

# DEVELOPMENT OF A RESEARCH-BASED TOOL FOR EVALUATING INTERPRETATION



*Sam Ham and Betty Weiler*

SUSTAINABLE  
TOURISM



CRC

## Technical Reports

The technical report series present data and its analysis, meta-studies and conceptual studies, and are considered to be of value to industry, government and researchers. Unlike the Sustainable Tourism Cooperative Research Centre's Monograph series, these reports have not been subjected to an external peer review process. As such, the scientific accuracy and merit of the research reported here is the responsibility of the authors, who should be contacted for clarification of any content. Author contact details are at the back of this report.

## Editors

Prof Chris Cooper	University of Queensland	Editor-in-Chief
Prof Terry De Lacy	Sustainable Tourism CRC	Chief Executive
Prof Leo Jago	Sustainable Tourism CRC	Director of Research

## National Library of Australia Cataloguing-in-Publication

Ham, Sam H.  
Development of a research-based tool for evaluating interpretation.

Includes index.  
ISBN 1 920965 00 9.

1. Tourism - Interpretive programs. I. Weiler, Betty. II. Cooperative Research Centre for Sustainable Tourism. III. Title.

910.01

## Copyright © CRC for Sustainable Tourism Pty Ltd 2006

All rights reserved. Apart from fair dealing for the purposes of study, research, criticism or review as permitted under the *Copyright Act*, no part of this book may be reproduced by any process without written permission from the publisher. Any enquiries should be directed to the Director of Communications & Industry Extension [brad@crctourism.com.au] or the Publishing Manager [trish@crctourism.com.au].

# CONTENTS

<b>PREFACE</b>	<b>V</b>
<b>SUMMARY</b>	<b>VII</b>
OBJECTIVES OF STUDY	VII
METHODOLOGY	VII
KEY FINDINGS	VIII
FUTURE ACTION	IX
<b>CHAPTER 1 INTRODUCTION AND OVERVIEW</b>	<b>1</b>
BACKGROUND TO THE STUDY	1
Interpretation as a Concept, Activity and Process	1
General Goals of Interpretation	1
SUMMARY	3
<b>CHAPTER 2 LITERATURE REVIEW</b>	<b>4</b>
WHY EVALUATE? THE PURPOSES OF EVALUATION	4
Assessing Performance of Individuals	4
Providing Accountability	5
Assessing Economic Efficiency	5
Determining Why A Communication Program Is or Is Not Effective	5
Measuring Audience Impacts	5
WHEN, WHERE AND HOW TO EVALUATE INTERPRETATION	6
Types of Evaluation	6
Timing and Frequency of Evaluation	7
Evaluative Techniques for Measuring Interpretation Outcomes	8
Observations of the Audience During an Interpretive Program	9
Data Collection Techniques	18
WHAT TO EVALUATE?	22
SUMMARY	23
<b>CHAPTER 3 METHODS</b>	<b>24</b>
IDENTIFICATION OF THE INDICATORS	24
Development of an Indicator Pool	24
Development and Refinement of Indicators	24
Selection of the Final Suite of Indicators	25
SELECTION AND DEVELOPMENT OF DATA COLLECTION TECHNIQUES AND INSTRUMENTS	28
DEVELOPMENT AND VALIDATION OF THE RESEARCH INSTRUMENTS	29
Translation of Indicators A Through J into a Visitor Questionnaire	29
Development and Testing of the Observation Indicator	31
Finalisation of the Instruments	32
SUMMARY	34
<b>CHAPTER 4 FINAL PRODUCT: THE INTERPRETATION EVALUATION TOOL KIT</b>	<b>35</b>
USES AND BENEFITS	35
LIMITATIONS	35
RECOMMENDATIONS FOR FOLLOW-UP AND FUTURE RESEARCH	36
A FINAL WORD	36
<b>REFERENCES</b>	<b>37</b>
<b>AUTHORS</b>	<b>44</b>

## List of Tables

Table 1: Indicators for each type of setting	ix
Table 2: Timing of evaluation	8
Table 3: Summary of impact studies that used observations	11
Table 4: Summary of impact studies that used Visitor Employed Photography	12
Table 5: Summary of impact studies that used questionnaires	14
Table 6: Summary of impact assessment studies that used formal interviews	16
Table 7: Summary of data collection techniques for assessing the impacts of interpretation on audience <i>cognition</i>	19
Table 8: Summary of data collection techniques for assessing the impacts of interpretation on visitor <i>affect</i>	20
Table 9: Summary of data collection techniques for assessing the impacts of interpretation on visitor <i>behaviour</i>	21
Table 10: Categorisation and Ranking of Visitor Response Indicators Shared by the Two Sites	26
Table 11: Ranking of Unshared Visitor Response Indicators	27
Table 12: Final Suite of Visitor Response Indicators	27
Table 13: Rating of Primary Data Collection Options	28
Table 14: Food and beverage interpretation indicators	30
Table 15: Heritage interpretation indicators	30
Table 16: Nature interpretation indicators	31
Table 17: Final Indicators and Corresponding Question Numbers	32

## PREFACE

The Sustainable Tourism Cooperative Research Centre, an Australian government initiative, funded this research.

This project is a product of discussions in 2003 and 2004 with staff at The Sovereign Hill Museums Association, Port Arthur Historic Site and Tourism Tasmania – and thus the early stages of the project in particular were very much informed and driven by these organisations and other members of the industry reference group (IRG).

We wish to thank the following IRG members in particular for their contribution to and enthusiasm for this project from start to finish:

- Deb Lewis – Manager, Cultural Heritage Tourism, Tourism Tasmania
- Julia Clark – Manager, Interpretation and Collections, Port Arthur Historic Site
- Tim Sullivan – Deputy CEO and Museums Director, The Sovereign Hill Museums Association
- Pamela Harmon-Price – Senior Interpretation Officer, Queensland Parks and Wildlife Service

We are particularly grateful for the support given by these organisations, and especially the enormous contribution of Tourism Tasmania for assisting with both the Tool Kit trial (November 2004) and the Tool Kit launch (October 2005).

Others who contributed significantly to this project include:

- Sue Drummond, who was Sovereign Hill's IRG representative until her departure from SHMA mid-way through the project;
- Dr Anne Hardy, University of Northern British Columbia, who was a member of the research team in the first year of the project and, until her departure from the University of Tasmania, contributed to the early stages of the process of developing the indicators used in the final Tool Kit;
- Dr Xin Yu, who assisted considerably with the development of the trial version of the Tool Kit Excel-based package;
- Naoko Yamada and Professor Sam Ham, University of Idaho, who generously shared portions of the English version of their literature review taken from Yamada and Ham (2005, 2004) and from Yamada's Masters thesis entitled *A Manual for Evaluating Interpretive Programs in Japanese Nature Parks*; and
- Dr Sue Broad and Liam Smith, Monash University, who assisted with the literature review.

We also wish to acknowledge the many industry people who collaborated with us on various stages of the project:

- The managers and staff who participated in the two facilitated workshops, one at Sovereign Hill and one at Port Arthur, in February 2004;
- The staff who piloted the original field methods and both the original visitor questionnaire and the observation instrument at Port Arthur and Sovereign Hill in May 2004;
- The seven operators, and their interpretive staff and visitors, who field-tested the revised visitor questionnaire and data collection methods in their operations in June and July 2004:
  - The Sovereign Hill Museums Association;
  - Port Arthur Historic Site;
  - West Coast Wilderness Railway;
  - Hobart Cruises;
  - Boags Centre for Beer Lovers;
  - Cascade Brewery; and
  - Mole Creek Caves;
- The 14 individuals who participated in the workshop to trial the draft Interpretation Evaluation Tool Kit project in November 2004:
  - Anna Housego – Consultant, Writing & Research;
  - Anne Kerr – Consultant, Red Inca;
  - Cathie Plowman – Tasmania Parks & Wildlife Service;
  - Clare McShane – Casaveen Woolgrowers;
  - David Quon – Consultant, Food & Hospitality Service;
  - Iain Lang – Tasmania South Regional Tourism Association;
  - Julia Clark – Port Arthur;
  - Julie Marshall – Federal Hotels and Resorts;
  - Maria Fletcher – Tidemark Design;
  - Meaghan Newson – Forestry Tasmania;
  - Natalie Hayes – Meadowbank Wines;

- Trish Melinz – Guide Manager / Trainer;
  - Grant Cock – Hobart Cruises; and
  - Nicole Smith – Cascade Brewery;
- Gerald Englebreetsen, Tourism Tasmania, who advised extensively on a number of technical issues associated with the Tool Kit software and Manual;
- STCRC staff Luke Latimer, Brad Cox and Professor Leo Jago, who assisted with the conversion of the Tool Kit to a database platform; and
- Annita Allman, for assistance on many aspects of this project including general project support and administration, refining the Tool Kit spreadsheet, editing and desktop publishing the Tool Kit Manual, and polishing and desktop publishing the final report.

The researchers also wish to acknowledge the support of their respective universities for administrative and other in-kind support on this project.

## SUMMARY

### Objectives of Study

The aim of this research project was to develop, test and refine an evaluation tool kit for assessing the impacts of interpretation on visitors at heritage, nature and food and beverage tourism sites. In particular, we sought to develop a package that was user-friendly and practical for operators to use, yet based on theory and rigorous research methods. We also wanted to make sure it measured outcomes consistent with the goals of tourism organisations, including enhancing the visitor experience and increasing the degree to which visitors think about the values of the features being interpreted, as these outcomes lead to positive word-of-mouth advertising and business success.

A review of evaluation literature, including but not limited to interpretation evaluation, underpinned the research. To ensure that the Interpretation Evaluation Tool Kit would be of use and relevance to a wide range of tourism operators, our research relied on continual input and feedback from small and large operators and government agencies in three States.

### Methodology

We undertook a series of steps to identify interpretation outcomes that reflected the priorities of our industry partners. We then used these desired outcomes to construct a set of indicators. In particular, we wanted the final suite of indicators in the Tool Kit to:

1. Reflect the types of outcomes that users of the Tool Kit actually want from their interpretive programs (for example, enhanced visitor enjoyment, positive visitor attitudes about conservation, positive word-of-mouth advertising, provoking visitors to think about the values inherent at the site, etc.);
2. Be theoretically valid based on what is known about interpretation's potential impacts on how visitors think, feel, and possibly behave with respect to the things being interpreted for them; and
3. Demand minimal effort, expense, and little or no social research expertise from operators to collect and analyse the data, yet produce results that are both valid (i.e. accurately measuring what each indicator is supposed to measure) and reliable (i.e. producing consistent measurements that can be used to make meaningful comparisons of an operator's results over time and lead to informed judgments about the status and progress of interpretive programs).

The process began with a structured facilitation process at two sites (Port Arthur in Tasmania and Sovereign Hill in Victoria) where a wide variety of face-to-face interpretive programs are offered, in order to learn from staff what they felt were the most important indicators of 'successful' or 'effective' interpretation at those sites. To ensure that the selected indicators would be relevant to a wide cross-section of industry needs, participants in these discussions included program managers, front-line interpreters and guides, sales and marketing staff, and mid- or executive-level administrators. A pool of 54 responses resulted from these sessions, including 24 at Sovereign Hill and 30 at Port Arthur. Since the purpose of the evaluation Tool Kit was to provide a practical method of evaluating visitor responses to interpretation, we later reduced this to a list of 26 indicators that were *visitor responses* to interpretation (e.g. visitors enjoyed or were impacted by the interpretation). Collectively, they focused on three main categories of outcomes:

1. *cognitive outcomes*: what visitors might *think*, know or believe as a result of interpretation (e.g. understanding something, having a new view or being provoked to thought)
2. *affective outcomes*: what visitors might *feel* as a result of interpretation (e.g. appreciation of something, satisfaction with something, an attitude about something)
3. *behavioural outcomes*: what visitors might *do* or be *motivated to do* as a result of interpretation (e.g. stay longer at the site, buy something, positive word-of-mouth advertising)

We then applied five additional criteria that determined whether:

- a single comprehensive indicator could capture the essence of multiple indicators (i.e. were multiple items measuring the same phenomenon);
- each indicator was applicable beyond one particular site to other heritage sites and, as discussed below, to nature-based and food and beverage attractions;
- the relative importance (ranking) of each indicator by the original participants;
- the indicator made sense in terms of contemporary communication theory and research (i.e. could interpretation realistically be expected to achieve this outcome?); and
- the indicator could be measured simply yet reliably by a non-social scientist.

We used other researchers to independently apply our criteria to select indicators, and then compared their decisions with ours. Applying these criteria produced the final suite of eleven indicators which, in turn, became the focus for the development of the Interpretation Evaluation Tool Kit.

Each of the eleven indicators was then independently rated by three social scientists according to the ease and precision with which it could be measured by (1) a questionnaire, (2) a formal interview, (3) an informal interview (including focus groups), and (4) observation. We also assessed each method using widely agreed upon criteria, including the method's ability to produce valid and reliable information and the burden imposed on staff (i.e. expertise, time and other resources required) who would be using it to collect data for each indicator. These assessments resulted in a decision to use a visitor questionnaire to measure 10 of the 11 indicators (A to J), and participant observation to measure Indicator K.

Development of the Tool Kit's quantitative indicators required two final steps. First, we designed a visitor questionnaire for each of the three types of settings (food and beverage, heritage and nature sites). Second, we collected data from real visitors at multiple locations representing each type of site to make sure the information being produced was valid and reliable.

To do this, we developed a draft questionnaire using a well-established method for measuring the kinds of responses sought by Indicators A to J. Then we used our industry reference group and other social scientists to help us adapt some of the questions to the needs and peculiarities of each of the three settings. Through this procedure, three slightly different versions of the questionnaire were eventually produced (one for food and beverage sites, one for heritage sites and one for nature sites).

Each questionnaire was field tested on at least 100 visitors at multiple sites corresponding to each type of setting. This produced three data sets that we could then use for assessing the validity and reliability of each of the indicators.

We then statistically analysed the results to ensure that each of the indicators is indeed measuring some aspect of the kind of outcome it is intended to measure (i.e. they are valid). In addition, we subjected the data from the field tests to rigorous reliability testing to make sure that the sub-indicators used to comprise some of the overall indicators were producing consistently meaningful data with respect to the indicator they were intended to make up. When there was any doubt, we deleted the sub-indicator from the questionnaire so that the items remaining had the strongest reliability they could have. For this reason, the number of items making up the same indicator in the three questionnaires sometimes varies, and so the questionnaires are not interchangeable.

Indicator K is assessed using participant observation, which involves a staff member essentially posing as a visitor and systematically observing the interaction between an interpreter or guide and the audience. We field tested a set of observation procedures and a recording form using real interpretive programs at Sovereign Hill and Port Arthur. In both cases, the procedures were found to be reliable.

## **Key Findings**

We have produced a set of methods and instruments for assessing the outcomes of interpretation with respect to the suite of eleven indicators listed in Table 1.

The final product, the Interpretation Evaluation Tool Kit, provides three evaluation packages (each consisting of a five-minute visitor survey and a form for observing visitor behaviour) customised for three types of setting and includes:

- An Interpretation Evaluation How-To Manual; and
- A Tool Kit CD – which includes the software and files needed to conduct evaluations of face-to-face interpretation programs and activities.

Other than access to a computer, printer and photocopier, the Tool Kit is self-contained. It contains easy-to-follow explanations of all aspects of an evaluation, including how to interpret the results and identify steps that need to be taken to enhance interpretation's outcomes. It enables an operator to take the 'pulse' of interpretation and, over a period of time, to monitor trends in what it is achieving. The Tool Kit is designed mainly to be used to evaluate overall site interpretation, but it can also be used to evaluate a specific interpretive product or program. It can also be used when trialling a new interpretation program or product, but this must be done with caution as a range of factors can impact on the outcome of a single interpretive product at a single point in time.

The results from any of these uses can help to determine the extent to which interpretation benefits an organisation and helps to advance its goals.



**Table 1: Indicators for each type of setting**

Indicator*	Type of Setting		
	Food & Beverage	Heritage	Nature
A		Impact on current world view via empathy with historic period & people	Impact on appreciation of indigenous connections to nature
B	Elaboration (provoked visitors to thought)	Elaboration (provoked visitors to thought)	Elaboration (provoked visitors to thought)
C		Positive attitude toward heritage preservation	Positive attitude toward nature conservation
D	Global evaluation of interpretation	Global evaluation of interpretation	Global evaluation of interpretation
E	Desire to participate in additional interpretive activities	Desire to participate in additional interpretive activities	Desire to participate in additional interpretive activities
F	Desire to purchase a product or memento related to the place	Desire to purchase a memento or souvenir related directly to the site story	Desire to purchase a memento or souvenir related directly to the site story
G	Desire to stay longer	Desire to stay longer	Desire to stay longer
H	Desire to return for a repeat visit	Desire to return for a repeat visit	Desire to return for a repeat visit
I	Positive word-of-mouth advertising	Positive word-of-mouth advertising	Positive word-of-mouth advertising
J	Interpretation was relevant & meaningful to visitors' lives	Interpretation was relevant & meaningful to visitors' lives	Interpretation was relevant & meaningful to visitors' lives
K	Visitors provoked to interact with the presenter	Visitors provoked to interact with the presenter	Visitors provoked to interact with the presenter

\* Indicator K is measured using the Observation Form. All other indicators are measured using the Visitor Questionnaire.

## Future Action

While the Tool Kit provides a valid, reliable, easy-to-use and rapid-response measure of 'how well' or 'how much' is being achieved, it does not provide information about the longer-term impacts of interpretation on visitors once they have returned home. Nor does it provide answers as to *why* interpretation is or is not achieving an organisation's desired outcomes. A more complex research design is required to assess cause-and-effect relationships.

The Tool Kit is designed to evaluate face-to-face interpretation, not self-guided or non-personal media such as signs, exhibits, websites, brochures etc. Thus there is scope for follow-up research to develop an Evaluation Tool Kit for non-personal interpretation media. Much of the process outlined in Chapter 3 could be replicated for evaluating other media, and the process can also provide a guide for researchers seeking to develop methods and instruments customised to a particular setting or operation not specifically covered by the current Tool Kit package (e.g. indigenous tourism operations, zoos, botanical gardens, events, etc.).



## **Chapter 1**

# **INTRODUCTION AND OVERVIEW**

While systematic evaluation of interpretation is something that nearly everyone agrees is a good idea, it has not become a routine part of interpretive program management in most places. This is partly because conducting evaluations is seen as complicated and time-consuming by most organisations. If evaluations are conducted at all, they are often one-off efforts conducted by outside evaluation experts at considerable expense to the organisation.

This research project was aimed at developing a rigorous yet practical tool for evaluating the outcomes of interpretation that could be used by any organisation wanting to monitor the ongoing performance of its interpretive products. It makes an important contribution in assisting operators to use interpretation more effectively and to evaluate its outcomes more regularly, to improve the contribution of interpretation not only to visitor awareness and appreciation of site values, but to the commercially important outcomes of visitor satisfaction, word-of-mouth advertising and competitive advantage. Thus the aim of the Interpretation Evaluation Tool Kit is to aid operators to effectively gauge how well they are doing in achieving their interpretive missions. Potential users include public, non-profit and private tourism operators, park and protected area management agencies, museums, attractions that provide interpretation as part of their product, and state tourism authorities who see their role as enhancing tourism product.

## **Background to the Study**

For readers unfamiliar with interpretation, this section provides a very brief introduction to interpretation. This is followed by a review of the goals of interpretation from the perspective of the organisations that provide or deliver it, whether protected area managers, heritage site managers, or tourism businesses such as tour operators and attractions.

## **Interpretation as a Concept, Activity and Process**

Interpretation in this report includes both face-to-face and non-personal interpretation delivered through a range of media such as printed materials, signs, exhibits, self-guided walks, pre-recorded tour commentaries on cassettes or videos, virtual tours, and other electronic media. Originally defined by Tilden (1957), interpretation is an educational activity aimed at revealing meanings and relationships to people about the places they visit and the things they see and do there, and many protected area management agencies identify strongly with the premise of interpretation as being 'educational'. For most commercial operators, interpretation is communication with visitors in entertaining and creative ways that add value to the product and thus improve their bottom line and sustainability as a business. Most professional interpreters and practitioners agree that good interpretation does both of these things, and increasingly needs to do even more, in the sense that interpretation is used to leverage many other outcomes for an organisation.

## **General Goals of Interpretation**

While there is not 100% agreement on the potential goals of interpretation, a number of authors (Lewis 1980; Sharpe 1982; Ham 1992; Regnier, Gross & Zimmerman 1994; Ham & Krumpe 1995; Knudson, Cable & Beck 1995; Beck & Cable 1998; Department of Natural Resources and Environment Victoria 1999; Ham, Housego & Weiler, 2005; Moscardo, Ballantyne & Hughes 2006; Ward & Wilkinson 2006) have argued that interpretation aims to do one or more of the following:

1. Enhance visitor experiences;
2. Protect resources at sites;
3. Protect visitors;
4. Increase public support for an agency and its management policies;
5. Add to or broaden visitors' perspectives about a place or idea; and
6. Enhance their knowledge and foster positive attitudes and behaviours with respect to the natural and cultural environment.

Each of these is discussed in a little more detail in the following paragraphs, with some examples of studies that have either attempted to or proposed the need to assess interpretation in relation to these goal categories.

