appropriate Sequence Degree to which the program followed a logical sequence (Beck & Cable, 2002; Ham, 1992; Jacobson, 1999; Larsen, 2003).	HIGH: This program was about the life cycle of a giant sequoia tree. The program itself followed a storyline that described the life of a tree and everything it saw during its lifespan. Each stop was related to the next stage of life and provided a clear example of that stage. We moved from an area full of cones and seeds, to a stop with several tiny saplings, to young trees, and on up to full size giants. We followed the growth of a sequoia from birth to death and understood everything it must overcome in the process. HIGH: The interpreter discussed several different animals that lived within the park, using the food chain to pair an animal to each corresponding stop on the walk. Transitions were provided between each stop that described how each animal had an impact on the next, giving the program a clear sequence and appropriate clarity and demonstrating the complexity and hierarchy of the food web. LOW: The talk provided a random assortment of facts and stories about both the War of 1812 and the Civil War. Each stop was disconnected from the next and jumped back and forth between the two wars. There was no logical sequence to the stops and seemed to be representative of whatever was on the interpreter's mind at the time. At a single stop we talked about iron clad battleships during the Civil War and a tavern that was located on the grounds during the War of 1812 with no connection drawn between them or any of the other stops.
Transitions Degree to which program used appropriate transitions that kept the audience engaged and did not detract from the program's sequence (Beck & Cable, 2002; Brochu & Merriman, 2002; Ham, 1992; Jacobson, 1999; Larsen, 2003).	HIGH: As we prepared to leave each stop, the interpreter said "I want you to be on the lookout for _____ as we head to our next stop and think about how it relates to _____." This kept the visitors curious, engaged, and thinking about the theme of the talk even while the interpreter wasn't talking. These transitions provided a logical flow from the topic of one stop to the next. LOW: At each stop, the interpreter would talk for a bit and then just stop. We would walk to the next stop in silence and then he would pick up right where he left off. It felt very much as if he were stopping halfway through a paragraph, waiting a bit, and then continuing without any explanation of why we had moved. It likely would have been more effective to just stay in one place and deliver a talk, as these long pauses left the audience bored and distracted from the program itself.

Holistic Story Degree to which the program aimed to present a holistic story (with characters and a plot) as opposed to disconnected pieces of information (Beck & Cable, 2002; Larsen, 2003; Tilden, 1957)	HIGH: This interpreter used the unique and sometimes valuable natural resources of the park to illustrate why native people originally settled here, why it inspired people to move westward, how they used these resources to settle and live off the land, how this led to their over-exploitation, and ultimately to their protection. Each stop taught us about a new resource (trees, rock, grazing fodder, minerals, water, etc.) that played a part in this story. As we moved along, so too did the plot of the story being told. HIGH: The interpreter made it very clear that he wanted to tell us a story during the program to help us understand the people who once lived here. He introduced different historical figures (generally fictionalized composites of people from the time period) and told us a bit about them. He then used them as vehicles to demonstrate the historical significance of what happened in the area and how the daily lives of people were affected by these events. The story progressed linearly through time and each stop represented a new time period. Every stop was tied back to the central theme and was relevant to the story being told. He used the repetition of certain ideas and interactions with the audience to build a story that came to its conclusion at our last stop. LOW: The talk was a jumble of dry facts about an otherwise interesting animal. There were several moments of "Hmm, what else can I tell you..." LOW: During the tour of a historical home, the interpreter listed off different facts and stories as we walked through each room. A piece of furniture or book would cause her to say "Oh, this reminds me about..." None of what she told us seemed to be connected, and although the facts were interesting, she did not tell us a story about the place or why it was worth preserving. The greatest focus was on which furniture pieces were original or reproductions rather than on the people who lived there and their stories. LOW: As we wandered along the path of our guided walk, the interpreter pointed out random trees, buildings, or objects. Each one was described in a manner unrelated to the last. There was no clear topic or point to the talk and visitors seemed disconnected and bored by the talk. LOW: The ranger provides a description of a native species that can be found in the park, detailing its appearance, unique traits, and status as a threatened species. The ranger continues working his way through species after species.
Clear Theme Degree to which the program had a clearly communicated theme(s). A theme is defined as a single sentence (not necessarily explicitly stated) that links tangibles, intangibles, and universals to organize and develop ideas (Beck & Cable, 2002; Brochu & Merriman, 2002; Ham, 1992; Jacobson, 1999; Knudson et. al, 2003; Larsen, 2003; Lewis, 2005; Moscardo, 1999; Sharpe, 1976; Veverka, 1998; Ward & Wilkinson, 2006)	HIGH: This program focused on the power of this particular site and the influence it has had in so many people's lives throughout time. The interpreter described how it had a spiritual power for native people, was a place of unrivaled beauty and reflection for early explorers, and a place of relaxation and escape for people today. Every stop supported the idea that the park is a unique and powerful place worth preserving, which he reinforced by reminding us that future generations have a right to experience and gain from this place. LOW: The interpreter on this program told us explicitly that he was going to tell us why a historical building was a unique place. We then walked around and through the hall. He told us where various treaties were signed and where historical figures sat. This was the extent of the program. He did not tell us how those documents have shaped our history, what role those figures played in founding our country, or why preserving the building itself should matter to us. The program was a collection of dates and names, but little more.
Intro/ Conclusion Linkage Degree to which program connected conclusion back to the introduction in an organized or cohesive way (i.e., program "came full circle") (Beck & Cable, 2002; Brochu & Merriman, 2002; Larsen, 2003)	HIGH: Before our first stop, the ranger told us a bit about what we were going to learn and why it was important to know. He taught us some basic facts about the war, how it came to the area, and some key players in the battles, but mostly he focused on the story of one young man and how the war affected him. We stopped at the house where the young man grew up, learned about the kind of education he received, and the trade he learned in his youth. Our final stop took us into a large cemetery, where the ranger pointed out all the other young men who had been buried there. Then he looked down at his feet and pointed out the grave we were standing around: the final resting place of the very man we had spent the past hour learning about. The sadness we all felt was very real and he had taken us full circle to truly connect us to the people and events here. LOW: The interpreter went so far past the designated end time of the program that he did not get the chance to wrap it up in any way. Visitors had to leave the program while he was still talking so they could catch the bus back to the visitor center. LOW: While it seemed like the interpreter was in the middle of his talk, he simply stopped, looked at the audience, and said "ok, well that's it." The program ended very abruptly, with no conclusion at all, leaving the audience wondering what the point of the program was. He had all the opportunity in the world to tie things together and leave us with a lasting message to think about.

