

# 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 et 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 <sup>a</sup>
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

<sup>a</sup> 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.<sup>1</sup> 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
<b>Totals</b>	<b>488</b>	<b>376</b>	<b>3,603</b>	<b>312</b>	<b>3,122</b>

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 *n*th 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.<sup>2</sup> 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.**

<b>Program characteristic</b>	<b>Definition</b>	<b>Scoring</b>
<b>Introduction quality</b> (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
<b>Appropriate logistics</b> (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
<b>Appropriate for audience</b> (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
<b>Appropriate sequence</b> (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
<b>Transitions</b> (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
<b>Links to intangible meanings and universal concepts</b> (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
<b>Multisensory</b> (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)
<b>Physical engagement</b> (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
<b>Verbal engagement</b> (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
<b>Cognitive engagement</b> (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
<b>Multiple activities</b> (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
<b>Props</b> (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

<p><b>Relevance to audience</b> (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)</p>	<p>Degree to which the program explicitly communicated the relevance of the subject to the lives of the audience.</p>	<p>5= Major focus of messaging 4= Well developed efforts 3= Moderate efforts 2= Minimal efforts 1= No efforts</p>
<p><b>Affective messaging</b> (Jacobson, 1999; Lewis, 2005; Madin and Fenton, 2004; Tilden, 1957; Ward and Wilkinson, 2006)</p>	<p>Degree to which the program communicated emotion (in terms of quantity, not quality).</p>	<p>5= Central programming element 4= Frequent and repeated messages 3= Occasional messages 2= Minimal effort to include messages 1= Messages absent</p>
<p><b>Fact-based messaging</b> (Frauman and Norman, 2003; Jacobson, 1999; Lewis, 2005; Tilden, 1957; Ward and Wilkinson, 2006)</p>	<p>Degree to which the program communicated factual information.</p>	<p>1 = Messaging was solely fact-based 0 = Messaging was not solely fact-based (incorporated affective messaging)</p>
<p><b>Surprise</b> (Beck and Cable, 2002; Moscardo, 1999)</p>	<p>Degree to which the program used the element of surprise in communication. This could include "aha" moments or unexpected or contrasting messages.</p>	<p>3= Major element 2= Minor element 1= Not used</p>
<p><b>Novelty</b> (Beck and Cable, 2002; Frauman and Norman, 2003; Knapp and Benton, 2004; Moscardo, 1999)</p>	<p>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.</p>	<p>3= Major element 2= Minor element 1= Not used</p>
<p><b>Provocation</b> (Beck and Cable, 2002; Brochu and Merriman, 2002; Knudson, et al., 2003; Tilden, 1957)</p>	<p>Degree to which the program explicitly provoked participants to personally reflect on content and its deeper meanings.</p>	<p>4= Powerful and explicit inclusion 3= Occasional inclusion 2= Isolated or vague inclusion 1= No attempt made</p>
<p><b>Multiple viewpoints</b> (Beck and Cable, 2002; Brochu and Merriman, 2002; Tilden, 1957)</p>	<p>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?)</p>	<p>3= Multiple viewpoints developed; none given clear priority 2= Primarily one viewpoint, with some focus on others 1= No effort NA = not applicable</p>
<p><b>Holistic storytelling</b> (Beck and Cable, 2002; Larsen, 2003; Tilden, 1957)</p>	<p>Degree to which the program aimed to present a holistic story (with characters and a plot) as opposed to disconnected pieces of information.</p>	<p>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.</p>
<p><b>Place-based messaging</b> (Beck &amp; Cable, 2002; Knudson, et al., 2003; Lewis, 2005; Moscardo, 1999; NPS Module 101; Sharpe, 1976)</p>	<p>Degree to which the program emphasized the connection between the visitor and the site/resource.</p>	<p>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</p>
<p><b>Introduction and conclusion linkage</b> (Beck and Cable, 2002; Brochu and Merriman, 2002; Larsen, 2003)</p>	<p>Degree to which program connected conclusion back to the introduction in an organized or cohesive way (i.e., program "came full circle.")</p>	<p>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</p>

<b>Clear theme</b> (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
<b>Central message</b> (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
<b>Consistency</b> (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
<b>Pace</b> (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
<b>Quality of the resource</b>	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
<b>Unexpected negative circumstance</b>	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
<b>Unexpected positive circumstance</b>	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
<b>Behavioral theory elements</b>		
The following were only measured for programs in which a behavioral change was expressed by the interpreter as a desired program outcome.		
<b>Benefits of action</b> (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
<b>Costs of action</b> (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
<b>Norms of action</b> (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
<b>Ease of action</b> (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
<b>Demonstrates action</b> (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.**

<b>Interpreter characteristic</b>	<b>Definition</b>	<b>Scoring</b>
<b>Professional appearance</b>	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
<b>Comfort of the interpreter</b> (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
<b>Responsiveness</b> (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
<b>Inequity</b> (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.
<b>Humor quality</b> (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
<b>Humor quantity</b>	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
<b>Sarcasm</b>	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
<b>Charisma</b> (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
<b>Sincerity</b> (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
<b>Passion</b> (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