### ***Enhancement of visitor experiences***

Interpretation can add value to a leisure experience. For instance, Pearce and Moscardo (1998) found interpretation at the Skyrail Rainforest Cableway near Cairns was positively linked to enhanced visitor satisfaction. Ham and Weiler (2000) found five interpretive services in Panama Canal protected areas contributed more to overall tourist satisfaction than any of six non-interpretive services. Hwang, Lee and Chen (2005) reported positive relationships between interpretation quality and tourists' satisfaction, and Hughes and Morrison-Saunders (2002) found visitors rated trailside interpretive signs as contributing to a positive learning experience. Indeed, making recreational experiences more enjoyable is almost always a goal of interpretation (Knudson, Cable & Beck 1995) because visitors come to recreational settings for relaxation or fun. Interpretation can also provide visitors with an escape from daily pressures (Lewis 1980). Since visitors engage in tourist experiences for enjoyment, assessing how well we entertain them is essential. Interpretation must also be enjoyable in order to hold visitors' attention (Sharpe 1982; Ham 1992). Although entertainment is not the only major goal of interpretation, particularly for managers of protected areas, it is one indicator of successful interpretation.

### ***Protection of resources***

Interpretation is commonly used for resource protection. As Armstrong and Weiler (2003) found in their 12-month study of guided interpretive tours in protected areas in Victoria, messages about minimising impacts on the natural environment were observed to be the most frequently communicated interpretive messages. A number of recent studies have focused on interpretation's capacity to mitigate visitor-induced management problems. For example, a three-year study by Ham and Weiler (2005) led to theory-based interpretive signs capable of reducing off-track walking at Port Campbell National Park; Mayes, Dyer and Richins (2004) researched the use of interpretation to enhance pro-environmental beliefs of visitors at dolphin sites; Beckmann (2003) used a research-based rationale for designing interpretive texts to promote minimal impact in the Australian Alps; Lackey and Ham (2003) researched the application of theory-based interpretive messages to reducing human-bear conflicts in Yosemite National Park; Cialdini (1996) examined norm activation as a strategy for deterring littering; Winter, Cialdini, Bator, Rhoads and Safarin (1998) explored normative approaches to developing interpretive signs aimed at reducing depreciative behaviour in US recreation areas; and research by Ward and Roggenbuck (2003) and Widner and Roggenbuck (2000) revealed message strategies for using interpretation to combat theft of petrified wood in Petrified Forest National Park, USA.

According to Sharpe (1982), interpretation can also attempt to control problems such as unauthorised fossil collecting, defacing archaeological resources, littering, driving vehicles off park roads and poor camper etiquette. It can assist people to become aware of, appreciate and understand the site they are visiting and lead them to responsible use. By educating people, proper use of park facilities or resources is promoted to reduce behaviours that threaten protected values. Kuo (2002) explains that interpretation is a 'soft visitor management strategy' that can be used to support 'hard visitor management strategies' such as regulations and restrictions, by explaining the necessity for the implementation of these 'hard' strategies and increasing visitor understanding of the impacts of inappropriate on-site behaviour. Thus, where 'problem' on-site behaviours occur, evaluating the effectiveness of communication targeted at these behaviours can be a measure of successful interpretation.

### ***Protection of visitors***

Visitors unknowingly behave in ways that threaten their own comfort and safety. Several recent studies have examined interpretation's role in protecting uninformed or misguided visitors from on-site hazards. For example, Ballantyne and Hughes (2006) used formative evaluation to guide development of interpretive signs aimed at warning Brisbane Forest Park visitors about the dangers associated with bird feeding; and Lackey, Ham and Hall (2002) and Ham and Weiler (2005) examined the effectiveness of interpretation at managing visitor behaviours that not only threaten the natural environment but also endanger visitors. Espiner (1999) evaluated interpretive warning signs aimed at alerting visitors to hazards in glacier settings. Howard, Lipscombe and Porter (2001) evaluated an interpretive program aimed at reducing inappropriate visitor interactions with dingoes on Fraser Island. Thus, the success of interpretation is often gauged by its contribution to protecting visitors.

### ***Increasing public support for the management policy and agency***

Interpretation can play an important role for a heritage or protected area management agency in terms of generating support from the public. Through interpreting the natural and cultural resources or explaining agency policies and goals, people come to better understand management goals and approaches. For example, some people are sceptical about controversial management issues such as timber harvesting or fire management. Employing interpretation to explain why the approach is taken helps those people better understand and may

reduce their negative attitude toward such management approaches. Consequently, this can improve public image and increase public support for the agency.

Ham and Krumpal (1995) argue that interpretation can contribute to producing a positive image and support toward the agency outside the protected area as well as from on-site visitors, provided the public is interested in learning about management policies or the agency and does not already hold a strong negative attitude toward them. In this light, Zuefle (2004) reported advantages of audience analysis in designing interpretation aimed at enhancing relations with stakeholders in wildlife management, and Ham, Housego and Weiler (2005) include constituency building as one of four key goals of interpretive planning. Thus, another measure of the success of interpretation is whether it increases public support for the management policy or the agency in general.

### ***Broadening visitors' perspectives, enhancing knowledge and influencing attitudes and behaviours***

Interpretation can also add other perspectives to visitors' lives. By inspiring and provoking, interpretation can reveal a world many may have never seen before (Ham 1992; Regnier, Gross & Zimmerman 1994; Larsen 2003; Pastorelli 2003; Ward & Wilkinson 2006). Cameron and Gatewood (2000) documented the effect of interpretation on visitors' perspectives at historic sites. "Most visitors to interpretive settings are interested and eager to learn the deeper truths of the places they have self-selected to experience" (Beck & Cable 1998, p.47). Often, interpretation aims to extend this awareness and concern about their natural and cultural environment beyond the area they are visiting. It can assist people to build relationships with or understand interrelationships between what they are observing or experiencing at the site and their lives (Larsen 2003; Lewis 1980). Interpretation can help visitors to better understand the idea of coexisting with the environment and can increase their awareness of their place in the total environment (Sharpe 1982; Tisdell & Wilson, 2001). A recent study by Madin and Fenton (2004) examining the effectiveness of visitor interpretation in the Great Barrier Reef Marine Park found that "the interpretive programme was an effective means of affecting visitors' perceived knowledge of the reef environment" (p.132). They found that "visitors who participated in greater numbers of activities felt that they had a greater knowledge of the reef environment and a greater understanding of the actual and potential anthropogenic impacts on the reef environment than those who participated in fewer activities" (p.132). Beck and Cable (1998, pp.55-56) state that interpreters have 'a responsibility to broaden horizons and provoke a more sensitive, more caring attitude toward our cultural and natural resources'. Beaumont (2001), on the other hand, found that attitudes of ecotour clients were largely already positive and that interpretation mainly reinforced or strengthened existing attitudes.

According to Knapp and Benton (2005) and Knudson, Cable and Beck (1995), a significant on-site interpretive experience may even lead people to an intelligent action to sustain their environment back home and to relate to the environment in which they live, or as Higham and Carr (2002) describe, interpretation programs can foster "behavioural change relating to domestic lifestyle that may contribute to the long-term benefit of the environment" (p.279). Evaluating whether such outcomes are actually achievable raises many questions, but as we discuss in Chapter 2, some studies have indeed tried to address them.

## **Summary**

As we shall see in the literature review (Chapter 2), evaluation of interpretation in Australia and around the world has tended to focus on visitor numbers, tourist satisfaction and factual recall, with less focus on interpretation's outcomes in relation to more long-term organisational goals. As mentioned above, where richer and more in-depth information has been collected from visitors, they tend to be one-off cross-sectional studies conducted by university researchers or students, as such studies require considerable research expertise and/or resources that make it too 'burdensome' for operators to do routinely. This report now turns to the published literature to extract what has been learned in previous research with regard to evaluating the outcomes of interpretation.

## Chapter 2

# LITERATURE REVIEW

Few attempts have been made to bring together the interpretation evaluation literature in a systematic way, and few published studies have comprehensively reviewed interpretation evaluation research in the context of the broader evaluation literature. Thus the picture is sometimes unclear as to what interpretation evaluation research *currently addresses* and the advances that interpretation evaluation *could* and *should* be achieving. Reviewing the research is thus a first step toward identifying gaps that an interpretation evaluation tool kit might fill.

This chapter aims to highlight, using examples from both overseas and Australia, the extensive body of evaluation research that has been brought to bear on interpretation in the past 30 years. The literature review first looks at the ‘why’ of interpretation evaluation (the purposes of evaluation), then briefly reviews ‘how’ (approaches to or types of evaluation, timing, frequency and data collection techniques) and finishes with ‘what’ (identification, selection and measurement of indicators).

As will be evident in Chapter 3, the concepts, resources and previous studies reviewed here have helped to inform the current study in terms of its purpose and focus. What emerges from this literature review is that there are three overarching constructs of interest that interpretation may try to influence: cognition, affect and behaviour. The chapter concludes by using these three broad constructs as a framework for determining ‘what’ to evaluate, which is where the current study makes its strongest methodological contribution.

## Why Evaluate? The Purposes of Evaluation

Over the past 30 years, many authors have reviewed evaluative concepts, including their application to interpretation programs (Wagar 1976; Roggenbuck & Propst 1981; Hodgson & Fritschen 1984; Pearce & Moscardo 1985; Ham 1986; Medlin 1992; Knudson, Cable & Beck 1995; Uzzell 1998a; Beckman 1999; Department of Natural Resources and Environment Victoria 1999; Jacobson 1999; Savage & James 2001; Yamada & Ham 2004). In addition to the field of interpretation, several authors studying recreational services and social programs have discussed the importance of evaluation (Suchman 1967; Lundegren & Farrell 1985; Theobald 1987; Henderson & Bialeschki 1995; Rossi, Freeman & Lipsey 1999). Many reasons for evaluation exist and they vary for each situation, but they include:

1. Assessing performance of individuals;
2. Providing accountability;
3. Assessing economic efficiency;
4. Determining reasons why a communication program is or is not effective; and
5. Measuring impacts or outcomes.

However, while it is clear that there are many components to evaluation, the Department of Natural Resources and Environment Victoria (1999) found that “most organisations do not address all components of the overall evaluation process” (p.44). The report concluded that evaluation is an area that requires improvement due to budget limitations, lack of staff training, difficulties in measuring outcomes such as knowledge, and standards and performance measures not being established during interpretive planning stages.

## Assessing Performance of Individuals

*Performance appraisal*, *performance evaluation* and *workplace assessment* are essentially synonymous, involving the assessment of an individual’s relative performance on the job, usually with a view to using that evaluation to provide feedback to the individual, identify the individual’s training and developmental needs, and/or assess their eligibility for some kind of reward or promotion. As such, performance evaluation needs to be sensitive to the type of employer (large or small operator; national park; non-profit organisation; local government authority), the organisational culture of the employer and the work conditions on the particular day(s) on which the appraisal is conducted (size and type of audience, weather etc.). This is the focus, for example, of the assessments done in conjunction with Australia’s EcoGuide Certification Program or as part of judging nominees for awards of excellence by professional interpretation and tour guiding associations. In the context of interpretation evaluation, the purpose of performance assessment should be to identify why and how performance is (and is not) enhancing interpretation’s impact, as a basis for improvement.

## Providing Accountability

Another primary reason for evaluation is providing accountability. Accountability means that professionals are responsible for explaining or demonstrating that a program makes beneficial contributions to management, the agency or society. According to Lundegren and Farrell (1985) and Henderson and Bialeschki (1995), there are three kinds of accountability:

1. Program accountability, which is the quality of work and whether or not it meets its objectives;
2. Process accountability, which addresses the adequacy of procedures and whether things are done as planned; and
3. Expenditure accountability, which assesses whether money is spent appropriately for the program.

Accountability also determines the legitimacy and value of the interpretation in relation to its burden on personnel and costs. Without accountability, it is difficult to convince administrators of the value of a particular interpretation program or product and get continued financial support from an agency.

Henderson and Bialeschki (1995) argue that evaluation thus helps inform decisions regarding whether to continue implementing a particular interpretation program, to rationalise or to invest more staff effort or money. The authors also point out that evaluation can be driven by the need to comply with external standards set by government or other professionals in order to obtain an external accreditation or license.

## Assessing Economic Efficiency

Assessing economic efficiency involves evaluating overall productivity and justifying an interpretive program's value beneficially and financially. Since an administrator is usually concerned with the task of choosing how to allocate scarce budgets to maximise benefits, it is important to evaluate the outcomes of a particular interpretive product or program compared to its costs. Measuring program impacts per unit cost can provide a reasonable indication of economic efficiency (Ham 1986). If an interpretive program has an impact but requires high cost relative to the impact, it may not be supported. As Uzzell (1998a) discussed, in the case of public sector interpretation there is a need to measure and quantify a program's costs and benefits because public investment will be reduced or terminated when a program is not cost effective or cannot demonstrate its cost-benefit. Thus, demonstrating benefits in relation to costs helps secure financial investment in both public and private sector interpretive programs and also increase budget allocations. Assessing economic efficiency also allows expenditure accountability.

## Determining Why A Communication Program Is or Is Not Effective

Determining the reasons why a program is or is not effective is important. Focusing only on impacts may provide little future direction unless evaluators can explain the reasons behind the program's success or failure. Thus, there is a need to identify specific factors that have contributed to its success or failure. The factor can be the topic, message, personal performance, medium, physical environment or settings, or something else (Ham 1986; Ham & Weiler 2005; Yamada & Ham 2004). This information allows interpreters to focus their resources on the most effective parts of the program and reduce other components (Beckmann & Savage 2003; Weinreich 1999). When interpreters know how a particular interpretive program or product works, it becomes easier to set future objectives (Henderson & Bialeschki, 1995). Knowing the reasons for success or failure allows interpreters to share what was learned with others (Jacobson 1999). This information is critical in terms of providing interpreters with practical information about which interpretive program or product is most likely to be effective (Belch & Belch 2001) and what action needs to be taken for improvement.

## Measuring Audience Impacts

Interpretation impacts or outcomes related to an audience provide critical information about program effectiveness. Henderson and Bialeschki (1995) state that "[d]etermining the impact, effects, outcomes, and results of a program, area or facility, or administrative procedure is the bottom line of evaluation" (p.11). Measuring what occurred as a result of a program determines its degree of success in reaching the targeted audience and in providing planned resources, services and benefits. Thus, it measures program effectiveness.

From a theoretical perspective, the desired outcomes of any communication program can be classified according to the kinds of responses they might elicit from an audience. Ham (1982) classified the outcomes of interpretation according to three main kinds of audience responses:

1. *Cognitive responses* (focused on learning, knowledge acquisition, or information processing);
2. *Affective responses* (focused on feelings, attitudes or emotions); and
3. *Behavioural responses* (focused on actions or inactions desired of the audience).

While others have used slightly different classification systems (see Dierking, Adelman, Ogden, Lehnhardt, Miller & Mellen 2004), there is consistency in terms of the responses captured.

Ham (1986) and Jacobson (1999) point out that assessing secondary or unanticipated outcomes is another aspect of impact assessment. These outcomes can be 'extra benefits' (Ham 1986, p.17) such as praise letters

from visitors, or can be undesirable outcomes such as visitors' complaints about too many rules after hearing them through interpretive programs. If the interpretation has little positive impact on the intended audiences, resources, or services, it is not worth doing. Therefore, measuring impacts or outcomes is necessary for convincing administrators of the value and legitimacy of an interpretation program or product and contributes to demonstrating program accountability.

In summary, various reasons for evaluation exist and they can overlap and contribute to each other. Indeed, evaluation is most useful when it is multi-focused. As Henderson and Bialeschki (1995) point out, if evaluation has a single purpose, evaluators may overlook certain outcomes and explanations derived from the evaluation. However, it is the final purpose – evaluating impacts or outcomes – that has the clearest links to the organisational goals of interpretation discussed in Chapter 1, and thus we return to these outcome categories in Chapter 3 as our basis for determining what should be measured in interpretation evaluation.

## **When, Where and How to Evaluate Interpretation**

As with the purposes of evaluation, choices and decisions about approaches and methods are informed by the broader evaluation and research methods literature, and can also be illustrated with examples from the interpretation evaluation literature.

### **Types of Evaluation**

Several types of evaluation exist, which can be classified according to their purposes as outlined in the previous section. For example, a question about interpretation outcomes is often “are the outcome goals and objectives being achieved?”, while a question about cost efficiency would be “are resources used efficiently?” (Rossi, Freeman & Lipsey 1999, p.88). Rossi, Freeman and Lipsey (1999: 62) further point out that “an important step in designing an evaluation... is determining the questions the evaluation must answer”. They categorised five major types of evaluation corresponding to the types of questions that are asked:

1. Needs assessment;
2. Program conceptualisation and design assessment;
3. Assessment of program process;
4. Efficiency assessment; and
5. Impact assessment.

As will be evident in the next section, some of these require implementation prior to or in the early stages of interpretive product or program development.

### ***Needs assessment***

Needs assessments determine whether there is a need for an interpretive product or program and what is most appropriate for the need (Rossi, Freeman & Lipsey 1999). One approach is to identify a problem that can be addressed through interpretation, its magnitude and extent, and the target population. Rossi, Freeman and Lipsey (1999) discuss the importance of needs assessments:

The primary rationale for initiating or maintaining a social program is a presenting or incipient social problem ... that legitimate social agents endeavour to remedy. ... If there is no significant problem or no perceived need for intervention, there is generally no basis for affirming the value of a program that purports to address this non-problem. (p.64)

Hobbs (1987) identifies five features of needs assessments strategies: who is the needs assessment intended to inform/influence; what purpose is intended; whose needs are to be assessed; what questions are asked; and what resources are available. Needs assessments produce supporting evidence (Hobbs 1987) and provide fundamental information for deciding whether an interpretive program should be offered or not. A program cannot be effective at ameliorating a problem if there is no problem or the program does not actually relate to the problem (Rossi, Freeman & Lipsey 1999). Of course, the ‘problem’ may be as straightforward as the desire to increase competitive advantage in the market place, attract more visitors, attract a different clientele, improve levels of satisfaction or increase repeat visitation.

### ***Program conceptualisation and design assessment***

This evaluation involves questions about the interpretive program's structure, functions and procedures that are appropriate to attain its objectives. Such assessments provide information about the concepts, assumptions, and expectations of the interpretive product or program. They are often called front-end evaluations or formative evaluations. Front-end evaluations take place in planning a program, to “learn about the audience before a program has been designed in order to predict how visitors will respond once the project has been developed” (Diamond 1999, p.16). According to Loomis (1996) and Screven (1990), they provide baseline information for



conceptualising and designing an interpretive program or product by understanding potential audiences' needs, expectations, prior knowledge and experience, and learning styles. Front-end evaluations are essential if little information about an audience is available (Ballantyne & Hughes 2006; Moscardo 1999a; Uzzell 1998a). Formative evaluations may examine the adequacy of the message, the interpretive program's organisation, and the media to be used in a program. Thus, needs assessment and formative evaluations help produce effective interpretation before a lot of time, cost, and staff effort are invested. They can consequently reduce the gap between what interpretation will be and what it should be.

### ***Assessment of program process***

According to Ham (1986), program process evaluations occur during the implementation of an interpretive program. They involve questions about how well the program is being conducted. Rossi, Freeman and Lipsey (1999) call them program monitoring evaluations and define their goals as assessing whether a program is reaching target audiences; whether it is being conducted as designed; and whether resources are being expended in the proper way in conducting the program. Process evaluations can also identify problems to be improved before an interpretive product or program is fully developed (Jurin, Danter & Roush 2000). In contrast to program conceptualisation and design assessments, process evaluations involve actual audiences.

### ***Efficiency assessment***

Efficiency assessments answer questions about program cost and cost-effectiveness. In interpretation (Ham 1986; Yamada 2002), they determine the degree to which an interpretive product or program produces benefits in relation to cost, using cost-benefit and/or cost-effectiveness analysis. Cost-benefit analysis measures the relationship between costs and benefits and is usually expressed in monetary terms. It can be used to compare programs that have different goals (Riddick & Russell 1999). For example, a program that costs \$200 and has benefits calculated at \$400 results in a net benefit of \$200. Cost-effectiveness measures costs in relation to both monetary and non-monetary outcomes derived from a program (Riddick & Russell 1999). It can compare the costs and evaluate programs with similar goals.

### ***Impact assessment***

Impact assessments or *summative* evaluations usually occur after implementing programs. They answer questions about program impacts or outcomes. These might include, as discussed earlier, enhancement of the experience, and/or influencing visitors at a cognitive (learning), affective (feeling) or behavioural (doing) level. It may even include interpretation's impact on resource protection and visitor management/protection. Ham (1986) points out that most authors pay attention only to intended outcomes and little to unexpected outcomes. Unintended outcomes can involve both positive and negative impacts and can help to recognise how an interpretive product or program affects audiences.

In summary, there are different types of evaluation and these are related to and indeed determined by the purpose of evaluation. Needs assessments and program conceptualisation and design assessments help to develop effective interpretation before implementing a product or program and thus without expending large amounts of resources. Program process evaluations identify needed improvements, ascertain efficiency, and provide accountability. Impact assessments are necessary to establish the interpretive product or program's effectiveness and value. Efficiency assessments examine program productivity and are critical for defending its value. All types of evaluation contribute to demonstrating program accountability.