Characteristics comprising "connection"	
Cognitive Engagement Degree to which the program cognitively engaged audience members in a participatory experience beyond simply listening; i.e. calls to imagine something, reflect, etc. (Knudson et al., 2003; Moscardo, 1999; Sharpe, 1976; Tilden, 1957; Veverka, 1998).	HIGH: The interpreter asked visitors to consider whether former inhabitants could have imagined what this valley is like today and whether the audience could imagine what it would be like in the future. The interpreter asked us to picture how the valley has changed over time and how strange and foreign it would look to us 100 or 1,000 years from now. HIGH: The walk focused much of the audience's cognitive abilities on imagining what the landscape used to look like, what features used to be there and how they played a role in the battle that took place there. At each stop and walking between them, the interpreter regularly reminded visitors to imagine themselves in the places of the soldiers who were there, walking the same lines that they did, and considering the emotions/decisions they faced during the battle. HIGH: The interpreter took time to describe what we would have seen if we were sitting with our family having a picnic and watching the battle, or what it would have looked like from the perspective of one of the soldiers.
Relevance to Audience Degree to which the program explicitly communicated the relevance of the subject to the lives of the audience (Beck & Cable, 2002; Brochu & Merriman, 2002; Ham, 1992, 2013; Jacobson, 1999; Knapp & Benton, 2004; Lewis, 2005; Moscardo, 1999; NPS Module 101; Sharpe, 1976; Tilden, 1957; Veverka, 1998).	HIGH: The interpreter clearly made it a priority to connect with and learn a bit about each program participant. He carried on conversations with various visitors between stops, using the information he gathered to shape what he talked about next. He related each story he told to something of particular interest to someone in the audience. HIGH: The interpreter compared people coming together in the 1800s after events at this historical site to people coming together after September 11, 2001 and other recent events. The interpreter described the Civil War as something that took place in back yards and town squares, had us imagine what life would be like now if war broke out in the United States. HIGH: The interpreter's main approach was connecting complex geology to something most people would understand: pizza. LOW: The interpreter provided massive amounts of factual information about the battle that took place here and the strategies used by either side to gain the upper hand. However, the program was entirely a lecture. The interpreter made no effort to connect the visitors to the resource, either through something of particular interest to them or by creating some relevance between what happened here and the lives of the audience. LOW: The interpreter attempted to connect black bears breaking into cars for food to how desperate we would be if we were hungry. If you've ever been starving hungry, you know that you'd be willing to break into a store or steal somebody's lunch....the audience's reactions suggested that this analogy did not connect at all.
Affective Messaging Degree to which the program communicated emotion (Jacobson, 1999; Lewis, 2005; Madin & Fenton, 2004; Tilden, 1957; Ward & Wilkinson, 2006).	HIGH: The interpreter discussed with us the heartache and suffering that went into sending a son off to war or finding out that a loved one had been killed in action. He spoke of the dedication to each other and to country that these soldiers displayed, the determination with which they fought, and the camaraderie on which they relied to keep their spirits up and keep fighting. He lowered his voice and explained the importance that their service should have to us. Rather than focusing on numbers or specific dates/battles, he focused on the emotional toll that war took on everyone. LOW: This interpreter relied solely on historical information to tell the story of FDR and his presidency. He told us the various offices FDR held, explained what polio was, and gave us descriptions of the design/construction of the monument itself. He told us about the impact that war and economic depression had on our country, but only in terms of money and power. He did not include any emotional connection to the struggles of poverty, the despair that people faced, the joy we felt after winning the war, or the emotional toll that polio must have taken on FDR and those around him.
Provocation Degree to which the program explicitly provoked participants to personally reflect on content and its deeper meanings (Beck & Cable, 2002; Brochu & Merriman, 2002; Knudson et al., 2003; Tilden, 1957)	HIGH: The interpreter told a very emotional story about how the coast Miwok tribes were torn away from their homes and lifestyle. He reminded us that their descendants are still alive today and that they can no longer visit the historic sites of their families. He asked us to think about the impact this must have on their culture and pride. HIGH: The ranger spent the majority of the program talking about different cultural groups that had populated the area throughout time. He gave us a glimpse into their daily life, their religions, and the things that were most important to them in life. He used vivid descriptions to get the audience to imagine the imagery of the time periods being described. He asked what we had in common with these people and how we were different. At the end of the program, we sat and watched the sunset, while the ranger asked us to think about our daily lives, what we are contributing to the world around us, and the things that make us feel truly alive. LOW: At one point during this program, the interpreter mentioned that urban sprawl is slowly taking over habitat and surrounding national parks in different places across the country. This was stated as a fact and then he moved on to the next subject. Rather than digging deeper or encouraging us to think about the effect that this might one day have, he just mentioned it and did nothing more with it.