<b>Personal sharing</b> (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)
<b>Apparent knowledge</b> (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
<b>Audibility</b>	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
<b>Eloquence</b> (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
<b>Impatience</b>	Did the interpreter show any explicit impatience toward audience members?	1 = Interpreter was explicitly impatient with the audience 0 = No issues noted
<b>Formality</b>	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
<b>False assumption of the audience</b>	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
<b>Character acting</b>	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
<b>Primary identity</b> (Ham and Weiler, 2002; Ham, 2002; Knapp and Yang, 2002; Larsen, 2003; Mills, 1920; Wallace and Gaudry, 2005)	<b>Friend:</b> outwardly friendly, casual, approachable, mingles informally	1 = primary identity; 0 = not
	<b>Authority figure:</b> emphasizes own role as a park ranger and focuses on rules, regulations, and/or authority to communicate <b>Walking encyclopedia:</b> Focused on conveying a large volume of facts	1 = primary identity; 0 = not
<b>Questionable information</b>	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
<b>Bias</b>	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  $\pm 3$  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.**

<b>OUTCOME INDEXES</b>
<p><b>Program outcome: Visitor Experience and Appreciation</b> (Cronbach's <math>\alpha = 0.89</math>)</p> <p>To what extent did the program you just attended influence any of the following for you?</p> <ul style="list-style-type: none"> <li>• Made my visit to this park more enjoyable</li> <li>• Made my visit to this park more meaningful</li> <li>• Enhanced my appreciation for this park</li> <li>• Increased my knowledge about the program's topic</li> <li>• Enhanced my appreciation for the National Park Service</li> </ul>
<p><b>Program outcome: Behavioral intentions</b> (Cronbach's <math>\alpha = 0.94</math>)</p> <p>To what extent did the program you just attended influence any of the following for you?</p> <ul style="list-style-type: none"> <li>• Changed the way I will behave while I'm in this park</li> <li>• Changed the way I will behave after I leave this park</li> </ul>

## 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  $\chi^2=338.41$ ; 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.**

<b>INDEPENDENT VARIABLE INDEXES</b>
<p><b>Interpreter characteristic: Confidence</b> (Cronbach's <math>\alpha = 0.70</math>)</p> <ul style="list-style-type: none"> <li>• Comfort of the Interpreter</li> <li>• Apparent knowledge</li> <li>• Eloquence</li> </ul>
<p><b>Interpreter characteristic: Authentic emotion and charisma</b> (Cronbach's <math>\alpha = 0.85</math>)</p> <ul style="list-style-type: none"> <li>• Passion</li> <li>• Charisma</li> <li>• Sincerity</li> </ul>
<p><b>Program characteristic: Organization</b> (Cronbach's <math>\alpha = 0.82</math>)</p> <ul style="list-style-type: none"> <li>• Quality of the introduction</li> <li>• Appropriate sequence</li> <li>• Effective transitions</li> <li>• Holistic story</li> <li>• Clarity of theme</li> <li>• Link between introduction and conclusion</li> </ul>
<p><b>Program characteristic: Connection</b> (Cronbach's <math>\alpha = 0.88</math>)</p> <ul style="list-style-type: none"> <li>• Links to intangible meanings and universal concepts</li> <li>• Cognitive engagement</li> <li>• Relevance to audience</li> <li>• Affective messaging</li> <li>• Provocation</li> </ul>

### *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
<b>Satisfaction (0 to 10)</b>	8.96 (0.68)	9.02 (0.89)
<b>Visitor experience and appreciation (1 to 5)</b>	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)
<b>Behavioral intentions (1 to 5)</b>	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.**

Variable (Scale)	Means (with standard deviations)	
	Five or more attendees	Fewer than five attendees
<b>Confidence index (1 to 4)</b>	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)
<b>Authentic emotion and charisma index (1 to 5)</b>	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) <sup>a</sup>	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)

<sup>a</sup> 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
<b>Organization index (1 to 5)</b>	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)
<b>Connection index (1 to 5)</b>	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) <sup>a</sup>	2.63 (0.51)	2.61 (0.50)
<b>Behavioral theory elements<sup>b</sup></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)

<sup>a</sup> 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).

<sup>b</sup> 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 <sup>A</sup>	4.27 <sup>AB</sup>	2.56 <sup>A</sup>
Too slow	8.43 <sup>A</sup>	4.23 <sup>A</sup>	2.84 <sup>AB</sup>
Appropriate	9.03 <sup>B</sup>	4.44 <sup>B</sup>	2.96 <sup>B</sup>
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).<sup>3</sup> 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 $\chi^2$ 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.**

Intended outcome	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
Increase 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:  $KR + KA \times AT = IO$

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: Manfredi, 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., & Cavanaugh, 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](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 Interpretive 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](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](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](http://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.

Copyright of Journal of Interpretation Research is the property of National Association for Interpretation and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.