## **Timing and Frequency of Evaluation**

As outlined in the previous section, evaluations can take place at all stages of interpretive product and program development: during planning or designing (front-end or formative), during implementation (process), and after completion (summative). The timing of an evaluation is determined by the kinds of questions it is designed to address (Ham 1986). In other words, the purpose of evaluation decides the timing as well as the type of evaluation.

Needs assessments, conceptualisation and design assessments and efficiency assessments occur during the planning and designing phase because they provide information for decision making and how a program should be created. Process evaluations must occur during the implementation since they require monitoring a program. Besides process evaluations, impact assessments and efficiency assessments can be undertaken at any time after implementation begins (Ham 1986). Impact assessment and efficiency assessment may also take place after the completion of programs. Rossi, Freeman and Lipsey (1999) describe efficiency assessment as appropriate at all stages but most common before program implementation or after the completion of impact assessment. Table 2 summarises when each type of evaluation typically occurs.

**Table 2: Timing of evaluation**

Before Program	During Program	After Program
Needs assessments	Program process assessments	Impact assessments
Program conceptualisation and design assessments	Impact assessments	Efficiency assessments
Efficiency assessments	Efficiency assessments	

Source: Ham (1986)

As Wagar (1976) and Roggenbuck and Propst (1981) point out, evaluation should occur as early as possible, to reduce wasting of time, cost and staff efforts. However, in the case of evaluating interpretation's impact on visitors' knowledge, attitudes and behaviours, evaluations of program accomplishment must occur after implementation begins, and extending ideally to when the visitors have returned home. Although such longitudinal studies are costly and time-consuming, a few evaluations have indeed tracked the long-term impacts of interpretation (e.g. Adelman, Falk & James 2000; Knapp & Yang 2002; Knapp & Benton 2005; Knapp & Benton 2006). In addition to the timing of evaluation, frequency of evaluation is also important. Roggenbuck and Propst (1981) point out that evaluations should occur routinely after implementation. Continuous feedback can provide new information in accordance with any change of audience, staff and other resources. According to these authors, any change to policies, goals, access to resources and visitor profiles can cause an interpretive program to become less effective.

Thus, assessing the appropriateness of the objectives themselves is a necessary function of evaluation in order to meet management goals and visitor interests (Roggenbuck & Propst 1981). In the case of protected areas, for example, this should occur periodically in accordance with changes in resource management systems, the resource itself, interpreters' knowledge of the resource and visitors' preferences through time (Roggenbuck & Propst 1981).

In summary, evaluations occur at all stages of program development; before, during, and after implementing a program. The timing is determined by the purpose of evaluation. Measuring an interpretive product or program's outcomes is just one type of evaluation, but it is an important focus. While other types of evaluation are more useful for feedback and updating purposes, impact assessments answer the questions of accountability and productivity that program administrators are increasingly called on to address.

Although impact assessments are only one of several forms of evaluation, they are the focus of this project because they address questions about interpretation's impact on audiences. Accordingly, they are undertaken after interpretation is implemented.

## **Evaluative Techniques for Measuring Interpretation Outcomes**

Many authors discuss evaluative techniques for communication programs (Wagar 1976; Roggenbuck & Propst 1981; Ham 1986; Wright & Wells 1990; Knudson, Cable & Beck 1995; Uzzell 1998b; Beckman 1999; Jacobson 1999; Jurin, Danter & Roush 2000; Yamada & Ham 2004). Wagar (1976) outlined 12 summative evaluative techniques such as direct audience feedback, auditing by an expert, observation of behaviour, length of viewing or listening time, questionnaires, self-testing devices and others. Some authors classify evaluative techniques based on Wagar's reports (Ham 1986; Roggenbuck & Propst 1981), adding techniques used in other types of evaluation such as front-end evaluation, formative evaluation, and program process assessment.

Based on these analyses and other evaluative studies, ten techniques are presented here as potentially useful for measuring interpretation's impacts or outcomes. They are:

1. Self-testing devices;
2. Observations of the audience during an interpretive program;
3. Visitor employed photography;
4. Observations of audience behaviour after an interpretive program;
5. Observations of an audience's behaviour traces;
6. Audience questionnaires;
7. Formal audience interviews;
8. Informal audience interviews;
9. Focus group interviews with audience members; and
10. Personal meaning mapping.

The following sections describe the procedures used for each technique, as well as outlining the strengths and weaknesses of each. Note that there are other evaluation techniques that may be more appropriate for evaluations undertaken for other purposes, as outlined in the first section of this chapter.

### ***Self-testing devices***

Usually this involves a recording quizboard (e.g. Wagar 1972), interactive audiovisual system, or computer to measure the audiences' factual knowledge gain. Thus, self-testing devices may be used after implementing programs to determine program effectiveness, if factual knowledge gain is an objective of the program. Self-testing devices ask audiences to choose a correct or incorrect answer from among multiple choices, and then the number of correct and incorrect responses to each question is recorded. Alternatively, audiences are asked to enter the correct answer. Besides measuring effects of a program, self-testing devices can be entertaining for audiences (Moses, Epstein & Wiseman Inc. 1977; Roggenbuck & Propst 1981). If they are well constructed, they can also reinforce the audience's learning (Moses, Epstein & Wiseman Inc. 1977). However, their accuracy is limited because they can include data that are not representative, due to the fact that respondents voluntarily participate and the same person may repeatedly respond to the questions. Other issues include inability to measure the depth of understanding, breakdown or vandalism of the device (Wagar 1976; Roggenbuck & Propst 1981), and the cost of installation and maintenance.

### **Observations of the Audience During an Interpretive Program**

Although observational methods are sometimes used in formative evaluations (e.g. Savage & James 2001), observing audiences is most commonly employed in evaluations of fully developed interpretive programs and devices. Observations typically monitor the behaviours or movements of audiences, measuring the length of time spent, and sometimes recording conversations. For personal interpretive programs, observations often assess the audiences' attention, while for non-personal interpretive programs, observations usually focus not only on the audiences' attention but other behaviours and learning as well.

To measure an audiences' attention to a personal interpretive program, an observer scans and records the audience at regular intervals, identifying how many people are paying attention to the program and what is being presented. This determines what part of the interpretive program or the interpreter's behaviour is associated with the audiences' response, allowing suggestions to be made to improve the content and delivery. Although an interpreter can directly observe his/her audience, the data is more accurate if someone else records them in a systematic way (Wagar 1976). Wagar recommended plotting the results in a graph to detect any differences in attention levels both within a presentation and between different presentations. Dick, Myklestad and Wagar (1975) found that using multiple observers produced consistent data, demonstrating the reliability of this technique.

Many if not most evaluations of non-personal interpretation (such as signs and exhibits) also involve measures of attention capture and holding power (e.g. Beer 1987; Bicknell & Mann 1993; Bitgood & Patterson 1993; Falk 1993; Phillpott 1996; Cole, Hammond & McCool 1997; Falk 1997; Sandifer 1997; Cole 1998; Bitgood 2000; Korn & Jones 2000; Chiozzi & Andreotti 2001; Lackey, Ham & Hall 2002). To measure how well a non-personal interpretive program attracts and holds an audience's attention, the behaviour and length of time visitors spend at an exhibit, audio program, or sign can be observed. To measure the attractiveness of an interpretive program, an observer counts how many visitors stop to look at an interpretive program and divides this by the number of visitors passing by (Loomis 1987). To determine how long an interpretive program can hold an audience's attention, an observer measures how long the audience actually spends viewing, reading or listening to the interpretive program. This information can be compared with the required or ideal time to complete hearing, reading, or viewing of the interpretive program (Wagar 1976; Loomis 1987; Donald 1991). This method can determine the effects of any improvements to the interpretive program by comparing the times before and after any changes are implemented (Bitgood & Patterson 1993; Harvey, Loomis, Bell & Marino 1998). Although most researchers agree with Donald (1991), who argued that the length of time spent may measure conditions for learning rather learning itself, Borun, Dritsa, Johnson, Peter, Wagner, Fadigan, Jangaard, Stroup and Wenger (1998) found that length of time spent did relate to learning level. They compared the time spent at an exhibit among families and concluded that families who spent more time at the exhibit had understood and connected to their prior knowledge more.

However, it should be cautioned that longer attention times do not always mean attractiveness or effectiveness of a message, but may mean the message is complicated or hard to understand (Roggenbuck & Propst 1981; Ham 1992). It should also be remembered that reading speed varies among people (Roggenbuck & Propst 1981; Donald 1991). Uzzell (1998b) therefore argues that using observation with interviews, questionnaires or attitude scales makes it more effective. Since recording time spent viewing an interpretive program involves only a single or few behaviours, it is relatively easy to achieve consistency and reliability among multiple observers (Henderson & Bialeschki 1995).

An auto camera can also be used for observing the audience's attention, instead of relying on personal observers. This usually involves a time-lapse photography system or trip-sensor system set in a hidden place to take pictures of people at regular intervals. Time-lapse photography places little burden on observers and produces a permanent visual record of people's behaviour (Wagar 1976; Roggenbuck & Propst 1981). Also, such mechanical observation ostensibly provides a less biased account of what occurred than human memory alone can provide (Grove & Fisk 1992). However, such systems are fairly expensive and require management for breakdown as well as film processing and analysis (Roggenbuck & Propst 1981). In addition, efforts must be made to preserve the visitor's privacy. Cole (1998) and Stevens and Hall (1997) are examples of studies using cameras to observe visitor behaviour.

Tracking visitors throughout a display area is another method to measure attention and behaviour, allowing comparisons to be made of the impact of different exhibits. An observer unobtrusively follows visitors and records their behaviour patterns or movements within a specified area. This can determine which displays attract and hold attention, and determines how visitors interact with an interpretive medium. Korn (1988), for example, used this technique to observe when and how visitors used brochures at a garden. Klein (1993) and Falk (1993) provide good examples of empirical studies utilising this technique and Loomis (1987) and Diamond (1999) explain the procedures involved in this approach. Visitor tracking places considerable burden on the observers because they must pay constant attention to visitors. In addition, tracking visitors is time-consuming and therefore expensive. However, it does provide considerable information about the relative attracting and holding power of all exhibits.

Recording the conversations and behaviour of visitors while they experience an interpretive program is one method of measuring certain types of behaviour and learning. Termed 'conversational elaboration' (e.g. Leinhardt & Crowley 1998; Leinhardt & Knutson 2004; Leinhardt, Crowley & Knutson 2002), the premise of this method is that the degree to which visitors are provoked to verbalise their thoughts while interacting with an interpretive device is both an outcome of interpretation and a causal factor in deeper learning. See Holbrook, Berent, Krosnick, Visser, and Boninger (2005) and Petty and Cacioppi (1986) for discussions of the theoretical rationale behind this research.

The outcome of elaboration and its impact of on learning and information processing has received significant research attention in various fields the past 25 years, including interpretation (see Ham & Weiler 2005; Moscardo 1999b). Often in these studies, a researcher quietly repeats a group's conversations and narrates what they are doing into a small micro cassette recorder without being noticed. Other times permission to record the conversation directly is given by the group. McManus (1987, 1988, 1989), Borun et al. (1998), Leinhardt, Crowley and Knutson (2002), Ash (2003) and Falk and Storksdieck (2005) are examples of studies employing this method. Borun et al. (1998) identified five behaviours that indicate family learning: asking a question, answering a question, commenting on the exhibit or explaining how to use the exhibit, reading text aloud and reading text silently. Thus, a researcher observes these five categories. Diamond (1999) argues that using a cassette recorder may affect visitors because they may become aware of a person talking into or using a recorder. Thus, it is suggested that the cassette recorder be small enough to be hidden in the hand or must have a small microphone connected to the hidden recorder. McManus (1989) and Borun et al. (1998) are good examples of this approach. Transcribing and analysing the data is time-consuming and therefore expensive. Also, observers and analysts need to be trained in order to attain acceptable reliability for both the observations and interpretation of them.

Observation of an audience during a program provides quick feedback and places no burden on the visitors. Observing a few categories of behaviour is not complicated and may require relatively low cost. Times for observation vary, ranging from short to long depending upon the duration of the behaviour of interest. For example, observing whether a visitor stops at an exhibit takes only a few seconds, while measuring how long each visitor spends reading an exhibit label may take less than a minute for each observation. However, tracking a visitor's movement or recording the conversations and behaviours require more time. The total time required for a comprehensive assessment depends on the per-observation time, but is also affected by the intensity of visitation. Depending on the amount of time required of observers, the cost of observations ranges from low to high.

Observations have some limitations. As Henderson and Bialeschki (1995) point out, since there are few standardised procedures for observations, an evaluator usually needs to develop and pilot test the instrument in the situation where it will be used. Behaviour that expresses attention, listening, reading and viewing must be defined clearly. The observer's role may affect the behaviour of the people being observed because when people know they are being watched they may behave differently. Observers must recognise their roles in their observation assessments, and consciously decide whether to be known or completely covert observers. Reliable and valid measurement needs to be established, especially for sophisticated observations. Observers must measure visitors' reactions in the same way across time or different programs. Inter-rater reliability, which is the consistency of the results across different observers, also needs to be established in order to achieve acceptable

reliability. Rigorous training of observers is required for complicated observation systems (Henderson & Bialeschki 1995) such as tracking visitors or recording their conversations. If multiple evaluators are hired for observations measuring attention to a personal program, the cost becomes higher. Finally, when observations are undertaken without informing those being observed, an ethical problem may arise (Grove & Fisk 1992). Table 3 presents a list of impact assessments that have relied on observations. The variables in the table are terms that were used by the authors.

**Table 3: Summary of impact studies that used observations**

Author(s)	Variable/s as described by authors	Setting
Ash (2003)	Duration and content of conversation	Museum
Beer (1987)	Attention capture	Museum
Bicknell & Mann (1993)	Viewing time	Museum
Bitgood, Patterson & Benefield (1988)	Stopping, viewing time	Museum
Bitgood & Patterson (1993)	Viewing and reading time	Museum
Bixler, Carlisle, Hammitt & Floyd (1994)	Fear	Field trip
Boisvert & Slez (1994)	Attraction, holding power, engagement	Museum
Borun et al. (1998)	Behaviour	Museum
Broad & Weiler (1998)	Time spent viewing	Zoo/theme park
Chiozzi & Andreotti (2001)	Viewing time	Museum
Cole (1998)	Attention	Trailside bulletin board
Diamond (1986)	Behaviour	Museum
Dick, Myklestad & Wagar (1975)	Attention	National park
Dufresne-Tasse & Lefebvre (1994)	Reaction	Museum
Falk (1993)	Behaviour	Museum
Falk & Storksdieck (2005)	Within and between group interaction	Science centre
Fernandez & Benlloch (2000)	Duration of conversation, reading behaviour	Museum
Goulding (2001)	Behaviour	Museum
Hall, Hockett & Smith-Jackson (2001)	Attention	National park
Harvey, Loomis, Bell & Marino (1998)	Attention, behaviour	Museum
Johnson (1989)	Trail use	Trail
Kanel & Tamir (1991)	Reading label Time spent	Science centre
Klein (1993)	Behaviour	Museum
Korn (1988)	Behaviour	Garden
Korn & Jones (2000)	Attention capture	Museum
Lackey, Ham & Hall (2002)	Attention and reading time	National park
Leinhardt & Crowley (1998)	Duration and content of conversation	Museums
Leinhardt & Knutson (2004)	Duration and content of conversation	
Leinhardt, Crowley & Knutson (2002)	Duration and content of conversation	
McManus (1987)	Duration and content of conversation	Museum
McManus (1989)	Duration and content of conversation	Museum
Phillpott (1996)	Attention paying	Reptile breeding centre
Sandifer (1997)	Duration of reading and other behaviours	Science museum
Schulhof (1990)	Behaviour	Botanical garden
Smith & Broad (in press)	Time spent viewing	Zoo exhibits
Stevens & Hall (1997)	Interaction	Museum
Tubb (2003)	Attention, holding time	Visitor centre
Tuckey (1992)	Activity patterns	Science centre
Wells (2000)	Attention, holding time	Visitor centre

### ***Visitor Employed Photography***

Photographs taken by visitors have been used for measuring their impression or attention to exhibits or perception of the natural environment. This method is called visitor employed photography (VEP) and asks visitors to take photographs of the features of exhibits they feel most interesting or important (Schulhof 1990; Camp, Koran & Koran 2000). This technique is also used to quantify common perceptions and recreational preferences in natural environments (Cherem & Driver 1983; Taylor, Czarnowski, Sexton & Flick 1995; Taylor, Sexton & Czarnowski 1995).

VEP provides a relatively unobtrusive measure of what visitors pay attention to (Taylor, Czarnowski, Sexton & Flick 1995; Taylor, Sexton & Czarnowski 1995; Camp, Koran & Koran 2000) as well as the social and physical contexts of the settings (Camp, Koran & Koran 2000). It requires relatively little time or skill for analysis (Cherem & Driver 1983) and using digital cameras, the photographs can be available immediately.

Cherem and Driver (1983) found that instructions given to the visitors by a researcher affected what they had photographed. This indicates the need for clear and consistent instructions. VEP cannot determine impacts on affect and cognition. Table 4 presents a list of impact assessments that have used VEP.

**Table 4: Summary of impact studies that used Visitor Employed Photography**

Author(s)	Variable/s as described by authors	Setting
Camp, Koran & Koran (2000)	Attention	Museum
Cherem & Driver (1983)	Perception	Nature trails
Markwell (1997)	Perceptions, preferences	Nature-based attractions
Schulhof (1990)	Impressions	Botanic garden
Taylor, Sexton & Czarnowski (1995)	Value	National park
Taylor et al. (1995)	Value	National park

### ***Observations of audience behaviour after an interpretive program***

Interpretive programs often try to influence or change visitors' behaviour on site. Direct observation of visitors after an interpretive program may determine how visitors behave in response to that interpretation. This may be an effective way to measure behaviour change (Jacobson 1999). For instance, if interpretation aims to increase the number of people who stay on a trail after hearing a program, evaluators can observe whether visitors attending to the interpretation subsequently stay on the trail. An auto camera can be used for these observations. For instance, Vander Stoep and Gramann (1987) used a time-lapse camera for their observations. Reid (2000) and Widner and Roggenbuck (2000) are additional examples of studies utilising observations of audiences' behaviour after interpretive programs.

Observations of audience behaviour after a program can provide accurate and explicit evidence of what the audience did at the site. They require no burden on the audience and provide prompt feedback. As long as evaluators are interested in only a few kinds of behaviours, this method requires little training and usually enjoys high inter-rater reliability (Henderson & Bialeschki 1995).

However, this type of observation has some limitations. By using only observations, it is hard to know whether an observed behaviour has resulted from the program or some other influence. For instance, people may stay on a trail not because an interpretive program convinced them to do so, but because they thought they might get dirty going off the trail. As Wolf (1980, p.40) stated, "it is often dangerous to draw inferences from behaviour without gaining insight into the person's own perception of what the behaviour means to him/her".

### ***Observations of audience behavioural traces***

Behavioural traces are indicators of past behaviour or accumulations of evidence providing clues to some past activity (Diamond 1999). Observations of behavioural traces include looking at physical evidence produced by an audience and linking it to an interpretive program. The observations are conducted after implementing an interpretive program and usually after visitors have left. Litter left on trail, fingerprints and nose prints on an exhibit glass are examples of behavioural traces (Roggenbuck & Propst 1981). For example, footwear on floors in front of an exhibit or nose prints on a display for children might mean it was popular.

Observations of behavioural traces usually do not require a lot of skill from observers, present no burden on audiences and provide prompt feedback. Furthermore, they are good for determining some behavioural impacts.

However, although they describe what actions took place, they cannot always tell whether the people who attended an interpretive program actually produced the traces. A trace must be carefully defined in order to make sure that it truly reflects the behaviours an observer wishes to measure (Weinreich 1999). Observing traces alone cannot determine the reason for the behaviour, nor can it explain an interpretive program's success. Since not all programs produce these kinds of observable traces, the usage of this technique is limited. As with any kind of observation, it is impossible to determine impacts on an audience's affect and cognition by relying exclusively on observation of their behavioural traces.

### ***Audience questionnaires***

A questionnaire is a data-gathering tool administered by mail, pick up, internet or personal distribution. Questionnaire surveys can collect data such as attributes of the respondents, facts relating to experiences, knowledge gain, attitudes, preferences and beliefs (Riddick & Russell 1999; Savage & James 2001). In impact assessments of interpretive programs, they can provide information on whether people enjoyed a program, what they felt or thought about the program, what behaviours they plan to engage in following a program, and/or what behaviours they report having done. Questionnaire surveys are the most widely used method for measuring learning in interpretive settings. Using different forms of questions, a questionnaire gathers quantitative or qualitative data. Closed questions, where respondents choose an answer from presented alternatives, provide quantitative data that are usually analysed statistically and may be generalised to wider visitor population if respondents were chosen randomly. Open-ended questions ask the respondents to answer in their own words and result in either qualitative or quantitative data.

A questionnaire can range from a short form with a few questions to a long survey with a few pages requiring respondents to write answers in sentences. Short questionnaires reduce the burden on the respondents while longer questionnaires provide richer and more complex information but more burden. Although it is generally agreed that "respondent fatigue sets in at approximately 15 to 20 minutes" (Lundegren & Farrell 1985, p.160), questionnaires in interpretive settings must usually be much shorter (requiring 10 minutes or less) because the interference in the visitor's experience that can result. As Diamond (1999, p.97) put it, "the longer it looks like it takes, the fewer people will choose to respond".