Connection to Universals Communication that connects tangibles to intangibles and universal concepts. Intangibles are stories, ideas, meanings, or significance that tangible resources represent. Universals are concepts with which most audience members can identify (NPS Module 101; Beck & Cable, 2002; Brochu & Merriman, 2002; Ham, 1992; Knudson et al., 2003; Larsen, 2003; Lewis, 2005; Moscardo, 1999; Tilden, 1957; Ward & Wilkinson, 2006).	HIGH: During the program, the ranger told stories about the daily lives of early native people. At each stop he asked the same poignant questions: "What did life mean to these people? Why was this place important to them? What made them feel alive?" As we worked our way to the last stop of the walk, the ranger pointed out that we (the visitors) were now the inhabitants of this park. As we quietly watched the sun set, he asked us those same questions: "Why were we here? Why was this place special to us? What made us feel alive?" He connected us on the deepest levels with the people who had once inhabited this park and with the very essence of what made it important to us as visitors. LOW: The ranger provided a description of a native species that can be found in the park, detailing its appearance, unique traits, and status as a threatened species. The ranger continued working his way through species after species and did not field any visitor's questions or try to connect the topics to them in any way. He did not seem particularly interested in the topic, but instead like he was reciting a series of facts he had memorized. No attempts were made to reveal deeper meanings or connect us with the wildlife found in the park.
Individual program characteristics	
Appropriate Logistics Degree to which basic audience and program needs were met (i.e., restrooms, weather, , accessibility, shade, etc.) (Jacobson, 1999; Knudson et al., 2003).	HIGH: The interpreter arrived before the program was scheduled to begin and announced several times what the program was and when it would be starting. This gave everyone the chance to get ready and know they were in the right place. Once the program began, the interpreter told the audience how long we would be gone, what we would be doing, and what supplies they should have. He reminded everyone to use the bathroom before we went out on the trail and to wear sunscreen. Once on the trail, he made sure to keep the group together and maintain a reasonable pace. We stopped at spots along the trail that were out of the way of other hikers, quiet, and cool. Once the program ended, he walked with the group back to where we had started. LOW: The interpreter kept the audience standing in the very hot sun for extended periods of time despite ample opportunity for shade. LOW: During the walk, we stopped at a historical structure and the interpreter allowed the group to explore inside the building and around the grounds for an extended period of time. This broke up the flow of the program and left 15-20 people behind as we moved on to the next spot. The interpreter made very little effort to round up the group and did not announce when we would be leaving. LOW: The interpreter showed up to this program three minutes after its designated start time. He told the group that it was his first time ever giving it and that he wasn't sure exactly what we were supposed to be doing. The program was scheduled for an hour, but only lasted 30 minutes. The tour only had two stops, one at the parking lot and one about 100 yards away, even though it was advertised as a walking tour.
Appropriate for the Audience Degree to which the program aligned with audience's ages, cultures, and level of knowledge, interest, and experience (Beck & Cable, 2002; Jacobson, 1999; Knudson et al., 2003).	HIGH: The ranger made an explicit effort to gear this campfire program toward the mix of families and older adults in attendance. The ranger included songs and activities that everyone could enjoy and made content relatable to children and adults alike. The content was relatable to children, but also included novel stories and facts that adults were unlikely to know. For parts of the program, adults were given specific roles helping to guide the kids through activities. LOW: There was only one woman with two very young children on the tour. The interpreter did not adapt the program at all to the kids and instead seemed impatient when one was running around. She dealt with the matter by picking up the child and holding her. LOW: Some gory descriptions of Civil War soldiers, their injuries, and medical treatments of the time period may have been too graphic for some of the younger children in the audience. LOW: Although the audience consisted of a dozen adults and only one child, the interpreter spent the entire program speaking only to the child. He used very basic language and got down on one knee to tell her certain things. This was certainly a great experience for the child, but left the rest of the group wanting more. The program was advertised as a history of FDR's life and his role in preserving the United States during war and economic depression, but everything was limited to a very basic level.
Multisensory Engagement Degree to which the program intentionally and actively engaged more than just basic sight and sound (Beck & Cable, 2002; Knudson et al., 2003; Lewis, 2005; Moscardo, 1999; Tilden, 1957; Veverka, 1998; Ward & Wilkinson, 2006).	HIGH: Visitors were actively engaged in the program in a number of different ways. Their hands and backs were used to complete tasks around the farm and help the ranger close up for the day. They could smell the fire in the fireplace, feel the roughness of the handles they were meant to use, and had to struggle to see certain things in the fading light. It truly immersed all of their senses in not just seeing, but also experiencing life on the farm and understanding where it has gotten us today. HIGH: The interpreter told people to stoop down and feel the sidewalk, because that's how smooth the carved faces of the presidents are. HIGH: The interpreter organized her talk around the five senses, providing opportunities throughout the talk to smell, see, hear, feel, and even taste.

Verbal Engagement Degree to which the program verbally engaged audience members in a participatory experience; i.e., two-way dialogue (Knudson et al., 2003; Moscardo, 1999; Sharpe, 1976; Tilden, 1957; Veverka, 1998).	HIGH: After sharing and explaining different sets of data on the giant video sphere, the rest of the program was treated like a discussion session with the audience members talking about what may be causing trends in climate change and how the trends may be reversed. HIGH: Visitors sang along with campfire songs, answered questions, and were allowed to tell stories of their experiences in the park. HIGH: Visitors participated in an exercise similar to what schoolchildren would have done in the schoolhouse where the program took place. We answered questions and repeated lessons back to the "teacher." LOW: The interpreter asked only rhetorical questions that didn't encourage visitor involvement. Eventually the audience stopped thinking about answers to her questions because we knew she'd answer them right away.
Central Message Degree to which the program's message(s) was clearly communicated; i.e., the "so what?" element of the program (Beck & Cable, 2002; Brochu & Merriman, 2002; Ham, 1992; Jacobson, 1999).	HIGH: This program focused on climate change and the impact that it can have on our lives. We were told over and over again throughout the program to think about why we should care. No matter what the science or politics say, the changes that have already occurred are something that will affect us and that we should be thinking about. The interpreter used powerful illustrations of flooding, storm damage, and drought to keep us thinking. HIGH: The interpreter used powerful emotional language ("the struggle for freedom," "the ultimate sacrifice," and "the value of our freedom") to remind us of why this monument exists and why it should matter to us. He convinced us that it deserves our respect and reverence, not because of what the monument is, but because of who it represents. LOW: During the course of this program, the interpreter talked about boats, earthquakes, sea life, and gold. He was very interesting to listen to and taught the audience a lot of things they likely didn't know before. However, these random topics together did not convey a central message. Rather, it left the audience with a feeling of "huh, that was interesting," but without any particular take-home message.
Consistency Degree to which the program's tone and quality were consistent throughout the program (Beck & Cable, 2002; Ham, 1992).	LOW: The program seemed oddly split; the first half was a very engaging, tactile program about buffalo, and the second half was an abrupt switch to plant identification, presented in a scientific manner on the hot prairie.
Fact-Based Messaging *(negative influence)* The program was exclusively factual (Jacobson, 1999; Lewis, 2005; Tilden, 1957; Ward & Wilkinson, 2006).	HIGH: This program, about the flora found within the park, provided an abundance of facts and scientific names. It did not touch upon why these plants mattered or what relevance they had to us. The interpreter simply listed fact after fact for the duration of an hour long program. After a point, everything began to blend together and lose its meaning.
Appropriate Pace Degree to which the pace of the program allowed for clarity and did not detract from the program (Jacobson, 1999).	TOO FAST: The ranger seemed hurried throughout the scheduled program. One visitor continued to ask detailed questions about the topic. The ranger responded with short, generally unhelpful answers, and even cut him off entirely on a few occasions. When a child in the group tried to ask a series of questions, he told the child he needed to hold his questions until the end so that he didn't "bother the other visitors." TOO SLOW: The interpreter kept the audience standing in the very hot sun while stumbling through long moments of silence punctuated by statements such as "Let's see," and "what else can I tell you?"

We conducted a stepwise binary logistic regression on all interpreter and program characteristics (Table 4) to determine how well the most parsimonious set of characteristics could predict an overall quality assessment of 8 or better. The characteristics in Table 4 predict with over 88% accuracy which programs scored above or below this threshold. We urge some caution in the interpretation of this model. Many of the characteristics observed in the field were highly correlated with each other. The absence of characteristics that were otherwise strongly related to our score of "excellent" does not lessen their importance. Rather, their covariance with the predictors that populated the final model precludes their inclusion. For example, connection is strongly correlated with authentic emotion and charisma, confidence, organization, and appropriate for the audience ($r > 0.4$ in each case). As such, these variables appear in its

place in the model. The primary value of the model, we believe, is in demonstrating the strength of interpreter and program characteristics in predicting membership in the "excellent" category of programming.