Developing a questionnaire requires a degree of both science and art (Stevens, Wrenn, Ruddick & Sherwood 1997) and there are many things that can compromise their reliability and validity. Interpreters designing a questionnaire are referred to Dillman (1978), Salant and Dillman (1994) and Stevens et al. (1997).

Questionnaire surveys can provide a large amount of data. The influence of researchers on respondents is less than with interviews because direct interaction between them often does not occur. Analysing quantitative data is usually not complicated. Little burden is placed on area personnel in implementing questionnaire surveys, except when they are personally distributed, but even this is less than required for observation or interviews.

However, disadvantages also exist for questionnaires. There is no chance to clarify a question unless the researcher is present, and the respondents may understand and respond differently than the researcher intended. Questionnaires often place a heavy burden on visitors. When administered on site, they have the potential to disrupt or negatively impact recreational experiences (Wagar 1976). Questionnaires also require the respondents to have reading and writing skills, which may cause inaccurate responses or eliminate people with low literacy levels such as children. Foreign visitors and visitors who are visually impaired may not be able to read them. Long, open-ended questions usually do not work well because people do not like to write long answers (Henderson & Bialeschki 1995). The form of the questions may also influence responses. Schuman & Scott (1987) conducted experiments to examine the difference in the way people answer open-ended and closed questions. They found that people tend to choose exclusively from alternatives offered in closed questions even though they are free to make other comments.

In order to ensure a question is properly asked and an appropriate answer can be obtained, a questionnaire must be piloted before being implemented. Time needed to fill out the questionnaire should also be tested to ensure a reasonable length. Measurements that are not pre-tested tend to reduce both the validity and reliability of an evaluation. Thus, careful wording and layout are needed to enhance the accuracy of responses. Developing a valid and reliable questionnaire requires skill and time. Analysing the data from open-ended questions also requires considerable time and skill.

Questionnaire surveys are often used for determining affective and cognitive responses to interpretation, as utilised, for example, by Cable, Knudson, Udd and Stewart (1987), Finson and Enochs (1987), Ryan and Dewar (1995), and Knapp and Barrie (1995). Questionnaires may be more effective at measuring sensitive or controversial issues compared to interviews because the presence of the researcher may influence responses in face-to-face techniques. A number of studies have used questionnaires to measure respondents' intentions with respect to various behaviours of interest (e.g. Cartmill & Day 1997; Knapp & Barrie 1998; Hendricks 2000; Beaumont 2001; Mayes, Dyer & Richins 2004) and a few have employed an experimental design, allowing greater confidence in establishing cause-and-effect relationships between interpretation and behavioural

intentions (e.g. Orams 1997; Howard 2000; Tubb 2003). Although some studies (e.g. Westphal & Halverson 1985-86; Taylor 1993) have attempted to measure actual behaviour change by using questionnaires, measuring behaviour by self-reporting alone may not produce accurate data because the respondents think they should reflect best on themselves (Weinreich 1999) or they may provide socially desired responses. Table 5 presents a list of interpretation impact assessments that have used questionnaires.

**Table 5: Summary of impact studies that used questionnaires**

Author(s)	Main variable(s) as described by authors	Setting
Allard, Boucher & Forest (1994)	Attitude Cognition	Museum
Armstrong & Weiler (2003)	Knowledge Satisfaction	Guided tours
Beaumont (2001)	Attitude Behavioural intention	Guided tours
Brown & Koran (1998)	Knowledge gain	Archaeological park
Cable, Knudson, Udd & Stewart (1987)	Attitude	Visitor centre
Cartmill & Day (1997)	Behavioural Intention	Museum
Crump, Bryne & Croucamp (2000)	Interest	Zoo
De White & Jacobson (1993)	Attitude Knowledge	Museum
Finson & Enochs (1987)	Attitude	Museum
Fisher (1997)	Retention	Planetarium
Geva & Goldman (1991)	Satisfaction	Guided tour
Hall, Hockett & Smith-Jackson (2001)	Knowledge	National park
Ham & Weiler (2002)	Satisfaction	Guided tours
Ham & Weiler (2005)	Beliefs and attitudes about a behaviour	National park
Harvey et al. (1998)	Perception	Museum
Hendricks (2000)	Attitude Behavioural intention	Personal program
Henriksen (2001)	Attitude Concepts Understanding	Museum
Hornfeldt, McAvoy & Schleien (1989)	Learning	Nature centre program
Horsley (1988)	Attitude Intention	Field
Howard (2000)	Knowledge Attitude Behavioural intention	Conservation park
Howard, Lipscombe & Porter (2001)	Knowledge Awareness	National Park
Jacobson (1987)	Attitude	Rural education program
Hughes & Morrison-Saunders (2002)	Knowledge	National park
Javlekar (1989)	Learning	Science centre
Knapp (2000)	Memories	Field trip
Knapp & Barrie (1998)	Attitude Knowledge Behavioural intention	Field program
Korn (1988)	Learning	Garden
Lee & Balchin (1995)	Attitude Learning	Visitor centre
Madin & Fenton (2004).	Knowledge, understanding	Marine park
Mayes, Dyer & Richins (2004)	Beliefs, attitudes, behavioural intentions Motivation, satisfaction	Dolphin attraction
Morgan, Absher, Loudon & Sutherland (1997)	Attitude Knowledge	National forest
Moscardo (1999a)	Brochure distribution system Site knowledge Image of site Messages extracted	Marine park tour boats



Nielsen & Buchanan (1986)	Attitude	Field program
Prentice (1991)	Recall	Field program
Orams (1997)	Knowledge Attitude Behavioural intention	Dolphin attraction
Reid (2000)	Knowledge	Field program
Rennie (1986)	Knowledge	Self guided trail
Roggenbuck, Williams & Bobinski (1992)	Knowledge Enjoyment Behavioural intention	Tour guide
Ryan & Dewar (1995)	Learning	Field program
Stevenson (1991)	Feeling Thought	Museum
Taylor (1993)	Behaviour change	Botanic garden
Tubb (2003)	Knowledge gain Attitude Behavioural intention	Visitor centre
Tuckey (1992)	Recall	Science centre
Westphal & Halverson (1985-86)	Attitude Behaviour modification	Workshop
Woods & Moscardo (2003)	Mindfulness	Tourism community
Woods & Moscardo (1998)	Understanding of pictorial symbols Knowledge of appropriate behaviour	Marine park
Young & Witter (1994)	Knowledge	Home
Zeidler & Surber (1999)	Understanding	Museum

### ***Formal audience interviews***

Formal interviews are similar to questionnaire surveys but are presented orally and responses are recorded by interviewers. Formal interviews, also called structured or semi-structured interviews, can be conducted face-to-face or by telephone. Like questionnaires, they can collect data on respondents' attributes, reported behavioural facts, attitudes, preferences and beliefs (Riddick & Russell 1999). Fowler and Mangione (1990) outline important issues associated with interview studies and their usage in evaluations. Salant and Dillman (1994) provide practical information about how to design and conduct interview surveys.

Using an interview schedule, interviewers read the questions to the respondents and record their answers. The interview schedule must be a clear guide for the interviewer, listing questions in the appropriate order and providing instructions and means to record respondents' answers (Graziano & Raulin 2000).

Formal interviews enable interviewers to ask more complicated questions than questionnaires can ask, clarify questions, seek more detailed answers, and get unexpected responses. Formal interviews can also involve people with lower literacy (Henderson & Bialeschki 1995) or visual impairments (Silverman & Masberg 2001). Interviewers can encourage respondents to answer, thereby increasing response rates compared to questionnaires. In addition, they can establish rapport with the respondents that leads to more cooperation (Stevens et al. 1997) or may be able to ask sensitive questions (Weinreich 1999). According to Slater (1998), a main advantage of interviews over questionnaires is that they allow a skilled interviewer to establish an atmosphere of confidentiality and sincerity.

Formal interviews have some disadvantages. They can be time-consuming, costly because of personnel involved, and place considerable burden on both the respondents and interviewers. In addition, respondents may try to please the interviewer and give answers they think the interviewer wants to hear (Diamond 1999; Jacobson 1999). There is also an influence of social desirability on validity wherein people 'over-report' because they think they should (Henderson & Bialeschki 1995). As Slater (1998) argued, one-on-one interviews require an extremely skilled interviewer and the interviewer needs to engage the respondent in a full and spontaneous expression of thoughts and feelings. The quality of information obtained from the interview and analysis of the data depends largely on the interviewer's skills and expertise (Slater 1998). Since interviewers tend to ask the same question differently across time (Roggenbuck & Propst 1981), training interviewers is needed in order to attain acceptable reliability. A list of interpretation impact assessments that have used formal interviews is shown in Table 6, including some involving post-visit interviews, usually telephone (Orams 1997; Broad & Smith 2004).

**Table 6: Summary of impact assessment studies that used formal interviews**

Author(s)	Main variable(s) as described by authors	Setting
Adelman, Falk & James (2000)	Attitude Concern Knowledge Understanding	Aquarium
Allen (1997)	Understanding	Museum
Anderson, Lucas & Ginns (2000)	Construct knowledge	Museum
Bass, Manfredo, Lee & Allen (1989)	Awareness Interest Participation	Boat trip
Beeho & Prentice (1995)	Experience Benefits	Museum
Blud (1990)	Understanding	Museum
Broad & Smith (2004)	Knowledge, understanding Behavioural intention	Zoo
Broad & Weiler (1998)	Learning	Zoo/theme park
Cartmill & Day (1997)	Behavioural Intention	Museum
Falk & Dierking (1997)	Affect Learning (memories)	Field trip
Falk (1997)	Understanding	Museum
Hall, Hockett & Smith-Jackson (2001)	Reaction to interpretive panels	National park
Hernandez & De Celis Herrero (1993)	Knowledge	Museum
Higham & Carr (2002)	Appreciation/Insight Environmental values Attitude, awareness Behavioural intention	Ecotourism sites Tourism operators
Jakobsen (1999)	Feeling Emotion	Visitor centre
Kanel & Tamir (1991)	Opinion Learning	Museum
Knapp & Poff (2001)	Attitude Knowledge Behavioural intention	Field trip
Light (1995)	Attention Awareness Interest	Heritage site
Mallick & Driessen (2003)	Attitude	National park
Ogden, Lindburg & Maple (1993)	Awareness/Appreciation Interest/Positive emotion Satisfaction/Learning	Zoo
Orams (1997)	Behaviour	Dolphin
Pearce & Moscardo (1998)	Satisfaction	Interpretive gondola
Prentice, Sinead & Stuart (1998)	Affect Knowledge Learning	Attraction
Schulhof (1990)	Appreciation Understanding	Botanical garden
Stevens & Hall (1997)	Understanding	Museum
Stevenson (1991)	Feeling Thought	Museum
Stewart, Hayward & Devlin (1998)	Appreciation	National park
Tuckey (1992)	Recall	Science centre

***Informal audience interviews***

In comparison to formal interviews, informal interviews involve casual conversation with respondents and typically are used to explore in-depth or new information. The findings are usually presented in a narrative or unstructured form that results in qualitative data that cannot be used for generalising to other situations (e.g.

Gurnhill & Vertigan 2002). Since informal interviews do not rely on formally structured questioning; visitors may not even notice they are being interviewed. They occur during the course of normal, natural conversation. The researcher may adopt the dress and play the role of a visitor and start a conversation with someone in the target audience. As the conversation flows, the researcher asks questions. According to Diamond (1999), both the interviewer and respondent play a role in directing questions and when an important or interesting idea appears, it may be pursued.

Since the respondents do not feel as though they are being formally questioned, informal interviews may be better for getting information on sensitive issues (Roggenbuck & Propst 1981). Diamond (1999) claims that informal interviews are the least threatening way of conducting interviews and can be a valuable source of ideas and general impressions. According to her, they are valuable tools for probing respondents' feelings because the conversational situation enables the respondents to gain clarity of their own beliefs. Informal interviews place less burden on respondents than do formal interviews or questionnaires.

Some disadvantages of informal interviews are that they are expensive and time-consuming (Roggenbuck & Propst 1981). Also, they require the interviewer to have a high level of skill, and they raise ethical issues about deception (making the visitor think that a private conversation is taking place). Since informal interviews occur in a casual conversation, it is difficult for interviewers to ask the same question in the same way across all situations, making them less reliable than formal interviews. It is also difficult to ask all the planned questions in actual conversation. Interviewer's influence on the conversation is almost certain to occur. Diamond (1999) points out that careful monitoring is needed because some respondents are highly suggestible and may incorporate their own ideas into what the interviewers seem to want to hear, or they may change their ideas throughout a conversation. Thus, interviewers must be highly trained and skilled to conduct informal interviews.

### ***Focus group interviews with audience***

Focus groups can be used as an evaluative technique that collects qualitative data from a group discussion focused on a specific topic. They are often used in marketing, advertising, and political research to gain understanding about how people feel about an issue, product, or candidate. Ideally, a focus group is composed of a skilled interviewer and six to eight people who have something in common with respect to a topic of interest (Krueger & Casey 2000). Participants are interviewed in detail about a single topic and usually receive a small gift or payment for their participation. When used for impact assessments of interpretive programs, focus group interviews can provide visitor feedback on whether a program achieves its objectives, how the program can be improved, what works well and what does not, and how visitors value an interpretive program.

Usually, focus group participants in leisure settings are volunteers, since they are recruited at the site and limited to those who are willing to attend (Braverman 1988; Medlin & Ham 1992; McDonald 2000). Once participants are gathered, a moderator asks open-ended questions in sequence from general to specific, usually using an interview guide. The guide outlines a series of questions in order but is less structured than a formal interview so that the moderator can react and probe the respondents' answers. Interviews are usually conducted with several groups (often three or four) until information obtained from interviews reaches saturation, meaning that no more new information emerges from the interviews (Krueger & Casey 2000). Having multiple focus groups allows data to be analysed and compared across groups to look for patterns and themes. Often focus group interviews are videotaped or tape-recorded for later review and analysis. Detailed strategies and procedures for focus group interviews are described in Morgan, Krueger and King's (1998) *Focus group kit* series and in Stewart and Shamdasani (1990). A recent guide to conducting focus groups in evaluations of interpretation is Savage and James (2001).

The purpose of focus groups is to gather a range of information, not necessarily to reach consensus. Thus, each member in a group must be encouraged to provide his/her opinion equally, avoiding letting a particular member dominate the conversation. Riddick & Russell (1999) argue that a small group of less than seven persons may be dominated by a certain member, while a larger group is more difficult to manage and inhibits participation by all members. In order to get candid responses from every participant, a moderator is usually responsible for creating a comfortable and permissive environment in a focus group. According to Krueger and Casey (2000), a focus group offers a more natural environment than an individual interview since participants influence each other as they do in real settings.

One of the main advantages of focus group interviews is that they stimulate interaction among members in a group. This encourages further responses and often leads to the expression of new ideas that may not come to a respondent's mind in an individual interview (Medlin & Ham 1992; Solomon 1992; Henderson & Bialeschki 1995; Jacobson 1999). Solomon (1992, p.84) points out, "the security of being in a group may also encourage more candour", especially when group members share some important characteristic. This sense of group cohesiveness may facilitate discussion of sensitive topics (Stewart & Shamdasani 1990). In addition, focus group interviews can include people of low literacy, and they are well-suited for children (Medlin & Ham 1992). A

skilled moderator is able to probe visitor responses so that diverse, complex, and unplanned information emerges. Unlike other forms of interviews, focus groups provide quick feedback (Jacobson 1999).

Focus group interviews also present some disadvantages. A moderator must be well-trained since they influence the types and quality of information obtained from the group (Riddick & Russell 1999), and must be able to act consistently over time and with different groups. Transcribing and analysing the data is time-consuming. Because the data consists of textual units, rigorous analysis is difficult. The findings of a focus group generally cannot be generalised. Because the participants are not a representative sample, the group members should be chosen to reflect the views that are most relevant in terms of the goals of the evaluation (Uzzell 1998b). Recruiting focus group participants requires planning and screening (Slater 1998). Solomon (1992) cautioned that focus group interviews usually do not occur in natural settings, which results in uncertainty about the accuracy of responses. He also cautioned about the influence that group judgments can have on an individual's decision-making. Group dynamics may influence an individual's responses because some participants simply agree with others rather than expressing their own thoughts (Slater 1998). The cost of focus group interviews has been variously rated as, for example, low (Wright & Wells 1990), relatively inexpensive (Jacobson 1999), cost-effective (Uzzell 1998b) and fairly expensive (Knudson, Cable & Beck 1995); however, they generally cost less than formal and informal interviews because fewer interviews and less time are needed.

### ***Personal meaning mapping***

Personal meaning mapping (PMM) is designed to quantifiably measure how an educational experience uniquely affects each individual's personal conceptual, attitudinal and emotional understanding (Falk, Moussouri & Coulson 1998; Adelman, Falk & James 2000). It measures the multi-dimensional aspects of learning in terms of extent, breadth, depth and mastery (Falk, Moussouri & Coulson 1998; Falk & Storksdieck 2005). PMM generally does not measure the attainment of learning objectives or compare people's knowledge gain; rather it determines *how* each person learned.

According to Adelman, Falk and James (2000) and Falk, Moussouri and Coulson (1998), PMM uses the following procedure. Visitors are asked to write words, ideas, images, phrases or thoughts that come to mind in relation to an interpretive program's topic. An interviewer clarifies those written words if they are unclear. PMM uses the same visitors before and after an interpretive program. After the program, visitors are asked to if they would add to, subtract from or modify any thoughts that they previously wrote. Pre-program PMM establishes baseline information about a visitor's understanding of a topic and is compared to the post-program PMM. Comparison of pre- and post-program PMM determines the extent of change in visitors' incoming and exiting knowledge, interests and attitudes.

PMM focuses on four dimensions of learning (Falk, Moussouri & Coulson 1998). First, it measures the extent of a visitor's awareness and understanding of a particular concept or experience by focusing on the visitor's vocabulary. This requires counting the number of relevant words written down on the map before a program and comparing them with those written after the program. Second, it measures the breadth of a visitor's understanding by focusing on the change in the quantity of appropriate concepts utilised. Third, it measures the depth of a visitor's understanding by determining how deeply and richly they understand a particular concept. Fourth, it measures mastery, the facility with which visitors use their understandings. Recent studies employing PMM methodology have been able to isolate several factors that underlie visitors learning (e.g. Falk 2003; Falk & Storksdieck 2005).

Adelman, Falk and James (2000) and Falk, Moussouri and Coulson (1998) showed high inter-rater reliability in their PMM studies. Falk, Moussouri and Coulson (1998) argue strongly that PMM is a versatile and reliable tool for assessing learning. However, like other individual interviews, the disadvantages of PMM include time and cost. It also places a heavy burden on the audience, and analysing the data may be difficult.

In summary, a range of techniques can be used for evaluating interpretation. Each technique has a specific purpose, advantages and disadvantages. Evaluations vary widely in terms of cost, time, expertise and burden on audiences and evaluators. As Ham (1986) pointed out, generally the costs and precision of the results are related, and the costs relate largely to the amount of time required to collect and analyse the data. He concludes that generally the precise measurement of program effects is expensive, but that less expensive and moderately precise measures could be used to provide a reasonable indication of program effectiveness (Ham 1986). Relying on only one evaluation technique usually will not accurately assess all aspects of an interpretive program's impacts and outcomes (Roggenbuck & Propst 1981). For example, no one technique by itself can properly measure all effects on an audience's cognition affect, and behaviour. Thus, the best evaluations use multiple techniques (Roggenbuck & Propst 1981; Jacobson 1999).

### **Data Collection Techniques**

This section assesses the usefulness and appropriateness of various data collection techniques in the context of interpretation evaluation. To compare the advantages, disadvantages and usefulness of each for specific

evaluation problems, and since our focus is on impacting visitors' thinking (cognition), feeling (affect) and action (behaviour), a separate matrix for each of these dimensions is provided. Table 7 lists data collection techniques appropriate for cognitive responses; Table 8 focuses on affective responses; and Table 9 focuses on behavioural responses.

**Table 7: Summary of data collection techniques for assessing the impacts of interpretation on audience cognition**

Techniques Criteria	Personal Meaning Mapping	Focus Group Interview	Informal Interview	Formal Interview	Questionnaire	Observation During Program (Recording)	Self-testing Devices
Cost	High	Moderate to high	High	High	Moderate to high	High	Moderate to high
Time	Long	Moderate to long	Long	Moderate to long	Moderate	Moderate	Short to moderate
Speed of Feedback	Moderate	Fast to moderate	Moderate	Moderate	Slow to moderate	Slow to moderate	Moderate to fast
Burden on Visitor	High	High	Moderate	High	High	None	Low
Burden on Personnel	High	High	High	High	Moderate	High	Low
Validity	High	Moderate	Low to moderate	High	High	Low to moderate	Moderate to high
Reliability	Moderate to high	Moderate to high	Low	Moderate to high	High	Low to moderate	High

Note: Each of the above criteria will be affected by the number of visitors contacted or observed.

**Table 8: Summary of data collection techniques for assessing the impacts of interpretation on visitor *affect***

Techniques Criteria	Observation During Program	Questionnaire	Formal Interview	Informal Interview	Focus Group Interview	Personal Meaning Mapping
Cost	Low to high	Moderate to high	High	High	Moderate to high	High
Time	Short	Moderate	Moderate to long	Long	Moderate to high	Long
Speed of Feedback	Fast	Slow to moderate	Moderate	Moderate	Fast to moderate	Moderate
Burden on Visitor	None	High	High	Moderate	High	High
Burden on Personnel	Moderate	Moderate	High	High	High	High
Validity	Low to moderate	Moderate to high	Moderate to high	Moderate	Moderate	Moderate to high
Reliability	Low to moderate	Moderate to high	Moderate to high	Low	Moderate to high	Moderate to high

Note: Each of the above criteria will be affected by the number of visitors contacted or observed.