We posit that the characteristics highlighted in the bivariate tests (shown in Tables 2 and 3), particularly those in bold italics, help to meaningfully differentiate programs that are adequate to satisfy visitors in a basic hedonic sense from those that may produce eudaimonic satisfaction. Our analyses suggest that each of these practices in various combinations may enhance outcomes across a majority of programs in which they were practiced. In other words, just like any other piece of art, there is no single recipe for success.

What do the practices look like?

Tables 5 and 6 provide definitions and examples from our field notes of the interpreter and program characteristics with the most powerful relationships to positive outcomes. We include only characteristics with strong statistically significant relationships ($p < .01$) with at least three measured outcomes (*satisfaction, visitor experience and appreciation, behavioral intentions,* and our own *overall quality assessment*). Positive examples in the tables reveal clear efforts to draw deeper connections to program attendees that go beyond mere entertainment and satisfaction of basic curiosity. The interpreters and programs exhibiting these traits seize the opportunity to go beyond the provision of basic hedonic satisfaction and move the visitor toward a more eudaimonic experience. This is not to say that all visitors to these programs experience life-changing moments, but rather the programs provide opportunities for visitors who are open to such provocation to make meaningful connections to the resources being interpreted.

We witnessed a number of brilliant programs over our three months of fieldwork. We've chosen one in particular to demonstrate the potential of interpretation to have meaningful longer-term influences on program attendees. This particular program scored an 8 on the overall quality measure.

> Following a thorough orientation to the program content and logistics, the ranger told us a little bit about what we were going to learn and why it was important to know. As we walked to the first stop, he also taught us some basic facts about the progression of the war, how it came to this site, and some key players in the battles that were fought here. This was the extent of the "history lesson" about the Civil War. The real meat of the program was the story of one young, unnamed man who lived in this town. We stopped at the house where he grew up, sat in the schoolhouse where he learned to read and write as a child, and visited the blacksmith shop where he learned his trade as a young man. At each place we learned about daily life during the time period: how meals were prepared in the oppressively hot family kitchen, the long walk to school and the cramped conditions inside the single room, the dangers of blacksmithing and the injuries that were regularly endured—all through the eyes of our main character. As such, we were able to frame the Civil War in a very tangible sense and see our character as a real person, similar to us, with real hopes, relationships, and struggles.
>
> As we moved onto the historic battlefield, the interpreter described how the young man saw the fight coming over the hill and rushed out his front door

> to join the Union, without enlisting in any official capacity. As we crossed the battlefield we saw the progression of the battle through the young man's eyes. We could feel his anxiety and excitement, his bravery and despair. As the tour neared its conclusion, we learned the young man's name. We also learned how he remained on the battlefield until the end, providing safe retreat for his Union Army comrades. His heroic actions saved the lives of many but cost him his own.
>
> We entered the National Cemetery, and the interpreter told us of many of the young men who had been buried here. We stopped. The ranger quietly paused and seemed to take it all in. Then he looked down at his feet and pointed out a grave stone near his feet—the final resting place of the young man we had spent the past hour coming to know. The audience's solemnity and sadness was palpable. The interpreter used few words to draw the connections between this young man's story and the magnitude of the Civil War's impact not only on our nation, but also on the people living so close to the battles. We had quite literally walked in this young man's footsteps as strong themes of sacrifice, beliefs, valor, and ordinary people unfolded. The audience stood in silence for quite some time after the program had ended.

This story, and many more like it, will stick with us for months and years to come. Like scenes from a great movie, a line from a song, or a favorite quote or poem, they arise in our minds and shape our decisions in ways that aren't always entirely tangible and for reasons we sometimes can't fully apprehend. Yet, they are there—a piece of our selves. Great interpretation provides this.

So what?

We've identified in both a statistical and qualitative sense throughout this special issue the characteristics of interpreters and their programs that appear to provide the most meaningful experiences for program attendees. We've attempted to demonstrate the difference between meeting basic expectations of the visitor and providing a truly exceptional experience. Sam Ham (2013) describes the endgame of interpretation as provocation, or "making people think and find personal meaning" (p. 62). Connection, stewardship, appreciation, understanding, revelation, inspiration, caring, motivation, and building support (or constituency) are other words commonly associated with the purpose of interpretation (Association for Heritage Interpretation, 2013; European Association for Heritage Interpretation, 2013; Interpretation Australia, 2013; Interpretation Canada, 2013; National Association for Interpretation, 2013; Stern & Powell, 2011; U.S. National Park Service, 2013). As such, satisfying the basic expectations of the visitor, such as orientation or entertainment, may be viewed not only as interpretive outcomes, but also as means to more meaningful and lasting ends (see Ham, 2013). Similar to Pine and Gilmore (1998), who urged the tourism industry to transition from a paradigm of *service delivery* to one of *experience creation,* we urge providers of interpretation to consider the potential of interpretation for meeting these more eudaimonic purposes in their planning and programming.

To meet these ultimate goals, we suggest that interpreters and interpretive organizations, such as the NPS, might consider the findings of this study in light of their hiring, training, and organizational cultures and practices. Many of the characteristics identified within the research effort are already clearly identified in training materials

and books used in classes on interpretation (U.S. National Park Service, 2013; Skibins et al., 2012). The influence of interpreters' expressed personalities and attitudes beg a deeper question, however, regarding how to train for, or otherwise influence, these characteristics.

Hiring and training

We focus in particular on the role of knowledge. We do this for two reasons. First, the hiring process for many interpretive agencies relies heavily on the self-reported knowledge, skills, and abilities (also known as KSAs) of potential hires. Second, we have witnessed interpretive training programs that we feel promote a potentially inappropriate role for facts and knowledge in communications with visitors. As discussed in Stern and Powell (this issue), the interpreter's knowledge of the subject matter is critical to the successful presentation of a program. However, knowledge should not necessarily be the focus of the communication itself. We rather posit that the knowledge of the interpreter serves a more important indirect role to successful communication through the development of confidence. This confidence frees the interpreter to be creative, emotive, and genuine in his or her communications instead of nervous or struggling to remember the correct facts and dates (Daly et al., 1989). Our data suggests that an over-emphasis on resource knowledge has the potential to hinder rather than promote positive visitor outcomes if it becomes the sole focus of the presentation (see also Stern & Powell, this issue).

Clearly, knowledge of the appropriate techniques and end goals of interpretation as well as knowledge of audiences and resources are critical for successful interpretation (Lacome, 2013). Our interviews with interpreters prior to their presentations revealed that those who aimed to provide visitors with new knowledge achieved less positive outcomes than those aiming to inspire visitors to gain a greater appreciation, change their attitudes, or desire to learn more (see also Stern & Powell, this issue). We argue that interpreters' understanding of these eudaimonic goals of interpretation may serve as a meaningful predictor of their success. As such, gauging beliefs about interpretation's appropriate outcomes in the hiring process might serve as reasonable predictors about how one might approach the task. Some assessment of general philosophies about the importance of story-telling and commitment to the mission of the organization might also be useful at this hiring stage. Each of these elements could also form the basis of meaningful training for all interpreters.

Knowledge of the resource, audience, and techniques can be further developed after hire on-site. Providing employees with the ability to spend time forming their own meaningful connections with the resources and stories they will be interpreting may be just as critical as time in the library or archives developing an understanding of the facts about the resource. Without these personal connections, it may prove quite challenging to provide similar connections for visitors. Without a holistic picture of a place or a resource, it may be quite difficult to develop compelling stories that reveal deeper meaning to audiences. Training can provide multiple versions of stories to interpreters, as they develop their own.

We have witnessed various approaches to training. Some have focused on accuracy and education through organizing facts into a coherent order for presentation, similar to what one learns in a college public speaking course—*tell 'em what you'll tell 'em, tell 'em, and then tell 'em what you told 'em*. While this approach can help with organizing information, it does not alone capture what is most important to interpretive

communication. We argue for a more hands-on approach that begins with demonstration of the practices uncovered in our study as well as demonstration of drier, more factual presentations. Without experiencing each, it may be difficult to grasp the difference between mediocre and excellent interpretation. As interpreters practice their own programs, the list of characteristics uncovered within this study can serve as a menu of elements for experimentation and constructive feedback. Perhaps most important would be to stress the desired outcomes of programs to interpreters. Currently, most park units' long-range interpretive plans in the NPS place emphasis on subject matter themes rather than desired outcomes for visitors. A slight shift in what is most prominently communicated to interpreters from the organization could make a meaningful impact.

Organizational support

Elements of organizational culture have been long identified as important drivers of employee performance (Gordon & DiTomaso, 1992; Judge et al., 2001; Schein, 2010). We focus on the concepts of employee empowerment and adaptability, critical task, and attitudinal organizational commitment (AOC). Our study revealed that interpreters tend to produce better outcomes for visitors when they are excited and positive about their work (Stern & Powell, this issue). Similarly, a large body of research suggests that happy employees tend to perform better (Judge et al., 2001). Organizational culture can have a strong influence on such feelings (Ouchi & Wilkins, 1985).

We posit that interpreters who feel empowered and supported by their organizations will be most successful in producing positive visitor outcomes. Our qualitative observations, interviews, and casual conversations with interpreters in the field strongly support this notion. The proposition is further supported in the management literature, where the empowerment of employees is equated to feelings of competence, self-determination (freedom to choose how to get the job done), a sense that the work is important, and a belief that the work will have a meaningful impact on the larger goals of the organization (Kirkman & Rosen, 1999; Spreitzer, 1996). Such empowerment, and the adaptability that is associated with it, has been empirically equated with better performance in multiple studies (e.g., Gordon & DiTomaso, 1992; Stern & Predmore, 2012). In our study, elements of confidence and authentic emotion served as critical ingredients of outstanding programs.

Multiple studies reveal that adaptability at the individual level is most predictive of success in organizations and work units that have a clear and consistent sense of mission and a strong organizational culture (Wilson, 1989). Wilson (1989) argues that a clear sense of mission emerges not necessarily from a mission statement, but from the articulation of a "critical task" that is widely accepted and endorsed by employees. A critical task involves the clear definition of the specific outcomes that employees can produce to accomplish the overall mission of the agency. A strong and healthy organizational culture can be defined as one where employees share consistent views about this critical task. They also share relatively consistent views that the organization emphasizes both its human resources and goal accomplishment (Hansen & Wernerfelt, 1989; Gordon & DiTomaso, 1992). This combination can influence high levels of AOC, which indicates the relative strength of an employee's commitment to and identification with an organization (Deery & Iverson, 2005; Mowday et al., 1982; Riketta, 2002). The stronger the AOC, the stronger the employee's motivation to pursue the agency's goals and improve its status (Riketta & Landerer, 2005).

With all this in mind, certain elements of organizational support may be particularly helpful in enhancing interpreter performance: a recognition and articulation of clear (and meaningful) objectives for interpretive outcomes for attendees, training and immersive time with the resource to enhance feelings of competency, freedom to develop programs creatively with organizationally important outcomes in mind, and appreciative support and recognition from supervisors and managers. In our study, each park unit appeared to have its own unique organizational culture. Our qualitative observations indicated strong influences in some cases of less than healthy organizational cultures upon interpreter performance. While the mood of individual interpreters on any given day may be largely independent of organizational culture, unhealthy cultures may predispose interpreters to falling short of providing the best programs within their abilities. Meanwhile, healthy, empowering cultures may influence higher levels of confidence, passion, and creativity in interpreters, enhancing their connections to both the resources they interpret and the audiences they engage.

Conclusions

The research reported within this special issue suggests that certain characteristics of interpreters and their programs may make the difference between mediocre, or adequate, experiences for visitors and exceptional experiences. In this article, we have tried to delineate the differences between the outcomes of each type of program. Most programs in the study attained positive levels of satisfaction from attendees, suggesting that basic expectations were typically met. Some programs, however, likely influenced attendees in far more meaningful ways, similar to the way a great work of art or movie might be revelatory or inspirational, or provide some new insight or viewpoint that remains long after the experience. We urge interpretive organizations to consider the findings presented within this manuscript and the rest of this special issue when developing and/or revising training for interpreters. We also urge interpretive organizations to reach toward more eudaimonic experiences for visitors by clearly articulating goals that go beyond merely satisfying visitors' basic expectations. Interpretation provides the opportunity to accomplish much more, not only in terms of visitor experiences, but also with regard to building constituencies for the interpreted resources and the organizations that protect them. Finally, we urge interpretive organizations to consider that training alone may be insufficient to create the conditions that produce great programs for visitors and that organizational culture may have powerful influences on visitor outcomes.

References

Association for Heritage Interpretation (2013). http://www.heritage-interpretation.org.uk. Accessed June 20, 2013.

Beck, L., & Cable, T. T. (2002). *Interpretation for the 21st century: Fifteen guiding principles for interpreting nature and culture* (2nd ed.). Champaign: Sagamore.

Brochu, L., & Merriman, T. (2002). *Personal interpretation: Connecting your audience to heritage resources.* InterpPress.

Deery, S. J., & Iverson, R. D. (2005). Labor-management cooperation: Antecedents and impact on organizational performance. *Industrial and Labor Relations Review,* 58(4): 588–608.

del Bosque, I. R., and H. San Martín (2008). Tourist satisfaction: a cognitive-affective model. *Annals of Tourism Research,* 35(2): 551–73.

European Association for Heritage Interpretation (2013). *What is heritage interpretation?* http://www.interpret-europe.net/?id=2. Accessed June 20, 2013.