**Table 9: Summary of data collection techniques for assessing the impacts of interpretation on visitor behaviour**

Techniques Criteria	Observation During Program	Observation During Program (Tracking/ Recording)	Observation After Program	Observation of Behavioural Traces	Visitor Employed Photography	Questionnaire	Formal Interview	Informal Interview	Focus Group Interview
<b>Cost</b>	Low	High	High	Low to moderate	High	Moderate to high	High	High	Moderate to high
<b>Time</b>	Short to moderate	Moderate	Short to moderate	Short to moderate	Short to moderate	Moderate	Moderate	Long	Moderate to long
<b>Speed of Feedback</b>	Fast	Fast to moderate	Fast	Fast	Slow to moderate	Slow to moderate	Slow to moderate	Moderate	Fast to moderate
<b>Burden on Visitor</b>	None	None	None	None	Low to moderate	High	High	Moderate	High
<b>Burden on Personnel</b>	Moderate to high	High	High	Moderate	Moderate	Moderate	High	High	High
<b>Validity</b>	Moderate to high	Moderate to high	Moderate to high	Moderate	Low to moderate	Moderate	Low to moderate	Low to moderate	Moderate
<b>Reliability</b>	Moderate to high	Moderate	Moderate to high	Moderate	Low to moderate	High	Moderate to high	Low	Moderate to high

Note: Each of the above criteria will be affected by the number of visitors contacted or observed.

In these matrices, each data collection method is evaluated based on several criteria including cost, time, speed of feedback, burden on visitors, burden on personnel, validity and reliability. Some of these criteria have been discussed in past studies (Roggenbuck & Propst 1981; Ham 1986). The following meanings for each criterion were applied in constructing the matrices used here.

- *Cost*: the relative expense of employing each data collection method, from design through writing a report, expressed as low, moderate or high.
- *Time*: how long it takes to implement the technique, expressed as short, moderate or long. The ratings shown in each matrix do not reflect the time necessary for data analysis or reporting.
- *Speed of feedback*: how fast results can be acquired after data-collection, expressed as slow, moderate or fast.
- *Burden on visitors*: the degree of effort required of audiences to provide data, expressed as none, low, moderate or high.
- *Burden on personnel*: the degree of work required of the evaluator(s), expressed as low, moderate or high.
- *Validity*: the ability of the technique to measure what it is supposed to measure, expressed as low, moderate or high.
- *Reliability*: the ability of the technique to produce consistent results across time or researchers, expressed as low, moderate or high.

Our assessment of evaluation methods was based in part on the application of these criteria by other authors mentioned in the previous section. When several authors rated the same method differently, the most common rating was chosen. For example, with regard to focus group interviews, some authors rated the cost as low (Wright & Wells 1990), relatively inexpensive (Jacobson 1999), cost-effective (Uzzell 1998b) and fairly expensive (Knudson, Cable & Beck 1995). Thus, in this review, moderate to high was chosen. All of the decisions about the criteria presented in the matrices are based on the assumption that an evaluation is done well and consistent with rules and procedures of sound research.

### ***Choosing a data collection method***

In conducting an evaluation, choosing an appropriate data collection method is critical to producing useful information. Machlis & McKendry (1989) suggest the use of the following criteria for field interpreters:

1. Practical;
2. Inexpensive;
3. Reliable;
4. Understandable and useful to field interpreters;
5. Flexible, and appropriate for various situations;
6. Useful to park operations, concessions, managers;
7. Timely;
8. Ethical and legal; and
9. Scientifically sound.

All the techniques described in the previous section are practical, understandable and useful for interpreters. However, considering the wide range of these criteria, only a few of the techniques discussed were found to be appropriate in all situations, usually because cost or soundness of results restricted their usefulness. As Ham (1986, p.24) mentions, “[g]enerally, the more precise and resistant to bias the data, the more costly the technique needed to collect them”. Thus, the evaluator’s priority among these criteria, especially with respect to expense and precision of results, determines which technique(s) should be used. Comparisons of known strengths and weaknesses, and compromises between them, will lead evaluators to make good choices of data collection techniques.

## **What to Evaluate?**

The foregoing review of the purposes and methods of interpretation evaluation serves as a useful guide for practitioners who must be able to weigh the relative trade-offs involved in selecting one method over another. It does not, however, provide more than general guidance about the *choice of the specific indicators* that might be used in any given evaluation. The broad classification of outcomes (cognitive, affective and behavioural) comprises a potentially enormous range of evaluative criteria, any one or combination of which might be used in a given evaluation. The question that remains unanswered is ‘which ones’ should be included in an Evaluation Tool Kit of the type we set out to develop in this project? That is, what are the most important indicators of effective interpretation? How can they be identified, selected and operationalised for the purposes of evaluating interpretation? As detailed in the following section, addressing these questions was the central purpose of the research conducted in this study.



## **Summary**

While the body of research evaluating the outcomes of interpretation is growing and maturing, until recently much of what has been published has lacked theoretical grounding, perhaps in part because interpretation evaluation is often funded and driven by industry where a need for quick answers understandably prevails over the need to explain any larger phenomenon. From an industry perspective, the research has often fallen short of linking interpretation outcomes to organisational goals, and therefore has had limited value in objectively gauging interpretation's success. While each study brings its own rationale for the evaluative criteria it employs, often a clear theoretical basis is lacking or unapparent. A conspicuous emphasis in past research has been its focus on 'how to' evaluate (i.e. making good decisions about the selection and design of field methods and data collection techniques), with less attention to 'what' to evaluate (i.e. what are the most important outcomes of an interpretive program, and how do we identify and refine indicators that effectively measure these outcomes?).

The current project recognised this as a gap that academic research could fill: to develop a suite of valid and reliable indicators that would be relevant to organisational goals. As such, the study focused on developing indicators and instruments that could assess interpretation's outcomes, using four important criteria:

1. Taken together, the suite of indicators needed to capture all three response categories (cognitive, affective and behavioural);
2. The indicators needed to be relevant to the real-world needs of interpretive practitioners;
3. The indicators needed to lend themselves to data collection techniques that practitioners (who are assumed to be non-social scientists) could employ in their workplace; and
4. The methods used to collect the data and measure the indicators needed to meet acceptable levels of validity and reliability, thus allowing practitioners to be confident that the data they collect are acceptably accurate and trustworthy.

## Chapter 3

# METHODS

As foreshadowed in the previous chapter, the significance of this project lies with its methodological contribution, specifically with respect to the processes and rationale for selecting, operationalising and validating indicators of interpretation outcomes. This chapter describes in detail the process and methods used to arrive at the particular indicators, select suitable data collection techniques and develop and validate the instruments.

## Identification of the Indicators

The process of selecting indicators sought to ensure they reflected the types of outcomes heritage and protected area managers and tourism operators actually want from their interpretation (see first section of Chapter 2: General Goals of Interpretation – for example, enhanced visitor experiences, positive visitor attitudes about conservation, positive word-of-mouth advertising about you as an agency or operator, provoking visitors to think about the values inherent at the site etc.). In addition, a number of steps were taken to ensure the indicators were theoretically valid based on what is known about interpretation's potential impacts on how visitors think, feel and possibly behave with respect to the things being interpreted for them. The preceding review of literature was useful in making this determination and the reader will find the final selection and development of indicators to be theoretically consistent with established evaluation precepts and concepts.

## Development of an Indicator Pool

A structured facilitation process (a modified nominal group process) was undertaken at two sites where a wide variety of face-to-face interpretive programs are offered (Port Arthur in Tasmania and Sovereign Hill in Victoria) in order to learn from staff what they felt were the most important indicators of 'successful' or 'effective' interpretation at those sites. To ensure the indicators selected would be desired outcomes of relevance to a wide cross-section of industry, participants in these discussions included program managers, front-line interpreters and guides, sales and marketing staff and mid- or executive-level administrators. Participants were asked to complete the following sentence in as many ways as they wanted:

'At the end of the day, I would know if the Sovereign Hill/Port Arthur interpretive program is doing a good job if ...'

Participants' completions of the sentence expressed the kinds of outcomes they felt were most important in determining the success or achievement of the respective interpretive program (e.g. 'visitors left with a positive attitude toward heritage preservation' or 'visitors enjoyed themselves' or 'visitors were inspired to learn more' etc.).

Following the nominal group technique, participants first individually reported their completions of the sentence and indicators of interest to them were extracted through discussion and clarification. Once all their ideas were arrayed on butcher's paper, the group was asked to prioritise the list by assigning points based on the importance they placed on each type of outcome. To do this, the facilitators asked them to distribute 100 points among the full list of outcomes, with larger numbers of points given to the ones they felt were more important. They were not required to assign points to all indicators, but rather only to those they felt were important to consider. Summing the points assigned to each outcome resulted in a prioritised list ranging from the most important (i.e. those receiving the most points) to the least important (i.e. those receiving fewer points or no points at all). Once the prioritised list was arrayed, the participants were allowed one opportunity to modify their point allocations. This final round of the process produced a pool of 54 responses, including 24 prioritised outcomes at Sovereign Hill and 30 prioritised outcomes at Port Arthur.

## Development and Refinement of Indicators

Since there were many overlaps and duplications between the two prioritised lists, the next steps involved a consolidation process as well as subjecting the 54 indicators in the pool to a range of other criteria, as outlined here.

First, responses were assigned to categories informed by human behaviour research (see final section of Chapter 2) as well as other categories that emerged from the data. Three researchers independently assigned each response to one of these categories and the responses were then analysed for consistency. Any item for which there was not 100% agreement among the three researchers was eliminated from the pool. In addition, three

categories of outcomes identified through this procedure were eliminated since they did not describe outcomes related to *visitor responses* to interpretation. These included a *staff-focused* category for outcomes that primarily benefited staff (e.g. it will be good for staff morale or make the guide's job easier), an *other-focused* category for responses that primarily benefited other stakeholder groups and an *input* category for responses that were inputs or features of interpretation design, infrastructure or delivery rather than outcomes. Taking into account the aims of this project, staff-focused, other-focused and input responses were then eliminated from the pool.

This resulted in a list of 37 potential indicators that were *visitor responses* to interpretation (i.e. ways in which visitors were impacted by the interpretation). To assess the level of *intra-site consensus*, the researchers then considered the relative importance (ranking) of each indicator by the original participants by re-examining the points assigned by individuals to each of the indicators in Step 1. A number of items from each site that received no points in the nominal group sessions were eliminated from each site's list, but 'archived' for re-analysing later, as outlined below.

A more complex task was to make *inter-site comparisons* and specifically to examine whether a single indicator could capture the essence of multiple indicators between the two sites. In several cases, responses from the two sites were found to be focused on the same phenomenon, and so were re-written so that a single indicator would capture the intent of the responses at both sites.

This consolidation procedure produced a reduced pool of 26 indicators that potentially provided a measure of visitor responses to interpretation. Collectively, they focused on the three main categories of outcomes discussed in the final section of Chapter 2:

1. *Cognitive outcomes*: what visitors might *think*, know or believe as a result of interpretation (e.g. understanding something, having a new view or being provoked to thought)
2. *Affective outcomes*: what visitors might *feel* as a result of interpretation (e.g. appreciation of something, satisfaction with something, an attitude about something)
3. *Behavioural outcomes*: what visitors might *do* or be *motivated to do* as a result of interpretation (e.g. stay longer at the site, buy something, positive word-of-mouth advertising)

A series of criteria was now applied to attempt to refine this list:

- Each indicator was assessed in terms of whether it had applicability beyond one particular site to other heritage sites and, as discussed in more detail below, to nature-based and food and beverage attractions;
- Each indicator was considered in the context of contemporary communication theory and research (i.e. could interpretation realistically be expected to achieve this outcome); and
- Each indicator was analysed with respect to whether it could be measured simply yet reliably by a non-social scientist.

Each of the 26 visitor response indicators that received points during the nominal group session at one or both sites was assigned to a category, with each category label describing the nature of the outcome suggested by the indicators it contained. If a category contained only one indicator, it was given the name of the indicator itself, and this occurred in three instances. However, in other cases, new category labels were created, as it was evident that a number of indicators were potentially measuring different dimensions of a single phenomenon or outcome. Ten categories resulted from this procedure.

Using these categories, two researchers not connected to this study independently assigned the same list of 26 indicators to the ten categories. Disagreement on the categorisation of five indicators led to three categories being more clearly defined and two new categories being created, making a total of twelve. A final independent categorisation by three 'judges' produced no disagreement, i.e. there was 100% reliability in the categorisation of each of the 26 indicators using the twelve categories.

Finally, the list of the ten 'archived' indicators (the ten indicators that had been listed by one or both sites but had received no points in the final rating process) was re-examined. The three judges independently assigned each of the archived indicators to one of the twelve categories based on how well it fit the category. If an archived indicator did not fit any category, it was put into a separate 'no-fit' category. Seven of the archived indicators were reliably assigned to one of the twelve categories (with 100% agreement). The three remaining indicators were left unassigned.

Together, these procedures resulted in two prioritised lists of indicators, 24 for Sovereign Hill and 30 for Port Arthur. Based on the points they had received during the respective nominal group processes, the indicators for each site were ranked from high to low (1-16 at Sovereign Hill and 1-20 at Port Arthur). Because some indicators were tied (i.e. received the same number of points in the nominal group process), the number of ranks is lower than the actual number of indicators.

## Selection of the Final Suite of Indicators

The twelve categories were then revisited with respect to the following criteria:

- Each indicator was reviewed to ensure it described a visitor's potential response to interpretation, broadly categorised as cognitive, affective or behavioural;

- The final suite of indicators was examined to ensure it included all indicators considered important by staff at *both* study sites (i.e. the ‘shared’ indicators), since this indicated a degree of consensus in their relevance as performance measures; and
- The final suite of indicators was reviewed to ensure each site’s top three choices (as indicated by their importance ratings during the nominal group process) were included, even if they were not shared by both sites. Since each site felt strongly about the relevance of its top-rated indicators as performance indicators, they should be among the indicators ultimately selected.

The results of this analysis are shown in Tables 10, 11 and 12. The ‘shared indicators’ (Table 10) were the indicators shared by the two sites. The ‘unshared indicators’ (Table 11) were the indicators listed by one site but not the other. Based on the three criteria described above, all nine of the shared indicators were included in the final suite. In addition, two of the unshared indicators shown in Table 11, SH #4 (Visitors found it relevant and meaningful) and PA #21 (Visitors interacted with the guide), were included because each is ranked in the top three by one of the sites.

**Table 10: Categorisation and Ranking of Visitor Response Indicators Shared by the Two Sites**

Indicator Name	Sovereign Hill Ranking (n=16) <sup>a</sup> [reference #s] <sup>b</sup>	Port Arthur Ranking (n=20) <sup>a</sup> [reference #s] <sup>b</sup>	Type of Response <sup>c</sup>
Impact on current world view via empathy with historic period and people	1, np [# 16, 19]	5 [# 18]	C/A
Elaboration (provoked to thought)	2, 3, 3 [# 15, 2, 3]	1, 4, np, np [# 17, 20, 15, 22]	C
Positive attitude toward heritage / heritage preservation	2, 3 [# 23, 11]	3 [# 3]	A
Positive global evaluation of interpretation at site	3, np [# 24, 18]	9, np [# 23, 24]	A
Desire to know more about / participate in additional interpretive activities	3 [# 10]	7 [# 16]	A/B
Desire to / do purchase souvenir related directly to site story	5 [# 6]	11 [# 6]	A/B
Desire to / do stay longer	3, np [# 13, 17]	12, 12 [# 26, 27]	A/B
Desire to / do return for repeat visit	np [#1]	6, 8 [# 5, 4]	A/B
Positive word-of-mouth advertising	4 [# 20]	2 [# 7]	A/B

**Notes:**

<sup>a</sup> Rankings based on overall points assigned to each indicator by nominal group process participants. Only visitor response indicators are included. Identical rankings resulted from ties. Due to ties, the indicator numbers exceed the rank values.

<sup>b</sup> Denotes indicator number assigned during the nominal group process

<sup>c</sup> C = cognitive; A = affective; B = behavioural.

<sup>d</sup> np = no points received during final rating in nominal group process

**Table 11: Ranking of Unshared Visitor Response Indicators**

Indicator Name	Sovereign Hill Ranking (n=16) <sup>a</sup> [reference #s] <sup>b</sup>	Port Arthur Ranking (n=20) <sup>a</sup> [reference #s] <sup>b</sup>	Type of Response <sup>c</sup>
J. Visitors found it relevant and meaningful to their lives	2 [# 4]		C
K. Visitors provoked to interact with the guide (interactive experience)		3 [# 21]	B
Visitors feel there was value for money		10 [# 8]	A
Visitors know what historic people did, when, why etc.		np [# 1]	C
Visitors appear to be interested		np [# 14]	B
Visitors take more photos		np [# 30]	B

**Notes:**

<sup>a</sup> Rankings based on overall points assigned to each indicator by nominal group process participants. Only visitor response indicators are included. Identical rankings resulted from ties. Due to ties, the indicator numbers exceed the rank values.

<sup>b</sup> Denotes indicator number assigned during the nominal group process

<sup>c</sup> C = cognitive; A = affective; B = behavioural.

<sup>d</sup> np = no points received during final rating in nominal group process

**Table 12: Final Suite of Visitor Response Indicators**

Indicator	Indicator Long Title	Category
A	Impact on current world view via empathy with historic period and people / impact on appreciation of indigenous connections to nature	Cognitive / affective
B	Elaboration (provoked to thought)	Cognitive
C	Positive attitude toward heritage / nature preservation	Affective
D	Positive global evaluation of interpretation at site	Affective
E	Desire to participate in additional interpretive activities	Affective / behavioural
F	Desire to purchase a memento or souvenir related directly to site story	Affective / behavioural
G	Desire to stay longer	Affective / behavioural
H	Desire to return for repeat visit	Affective / behavioural
I	Positive word-of-mouth advertising	Affective / behavioural
J	Visitors found it relevant and meaningful to their lives	Cognitive
K	Visitors provoked to interact with the guide / interpreter (interactive experience)	Behavioural

A final set of criteria was then applied to the eleven indicators as follows:

- Each indicator was examined to ensure it lent itself to an ex post facto or correlational study design, since it is unfeasible to expect an interpretive program to be suspended in order to create a control condition necessary for an experimental design;
- Each indicator was assessed as to its theoretical basis (i.e. whether it was logical to expect that there would be some connection between it and the quality of on-site face-to-face interpretation); and
- The final suite of indicators needed to include at least one cognitive, one affective and one behavioural outcome.

No changes were made to the final suite of indicators as a result of applying this final set of criteria.

Thus, applying these criteria produced the final suite of eleven indicators which, in turn, became the focus for the development of the Interpretation Evaluation Tool Kit. These eleven indicators are shown in Table 12.

## Selection and development of data collection techniques and instruments

In theory, data can be collected on many of these indicators in several ways. However, as was discussed in Chapter 2 and in Tables 7, 8 and 9, some of these techniques require considerable research expertise, time and/or resources, or they are demanding of visitors. As shown in Table 13, four data collection options (a questionnaire, a formal interview, an informal interview including focus groups and observation) were reviewed with respect to validity, reliability and the criteria outlined in Chapter 2 (captured in the ‘burden on staff’ column), and then adjusted depending on the particular indicator in question.

These assessments resulted in a decision to use a visitor questionnaire to measure 10 of the 11 indicators (A to J), and participant observation to measure Indicator K.

**Table 13: Rating of Primary Data Collection Options**

Indicator Name	Type of Response <sup>a</sup>	Data Collection Option <sup>b</sup>	Reliability Rating <sup>c</sup>	Validity Rating <sup>c</sup>	Burden on Staff Rating <sup>d</sup>
A. Impact on current world view via empathy with historic period and people / impact on appreciation of indigenous connections to nature	Cognitive / Affective	Q	G	G	M
		FI	G	G	H
		II	P	M	H
		O	P	P	H
B. Elaboration (provoked to thought)	Cognitive	Q	M-P	M	M
		FI	G	G	H
		II	M-P	P	H
		O	M	P	H
C. Positive attitude toward heritage / heritage preservation	Affective	Q	G	G	M
		FI	G	G	H
		II	M	G	H
		O	P	P	H
D. Positive global evaluation of interpretation at site	Affective	Q	G	G	M
		FI	G	G	H
		II	M	G	H
		O	P	P	H
E. Desire to participate in additional interpretive activities	Affective / Behavioural	Q	G (A/B)	G (A/B)	M
		FI	G (A/B)	G (A/B)	H
		II	M (A/B)	G (A/B)	H
		O	P (A)	P (A)	VH
			G (B)	G (B)	VH
F. Desire to purchase souvenir related directly to site story	Affective / Behavioural	Q	G (A/B)	G (A/B)	M
		FI	G (A/B)	G (A/B)	H
		II	M (A/B)	G (A/B)	H
		O	P (A)	P (A)	VH
			G (B)	G (B)	VH

G. Desire to stay longer	Affective / Behavioural	Q FI II O	G (A/B) G (A/B) M (A/B) P (A/B))	G (A/B) G (A/B) G (A/B) P (A/B)	M H H VH
H. Desire return for repeat visit	Affective / Behavioural	Q FI II O	G (A/B) G (A/B) M (A/B) P (A/B))	G (A/B) G (A/B) G (A/B) P (A/B)	M H H VH
I. Positive word-of-mouth advertising	Affective / Behavioural	Q FI II O	G G M P	G G G P	M H H H
J. Visitors found it relevant and meaningful to their lives	Cognitive	Q FI II O	G G M P	M G G P	M H H H
K. Visitors provoked to interact with the guide (interactive experience)	Behavioural	Q FI II O	G G M G	M G G G	M H H H

**Notes:** Rating system based on Wagar (1976), Roggenbuck & Propst (1981), Ham (1986), and Yamada & Ham (2004).

<sup>a</sup> C = cognitive; A = affective; B = behavioural.

<sup>b</sup> Q = questionnaire; FI = formal interview; II = informal interview (incl. focus group), O = observation

<sup>c</sup> G = good; M = moderate; P = poor

<sup>d</sup> L = light; M = moderate; H = heavy; VH = very heavy

## Development and Validation of the Research Instruments

### Translation of Indicators A Through J into a Visitor Questionnaire

Development of the indicators into a research instrument required two main steps. First was designing a visitor questionnaire for each of the three types of settings (food and beverage, heritage and nature sites). Second was collecting data from real visitors at multiple locations representing each type of setting to make sure the information being produced was valid and reliable.

To do this, a questionnaire was drafted for measuring the kinds of responses sought by Indicators A to J, using Dillman's (1978) *Total Design Method*. Then the industry reference group and other social scientists contributed to ensuring that the wording of some of the questions was appropriate to each of the three settings. Through this procedure, three slightly different versions of the questionnaire were eventually produced (one for food and beverage sites, one for heritage sites and one for nature sites).

In designing the visitor questionnaire, an effort was made to use multiple measures for the more complex (possibly multi-dimensional) indicators (such as elaboration, attitudes and global evaluations) so the results might provide a more comprehensive picture about the impact of interpretation than a single measure alone could provide. This was possible for indicators A to D and I to J. Finally, in sequencing the questions in the survey form, sub-indicators were separated so that visitors would not recognise them as 'belonging together'.

Each questionnaire was subsequently field tested on 80 to 150 visitors at two or three different sites corresponding to each type of setting. The questionnaires typically took visitors between three and five minutes to complete. Although there were some refusals, there was a high rate of response at all sites, producing three data sets that could then be used for assessing the validity and reliability of each of the indicators.

Tables 14, 15 and 16 show the indicators used in each type of setting (food and beverage, heritage and nature). In each case, the table indicates whether the indicator was measured by a single item or by multiple sub-indicators. Validation of the multiple-item indicators was established using confirmatory factor analysis (CFA) in which each group of the sub-indicators was found to comprise a single factor. (In CFA terms, the criteria applied were that each sub-indicator needed to 'load' on the overall indicator, as determined by an eigenvalue of at least 1.0 and a Kaiser-Meyer-Olkin (KMO) score of at least 0.6. All the overall indicators and sub-indicators satisfied these conditions.)

**Table 14: Food and beverage interpretation indicators**

Indicator*	Indicator Long Title	Question Numbers**	Reliability (alpha)
B	Elaboration (provoked to thought)	1a, 1d, 1f, 1i	.75
D	Positive global evaluation of interpretation at site	1b, 1g	.85
E	Desire to participate in additional interpretive activities	3a	NA
F	Desire to purchase a product or memento related to place	3d	NA
G	Desire to stay longer	3b	NA
H	Desire to return for repeat visit	3c	NA
I	Positive word-of-mouth advertising	2a, 2b, 2c, 2d, 2e	.86
J	Visitors found it relevant and meaningful to their lives	1c, 1e, 1h	.62
K	Visitors provoked to interact with the guide (interactive experience)	Observation instrument	NA

\* Items making up indicators A and C were not reliable and these have been deleted. Selected items with low alpha scores were also deleted from indicators B, D and J.

\*\* These are the question numbers in the visitor questionnaire (Food & Beverage package) and in the evaluation template for food and beverage sites (in Data Entry View)

**Table 15: Heritage interpretation indicators**

Indicator	Indicator Long Title	Question Numbers*	Reliability (alpha)
A	Impact on current world view via empathy with historic period and people	1e, 1j, 1o, 1s, 1u	.77
B	Elaboration (provoked to thought)	1c, 1h, 1m, 1r, 1t	.83
C	Positive attitude toward heritage / heritage preservation	1d, 1i, 1n	.85
D	Positive global evaluation of interpretation at site	1a, 1f, 1k, 1p	.84
E	Desire to participate in additional interpretive activities	3a	NA
F	Desire to purchase a memento or souvenir related directly to site story	3d	NA
G	Desire to stay longer	3b	NA
H	Desire to return for repeat visit	3c	NA
I	Positive word-of-mouth advertising	2a, 2b, 2c, 2d, 2e	.89
J	Visitors found it relevant and meaningful to their lives	1b, 1g, 1l, 1q	.67
K	Visitors provoked to interact with the guide (interactive experience)	Observation instrument	NA

\* These are the question numbers in the visitor questionnaire (Heritage package) and in the evaluation template for heritage sites (in Data Entry View).



**Table 16: Nature interpretation indicators**

Indicator*	Indicator Long Title	Question Numbers**	Reliability (alpha)
A	Impact on appreciation of indigenous connections to nature	1e, 1n, 1r, 1t	.95
B	Elaboration (provoked to thought)	1c, 1h, 1l, 1q, 1s	.88
C	Positive attitude toward nature conservation	1d, 1i, 1m	.73
D	Positive global evaluation of interpretation at site	1a, 1f, 1j, 1o	.84
E	Desire to participate in additional interpretive activities	3a	NA
F	Desire to purchase a memento or souvenir related directly to site story	3d	NA
G	Desire to stay longer	3b	NA
H	Desire to return for repeat visit	3c	NA
I	Positive word-of-mouth advertising	2a, 2b, 2c, 2d, 2e	.90
J	Visitors found it relevant and meaningful to their lives	1b, 1g, 1k, 1p	.67
K	Visitors provoked to interact with the guide (interactive experience)	Observation instrument	NA

\* One item for indicator A was deleted due to low reliability.

\*\* These are the question numbers in the visitor questionnaire (Nature package) and in the evaluation template for nature-based sites (in Data Entry View).

The column labelled 'Question Numbers' lists the questions retained in the final questionnaire(s). The column labelled 'Reliability' shows (using 'Cronbach's alpha') the degree to which the multiple-item indicators vary consistently with each other and with the overall indicator they comprise. Reliability was determined using the multi-stage, item-to-whole method in which the contribution of each sub-indicator to the reliability of the overall indicator was assessed. In cases when a sub-indicator did not contribute a positive amount of reliability to the overall indicator, it was removed from the group. In this way, the resulting multiple-item indicators contained only the most reliable sub-indicators.

Alpha can vary from 0 to 1.0 (perfect reliability). An alpha of .60 is considered the lowest acceptable level of reliability for the types of indicators included here. Indicators with alphas larger than .70 are considered to have at least moderately strong reliability, and those approaching or exceeding .80 are exceptionally reliable. Those below .60 were therefore eliminated as a result of this process, and thus all indicators retained in the final Tool Kit instruments have an acceptable level of reliability, and most are strongly reliable. The reason some multiple-item indicators have fewer items than the same indicator has in other questionnaires is that they were found to be more reliable when they contained only some of the items.

In summary, the combination of procedures used to develop and test the methods, instruments and indicators ensured they are valid (i.e. they measure what they are supposed to measure). In addition, the rigorous reliability testing ensured that the sub-indicators used to comprise some of the overall indicators were producing consistently meaningful data with respect to the relevant indicator. As mentioned above, where there was any doubt, the sub-indicator was deleted from the questionnaire so the items remaining had the strongest reliability they could have. As mentioned previously, for this reason the number of items making up the same indicator in the three questionnaires sometimes varies, and so the questionnaires are not interchangeable.

These procedures resulted in the three versions of the visitor questionnaire that appear in the Interpretation Evaluation Tool Kit, each able to produce usable data on the performance of face-to-face interpretive programs for the type of setting for which it was designed. Of course, the validity and reliability of the three questionnaires is assured only if used in their current form, without changes to the order or wording of the questions and response categories.

### Development and Testing of the Observation Indicator

Indicator K (visitors were provoked to interact with the guide / interpreter) was assessed using participant observation, which involved a staff member essentially posing as a visitor and systematically observing the

interaction between an interpreter or guide and the other visitors. Following a set of observational ‘rules’ that have since been refined and included in the final Tool Kit, the observation procedures and recording form were field tested using real interpretive programs at Sovereign Hill and Port Arthur. In both cases, the procedures were found to be reliable. At Sovereign Hill, four independent observers recorded the same interactions in the same way almost all of the time (90 percent of interactions coded consistently) and at Port Arthur three observers were consistent in their coding 100 percent of the time.

## Finalisation of the Instruments

For all of the reasons outlined in the preceding sections, not all of the indicators appear in the instruments for all three settings and there are some differences in how a given indicator is described in the various settings (see Table 17). For example, indicator A in the Heritage package is labelled as ‘Impact on current world view via empathy with historic period and people’, but it is called ‘Impact on appreciation of indigenous connections to nature’ in the Nature package. This is because the items used to measure indicator A were different in the two cases and the title of the indicator needs to reflect this difference. Likewise, the wording of indicator C is slightly different in the Heritage and Nature packages, and indicator E in the Food & Beverage package is slightly different from its usage in the Heritage and Nature packages. In addition, neither indicators A or C are part of the Food & Beverage package.

Of the eleven overall indicators, ten (A-J) are measured using a visitor questionnaire. Indicator K focuses on how much verbal and physical interaction occurs between a presenter and an audience, an outcome best assessed by observing and listening to actual presentations. For this reason, a separate observation form is used for indicator K.

**Table 17: Final Indicators and Corresponding Question Numbers**

Indicator	Type of Setting		
	Food & Beverage (question numbers in Table 14)	Heritage (question numbers in Table 15)	Nature (question numbers in Table 16)
A*		Impact on current world view via empathy with historic period and people	Impact on appreciation of indigenous connections to nature
B	Elaboration (provoked visitors to thought)	Elaboration (provoked visitors to thought)	Elaboration (provoked visitors to thought)
C*		Positive attitude toward heritage preservation	Positive attitude toward nature conservation
D	Global evaluation of interpretation	Global evaluation of interpretation	Global evaluation of interpretation
E	Desire to participate in additional interpretive activities	Desire to participate in additional interpretive activities	Desire to participate in additional interpretive activities
F	Desire to purchase a product or memento related to the place	Desire to purchase a memento or souvenir related directly to the site story	Desire to purchase a memento or souvenir related directly to the site story
G	Desire to stay longer	Desire to stay longer	Desire to stay longer
H	Desire to return for a repeat visit	Desire to return for a repeat visit	Desire to return for a repeat visit
I	Positive word-of-mouth advertising	Positive word-of-mouth advertising	Positive word-of-mouth advertising
J	Interpretation was relevant and meaningful to visitors’ lives	Interpretation was relevant and meaningful to visitors’ lives	Interpretation was relevant and meaningful to visitors’ lives
K**	Visitors provoked to interact with the presenter	Visitors provoked to interact with the presenter	Visitors provoked to interact with the presenter

\* Wording of questions differs depending on the setting (Heritage or Nature), and these indicators are not part of the Food & Beverage setting.

\*\*Indicator K is measured using the Observation Form. All other indicators are measured using the Visitor Questionnaire.

The following excerpt from the Interpretation Evaluation Tool Kit describes for the user what the various indicators measure:

- **Indicator A:** This indicator captures the degree to which your interpretation impacted visitors’ point of view about their own lives (Heritage package) and their appreciation of indigenous feelings of connectedness to nature (Nature package). Results for this indicator give you a broad indication of whether your interpretation helps visitors to make connections and draw conclusions about these issues. This is accomplished with five

strongly reliable measurements (sub-indicators) in the Heritage questionnaire and with four strongly reliable measurements in the Nature questionnaire.

- **Indicator B:** This indicator is a measure of ‘elaboration,’ which is the amount of thinking your interpretation provokes visitors to engage in. ‘Provocation’ is considered by many experts to be the single most important outcome of interpretation. When visitors are provoked to thought, they have been moved to make connections between the topic being interpreted and what they already know and feel. Therefore, results for this indicator give you an indication of how much your interpretation has provoked people to process new thoughts about the things you interpret. This is accomplished with five strongly reliable measurements (sub-indicators) in the Heritage and Nature questionnaires and with four strongly reliable measurements in the Food & Beverage questionnaire.
- **Indicator C:** This indicator captures the degree to which your interpretation led visitors to have a stronger positive attitude to heritage preservation (Heritage package) and nature conservation (Nature package). Results for this indicator give you a broad indication of whether your interpretation is leading visitors to have a stronger positive attitude toward long-term protection of the kinds of values your site represents. This is accomplished with three strongly reliable measurements (sub-indicators) in both the Heritage questionnaire and Nature questionnaire.
- **Indicator D:** This indicator measures visitors’ overall (global) evaluation of interpretation at your site. Specifically, it captures whether they found the interpretive activities they attended to be enjoyable, good, interesting and satisfying. Results for this indicator give you a broad sense of visitors’ overall enjoyment and satisfaction with the interpretation you offer. This is accomplished with two strongly reliable sub-indicators in the Food & Beverage questionnaire, and with 4 strongly reliable sub-indicators in both the Heritage and Nature questionnaires.
- **Indicator E:** This indicator measures whether your interpretation is good enough to make visitors want to have even more. Results for this indicator give you a broad indication of whether your interpretation is stimulating visitors to want to immerse themselves more deeply in the things you are interpreting. This is accomplished with a single question in all three questionnaires.
- **Indicator F:** This indicator measures a commercially important outcome of interpretation – whether visitors were *stimulated to buy* a product, memento, or souvenir that is directly related to the story interpreted at your site (such as a bottle of wine at your cellar door, box of chocolate at your factory, or postcard, clothing item or other keepsake that the visitor feels is directly related to the place). Results for this indicator give you a sense of whether your interpretation is stimulating a ‘buying impulse’ in visitors. This is accomplished with a single question in all three questionnaires.
- **Indicator G:** This indicator measures whether your interpretation is good enough to make visitors want to stay longer at your site or operation. This may have commercial implications since visitors who actually extend their stay may spend additional money on food or in a gift shop. Results for this indicator give you a sense of whether your interpretation is contributing to visitors spending more time at the site than they had initially planned to spend. This is accomplished with a single question in all three questionnaires.
- **Indicator H:** This indicator measures the degree to which your interpretation stimulates visitors to want to return for a repeat visit. Of course, there is no way of knowing whether visitors actually return to your site, but scores on this indicator would give you a broad indication of whether visitors have the desire to do so. Measurement is accomplished with a single question in all three questionnaires.
- **Indicator I:** This indicator measures the degree to which your interpretation stimulates visitors to want to say positive things to another person about your site or operation (positive word-of-mouth advertising). Results for this indicator give you a broad indication of how inclined your visitors are to tell other people that your site or operation is interesting, enjoyable, worth the money and time to visit and whether other people should visit you. This is accomplished with five strongly reliable sub-indicators (questions 2A-2E) in all three questionnaires.
- **Indicator J:** This indicator measures the degree to which visitors think your interpretation is relevant and meaningful to their lives. Results for this indicator give you a broad indication of whether visitors felt your interpretation connected to things they already know and care about. This is accomplished with three modestly reliable sub-indicators in the Food & Beverage questionnaire, and with four modestly reliable sub-indicators in the Heritage and Nature questionnaires.
- **Indicator K:** This observational indicator allows you to assess how much your activities are provoking visitors to interact with the people presenting your interpretive program (i.e. guides, interpretive staff, etc.). In a sense, this is a measure of how one-way or two-way their communication with visitors is. Results for this indicator give you a detailed indication of the amount and kinds of interaction that is going on between presenters and visitors, who is initiating it, and what kinds of responses it is causing. This is accomplished using the observation form and procedures detailed in the Interpretation Evaluation Tool Kit.

## **Summary**

As outlined in this chapter, a rigorous process was followed to identify, select, validate and refine the methods, instruments and set of indicators that are the basis for assessing interpretation's outcomes. While there may well be other indicators of effectiveness or success, the indicators developed here were identified by a range of tourism industry representatives and heritage/protected area managers as the most important ones. It is the combination of this industry-driven process for identifying and selecting important indicators and the researcher-driven process used to develop valid and reliable ways of measuring them that makes the final set of indicators and the Tool Kit itself significant. The detail provided in this chapter is aimed at substantiating the claim that each instrument must be used only in its current form, retaining the structure, order and wording of the instrument and of the individual questions.

## Chapter 4

# FINAL PRODUCT: THE INTERPRETATION EVALUATION TOOL KIT

The most important achievement of this study has been the development and validation of a set of methods and instruments for assessing the outcomes of interpretation. The final product, the Interpretation Evaluation Tool Kit itself, can be purchased as a package from the STCRC. The Tool Kit consists of:

- An Interpretation Evaluation How-To Manual; and
- A Tool Kit CD – which includes all the applications and files needed to conduct evaluations of face-to-face interpretation programs and activities.

On the CD there is an Evaluation Tool Kit folder containing three versions of the visitor questionnaire, each designed for use in one of the three types of tourism operations: food and beverage interpretation settings, heritage interpretation settings, and nature interpretation settings, plus an observation form. There is also a software application that is custom-designed to align with the questionnaires, with a data entry view (similar to a spreadsheet) for typing responses from the self-completed questionnaires. Once the data has been entered, the application automatically produces tables and charts that can be printed directly or exported to documents and other applications such as Powerpoint slides.

The observation method allows the evaluator to use listening and observation to determine the extent to which visitors were provoked to interact with the presenter (interpreter, guide etc.). The Tool Kit includes detailed instructions on how to use the observation form to conduct unobtrusive observations.

Other than access to a computer, printer and photocopier, the Tool Kit is self-contained, including how to interpret the results and identify steps that need to be taken to enhance interpretation's outcomes.

## Uses and Benefits

The Tool Kit was designed as a tool for taking the 'pulse' of interpretation and, over a period of time, to monitor trends in what it is achieving. The Tool Kit is designed mainly to be used to evaluate overall site interpretation, but it can also be used to evaluate a specific interpretive product or program. It can also be used when trialling a new interpretation program or product, but this must be done with caution as a range of factors can impact on the outcome of a single interpretive product at a single point in time.

The results from any of these uses can help to determine the extent to which interpretation benefits an organisation and helps to advance the organisation's goals.

## Limitations

While the Tool Kit provides a valid, reliable, easy-to-use and rapid-response measure of 'how well' or 'how much' is being achieved, it does not provide information about the longer-term impacts of interpretation on visitors once they have returned home. Nor does it provide answers as to *why* interpretation is or is not achieving an organisation's desired outcomes. A more complex research design is required to assess cause-and-effect relationships and to establish what specific actions need to be taken to improve interpretation. Nevertheless, patterns in results can be used to suggest broad avenues for improvement or correction, and the Tool Kit highlights these.

Since the Tool Kit's focus is strictly on interpretation, it is not meant to assess the performance of other types of visitor services (such as how satisfied people are with the cleanliness and appearance of the setting, their opinions about the courtesy of staff, whether they liked the food etc.). In addition, the Tool Kit is designed to evaluate face-to-face interpretation, not self-guided or non-personal media such as signs, exhibits, websites, brochures etc. However, it needs to be recognised that, in an exit survey, a visitor's responses are likely to reflect their experience with the entire interpretive program, not necessarily just the face-to-face elements of it.

The Tool Kit is *not* intended as a method or set of instruments for assessing individual staff performance.

Finally, the visitor questionnaires in the Tool Kit are intended to be used intact to measure multiple indicators of interpretation outcomes. Any editing, addition or deletion to individual indicators will lead to unavoidable measurement errors that will undermine the accuracy and usefulness of an evaluation. Users are encouraged,

however, to add additional questions of interest (e.g. sociodemographic questions) to the end of the questionnaire as long as they do not reword or reorder the existing questions.

## **Recommendations for Follow-up and Future Research**

The Interpretation Evaluation Tool Kit lends itself not only to multiple evaluations over time, but also to cross-site comparisons. It is the intention of the researchers and the STCRC to encourage this kind of comparison by creating a website where operators can download their results. We anticipate that this will be accommodated at no charge to the user, and that online submissions will be anonymous, so that the outcomes of an organisation's interpretation could be benchmarked against other comparable operations without embarrassment, risk or loss of competitive advantage.

There is also scope for a follow-up project to develop an Evaluation Tool Kit for non-personal interpretation media. Much of the process outlined in Chapter 3 could be replicated for evaluating other media, and the process can also provide a guide for researchers seeking to develop methods and instruments customised to a particular setting or operation not covered by the current Tool Kit package (e.g. indigenous tourism operations).

Because development of the Tool Kit indicators was based on facilitated sessions at two heritage tourism sites, it is possible that additional indicators of interest might have emerged had separate sessions been conducted at nature or food and beverage sites. Although the 11 indicators included in the Tool Kit were found to have face validity according to users from these sites, there is scope to separately workshop indicators not only at nature and food and beverage destinations, but also at other sites with specific content foci (e.g. zoos, botanical gardens, aquaria, museums etc.).