Gordon, , G.G., & DiTomaso, N. (1992). Predicting corporate performance from organizational culture. *Journal of Management Studies* 29(6): 78–798.

Ham, S. H. (1992). *Environmental interpretation: A practical guide for people with big ideas and small budgets*. Golden: Fulcrum.

Ham, S. H. & Weiler, B. (2002). Toward a theory of quality in cruise-based nature guiding. *Journal of Interpretation Research*, 7(2): 29–49.

Ham, S. H. (2009). From interpretation to protection: is there a theoretical basis? *Journal of Interpretation Research,* 14(2): 49–57.

Ham, S. H. (2013). *Interpretation: Making a difference on purpose*. Golden: Fulcrum.

Hansen, G.S., & Wernerfelt, B. (1989). Determinants or firm performance: the relative importance of economic and organizational factors. *Strategic Management Journal* 10: 399–411.

Interpretation Australia (2013). *What is interpretation?* http://www.interpretationaustralia.asn.au/about-ia/what-is-interpretation#. Accessed June 20, 2013.

Interpretation Canada (2013). *Our work defined*. http://www.interpscan.ca/our-work-defined. Accessed June 20, 2013.

Jacobson, S. K. (1999). *Communication skills for conservation professionals*. Washington, D.C.: Island Press.

Judge, T.A., Thoresen, C.J., Bono, J.E., & Patton, G.K. (2001). The job satisfaction-job performance relationship: a qualitative and quantitative review. *Psychological Bulletin,* 127(3): 376–407.

Kirkman, B.L., & Rosen, B (1999). Beyond self-management: antecedents and consequences of team empowerment. *Academy of Management Journal,* 42(1): 58–74.

Knapp, D., & Yang, L. (2002). A phenomenological analysis of long-term recollections of an interpretive program. *Journal of Interpretation Research,* 7(2): 7–17.

Knapp, D., & Benton, G. M. (2004). Elements to successful interpretation: A multiple case study of five national parks. *Journal of Interpretation Research,* 9(2): 9–25.

Knudson, D. M., Cable, T. T., & Beck, L. (2003). *Interpretation of cultural and natural resources* (2nd ed.). State College: Venture Publishing.

Lacome, B. (2013). *The Intepretive Equation*. http://idp.eppley.org/IDP/sites/default/files/EquationEssay.pdf. Accessed April 6, 2013.

Larsen, D. L. (2003). *Meaningful Interpretation: How to Connect Hearts and Minds to Places, Objects, and Other Resources*. Eastern National.

Lewis, W. J. (2005). *Interpreting for park visitors* (9th ed.). Fort Washington: Eastern National.

Madin, E. M. P., & Fenton, D. M. (2004). Environmental interpretation in the Great Barrier Reef Marine Park: An assessment of programme effectiveness. *Journal of Sustainable Tourism*, 12(2): 121–137.

Maslow, A.H. (1943). A theory of human motivation. *Psychological Review*, 50(4): 370–96.

Moscardo, G. (1999). *Making visitors mindful: Principles for creating quality sustainable visitor experiences through effective communication*. Champaign: Sagamore.

Mowday, R. T., Porter, L.W., & Steers, R. M. (1982). *Employee organization linkages: The psychology of commitment, absenteeism, and turnover.* New York: Academic.

National Association for Interpretation (2013). *Mission, Vision, and Core Values*. http://www.interpnet.com/nai/About/What_We_Believe/nai/_About/Mission_Vision_and_Core_Values.aspx?hkey=ef5896dc-53e4-4dbb-929e-96d45bdb1cc1. Accessed June 20, 2013.

National Park Service 2003. *Module 101: Fulfilling the NPS mission: The process of interpretation*. National Park Service, Washington, DC.

Oliver, M.B., & Bartsch, A. (2010). Appreciation as audience response: exploring entertainment gratifications beyond hedonism. *Human Communication Research* 36: 53–81.

Ouichi, W.G., & Wilkins, A.L. (1985). Organizational culture. *Annual review of Sociology*, 11: 457–483.

Pearce, C.L. & Sims, H.P., (2002). Vertical versus shared leadership as predictors of the effectiveness of change management teams: an examination of aversive, directive, transactional, transformational, and empowering leader behaviors. *Group dynamics: theory, research, and practice,* 6 (2): 172–197.

Petty, R., & Cacioppo, J. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, NY: Springer-Verlag.

Pine, J., & Gilmore, J.H. (1998). *Welcome to the experience economy.* Harvard Business Review, 76(4): 97–105.

Regnier, K., Gross, M., & Zimmerman, R. (1992). *The Interpreter's Guidebook: Techniques for Programs and Presentations*. Interpreter's Handbook Series. UW-SP Foundation Press.

Riketta, M. (2002). Attitudinal organizational commitment and job performance: A meta-analysis. *Journal of Organizational Behavior*, 23: 257–266.

Riketta, M., & Landerer, A. (2005). Does perceived threat to organizational status moderate the relation between organizational commitment and work behavior? *International Journal of Management*, 22: 193–200.

Ryan, R.M., & Deci, E.L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology,* 52: 141–160.

Schein, E.H. (2010). *Organizational culture and leadership*. San Francisco: Josey-Bass.

Sharpe, G. W. (1976). *Interpreting the environment*. New York: John Wiley & Sons.

Sherif, M., & Hovland, C. (1961). *Social judgment: assimilation and contrast effects in communication and attitude change*. New Haven: Yale University Express.

Skibins, J.C., Powell, R.B., & Stern, M.J. (2012). Exploring empirical support for interpretation's best practices. *Journal of Interpretation Research,* 17(1): 25–44.

Spreitzer, G.M., 1996. Social structural characteristics of psychological empowerment. *Academy of Management Journal,* 39(2): 483–584.

Stern, M.J. , & Powell, R.B. (2011). *The Views of National Park Service Superintendents and Interpretation and Education Supervisors on Interpretation and Education in the National Park Service*. U.S. National Park Service. October, 2011.

Stern, M.J., Powell, R.B., & Hockett, K.S. (2011). Why do they come? Understanding interpretive program attendance at Great Smoky Mountains National Park. *Journal of Interpretation Research,* 16(2): 35–52.

Stern, M.J., & Predmore, S.A. (2012). The importance of team functioning to natural resource planning outcomes. *Journal of Environmental Management,* 106: 30–39.

Szymanski, D., & Henard, D. (2001).Customer satisfaction: a meta-analysis of the empirical evidence. *Journal of the Academy of Marketing Science,* 29: 16–35.

Tilden, F. (1957). *Interpreting our heritage* (3rd ed.). Chapel Hill: The University of North Carolina Press.