## **A Final Word**

This project has developed, tested and refined a theoretically sound and data-driven methodology and practical evaluation tool for assessing the outcomes of interpretation for use in heritage settings, nature settings, and food and beverage operations. The most important outcomes of interpretation were identified by industry representatives in the early stages of the research and formed the basis of the final set of eleven indicators. The final Tool Kit includes three different evaluation packages (each consisting of a five-minute visitor survey and a form for observing visitor behaviour) that are customised for each type of setting. The indicators in each package reflect the types of outcomes that users of the Tool Kit actually want from their interpretive programs and they measure outcomes that are achievable in terms of what theory tells us about interpretation's potential impacts on how visitors think, feel and possibly behave, with respect to the things being interpreted for them.

Use of the Tool Kit itself requires minimal effort and expense, and no social research expertise, yet produces results that are both valid (i.e. accurately measuring what each indicator is supposed to measure) and reliable (i.e. producing consistent measurements so meaningful comparisons can be made over time). The Tool Kit is suitable for use by individual site managers, tourism operators, protected area managers and anyone interested in evaluating or monitoring the outcomes of its face-to-face interpretation.

## REFERENCES

- Adelman, L., Falk, J. & James, S. (2000). 'Impact of national aquarium in Baltimore on visitors' conservation attitudes, behavior, and knowledge'. *Curator*, 43(1), pp.33-60.
- Allard, M., Boucher, S. & Forest, L. (1994). 'The museum and the school'. *McGill Journal of Education*, 29(2), pp.197-212.
- Allen, S. (1997). 'Using scientific inquiry activities in exhibit explanations'. *Science Education*, 81(6), pp.715-734.
- Anderson, D., Lucas, K. & Ginns, I. (2000). 'Development of knowledge about electricity and magnetism during a visit to a science museum and related post-visit activities'. *Science Education*, 84(5), pp.658-679.
- Armstrong, E. & Weiler, B. (2003). 'They said what to whom?!' In Black, R. & Weiler, B. (Eds.), *Interpreting the land down under – Australian heritage interpretation and tour guiding*. Fulcrum Applied Communication Series, Fulcrum Books, Golden, Colorado, pp.109-127.
- Ash, D. (2003). 'Dialogic inquiry in life science conversations of family groups in a museum'. *Journal of Research in Science Teaching*, 40(2), pp.138-162.
- Ballantyne, R. & Hughes, K. (2006). 'Using front-end and formative evaluation to design and test persuasive bird feeding warning signs'. *Tourism Management*, 27(2), pp.235-246.
- Bass, J., Manfredo, M., Lee, M. & Allen, D. (1989). 'Evaluation of an informational brochure for promoting charter boat trip opportunities along the Oregon Coast'. *Journal of Travel Research*, 27(3), pp.35-37.
- Beaumont, N. (2001). 'Ecotourism and the conservation ethic: recruiting the uninitiated or preaching to the converted?' *Journal of Sustainable Tourism*, 9(4), pp.317-341.
- Beck, L. & Cable, T. (1998). *Interpretation for the 21st Century*. Sagamore Publishing, Champaign, Illinois.
- Beckmann, E. (2003). 'Listening between the lines – using research-based knowledge to support the development of interpretive text'. In Black, R. & Weiler, B. (Eds.), *Interpreting the land down under – Australian heritage interpretation and tour guiding*. Fulcrum Applied Communication Series, Fulcrum Books, Golden, Colorado, pp.128-150.
- Beckmann, E. & Savage, G. (2003). 'Evaluation of dingo education strategy and programs for Fraser Island'. Queensland Parks & Wildlife Service, Brisbane.
- Beeho, A. & Prentice, R. (1995). 'Evaluating the experiences and benefits gained by tourists visiting a socio-industrial heritage museum: An application of ASEB Grid analysis to Blists Hill Open-air Museum, the Ironbridge Gorge Museum, United Kingdom'. *Museum Management and Curatorship*, 14(3), pp.229-251.
- Beer, V. (1987). 'Great expectations: do museums know what visitors are doing?' *Curator*, 30(3), pp.206-215.
- Belch, G. & Belch, M. (2001). *Advertising and promotion: An integrated marketing communications perspective* (4<sup>th</sup> ed.). McGraw-Hill Companies, New York.
- Bicknell, S. & Mann, P. (1993). 'A picture of visitors for exhibition developers'. In Thompson, D., Benefield, A., Bitgood, S., Shettel, H. & Williams, R. (Eds.), *Visitor studies – theory, research and practice* (Volume 5). Visitor Studies Association, Jacksonville, Alabama, pp.88-98.
- Bitgood, S. & Patterson, D. (1993). 'The effects of gallery changes on visitor reading and object viewing time'. *Environment and Behavior*, 25(6), pp.761-781.
- Bitgood, S. (2000). 'The role of attention in designing effective interpretive labels'. *Journal of Interpretation Research*, 5(2), pp.31-45.
- Bitgood, S., Patterson, D. & Benefield, A. (1988). 'Exhibit design and visitor behaviour: empirical relationships'. *Environment and Behavior*, 20(4), pp.474-491.
- Bixler, R., Carlisle, C., Hammitt, W. & Floyd, M. (1994). 'Observed fears and discomforts among urban students on field trips to wildland areas'. *The Journal of Environmental Education*, 26(1), pp.24-33.
- Blud, L. (1990). 'Social interaction and learning among family groups visiting a museum'. *Museum Management and Curatorship*, 9, pp.43-51.
- Boisvert, D. & Slez, B. (1994). 'The relationship between visitor characteristics and learning-associated behaviors in a science museum discovery space'. *Science Education*, 78(2), pp.137-148.
- Borun, M., Drita, J., Johnson, J., Peter, N., Wagner, K., Fadigan, K., Jangaard, A., Stroup, E. & Wenger, A. (1998). *Family learning in museums: The PISEC perspective*. Philadelphia/Camden Informal Science Education Collaborative, The Franklin Institute, Philadelphia.
- Braverman, B. (1988). 'Empowering visitors: Focus group interviews for art museums'. *Curator*, 31(1), pp.43-52.
- Broad, S. & Smith, L. (2004). 'Who educates the public about conservation issues? Examining the role of zoos and the media'. Paper presented to the International Tourism and Media Conference, Melbourne, 24-26 November.
- Broad, S. & Weiler, B. (1998). 'Captive animals and interpretation – A tale of two tiger exhibits'. *The Journal of Tourism Studies*, 9(1), pp.14-27.
- Brown, F. & Koran, J. (1998). 'Learning from ruins: A visitor study at Uxmal'. *Curator*, 41(2), pp.121-131.
- Cable, T., Knudson, D., Udd, E. & Stewart, D. (1987). 'Attitude changes as a result of exposure to interpretive messages'.

- Journal of Park and Recreation Administration*, 5(1), pp.47-60.
- Cameron, C. & Gatewood, J. (2000). 'Excursions into the un-remembered past: what people want from visits to historical sites'. *The Public Historian*, 22(3), pp.107-127.
- Camp, B., Koran, J. & Koran, M. (2000). 'Photographs as a research tool in visitor studies'. *Journal of Interpretation Research*, 5(2), pp.47-58.
- Cartmill, R. & Day, L. (1997). 'Prevention of substance abuse: Can museums make a difference?' *Curator*, 40(3), pp.197-210.
- Cherem, C. & Driver, B. (1983). 'Visitor employed photography: A technique to measure common perceptions of natural environments'. *Journal of Leisure Research*, 15(1), pp.65-83.
- Chiozzi, G. & Andreotti, L. (2001). 'Behavior vs. time: understanding how visitors utilize the Milan Natural History Museum'. *Curator*, 44(2), pp.153-165.
- Cialdini, R. (1996). 'Activating and aligning two kinds of norms in persuasive communications'. *Journal of Interpretation Research* 1(1), pp.3-10.
- Cole, D., Hammond, T. & McCool, S. (1997). 'Information quantity and communication effectiveness: low-impact messages on wilderness trailsides bulletin boards'. *Leisure Sciences*, 19, pp.59-72.
- Cole, D. (1998). 'Written appeals for attention to low-impact messages on wilderness trailsides bulletin boards: Experimental evaluations of effectiveness'. *Journal of Park and Recreation Administration*, 16(1), pp.65-79.
- Crump, C., Bryne, M. & Croucamp, W. (2000). 'A South African interactive arthropod exhibition'. *Journal of Biological Education*, 35(1), pp.12-16.
- De White, T. & Jacobson, S. (1993). 'Evaluating conservation education programs at a South American zoo'. *Visitor studies: theory, research and practice* (Vol. 6). The Visitor Studies Association, Jacksonville, Alabama, pp. 207-210.
- Department of Natural Resources and Environment Victoria (1999). *Best practice in park interpretation and education – a report to the ANZECC working group on national park and protected area management benchmarking and best practice program*. Available at <http://www.deh.gov.au/parks/best-practice/reports/interpretation/index.html>
- Diamond, J. (1986). 'The behavior of family groups in science museums'. *Curator*, 29(2), pp.139-154.
- Diamond, J. (1999). *Practical evaluation guide: Tools for museums & other informal educational settings*. AltaMira Press, Walnut Creek, Calif.
- Dick, R., Myklestad, E. & Wagar, J. (1975). *Audience attention as a basis for evaluating interpretive presentations* (Research Paper PNW-198). USDA Forest Service, Portland.
- Dierking, L., Adelman, L.M., Ogden, J., Lehnhardt, K., Miller, L. & Mellen, J. (2004). 'Using a behaviour change model to document impacts of visits to Disney's Animal Kingdom: A study of investigating intended conservation action'. *Curator*, 47(3), pp.322-43.
- Dillman, D. (1978). *Mail and telephone surveys: the total design method*. Wiley, New York.
- Donald, J. (1991). 'The measurement of learning in the museum'. *Canadian Journal of Education*, 16(3), pp.371-382.
- Dufresne-Tasse, C. & Lefebvre, A. (1994). 'The museum in adult education: A psychological study of visitor reactions'. *International Review of Education*, 40(6), pp.469-484.
- Espinier, S. (1999). *The use and effect of hazard warning signs: managing visitor safety at Franz Josef and Fox Glaciers*. Science for Conservation, 108. Department of Conservation, Wellington, New Zealand.
- Falk, J. (1993). 'Assessing the impact of exhibit arrangement on visitor behavior and learning'. *Curator*, 36(2), pp.133-146.
- Falk, J. (1997). 'Testing a museum exhibition design assumption: Effect of explicit labeling of exhibit clusters on visitor concept development'. *Science Education*, 81(6), pp.697-687.
- Falk, J. (2003). 'Personal meaning mapping'. In Caban, G., Scott, C., Falk, J. & Dierking, L. (Eds.), *Museums and creativity: a study into the role of museums in design education*. Powerhouse Publishing, Sydney.
- Falk, J. & Dierking, L. (1997). 'School field trips: Assessing their long-term impact'. *Curator*, 40(3), pp.211-218.
- Falk, J. & Storksdieck, M. (2005). 'Using the contextual model of learning to understand visitor learning from a science center exhibition'. *Science Education*, 89, pp.744-778.
- Falk, J., Moussouri, T. & Coulson, D. (1998). 'The effect of visitors' agendas on museum learning'. *Curator*, 41(2), pp.107-120.
- Fernandez, G. & Benloch, M. (2000). 'Interactive exhibits: how visitors respond'. *Museum International*, 52(4), pp.53-59.
- Finson, K. & Enochs, L. (1987). 'Student attitudes toward science-technology-society resulting from visitation to a science-technology museum'. *Journal of Research in Science Teaching*, 24(7), pp.593-609.
- Fisher, M. (1997). 'The effect of humor on learning in a planetarium'. *Science Education*, 81(6), pp.703-713.
- Fowler, F. & Mangione, T. (1990). *Standardized survey interviewing*. SAGE Publications, Newbury Park, Calif.
- Geva, A. & Goldman, A. (1991). 'Satisfaction measurement in guided tours'. *Annals of Tourism Research*, 18, pp.177-185.
- Goulding, C. (2001). 'Romancing the past: heritage visiting and the nostalgic consumer'. *Psychology & Marketing*, 18(6), pp.565-592.
- Graziano, A. & Raulin, M. (2000). *Research methods: a process of inquiry* (4<sup>th</sup> ed.). Allyn & Bacon, Needham Heights, Massachusetts.
- Grove, S. & Fisk, R. (1992). 'Observational data collection methods for services marketing: An overview'. *Journal of the*



- Academy of Marketing Science*, 20(3), pp.217-224.
- Gurnhill, A. & Vertigan, C. (2002). *Visitor evaluation of ghost tours and proposed audio tours at the Port Arthur Historic Site*. Port Arthur Historic Site, Port Arthur, Tasmania.
- Hall, T., Hockett, K. & Smith-Jackson, T. (2001). *Tests of small panel signs in Yosemite National Park*. Report prepared for Yosemite National Park, Blacksburg, Virginia.
- Ham, S. (1982). 'Familiarity and adaptation: a study of family attendances at interpretive activities'. Ph.D. thesis, University of Idaho, Moscow, Idaho.
- Ham, S. (1986). 'Social program evaluation and interpretation: A literature review'. In G. E. Machlis (Ed.), *Interpretive Views*. National Parks and Conservation Association, Washington, pp.9-38.
- Ham, S. (1992). *Environmental interpretation: a practical guide for people with big ideas and small budgets*. North American Press, Golden, Colorado.
- Ham, S. & Krumpke, E. (1995). 'How can site-based interpretation be used to change behaviors and contribute to biodiversity and ecosystem conservation in Central America?' Paper presented at Academy for Educational Development (GreenCom Project ), Technical Advisory Group Meeting, October 16.
- Ham, S. & Weiler, B. (2000). 'Profile of tourists in the Panama Canal watershed protected areas: Final research report'. Academy for Educational Development / GreenCOM Project, Washington.
- Ham, S. & Weiler, B. (2002). 'Toward a theory of quality in cruise-based nature guiding'. *Journal of Interpretation Research*, 7(1), pp.29-49.
- Ham, S. & Weiler, B. (2005). *Experimental evaluation of persuasive communication strategies aimed at influencing problem visitor behaviour at Port Campbell National Park*. Final research report to Parks Victoria. Monash University, Tourism Research Unit, Melbourne.
- Ham, S., Housego, A. & Weiler, B. (2005). *Tasmanian thematic interpretation planning manual*. Tourism Tasmania, Hobart.
- Harvey, M., Loomis, R., Bell, P. & Marino, M. (1998). 'The influence of museum exhibit design on immersion and psychological flow'. *Environment and Behavior*, 30(5), pp.601-627.
- Henderson, K. & Bialeschki, D. (1995). *Evaluating leisure services: making enlightened decisions*. Venture Publishing, State College, Pennsylvania.
- Hendricks, W. (2000). 'Attitudes toward roles in a wilderness education program'. *USDA Forest Service Research Conference Proceedings RMRS-P-15*, 4, pp.203-207.
- Henriksen, E. (2001). 'High school students' understanding of radiation and the environment: Can museums play a role?' *Science Education*, 85(2), pp.189-207.
- Hernandez, S. & De Celis Herrero, J. (1993). 'Beliefs in the science museum: An exhibit evaluation', *Visitor studies: theory, Research and practice* (Vol. 6). The Visitor Studies Association, Jacksonville, Alabama, pp.116-128.
- Higham J. & Carr, A. (2002) 'Ecotourism visitor experiences in Aotearoa/New Zealand: challenging the environmental values of visitors in pursuit of pro-environmental behaviour', *Journal of Sustainable Tourism*, 10 (4), pp.277-294.
- Hobbs, D. (1987). 'Strategy for needs assessments'. In Johnson, D., Clancy, M. & Summers, G. (Eds.), *Needs assessment: theory and methods*. Iowa State University Press, Ames, Iowa, pp.20-34.
- Hodgson, R. & Fritschen, J. (1984). 'Evaluating interpretation'. In Hodgson, R. (Ed.), *Supplements to a guide to cultural and environmental interpretation in the U.S. Army Corps of Engineers*. Army Engineers Waterways Experiment Station, Vicksburg, Mississippi, pp.57-85.
- Holbrook, A., Berent, M., Krosnick, J., Visser, P. & Boninger, D. (2005). 'Attitude importance and the accumulation of attitude-relevant knowledge in memory'. *Journal of Personality and Social Psychology*, 88(5), pp.749-769.
- Hornfeldt, D., McAvoy, L. & Schleien, S. (1989). 'Influences of integration on learning of natural history concepts by children with and without disabilities'. *Proceedings of the National Interpreters Workshop*, pp.408-415.
- Horsley, D. (1988). 'The unintended effects of a posted sign on littering attitudes and stated intentions'. *Journal of Environmental Education*, 19(3), pp.10-14.
- Howard, J. (2000). 'Research in progress: does environmental interpretation influence behaviour through knowledge or affect?' *Australian Journal of Environmental Education*, 15/16, pp.153-156.
- Howard, J., Lipscombe, N. & Porter, A. (2001). 'The tourist, the dingo, and interpretation on Fraser Island'. Paper presented to Interpretation Australia Association, pp.98-106, available at <http://www.interpretationaustralia.asn.au/confproceedings/109-117.pdf>
- Hughes, M. & Morrison-Saunders, A. (2002). 'Impact of trail-side interpretive signs on visitor knowledge'. *Journal of Ecotourism*, 1(2&3), pp.122-132.
- Hwang, S., Lee, C. & Chen, H. (2005). 'The relationship among tourists' involvement, place attachment and interpretation satisfaction in Taiwan's national parks'. *Tourism Management*, 26, pp.143-156.
- Jacobson, S. (1987). 'Conservation education programs: Evaluate and improve them'. *Environmental Conservation*, 14(3), pp.201-206.
- Jacobson, S. (1999). *Communication skills for conservation professionals*. Island Press, Washington DC.

- Jakobsen, C. (1999). *Evaluation of interpretive exhibits in Rocky Mountain National Park*. University of Idaho, Moscow, Idaho.
- Javlekar, V. (1989). 'Learning scientific concepts in science centers', *Visitor Studies: theory, research and practice* (Vol. 2), Visitor Studies Association, pp.168-179
- Johnson, B. (1989). 'Interpretive signs increase effectiveness of brush-pile barriers (North Carolina)'. *Restoration and Management*, 7(2), pp.103.
- Jurin, R., Danter, J. & Roush Jr., D. (2000). 'Evaluating your message effects', *Environmental communication: skills and principles for natural resource managers, scientists, and engineers*. Pearson Custom Publishing, Boston, pp.71-82.
- Kanel, V. & Tamir, P. (1991). 'Different labels – different learnings'. *Curator*, 34(1), pp.18-30.
- Klein, H. (1993). 'Tracking visitor circulation in museum settings'. *Environment and Behavior*, 25(6), pp.782-800.
- Knapp, D. (2000). 'Memorable experiences of a science field trip'. *School Science and Mathematics*, 100(2), pp.65-72.
- Knapp, D. & Barrie, D. (1998). 'Ecology versus issue interpretation – the analysis of two different messages'. *Journal of Interpretation Research*, 3(1), pp.21-38.
- Knapp, D. & Benton, G. (2005). 'Long-term recollections of an environmental interpretive program'. *Journal of Interpretation Research*, 10(1), pp.51-53.
- Knapp, D. & Benton, G. (2006). 'Episodic and semantic memories of a residential environmental education program'. *Environmental Education Research*, 12(2), pp.165 - 177.
- Knapp, D. & Poff, R. (2001). 'A qualitative analysis of the immediate and short-term impact of an environmental interpretive program'. *Environmental Education Research*, 7(1), pp.55-65.
- Knapp, D. & Yang, L. (2002). 'A phenomenological analysis of long-term recollections of an interpretive program'. *Journal of Interpretation Research* 7(2), pp.7-17.
- Knudson, D., Cable, T. & Beck, L. (1995). 'Evaluation', *Interpretation of cultural and natural resources*. Venture Publishing, State College, Pennsylvania, pp.441-465.
- Korn, R. (1988). 'Self-guiding brochures: An evaluation'. *Curator*, 31(1), pp.9-19.
- Korn, R. & Jones, J. (2000). 'Visitor behaviour and experiences in the four permanent galleries at the Tech Museum of Innovation'. *Curator*, 43(3), pp.261-281.
- Krueger, R. & Casey, M. (2000). *Focus groups: A practical guide for applied research* (3<sup>rd</sup> ed.). Sage Publications, Thousand Oaks, Calif.
- Lackey, B., Ham, S. & Hall, T. (2002). *Tests of perceived risk and attention paying to bear safety signs in Yosemite National Park*. Research report to the National Park Service. University of Idaho, Moscow, Idaho.
- Larsen, D. (Ed.) (2003). *Meaningful interpretation – how to connect hearts and minds to places, objects and other resources*. U.S. National Park Service, Eastern National, Fort Washington, Pennsylvania.
- Lee, T. & Balchin, N. (1995). 'Learning and attitude change at British Nuclear Fuel's Sellafield Visitors Centre'. *Environmental Psychology*, 15(4), pp.283-298.
- Leinhardt, G. & Crowley, K. (1998). *Conversational elaboration as a process and an outcome of museum learning*. Museum Learning Collaborative Technical Report (MLC-01). University of Pittsburgh, Pittsburgh, Pennsylvania.
- Leinhardt, G. & Knutson, K. (2004). *Listening in on museum conversations*. AltaMira Press, Walnut Creek, Calif.
- Leinhardt, G., Crowley, K. & Knutson, K. (Eds.) (2002). *Learning conversations in museums*. Lawrence Erlbaum Associates Mahwah, New Jersey.
- Lewis, W. (1980). *Interpreting for park visitors*. Eastern Acorn Press, Philadelphia.
- Light, D. (1995). 'Visitors' use of interpretive media at heritage sites'. *Leisure Studies*, 14(2), pp.132-149.
- Loomis, R. (1987). *Museum visitor evaluation: New tool for management*. American Association for State and Local History, Nashville, Tennessee.
- Loomis, R. (1996). 'How do we know what the visitor knows? – learning from interpretation. *Journal of Interpretation Research* 1(1), pp.39-45.
- Lundegren, H. & Farrell, P. (1985). *Evaluation for leisure service managers*. Saunders College Publishing, New York.
- Machlis, G. & McKendry, J. (1989). *Visitor Service Project Phase III: Techniques for evaluating interpretation*. University of Idaho, Moscow, Idaho.
- Madin, E. & Fenton, D. (2004). 'Environmental interpretation in the Great Barrier Reef Marine Park: an assessment of programme effectiveness'. *Journal of Sustainable Tourism*, 12(2), pp.121-137.
- Mallick, S. & Driessen, M. (2003). 'Feeding of wildlife: how effective are the 'keep wildlife wild' signs in Tasmania's national parks?' *Ecological Management & Restoration*, 4, pp.199-204.
- Markwell, K. (1997). 'Dimensions of photography in a nature-based tour'. *Annals of Tourism Research*, 24(1), pp.131-155.
- Mayes, G., Dyer, P. & Richins, H. (2004). 'Dolphin-human interaction: pro-environmental attitudes, beliefs, and intended behaviour and actions of participants in interpretation programs: a pilot study'. *Annals of Leisure Research*, 7(1), pp.34-53.
- McDonald, L. (2000). *Aiming for excellence: an evaluation handbook for interpretive services in California State Parks*.