U.S. National Park Service (2013). *Interpretive development program*. http://idp.eppley.org/. Accessed June 20, 2013.

Veverka, J. A. (1998). *Interpretive master planning: the essential planning guide for interpretive centers, parks, self guided trails, historic sites, zoos, exhibits and programs* (2nd ed.). Tustin: Acorn Naturalists.

Ward, C.W., & Wilkinson, A. E. (2006). *Conducting meaningful interpretation: A field guide for success*. Golden: Fulcrum.

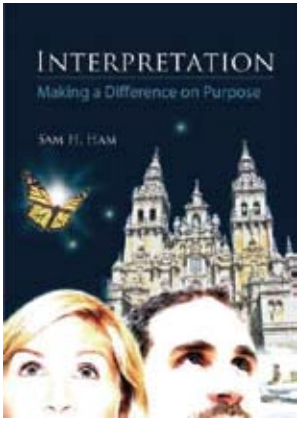
Wirth, W. Hofer, M., & Schramm, H. (2012). Beyond pleasure: exploring the eudaimonic entertainment experience. *Human Communication Research*, 38: 406–428.

Book Review

Interpretation: Making a Difference on Purpose

Carolyn Ward
Editor, *Journal of Interpretation Research*
CEO, Blue Ridge Parkway Foundation
322 Gashes Creek Road
Asheville NC 28803
cward@brpfoundation.org

It has been almost 10 years since I called Sam Ham about writing *Conducting Meaningful Interpretation* with my long-time colleague and friend Alan Wilkinson. I was looking for guidance and for advice about publishing, but mostly I wanted to talk to the person whom I thought had written the most definitive book on interpretation in more than 30 years, *Environmental Interpretation*. I have two copies of his book, one so worn with the dirt-covered fingers of a field interpreter that the spine had long fallen apart and another so highlighted and marked up with the pen of a lecturer that reading it unimpaired was impossible.



When I got my first job as an interpreter at Hungry Mother State Park in Virginia, before I even knew what the term *interpretation* meant, I was handed Ham's book and told that I needed to prepare a program in two weeks. I used it to walk me through the process of program development, to understand techniques and strategies for dealing with a group and to create a program with some meaning. Years later, when I became a professor of interpretation at Humboldt State University, I used his book to teach my students. My students loved it, and dog-eared copies adorned the classroom until the day I left. As my career evolved in the profession of interpretation, I become more immersed in research, evidence-based practice and "proof" of what worked and why. The further away from practicing in the field I moved, the more I wanted to explain and understand that "light in the eyes" of my visitors that I based my own determinations of "success" upon. How

could we be sure that programs were successful? What was success? How could we do better?

So when I picked up the phone that day 10 years ago to talk to Ham about writing a new book on the practice and science of interpretation, it was certainly with some trepidation. Although I did not believe our text would replace Ham's seminal book, I did want to convey more of the science behind the practice and to bound practice in current theory as much as possible. My trepidation fell quickly away, when Ham professed that he not only thought it was time for a new book, but that he would like for it be a part of Fulcrum's acclaimed Applied Communication Series for which he served as the executive editor. He also confessed that he thinking about writing a new book and that he wanted it to be different from his first one. He wanted to springboard from where he left off in *Environmental Interpretation* and assured me that our two "new" books would fit nicely together in the series.

Now, almost 10 years after my phone call, Ham's long-awaited *Interpretation: Making a Difference on Purpose* has been released. Although Ham could have released this book years ago (I remember seeing drafts of a chapter almost three years before actual release), his tenacity and dogged persistence in seeking his colleagues' feedback and refining his work has paid off. After being in the field for so long, in so many different capacities, it is hard to think of reading anything about interpretation that would take me by surprise. But this book made me stop, made me think and made me re-think some of those things which are assumptions and quickly passed over as "basics."

I spent many days pouring over his book in preparation for writing this review and soon became lost in the work itself instead of focusing on how I would covey it in a review. This was to me the mark of yet another seminal work from Sam Ham. I wasn't reading a textbook, as much as having an intimate conversation with the author about the profession of interpretation. He drew me in with his relaxed conversational style of writing and set the researcher in me to rest with the detailed annotated notes that followed each chapter.

Ham is the embodiment of an interpreter. His TORE model (Chapter 2) is not just one he preaches; he practices it in his writing. He does not bore the reader with the citations throughout the text which would deter from the conversation, but instead includes them, along with a glossary, after each chapter for those interpretation nerds like me who want to delve deeper. Although I am sure many captive audiences will read this book in a classroom preparing for a pending test, many more readers like me will enjoy the conversation that unfolds within the pages and will read it for sheer *enjoyment*. The writing is light, funny and engaging.

The *relevance* of the book is without question. Whether the reader is a student, practitioner, researcher, or manager, Ham's book has something relevant to offer. I learned new concepts and terms which are sure to have lasting value and application in the profession for years to come. Ham's introduction and discussion of the Zone of Tolerance (Chapter 8) is a concept that has been missing in the field and provides a real-world solution to practicing interpreters for judging whether or not they are "successful." It is elegantly simplistic and yet well-grounded in theory and reality. Discussions of Thought-Listing (Appendix 3) and the End Game (Chapter 3) are other relevant concepts revealed in the book that are sure to keep readers engaged and leave them enlightened.

Ham's *organizational* style follows that of a honed interpretive program with main concepts repeated throughout and new ideas carefully woven into our schemata. In fact,

Chapter 4 allows readers to digest and reflect on the three previous chapters, assisting readers in building their own scaffolding before moving on to the heart of the matter in Chapter 5, Making a Difference on Purpose. Ham's organizational mastery is not only reflected in repetition but in his frequent use of foreshadowing which is applied with the skill of a great storyteller keeping readers captured with the excitement of what the next page will bring.

Although according to Ham the *theme* comes first, I have chosen to cover it last in this review because it is the "So What?" that made me stop and think. Like a good interpretive program, the theme discussions in Ham's book are what bind it all together. And with four chapters, 6, 7, 9, and 10, dedicated to discussing thematic interpretation, Ham clearly supports the notion that it is the critical component of interpretation. The discussion of thematic interpretation is dynamic and detailed and takes the reader on a journey from simplistic understanding to sophisticated application. The theme helps us define the "So What?" and should assist us in determining the Zone of Tolerance (Chapter 8).

Ham's *Interpretation: Making a Difference on Purpose* is the ultimate "end game." It made me think, it made me elaborate on what I already knew, and it provoked me to want to know more. As Ham said on page 65, "making a difference on purpose is both the premise and promise of interpretation," and this book is both.

Manuscript Submission

Instructions to Authors

Purpose

The purposes of the *Journal of Interpretation Research* are to communicate original empirical research dealing with heritage interpretation and to provide a forum for scholarly discourse about issues facing the profession of interpretation. The *Journal* strives to link research with practice. The *Journal of Interpretation Research* is published by the National Association for Interpretation, the preeminent professional association representing the heritage interpretation profession.

General Information

The primary function of the *Journal* is to disseminate original empirical research regarding interpretation. However, the *Journal of Interpretation Research* takes a broad view of the field of interpretation and publishes manuscripts from a wide-range of academic disciplines. The primary criteria for deeming a manuscript appropriate for the *Journal* are whether it adds to the current state-of-knowledge for practitioners, researchers, academics, or administrators who work in the field of interpretation.

In recognition of how diverse the relevant literature is, the *Journal* will also publish reviews of recent books, government publications, original literature reviews, and bibliographies dealing with interpretation. Abstracts from dissertations, private consultant materials, and reports from public agencies will be published in the *Journal* in a section called "In Short: Reports and Reviews." This section will also provide an outlet for summaries of research studies with limited scope. Interpretation research often consists of small "in-house" program evaluations and basic visitor studies. The purpose of this section is to communicate current research activities, allow readers to identify colleagues with similar interests, and provide practitioners and administrators with useful information and direction for conducting their own mini-research projects. Submissions for the "In Short: Reports and Reviews" section should be limited to 800 to 1,000 words and will be reviewed by the editor and two associate editors.

Additionally, the *Journal* will publish thought pieces that exhibit excellence and offer original or relevant philosophical discourse on the state of heritage interpretation. The "In My Opinion" section of the *Journal* encourages the development of the profession and the practice of interpretation by fostering

discussion and debate. Submissions for the "In My Opinion" section should be limited to 1,000 to 1,200 words and will be reviewed by the editor and two associate editors.

Research Manuscript Submission Guidelines

All research manuscripts will be reviewed anonymously by an associate editor and by at least two other reviewers. Based on the nature of the manuscript, special efforts will be made to identify well-qualified associate editors and reviewers to evaluate the manuscripts. From the recommendations of the associate editor, the editor will make the final decision of the manuscript's disposition and communicate this information to the author.

Manuscripts

Manuscripts will be accepted with the understanding that their content is unpublished and not being submitted elsewhere for publication.

- All parts of the manuscript, including title page, abstract, tables, and legends, should be typed in 12-point font, and double-spaced on one side of 8.5" x 11" or A4 white paper.
- Margins should be 1" on all sides.
- Manuscript pages should be numbered consecutively in the top right corner.
- All papers must be submitted in English. Translations of papers previously published in other languages will be considered for publication, but the author must supply this information when the manuscript is submitted.
- Maximum length of manuscripts shall be 30 double-spaced pages (including all text, figures, tables, and citations). The editor will consider longer manuscripts on an individual basis.

Titles

Must be as brief as possible (six to 12 words). Authors should also supply a shortened version of the title, suitable for the running head, not exceeding 50 character spaces.

Affiliation

On the title page include full names of authors, academic, and/or other professional affiliations, and the complete mailing address of the author to whom proofs and correspondence should be sent. An e-mail address and phone and fax numbers should also be included. As all manuscripts will be reviewed anonymously; the name(s) of the author(s) should only appear on the title page.

Abstract

Each paper should be summarized in an abstract of no more than 150 words. The abstract will preface the paper and should be a comprehensive summary of the paper's content, including the purpose or problem, methods, findings, and implications or applications. It should enable the reader to determine exactly what the paper is about and make an informed decision about whether to read the entire paper. Abbreviations and references to the text should be avoided. All abstracts shall be listed on the *Journal of Interpretation Research* Web site (www.interpnet.com/JIR).

Keywords
Authors must supply five to 10 key words or phrases that identify the most important subjects covered by the paper.

References and Citations
Include only references to books, articles, and bulletins actually cited in the text. All references must follow the *Publication Manual of the American Psychological Association* (APA), version 6.2. References in the text should cite the author's last name, year of publication, and page (if appropriate). All references used in the text should appear at the end of the typed script in alphabetical order using APA version 6.2 style.

Examples of references:

McCool, S. & Braithwaite, A. (1992). Persuasive Messages and Safety Hazards in Dispersed and Natural Recreation Settings. In M. Manfredo (Ed.), *Influencing Human Behavior.* Champaign, IL: Sagamore Publishing.

Ryan, C. & Dewar, K. (1995). Evaluating the Communication Process Between Interpreter and Visitor. *Tourism Management, 16*(4): 295-303.

Tilden, F. (1977). *Interpreting Our Heritage* (2nd ed.). Chapel Hill: University of North Carolina Press.

Figures
All figures must be discussed in the text and numbered in order of mention. *Each figure must be submitted as a print-ready digital file*. Label each figure with article title, author's name, and figure number by attaching a separate sheet of white paper to the back of each figure. Each figure should be provided with a brief, descriptive legend. All legends should be typed on a separate page at the end of the manuscript.

Tables
All tables must be discussed in the text and numbered in order of mention. Each table should have a brief descriptive title. Do not include explanatory material in the title: use footnotes keyed to the table with superscript lowercase letters. Place all footnotes to a table at the end of the table. Define all data in the column heads. Every table should be fully understandable without reference to the text. Type all tables on separate sheets; do not include them within the text.

Permissions
If any figure, table, or more than a few lines of text from a previously published work are included in a manuscript, the author must obtain written permission for publication from the copyright holder and forward a copy to the editor with the manuscript.

Copyright
Under U.S. *copyright* law, the transfer of copyright from the author to the publisher (National Association for Interpretation, DBA *Journal of Interpretation Research*) must be explicitly stated to enable the publisher to ensure maximum dissemination of the author's work. A completed copyright form sent to you with the acknowledgment must be returned to the publisher before any manuscript can be assigned an issue for publication.

Proofs
All proofs must be corrected and returned to the publisher within 48 hours of receipt. If the manuscript is not returned within the allotted time, the editor will proofread the article, and it will be printed per his/her instruction. Only correction of typographical errors is permitted. The author will be charged for additional alterations to text at the proof stage.

Submission
Please submit a either a digital file (PDF or Microsoft Word) or an original hard copy and three copies of your manuscript to Carolyn J. Ward at the address below. Authors whose manuscripts are accepted for publication must submit final manuscripts electronically or on computer disk.

Contact

If you have comments or questions regarding the *Journal of Interpretation Research*, please contact the editor:

Carolyn J. Ward, Ph. D.
CEO, Blue Ridge Parkway Foundation
322 Gashes Creek Road
Asheville NC 28803

cward@brpfoundation.org
828-776-4547

Subscriptions

If you have questions regarding subscription rates or delivery services, please contact the National Association for Interpretation toll-free at 888-900-8283, online at www.interpnet.com, or by mail at P.O. Box 2246, Fort Collins, CO 80522.