- California State Parks Interpretation Section, Sacramento, Calif.
- McManus, P. (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(3), pp.263-270.
- McManus, P. (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, pp.37-44.
- McManus, P. (1989). 'Oh, yes, they do: How museum visitors read labels and interact with exhibit texts'. *Curator*, 32(3), pp.174-189.
- Medlin, N. (1992). *Development of a handbook for evaluating interpretive services*. Unpublished MS, University of Idaho, Moscow, Idaho.
- Medlin, N. & Ham, S. (1992). *A handbook for evaluating interpretive services*. University of Idaho, Moscow, Idaho.
- Morgan, D., Krueger, R. & King, J. (1998). *Focus group kit*. Sage Publications, Thousand Oaks, Calif.
- Morgan, M. J., Absher, J., Loudon, B. & Sutherland, D. (1997). 'The relative effectiveness of interpretive programs directed by youth and adult naturalists in a national forest'. *Journal of Interpretation Research*, 2(1), pp.13-26.
- Moscardo, G. (1999a). 'Communicating with two million tourists – a formative evaluation of an interpretive brochure'. *Journal of Interpretation Research*, 4(1), pp.21-37.
- Moscardo, G. (1999b). *Making visitors mindful – principles for creating sustainable visitor experiences through effective communication*. Sagamore Publishing, Champaign, Illinois.
- Moscardo, G., Ballantyne, R. & Hughes, K. (in press). *Interpretive signs and exhibits – principles in practice*. Fulcrum Applied Communication Series, Fulcrum Books, Golden, Colorado.
- Moses, Epstein & Wiseman Inc. (1977). *Assessing the impact of interpretive programs*. Moses, Epstein & Wiseman, Washington.
- Nielsen, C. & Buchanan, T. (1986). 'A comparison of the effectiveness of two interpretive programs regarding fire ecology and fire management'. *Journal of Interpretation*, 2(1), pp.1-10.
- Ogden, J., Lindburg, D. & Maple, T. (1993). 'The effects of ecologically-relevant sounds on zoo visitors'. *Curator*, 36(2), pp.147-156.
- Orams, M. (1997). 'The effectiveness of environmental education: can we turn tourists into 'greenies'?' *Progress in Tourism and Hospitality Research*, 3, pp.295-306.
- Pastorelli, J. (2003). *Enriching the experience: an interpretive approach to tour guiding*. Hospitality Press, Frenchs Forest, New South Wales.
- Pearce, P. & Moscardo, G. (1985). 'Visitor evaluation: an appraisal of goals and techniques'. *Evaluation Review*, 9(3), pp.281-306.
- Pearce, P. & Moscardo, G. (1998). 'The role of interpretation in influencing visitor satisfaction: a rainforest case study'. In Faulkner, W., Tidswell, C., & Weaver, D. (Eds.), *Progress in tourism and hospitality research*, 1998, Part 1. Proceedings of the Eighth Australian Tourism and Hospitality Research Conference, Gold Coast. Bureau of Tourism Research, Canberra, pp.309-319.
- Petty, R. & Cacioppo, J. (1986). *Communication and persuasion: central and peripheral routes to attitude change*. Springer-Verlag, New York.
- Phillpott, P. (1996). 'Visitor viewing behaviour in the Gaherty Reptile Breeding Centre, Jersey Wildlife Preservation Trust: a preliminary study'. *Dodo: Journal of Wildlife Preservation Trusts*, 32, pp.193-202.
- Prentice, R. (1991). 'Measuring the educational effectiveness of on-site interpretation designed for tourists: An assessment of student recall from geographical field visits to Kidwelly Castle, Dyfed'. *Area*, 23(4), pp.297-308.
- Prentice, R., Sinead, G. & Stuart, M. (1998). 'Visitor learning at a heritage attraction: A case study of *Discovery* as a media product'. *Tourism Management*, 19(1), pp.5-23.
- Regnier, K., Gross, M. & Zimmerman, R. (1994). *Techniques for programs and presentations* (3<sup>rd</sup> ed.). UW-SP Foundation Press, Stevens Point, Wisconsin.
- Reid, S. (2000). 'From dog waste to daisy pickers: 'Leave No Trace' educates outdoor recreators'. *Parks and Recreation*, 35(7), pp.70-78.
- Rennie, G. (1986). 'Development of a methodology for evaluating self guided trails'. In Marsh, J.S. (Ed.), *Natural and cultural heritage interpretation evaluation*, Interpretation Canada, Ontario, Canada, pp. 57-62.
- Riddick, C. & Russell, R. (1999). *Evaluative research in recreation, park, and sport settings: searching for useful information*. Sagamore Publishing, Champaign, Illinois.
- Roggenbuck, J. & Propst, D. (1981). 'Evaluation of interpretation'. *The Journal of Interpretation*, 6(1), pp.13-23.
- Roggenbuck, J., Williams, D. & Bobinski, C. (1992). 'Public-private partnership to increase commercial tour guides' effectiveness as nature interpreters'. *Journal of Park and Recreation Administration*, 10(2), pp.41-50.
- Rossi, P., Freeman, H. & Lipsey, M. (1999). *Evaluation: a systematic approach* (6<sup>th</sup> ed.). Sage Publications, Newbury Park, Calif.
- Ryan, C. & Dewar, K. (1995). 'Evaluating the communication process between interpreter and visitor'. *Tourism Management*, 16(4), pp.295-303.

- Salant, P. & Dillman, D. (1994). *How to conduct your own survey*. John Wiley & Sons, New York.
- Sandifer, C. (1997). 'Time-based behaviors at an interactive science museum – exploring the differences between weekday/weekend and family/nonfamily visitors'. *Science Education*, 81(6), pp.689-701.
- Savage, G. & James, J. (2001). *A practical guide to evaluating natural and cultural heritage interpretation: exit interviews, observation methods, focus group discussions*. Flinders University Press, Adelaide.
- Schulhof, R. (1990). 'Visitor response to a native plant habitat exhibit'. *Visitor Behavior*, 5(4), pp.11-12.
- Schuman, H. & Scott, J. (1987). 'Problem in the use of survey questions to measure public opinions'. *Science*, 236(22), pp.957-959.
- Screven, C. (1990). 'Uses of evaluation before, during, and after exhibit design'. *ILVS Review: Journal of Visitor Behavior*, 1(2), pp.36-67.
- Sharpe, G. (1982). 'An overview of interpretation'. In Sharpe, G. (Ed.), *Interpreting the environment* (2<sup>nd</sup> ed.), John Wiley & Sons, New York, pp.2-27
- Silverman, L. & Masberg, B. (2001). 'Through their eyes: the meaning of heritage site experiences to visitor who are blind or visually impaired'. *Journal of Interpretation Research*, 6(1), pp.31-47.
- Slater, J. (1998). 'Qualitative research in advertising'. In J. Jones (Ed.), *How advertising works: The role of research*. SAGE Publications, Thousand Oaks, Calif., pp.121-135.
- Smith, L. & Broad, S. (in press). 'Examining zoo visitors' attention to conservation messages: a case study of an elephant exhibit'. *Tourism Review International*.
- Solomon, M. (1992). *Consumer behavior: buying, having, and being*. Allyn and Bacon, Needham Heights, Massachusetts.
- Stevens, R. & Hall, R. (1997). 'Seeing *Tornado*: How video traces mediate visitor understandings of natural phenomena in a science museum'. *Science Education*, 81(6), pp.735-747.
- Stevens, R., Wrenn, B., Ruddick, M. & Sherwood, P. (1997). *The marketing research guide*. The Haworth Press, Binghamton, New York.
- Stevenson, J. (1991). 'The long-term impact of interactive exhibits'. *International Journal of Science Education*, 13(5), pp.521-531.
- Stewart, D. & Shamdasani, R. (1990). *Focus groups: theory and practice*. Sage Publications, Newbury Park, Calif.
- Stewart, E., Hayward, B. & Devlin, P. (1998). 'The 'place' of interpretation: A new approach to the evaluation of interpretation'. *Tourism Management*, 19(3), pp.257-266.
- Suchman, E. (1967). *Evaluative research*. Russell Sage Foundation, New York.
- Taylor, J., Sexton, N. & Czarnowski, K. (1995). 'Visitor employed photography at Rocky Mountain National Park: A valuation technique'. *Park Science*, 15(1), pp.10-12.
- Taylor, J., Czarnowski, K., Sexton, N. & Flick, S. (1995). 'The importance of water to Rocky Mountain National Park visitors: an adaptation of visitor-employed photography to natural resources management'. *Journal of Applied Recreation Research*, 20(1), pp.61-85.
- Taylor, R. (1993). 'The influence of a visit on attitude and behavior towards nature conservation'. In Thompson, D., Bitgood, S., Benefield, A., Shettel, H. & Williams, R. (Eds.), *Visitor Studies: Theory, Research and Practice Volume 6*. The Visitor Studies Association, Jacksonville, Alabama.
- Tisdell, C. & Wilson, C. (2001). 'Wildlife-based tourism and increased support for nature conservation financially and otherwise: evidence from sea turtle ecotourism at Mon Repos'. *Tourism Economics*, 7(3), pp.233-249.
- Theobald, W. (1987). 'Historical antecedents of evaluation in leisure programs and services'. *Journal of Park and Recreation Administration*, 5(4), pp.1-9.
- Tilden, F. (1957). *Interpreting our heritage*. University of North Carolina Press, Chapel Hill, North Carolina.
- Tubb, K. (2003). 'An evaluation of the effectiveness of interpretation within Dartmoor National Park in reaching the goals of sustainable tourism development'. *Journal of Sustainable Tourism*, 11(6), pp.476-498.
- Tuckey, C. (1992). 'School children's reactions to an interactive science centre'. *Curator*, 35(1), pp.28-38.
- Uzzell, D. (1998a). 'Strategic considerations and practical approaches to the evaluation of heritage interpretation'. In Uzzell, D. & Ballantyne, R. (Eds.), *Contemporary issues in heritage and environmental interpretation: Problems and prospects*. Stationery Office, London, pp.185-202.
- Uzzell, D. (1998b). 'Evaluating the effectiveness of heritage and environmental interpretation'. In Uzzell, D. & Ballantyne, R. (Eds.), *Contemporary issues in heritage and environmental interpretation: problems and prospects*. Stationery Office, London, pp.203-231.
- Vander Stoep, G. & Gramann, J. (1987). 'The effect of verbal appeals and incentives on depreciative behavior among youthful park visitors'. *Journal of Leisure Research*, 19(2), pp.69-83.
- Wagar, J. (1972). *The recording quizboard – a device for evaluating interpretive services*. USDA Forest Service Research Paper PNW-139. Pacific Northwest Forest and Range Experiment Station, Portland, Oregon.
- Wagar, J. (1976). 'Evaluating the effectiveness of interpretation'. *Journal of Interpretation*, 1(1), pp.1-8.
- Ward, C. & Roggenbuck, J. (2003). 'Understanding park visitors' responses to interventions to reduce petrified wood theft'. *Journal of Interpretation Research*, 8(1), pp.67-82.
- Ward, C. & Wilkinson, A. (2006). *Personal interpretation – a field guide for success*. Fulcrum Applied Communication Series, Fulcrum Books, Golden, Colorado.

- Weinreich, N. (1999). *Hands-on social marketing: A step-by-step guide*. SAGE Publications, Thousand Oaks, Calif.
- Wells, M. (2000). 'Summative evaluation of Golden Gate Canyon State Park visitor centre'. *Journal of Interpretation Research*, 5(2), pp.59-64.
- Westphal, J. & Halverson, W. (1985-86). 'Assessing the long-term effects of an environmental education program: A pragmatic approach'. *Journal of Environmental Education*, 17(2), pp.26-30.
- Widner, C. & Roggenbuck, J. (2000). 'Reducing theft of petrified wood at Petrified Forest National Park'. *Journal of Interpretation Research*, 5(1), pp.1-18.
- Winter, P., Cialdini, R., Bator, R., Rhoads, K. & Safarin, B. (1998). 'An analysis of normative messages in signs at recreation settings'. *Journal of Interpretation Research*, 3(1), pp.39-47.
- Wolf, R. (1980). 'A naturalistic view of evaluation'. *Museum News*, 58(1), pp.39-45.
- Woods, B. & Moscardo, G. (1998). 'Researching interpretive techniques in tourism: an evaluation of the effectiveness of pictorial symbols in reef tourist education'. In Faulkner, W., Tidswell, C. & Weaver, D. (Eds.), *Progress in tourism and hospitality research, 1998, Part 1*. Proceedings of the Eighth Australian Tourism and Hospitality Research Conference, Gold Coast. Bureau of Tourism Research, Canberra, pp.320-333.
- Woods, B. & Moscardo, G. (2003). 'Enhancing wildlife education through mindfulness'. *Australian Journal of Environmental Education*, 19, pp.97-108.
- Wright, B. & Wells, M. (1990). *A field guide for evaluating National Park Service interpretation*. Centre for Recreation Resources Policy George Mason University, Richmond, Virginia.
- Yamada, N. (2002). 'Development of a manual for evaluating interpretive programs in Japanese nature parks'. Masters thesis, University of Idaho, Moscow, Idaho.
- Yamada, N. & Ham, S. (2004). 'Assessing interpretation: evaluative concepts and application to natural parks'. *Japanese Journal of Outdoor Education* 7(2), pp.1-12.
- Young, C. & Witter, J. (1994). 'Developing effective brochures for increasing knowledge of environmental problems: The case of the gypsy moth'. *Journal of Environmental Education*, 25(3), pp.27-34.
- Zeidler, A. & Surber, J. (1999). 'Understanding topic, structure, and importance of information in a visual and verbal display'. *The Journal of Experimental Education*, 67(2), pp.114-134.
- Zuefle, D. (2004). 'Processing and utilizing counterintuitive information in interpretation and management: a case study'. *Journal of Interpretation Research*, 9(1), pp.49-53.

## **AUTHORS**

### **Prof Sam H. Ham**

Sam Ham is Director of the Center for International Training and Outreach and Professor of environmental communication and international conservation in the University of Idaho's College of Natural Resources, Department of Resource Recreation and Tourism. He also holds a courtesy appointment as Adjunct Professor at Monash University. Dr Ham teaches graduate courses in interpretation, environmental communication, international issues in nature conservation, and recreation and tourism management. His research has focused on interpretation and strategic communication. He has authored more than 300 publications including two widely acclaimed books on interpretive methods. Email: sham@uidaho.edu

### **Prof Betty Weiler**

Betty Weiler is Director of the Monash Tourism Research Unit and Professor of Tourism at Monash University and has been teaching, researching and writing in the area of tourism marketing, management and planning for twenty years. She has published over one hundred journal articles and book chapters, presented 15 invited addresses and plenaries, and delivered dozens of symposia papers and workshops. Dr Weiler has managed or co-managed over a dozen major research projects and seven international and national consultancy projects related to ecotourism, interpretation and tour guiding, and is known for her industry-relevant and applied research focus. Email: betty.weiler@buseco.monash.edu.au



The Sustainable Tourism Cooperative Research Centre (STCRC) is established under the Australian Government's Cooperative Research Centres Program. STCRC is the world's leading scientific institution delivering research to support the sustainability of travel and tourism - one of the world's largest and fastest growing industries.

## Research Programs

Tourism is a dynamic industry comprising many sectors from accommodation to hospitality, transportation to retail and many more. STCRC's research program addresses the challenges faced by small and large operators, tourism destinations and natural resource managers.

**Areas of Research Expertise:** Research teams in five discipline areas - modelling, environmental science, engineering & architecture, information & communication technology and tourism management, focus on three research programs:

**Sustainable Resources:** Natural and cultural heritage sites serve as a foundation for tourism in Australia. These sites exist in rural and remote Australia and are environmentally sensitive requiring specialist infrastructure, technologies and management.

**Sustainable Enterprises:** Enterprises that adhere to best practices, innovate, and harness the latest technologies will be more likely to prosper.

**Sustainable Destinations:** Infrastructural, economic, social and environmental aspects of tourism development are examined simultaneously.

**Postgraduate Students:** STCRC's Education Program recruits high quality postgraduate students and provides scholarships, capacity building, research training and professional development opportunities.

**THE-ICE:** Promotes excellence in Australian Tourism and Hospitality Education and facilitates its export to international markets.

## Commercialisation

STCRC uses its research network, spin-off companies and partnerships to extend knowledge and deliver innovation to the tourism industry. STCRC endeavours to secure investment in the development of its research into new services, technologies and commercial operations.



## Australia's CRC Program

The Cooperative Research Centres (CRC) Program brings together researchers and research users. The program maximises the benefits of research through an enhanced process of utilisation, commercialisation and technology transfer. It also has a strong education component producing graduates with skills relevant to industry needs.

# Sustainable Tourism Cooperative Research Centre

## DARWIN

### NT Coordinator

Ms Alicia Boyle  
Tel: + 61 8 8946 7267  
alicia.boyle@cdu.edu.au

## PERTH

### WA Coordinator

Dr Diane Lee  
Tel: + 61 8 9360 2616  
d.lee@murdoch.edu.au

## ADELAIDE

### SA Coordinator

Prof Graham Brown  
Tel: +61 8 8302 0313  
graham.brown@unisa.edu.au

## MELBOURNE

### VIC Coordinator

A/Prof Sue Beeton  
Tel: +61 3 9479 3500  
s.beeton@latrobe.edu.au

## HOBART

### TAS Coordinator

Adjunct Prof Malcolm Wells  
Tel: + 61 3 6226 7686  
Malcolm.Wells@utas.edu.au

## CANBERRA

### ACT Coordinator

Dr Brent Ritchie  
Tel: +61 2 6201 5016  
Brent.Ritchie@canberra.edu.au

## CAIRNS

### NQ Coordinator

Prof Bruce Prideaux  
Tel: +61 7 4042 1039  
bruce.prideaux@jcu.edu.au

## BRISBANE

### Managing Director - STS

Mr Stewart Moore  
Tel: +61 7 3321 4726  
stewart@crctourism.com.au

## BRISBANE

### QLD Coordinator

Mr Noel Scott  
Tel: +61 7 3381 1024  
noel.scott@uq.edu.au

## LISMORE

### NSW Coordinator

**Regional Tourism Research**  
Dr Jeremy Buultjens  
Tel: +61 2 6620 3382  
jbuiltje@scu.edu.au

## SYDNEY

### Sustainable Destinations

Mr Ray Spurr  
Tel: +61 2 9385 1600  
r.spurr@unsw.edu.au



## NATIONAL NETWORK

## INDUSTRY PARTNERS



## UNIVERSITY PARTNERS



THE UNIVERSITY OF  
NEW SOUTH WALES



## SPIN-OFF COMPANIES



## SUSTAINABLE TOURISM



## CRC

CRC for Sustainable Tourism Pty Ltd ABN 53 077 407 286  
PMB 50 Gold Coast MC  
Queensland 9726 Australia  
Telephone: +61 7 5552 8172 Facsimile: +61 7 5552 8171  
Chairman: Sir Frank Moore AO  
Chief Executive: Prof Terry De Lacy  
Director of Research: Prof Leo Jago

Website: [www.crctourism.com.au](http://www.crctourism.com.au)  
Bookshop: [www.crctourism.com.au/bookshop](http://www.crctourism.com.au/bookshop)  
Email: [info@crctourism.com.au](mailto:info@crctourism.com.au)



An Australian Government Initiative